LNCS 3675

Yuhua Luo (Ed.)

Cooperative Design, Visualization, and Engineering

Second International Conference, CDVE 2005
Palma de Mallorca, Spain, September 2005
Proceedings

Springer

Lecture Notes in Computer Science 3675

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Yuhua Luo (Ed.)

Cooperative Design, Visualization, and Engineering

Second International Conference, CDVE 2005
Palma de Mallorca, Spain, September 18-21, 2005
Proceedings

Springer

Volume Editor

Yuhua Luo
University of Balearic Islands
Department of Mathematics and Computer Science
07122 Palma de Mallorca, Spain
E-mail: dmilyu0@uib.es
http://www.cdve.org

Library of Congress Control Number: 2005932201

CR Subject Classification (1998): H.5.3, H.5.2, H.5, H.4, C.2.4, D.2.12, D.4, H.2.8

ISSN 0302-9743
ISBN-10 3-540-28948-8 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-28948-7 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springeronline.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11555223 06/3142 5 4 3 2 1 0

Preface

After one year, the major actors in the field of cooperative design, visualization and engineering gathered together again by the side of the beautiful Mediterranean Sea to exchange the research and development experience in the field. CDVE 2005 served as a forum to promote the research in the field and it attracted great attention from the CDVE community. This year, we received contributions from over 100 authors, 5 continents and more than 20 countries.

As we can see, great progress in research and development has been achieved since the last conference. We received papers on cooperative design, cooperative visualization, cooperative engineering and other cooperative applications. As an important trend, the researchers have started to attack the problems in CDVE from a more generic base. We are happy to see contributions such as constraint maintenance, decision support, and security enforcement for CDVE. Case studies and application-specific developments are among the cooperative visualization papers. Along the line of cooperative engineering, knowledge management, reconfigurability, and concurrency control are major issues being addressed.

Cooperative working in design, visualization, engineering and other areas has different degrees of cooperation. I classify them as strong cooperation, intermediate cooperation, and light cooperation. Strong cooperation involves real-time multiple-user multiple-location modification to the same workspace. Light cooperation exists in the applications where the basic working relationship is only information or workspace sharing among the cooperative entities, no modification to the workspace is involved. Therefore, any application that is shared by more than one single user can be considered as a light-degree cooperative application. Any application between these two extremes can be considered as intermediately cooperative. Our conference addressed the common and specific issues of all of them.

I would like to express my thanks to the Program Committee and the numerous volunteer paper reviewers for their generous contribution to the conference which made it at a high-quality event. Without their support, the CDVE 2005 conference would not have been successful.

September 2005 Yuhua Luo

Organization

Conference Chair:	Professor Yuhua Luo University of Balearic Islands Spain

International Program Committee

Program Chair:	Professor Dieter Roller Institut für Rechnergestütz Ingenieursysteme University of Stuttgart Germany

Members

Katy Böner
Susan Finger
Gu Ning
Kurt Kosanke
Francis C.M. Lau
Moira C. Norrie
Li Bohu
Mikael Jern
Bjørn Erik Munkvold
Zhang Heming
Mary Lou Maher
Benoît Otjacques
Weiming Shen
Ram Sriram
Chengzheng Sun
Jos P. van Leeuwen
Carlos Vila
Nobuyoshi Yabuki

Table of Contents

Building a CSCW Infrastructure Utilizing an M&S Architecture and XML

Jenny Ulriksson[1], Farshad Moradi[1], Mattias Liljeström[2], and Nicholas Montgomerie-Neilson[2]

[1] Swedish Defense Research Agency, Department of Systems Modeling, 172 90 Stockholm, Sweden
{jenny.ulriksson, farshad.moradi}@foi.se
http://www.foi.se
[2] Royal Institute of Technology, Stockholm, Sweden

Abstract. CSCW applications provide virtual spaces for human-human cooperation. Within Modelling and Simulation (M&S), CSCW support is highly beneficial but, as in most single-user applications, not natively provided. External support, such as infrastructures for creating or adapting applications for collaborative work, is generally missing in the CSCW community. At the Swedish Defence Research Agency we study the possibilities of defence CSCW, for the primary intention of collaborative M&S. Since most of our M&S complies with the HLA (a distributed simulation architecture), HLA appeared as a candidate for building a CSCW infrastructure upon. Thus we have designed and developed a prototype of a CSCW infrastructure, based on a combination of HLA and XML. The goal is to provide a foundation for developing CSCW applications, and for adapting existing applications to collaborative work. This paper presents the design and development, and experiments conducted for verifying utilization of HLA and XML for the purpose.

1 Introduction and Motivation

CSCW *(Computer Supported Cooperative Work*) is a term often used for human collaboration in computer environments. CSCW enables people to share applications and tools, and at the same time communicate, to collaboratively perform or solve tasks. Hence virtual spaces for human-human cooperation can be created to support people and organizations to overcome geographical distances. The problem is that most single-user applications are specialized for their purpose, and do not naturally provide CSCW support. External support, such as infrastructures for developing CSCW software or adapting existing applications to collaborative work, is generally a missing part within the CSCW community.

At the Swedish Defense Research Agency (FOI) a project called NetSim conducts research activities concerning among other things the possibilities of collaborative M&S within the defense. The need for a CSCW infrastructure was identified within this work, and the lack of such existing ones determined. Consequently, the decision was made to develop our own. Most M&S within our research complies with an IEEE

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 1 – 13, 2005.

standardized simulation architecture called the *High Level Architecture* (HLA), and the architecture came to mind as a potential candidate technology also for CSCW. The technology requirements for building a CSCW infrastructure are many, and HLA not only provides some of them, but also offers additional valuable features. But the question that rises is: is HLA really suitable for use in distributed CSCW?

We propose an infrastructure for CSCW that is based on a combination of HLA and XML (*Extensible Markup Language*). The design has partly been implemented in a prototype called *Collaborative Core* (CC), which is based on a replicated distributed architecture. All information transmitted in the system is structured according to XML information models, and HLA services are used for transmission and other functions. Performance experiments have been conducted to verify both the use of XML for the purpose, and for verifying the combination of HLA and XML as the foundation of a CSCW infrastructure. The result confirmed the suitability of our proposal, but also revealed some HLA limitations that must be considered.

This paper presents the proposed design and the developed CC prototype. It also presents experiment results, our conclusions, some recommendations, and future work.

Clarifications for the reader: when further on discussing *nodes*, we mean client computer environments (i.e. PCs with a client using it). When referred to, one *node* hosts only one *client* (though in reality a node may host several clients). When discussing *applications*, what are intended are the applications used for collaboration (i.e. shared tools) within a *collaboration session,* and not the application providing CSCW services. A *collaboration session* is a session that transpires whenever two or more users jointly and concurrently perform tasks using interconnected software.

2 NetSim and HLA

At FOI we have performed research within network based M&S for several years. Recently we initiated activities concerning development of a common platform for defense M&S, and we believe CSCW support to be a very valuable service in such an environment. In this work we quickly identified the lack of such infrastructures that could be easily integrated with M&S applications and within the proposed common platform, and decided to develop our own.

2.1 The NetSim Project and CSCW

The NetSim project was initiated in 2003, at the Department of Systems Modeling at FOI. The primary project goal is to study, develop and modify methods and techniques for the purpose of network based M&S, and is described in [1]. One of the main activities aims at constructing an architecture, and develop a prototype, that can act as a common platform for defense M&S – the NetSim Platform. Apart from issues within distributed simulation such as fault-tolerance and distributed execution, one of the major project actions has concerned exploiting the potential of CSCW for defense applications. Since defense related operations engage actors which are often spread over long distances, cooperation is not easily achieved and the correct knowledge and support are not accessed without difficulty. A common defense platform that provides CSCW support may not replace real human-human cooperation, but can constitute an essential alternative.

2.2 Collaborative M&S Within NetSim

The area of M&S can substantially benefit from collaborative services, since CSCW can facilitate the provision of support that is required in the M&S process. Here it does not only make M&S expertise and knowledge more easily available, but also activities such as distance training and education more easily accessed. Moreover, the quality of M&S activities and products can be secured and controlled. The problem is that M&S applications are specialized for their purpose, and do not naturally provide support for collaborative services.

As a first attempt to address the problem, a simple prototype of a collaborative M&S environment was designed and developed [2]. This was based on JXTA[1] Peer-to-peer technology. The prototype was dismissed but gave us valuable experiences. As an example, CSCW support should not be based on an application specific approach, but rather constitute a more general infrastructure for various applications and purposes. More advanced support for time mechanisms were also needed. A final lesson was that it is preferable to use a more mature technology, one that supports more functionality than the one used. But we identified no infrastructures that suited above functionality. Various applications support some of the functions we desire, but none have been successful, and few support all of the desired functionality.

A general problem that developers of CSCW applications face is the complexity of integrating management of collaboration groups and activities within an application. As a result, the superior software for most computer-related professional tasks are single-user local desktop applications, and the dedicated CSCW software tend to lag behind in other than CSCW functionality [3]. Especially within domain specific applications, such as M&S applications, this kind of support is rarely seen. Thus we decided to address the problem through in-house design and development of the needed infrastructure, a foundation we call *Collaborative Core* (CC). CC will provide collaborative services, and should due to the nature of defense operations be distributed. CC is designed and has so far partly been implemented (see Section 3.1 for more detail). The development has been performed in close cooperation with two students from the Royal Institute of Technology [4, 5].

2.3 The High Level Architecture

Whilst parallel simulation aims at reducing the total simulation execution time (so called *speed-up*), the science of distributed simulation (DS) has different goals [6]. DS is the M&S answer to the Object Oriented thinking in the community of Computer Science. In DS, simulation components are executed on different computing nodes and coordinated in a joint simulation. This makes it possible for simulations too

---

[1] JXTA is a Java based, open-source Peer-to-peer framework distributed at http://www.jxta.org/ [accessed February 2005].

large for execution on one single computer to be distributed and executed on several nodes. Further it facilitates and makes it possible for different kinds of simulation components to interact, despite residing in different computer environments.

Above features with DS assume simulation component interoperability, namely that all components utilize a common standard. The *High Level Architecture* (HLA) is a framework for component-based simulation systems, originally developed within the US Defense Modeling and Simulation Office (DMSO) [7] and is the proposed simulation standard within the Swedish Defense. HLA provides a common communication structure, and rules for simulation components (so called *federates*) to follow, assuring them to be interoperable and able to together act in a distributed simulation (so called *federation*) [8]. The federation is managed and communication is transmitted through an implementation of the HLA services, a distributed operating system, called RTI (*Runtime Infrastructure*). RTI provides required services, such as:

- *Federation Management* – creating and controlling the federation execution etc.
- *Time Management* (TM) – flexible and advanced means of federation time management
- and *Data Distribution Management* (DDM) – mechanisms for efficient routing of information among federates

3 CC Design

When the CC infrastructure was developed, a concept was first designed and proposed. Thereafter technologies for implementation were chosen. The developed prototype was a partial implementation of the CC design, and is described in Chapter 4.

3.1 CC Infrastructure: The Concept

The CSCW service in the NetSim platform will be provided as a module that supports the user transparently with group management and tool sharing services. Applications using the services will do so without user specific complementary action. CC leverages support and services required for a user to be able to start, administrate, and participate in computer based collaboration groups, and hence also for sharing tools and other functions in NetSim. Besides, CC will offer development support for integrating new tools for CSCW. In general, the CC design is comprised by three main components, presented in Table 1.

Our research and development has focused on the two latter issues. Concerning the second component, the CC framework will provide a pluggable interface for collaborative applications, and that relieves tools of responsibility for managing user groups, and most of communication responsibility (discussed in Section 4.3). This way, developers will be supported in developing new CSCW tools, and in modification of existing applications to become collaborative. The third component represents straight-forward services for CSCW group management and administration, including shared group status and shared areas.

Table 1. The three CC main components, of which number 2 and 3 are treated in this paper

CC Service	Content
1. Communication Means	Integrated tools for communication
2. Application Interface	API, interface and software as development support for integrating new tools into CC
3. Group Management Services	Creating and maintaining collaboration groups, group administration etc.

3.2 Requirements and Design Choices

The overall requirements for CSCW in the NetSim platform, and consequently for the CC module, are described in Table 2. Considering these, some design choices were made. The last requirement lead us to choose a replicated architecture for the CC, i.e. all clients execute their own equal set of applications, and are themselves responsible for taking the correct action, to the retrieved changes from other clients. Replication was chosen since various client types may have different requirements in terms of GUI complexity and limited network connections etc. Transmitting only data, and letting the client specific application process it, is here a suitable solution.

The aim of the CC application interface is to provide support for plugging tools into the CC. There are two ways of accomplishing this, either by letting CC take all responsibility, or through requiring modification and responsibility from the application. Allowing no tool modifications at all requires a very generic CC interface, which is hardly possible. Thus we decided that CC can set some application requirements, with the primary aim to keep modifications at a minimum.

Table 2. Primary requirements for CC

Requirement	Significance
Distributed CSCW	The environment should be distributed. But this does not exclude future architecture combinations
Synchronous work	The CSCW intended here is immediate and synchronous
Short Persistence Collaboration Groups	Collaborative activities within NetSim are assumed to most often be directly task oriented, i.e. life times of collaboration sessions are assumed short
Small group sizes	Groups are assumed small, 2-8 persons. Considering scalability and regarding HLA as technology larger groups are possible. However, that requires social support to address virtual conflicts etc., an issue not handled here
Various client types	Different client types are considered, such as thin clients with poor network connections. In this regard, we assume virtual worlds to be too complex for the purpose

3.3 HLA for CSCW?

Since the acceptance as an IEEE standard (IEEE 1516), the area of application for HLA has been broadened to both non-military and non-simulation areas. An example is the multiplayer online gaming framework by Vuong et al. [9], and the collaborative virtual shopping mall by Zhao & Georganas [10]. The latter made an evaluation of the HLA as an enabler of collaborative virtual environments, with overall affirmative results [10].

Using an already existing architecture to build a CSCW infrastructure upon, may contribute to avoiding unnecessary development. HLA provides an infrastructure with essential services that beneficially could be used for CSCW, such as time management, group management (*Federation Management*), efficient information filtering (DDM), and communication management. Moreover it is a mature technology and standardized. An issue is that the HLA originally was not developed for real-time applications, something that CSCW applications highly are. Thus, the suitability of HLA for the purpose must be evaluated. And as stated, one of the challenging issues for CSCW is consistency management. Advanced, flexible consistency management has been declared a lacking part in current implementations of CSCW and in existing systems [11]. Utilizing HLA and the RTI, which provide advanced, flexible time management, this issue can be addressed appropriately. Moreover, since most simulation within our research complies with the HLA, we assumed it to be a candidate technology for our purpose.

3.4 XML for the Purpose?

XML (the *Extensible Modeling Language*) provides a way of structuring information in a platform independent, human-readable way. XML allows developers to create their own markup languages, and templates and rules (defined in *XML Schemas*), that help assuring the interoperability between data. A very beneficial feature is that XML efficiently separates data from presentation. In a CSCW infrastructure such as described above a lot of information is managed, such as tool specific information, collaboration group information and client information. A design choice could be to strictly follow the technology chosen (here HLA), but to accomplish a more generic structure and less technology dependent, XML can be used for structuring and handling information. Another reason for using XML is that here various client types are expected. Using XML, the same information is provided to all participants. At the client side, parsing of the XML formatted information can allow for user specific utilization and presentation.

3.5 Combining HLA and XML for CSCW

Concluding the above discussion and our experiences from earlier work, we propose a combination of HLA and XML to constitute the foundation for a CSCW infrastructure. It uses XML-based group definitions and a communication infrastructure built on the HLA.

4 Implementation

4.1 The Prototype

A first CC prototype has been implemented, of which the two main components were implemented in separate, but closely coordinated tracks. They are described in

Section 4.3 and 4.4 and in detail in [4, 5]. Besides these, a prototype for the NetSim environment was developed. A simple user GUI was created along with three simple, but demonstrative, prototype applications (shown in Figure 1). These were used for practically evaluating the implemented CC functionality.

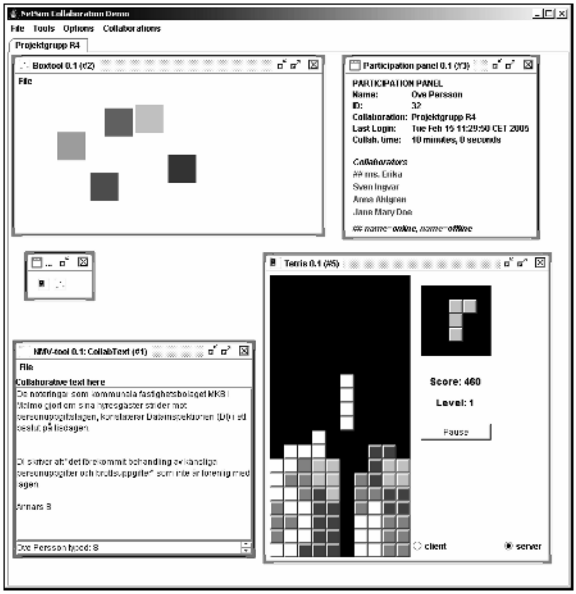


Fig. 1. CC screenshot. Depicted are three simple tools; a text editor, a box drawing tool and a game, that are used within a collaboration session. Additional CC features are shown, such as a Participation Panel (group specific information) and an Activity Panel (alerts tool activity etc.).

4.2 HLA and CC

Two models were considered for organizing the RTI communication at each client, the first one using only one general-purpose federate[2] in each node, that managed both application specific information and group management communication. The second model, which is the one we chose, considered using one group service federate and one tool specific federate for each application used within the collaboration. This implementation leads to several federates within the same client. But having a single general-purpose federate at each client, handling traffic from an unknown number of federates while at the same time managing group functions and also tool specific functions, would lead to a complex design, without significant associated gains in performance. Besides, applications can have different requirements on time management, an issue that is more easily managed if each application has its own HLA federate.

---

[2] Recall that a federate is an individual simulation component (see Section 2.3) that jointly is executed in a *federation* (simulation).

The HLA RTI was utilized as communication medium and for essential facilities such as Time Management, Federation Management and also DDM mechanisms. The specific RTI used was the Pitch developed pRTI. DDM provides services for efficient routing of information, i.e. provides means of traffic filtering among federates. The DDM functionality is in our solution for example used for directing information (*instant messaging*), and forming subgroups for communication within collaboration groups. Furthermore, a distributed lock was implemented using HLA that is used for restricting mutual exclusion of some methods.

4.3 CC Services

Implemented services of CC address four areas of functionality:

- *Group services* – searching, creating, joining and resigning collaboration groups
- *Settings* – administration services for reading and writing information about tools and members in groups and other properties
- *Communication features* – data transmission and data filtering
- *Other features* – instant messaging, desktop layout sharing, presence awareness and activity awareness etc.

The central mechanism of the CC services is the *Collaboration Description* (CD). This is the vehicle of all relevant information within a collaboration group, including settings, member information and status etc. It does not hold tool specific settings, which are part of the application interface component (see Section 4.3). The CD takes two forms within CC. Distributed and propagated amongst CC clients, the CD is implemented in XML. And for facilitated management purposes, when loaded into client computer memory, the CD is implemented as a DOM document wrapped in a class called `CollaborationDescription`. The `CollaborationDescription` provides valuable methods for accessing the data, and parsing services to and from XML.

CC clients collaborate in groups and sessions, described in the CD, which can be seen as a replicated state XML information model. A user may create and join several groups. The summarized group state is stored in a persistent storage space, when the last member leaves the group. As a member rejoins the group, he or she initiates the last saved state and session of the group, through retrieving the last update of the group specific CD. In the same manner a late-joiner is updated and initiated, using the CD.

4.4 CC Application Interface

As stated before, it should be possible to plug tools into CC, without much modification. This requires a generic interface, which provides communication facilities among the similar tools used within the same collaboration session. Each application implements a thin CC interface and the few additional required methods that will be used by CC.

When an application is started in collaborative mode, CC assures a federate is started on behalf of the application, and with eventual required settings appended from the tool, such as Time Management (TM) mode. The TM decides in what fashion communication between applications is synchronized and handled, and consequently the consistency mechanism used. The federate's TM mode can be changed at any time during a session, without restarting the application. So far two modes have been implemented, conservative and relaxed, but further development will consider other types as well.

Applications communicate in the form of messages, and all messages are formatted in XML and validated towards an XML Schema created for this reason. To support applications in using, parsing and reading XML, the Application Interface provides such functionality. There exist three different types of messages, described in Table 3.

Table 3. The communication between CC applications is held through three different kinds of messages, as described here

Message Type	Description
Event	General application event. Used for communication among applications; update events, state changes etc.
Internal Event	Federate internal events communicated only between federates, for example distributed locks
Status Event	The collected status of an application at a specific timestamp. Used for updating newcomers, restarting a sleeping collaboration etc.

5 Experiments and Results

As presented, HLA and XML were successfully used for developing an infrastructure for CSCW. The question that remains is whether the performance of the implemented solution is sufficient for our needs. Two kinds of experiments were performed. Primary parameter of our interest here was time, i.e. message transmission times, since this is an important factor for update time (and that causes delay) and consistency management in CSCW.

5.1 XML Conversion Tests

The first experiment concerned evaluation of XML serialization. The time consumption of conversion tests in both directions was measured, i.e. to and from XML. The converter used was XStream[3], to comprise information from and to XML within the applications, when receiving updates and information. It is essential that the process of doing so is not too time consuming. The test result is presented in Figure 2 below.

Test bed: 1 PC with 256 Mb RAM, Intel Pentium IV 1.5 GHz processor, running Windows XP.

[3] XStream is an open-source package developed by Codehaus. Homepage: http://xstream.codehaus.org/ [accessed February 2005].

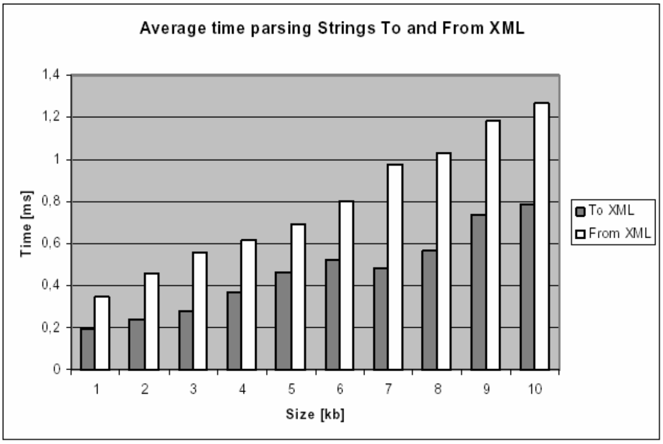


Fig. 2. Conversion tests of Java String to and from XML. The X-axis displays the size of Strings parsed, and the Y-axis the parsing time in milliseconds. As is seen, total parsing time shows good performance, displaying an even linearity, and very low average parsing values.

5.2 HLA Versus Sockets

Utilizing the already developed and mature facilities of the HLA framework was successful for the purpose. However, before determining the overall HLA suitability HLA, some performance tests had to be conducted in order to conclude if using the M&S architecture conveyed too much overhead compared to more light-weight architectures. First a pure socket implementation was measured, where chosen message sizes were evaluated towards message transmission times (transfer rate). Then the same set of experiments was conducted using communication through the RTI instead. Figure 3 and 4 presents the result of these experiments respectively in two histograms. The same notations are used in both, and results are here displayed for twenty tests – the 50 kB test is shown front left, and the 1Mb test is shown furthest back into the diagram. Horizontally, the diagram is divided into millisecond-wide intervals of transfer-time. Vertical bars show the number of messages for a particular test that were transferred within a particular millisecond interval.

Test bed: 2 PCs equally configured Pentium III 1GHz, with 256 Mb RAM, running XP, connected in a LAN with a 100Mbit switch..

The result displayed severe differences between the two experiments. First of all the HLA was on average slower, but looking merely at the average, HLA compared with the socket-based solution in an acceptable way. However, the RTI transfer rates presented strange behavior when sending larger packages. When message sizes increased from 550 kilobytes to 600 kilobytes, a huge increase in transmission times was observed. When exposing these observations for the developer, Pitch, they explained that this discrepancy may relate to the complex implementation of memory buffers that pRTI possess. A number of suggestions for running the experiments, that for example refer to the use and functionality of the JVM, and trying again were proposed, but have not yet been conducted.

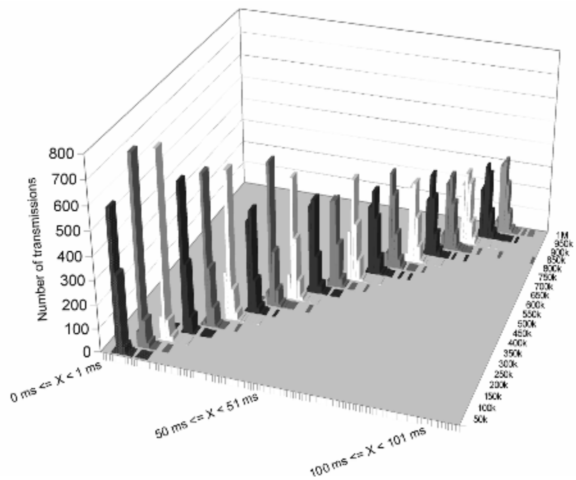


Fig. 3. Socket-communication. Data follows a normal distributed pattern along a linear curve.

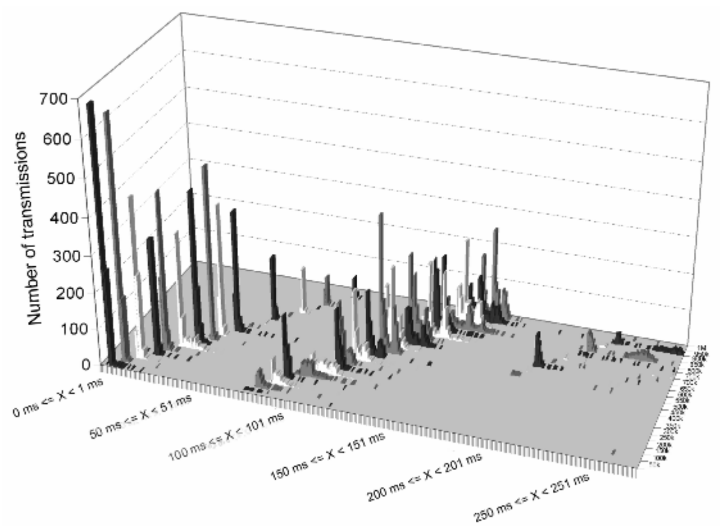


Fig. 4. pRTI-communication. Here, result shows data is distributed in a multi-modal way.

5.3 Conclusions

Considering the XML tests, the main conclusion drawn was that XML fulfills the requirements of our purpose, and that the XStream converter successfully was able to convert all custom classes given to it.

HLA performed well compared to a socket-based system, which verified the use of it compared to other architectures. However, RTI showed some disturbances when using large message packets, with sizes above 550 kilobytes, and did not display the same linear behavior as the socket-based solution. These diverging transmission times were not acceptable for our solution, and consequently we concluded that those message sizes are not recommended if using HLA for the purpose of CSCW and real-time applications, unless the problem is solved. Despite this, the essential conclusion was that HLA provided useful functionality that could be used instead of developing these ourselves, and additional features that were beneficially utilized, such as DDM.

6 Conclusions and Future Work

Results of performance tests show that the HLA communication architecture compares reasonably well with a socket-based. Overall, results demonstrate feasibility of the CC infrastructure, and of the objective of extending the use of HLA to non-simulation applications. HLA proved well suited as communication structure for real-time user interactive applications, and the already built-in advanced functionality in HLA was beneficially used for CSCW services. Furthermore, XML conversion tests verified efficient use of XML for the purpose and combination with HLA for CSCW.

Recommendation: Pending resolution of the issues around large-message transmission, HLA is not recommended as candidate technology for utilization in the purpose regarded in this paper if typical message package sizes exceed 550 kilobytes.

Future work to adapt full-scale applications to the collaborative framework is invited and planned. Furthermore, more advanced synchronization (and consistency management) will be considered, and the efficiency of respective time management mechanisms for the purpose of CSCW will be evaluated. And last, more extensive HLA experiments, to resolve the large-message transmission issue, are proposed.

References

1. Eklöf, M., Ulriksson, J., Moradi, F: NetSim – A Network Based Environment for Modelling and Simulation. NATO Modeling and Simulation Group, Symposium on C3I and M&S Interoperability, Antalya, Turkey, (2003)
2. Ulriksson, J., Moradi, F., Ayani, R.: Collaborative Modeling and Simulation in a Distributed Environment. Proceedings of the European Simulation Interoperability Workshop (ESIW´03). Stockholm Sweden (2003)
3. Li, D., Li, Rui.: Bridging the Gap Between Single-User and Multi-user Editors: Challengers, Solutions, and Open Issues. Session III of the Fourth International Workshop on Collaborative Editing Systems, New Orleans, LA, USA (2002)
4. Montgomerie-Neilson, N.: A collaborative software infrastructure based on the High Level Architecture and XML. Master's Thesis, Royal Institute of Technology (2005)
5. Liljeström, M.: Generic XML-based Interface for Computer Supported Collaborative Work in an HLA Environment. Master's Thesis, Royal Institute of Technology (2005)
6. Fujimoto, R.: Parallel and Distributed Simulation Systems. Johnn Wiley & Sons, Inc. USA. ISBN: 0-471-18383-0 (2000)
7. Defence Modelling and Simulation Office [online]. Available via: http://www.dmso.mil/. Last accessed February 2005.
8. Kuhl, F., Weatherly, R., Dahmann, J.: Creating Computer Simulation Systems – An Introduction to the High Level Architecture. Upper Saddle River, NJ: Prentice Hall. USA. ISBN: 0-13-022511-8 (1999)

9. Vuong, S., Scratchley, C., Le, C., Cai, X.J., Leong, I., Li, L., Zeng, J., and Sigharian, S.: Towards a Scalable Collaborative Environment (SCE) for Internet Distributed Application: A P2P Chess Game System as an Example [online]. Available via http://www.magnetargames.com/Technology/DAIS-Vuong-Chess-230603R.doc. Last accessed February (2005)
10. Zhao, H., Georganas, N.D.: Collaborative Virtual Environments: Managing the Shared Spaces. Networking and Information Systems Journal, 3(2), 1-23 (2003)
11. Chung, G. & Dewan, P. Towards Dynamic Collaboration Architectures. Proceedings of ACM conference on Computer Supported Cooperative Work, Chicago, USA, November (2004)

Modelization of a Communication Protocol for CSCW Systems Using Coloured Petri Nets

Eric Garcia, Julien Henriet, and Jean-Christophe Lapayre

LIFC, Laboratoire d'Informatique de l'université de Franche-Comté,
16, route de Gray, 25030 Besançon, Cedex, France
tel : +33 (0) 3 81 66 20 92, fax : +33 (0) 3 81 66 64 50
{garcia, henriet, lapayre}@lifc.univ-fcomte.fr

Abstract. Concurrency management protocols are implemented over collaborative platforms in order to manage the shared object modifications. We have defined the *Chameleon* protocol which allows us to obtain a dynamic management of collaborative architecture. Now, we want to use this algorithm over a telemedicine platform within a European project on tele-diagnosis for neurology: *TeNeCi*. In such a domain, a robust (no errors, no losses...) management of interactions between cooperative members is crucial. Indeed, *TeNeCi* must provide an efficient and infallible tool to doctors in order to establish a collaborative diagnosis. So, we must validate and prove the *Chameleon* robustness. Coloured Petri Nets allowed us to correct the *Chameleon* weaknesses we have discovered during this modeling. Indeed, a time-out and an automatic forwarding mechanisms have been implemented in the first version of our communication protocol. In addition, a timestamp has been added to some messages sent to inactive sites. Finally, a lock mechanism during a particular step has been also implemented in the *Chameleon*. Due to those results, we have proved that our protocol is robust. Thus, it can be used in medicine applications.

1 Introduction

In collaborative environments, sharing notion is everywhere. Concurrency management protocols become very important. We have defined the *Chameleon* protocol [1] which allows us to obtain a dynamic management of the collaborative architecture. This protocol has been tested [1] and integrated in an e-maintenance platform to support tele-diagnosis tools. Our aim is to integrate the *Chameleon* algorithm in a telemedicine application developed within the framework of a European project on tele-diagnosis for neurology. Such an application imposes a great robustness to be used in an emergency context. In the first part of this paper, we present this new framework of utilization: the collaborative tele-neurology project *TeNeCi* [2]. The second part of this paper allows us to define the way to modelize our distributed protocol through coloured Petri nets. The third part presents the *Chameleon* specifications and the coloured Petri net we draw to modelize it. The results and the improvements we did to the *Chameleon* regarding the properties of the obtained model are discussed in the fourth part of this paper.

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 14–21, 2005.

2 The TeNeCi Project

With the support of European/INTERREGIII funding (France - Switzerland), our team has been working in close collaboration with the Federation of Neurosciences in Besançon on the creation of the *TeNeCi* project (cooperative teleneurology) [2]. This project will enable neurological expertise to be improved through the Emergency Neurology Network (RUN) which is already established in Franche-Comté hospitals and coordinated by the University Hospital in Besançon. It also seeks to promote a collaborative network structure coordinated by the Vaud University Hospital in Lausanne which will gradually develop in hospitals in the Swiss cantons of Vaud and Neufchatel. *TeNeCi* will ensure improved neurological expertise through the activities of the RUN network. This network, coordinated by the University Hospital of Besançon established between the hospitals in the Franche-Comté region, is dedicated to facilitating the diagnosis and treatment of neurological emergencies. In the long term a veritable centre of cross-border and regional excellence will be created. *TeNeCi* collaborative tele-neurology application works with two complementary modes. In asynchronous mode: one person performs the diagnosis with the tools provided by the application. This mode can be composed of an image search stage to establish a diagnosis, followed by an emission stage to ask for an opinion, for example. This opinion can be given a few days after reception. For this mode, the interface must provide the necessary tools to create a package which must allow the expert to view images and final diagnosis and also to reconstruct how the diagnosis was posed (annotations, image settings, treatments). In Synchronous mode: it uses the same tools allowing real-time collaboration among several people. This mode provides mechanisms to manage concurrent access to particular commands, e.g. contrast and zoom etc. It is composed of an image search stage, not in order to pose the diagnosis directly and to broadcast information to each person involved in the real-time collaborative diagnosis. In this project, a perfect (no errors, no losses...) management of interactions between cooperative members is crucial. Indeed, *TeNeCi* must provide an efficient and infallible tool to doctors in order to establish a collaborative diagnosis. In a medical environment (emergency for example), errors or looses are not permitted, so we must validate and prove the *Chameleon* robustness. The applications that are working according to the synchronous mode are based on our communication protocol *Chameleon*. The specifications of this protocol are available in 4.

3 Bref Introduction to Petri Nets

A *Petri net* is a mathematical modeling tool. It is possible to modelize and verify properties and functionalities of different types of system. Petri nets can be used for production systems but also CSCW systems [3,4,5]. In CSCW systems, Petri nets can modelize behaviors as parallelism, synchronization, semaphore and resource sharing. There are different types of Petri nets : classical [6], timed [7] (with discrete delays), stochastic [3,8,5] (with continuous delays) and coloured [9,10,4] (with data on the marks). Actually, the aim of this paper is to fulfill a qualitative study of our communication protocol in order to verify its robustness. Thus, we decided to use a coloured Petri net to modelize and evaluate the *Chameleon*.

4 Chameleon Specifications Using a Coloured Petri Nets

The first subsection of this part presents the first specifications of the *Chameleon*. The second subsection presents its lacks. The last one the used natations and the designed Petri net.

4.1 The Chameleon Speficiations

Our CSCW Framework called CAliF Multimedia uses distributed shared memory to manage the consistency of discrete media. The main part of the platform is the communication service. It allows cooperating members to broadcast their information. We chose to develop a new token strategy based algorithm. The originality of these algorithm, named *Chameleon*, is to allow the virtual topology to be reconfigured using two communication techniques: a rotating sequencer for one site and a symmetric approach for another, for example. Thus, the token only visits the active sites. Another important characteristic of this algorithm is that the representation of the virtual topology is a distributed shared object and is transported by the token. Ring topology is more efficient than a full connected network topology for a multi-broadcast operation; however this tendency is reversed when the number of active sites decreases. So, we use the second method (symmetric) for the sites that do not participate actively in cooperative work to complete the ring which links the active sites. Actors of a cooperative application can have different roles and rights, thus implying that sites can have different states. *Producer-Consumer* (**PCo**): The site modifies the shared data. It receives information from the other producer sites and sends the result of its operations to the next producer through the token; *Tutor Producer-Consumer* (**TPCo**): The site modifies the shared data and sends the results of its operations to its next producer and to the sites for which it is the tutor. A tutor is a member of the virtual ring, whereas a *tutored* site is not a member of the ring but depends on a tutor; *Simple-Consumer or Tutored site* (**SCo**): The site does not modify the shared data; it only receives information from its tutor via an inactive copy of the token. It never broadcasts a token; *Simple Producer* (**SP**): This site is the only producer. If a *PCo* site is inactive for a given time it becomes *SCo*. The evolution of virtual topology is dynamic and does not require reconstruction, so it is not very costly. It is based on a ring topology, but it can evolve either towards a centralized system if there is a *Simple Producer*, or towards the use of several techniques: a ring between *n* sites and a centralized system from a node. The Chameleon algorithm manages the membership of a site whatever its state may be. The token is not a simple tool allowing a site to broadcast, it contains the virtual topology representation. The consistency of this topology has to be maintained, to ensure that all the sites have the same view of the system. The nature and the treatment of the token depend on the state of the sites that receive it. When a *TPCo* site broadcasts the token, it begins by sending it to the next producer. As broadcast is a blocking operation, the *TPCo* site has to wait until it is terminated. After that, the token is sent in parallel by the *TPCo* site to its tutored site(s) and by its next producer to another producer. So, at the same time, two sites are treating the token. There is no problem of consistency management because one of these two sites is *SCo* and cannot modify the distributed shared memory. Several events can lead *SCo* sites to send messages: for example, they may have to become *PCo* if they have a modification to carry out, or they may have to manage the membership of a new site. In such a case, the *SCo* site can send messages to its tutor which stocks and treats them when it is in possession of the token. In this way, the consistency of the virtual topology as well as the order and integrity of sent messages are maintained. If a *PCo* site is inactive for a given number of token revolutions it becomes *SCo*. When the change of state criterion is reached, the virtual topology has to be modified. This modification depends on the state of the site that causes the change. When a *PCo* site has to become *SCo*, it waits for the token, then it modifies the virtual topology contained in the token and routes this token, which will broadcast this change of state. The modification of the virtual topology on the token is performed by a calculation function, which builds a new topology according to the current topology and to all waiting requests (membership, departure...). If this site wants to become a producer again, it has to send a request that will not be transported by the token (an *SCo* site cannot send a token); this request will be treated by its tutor. Then, if a *TPCo* (tutor) site becomes *Simple-Consumer*, it has to find new tutor(s) for its tutored site(s), find a tutor for itself, modify the virtual topology on the token, send the token to the next producer site, send the token to its tutored site(s).

4.2 Lacks of the Chameleon Specifications

The description of the way *Chameleon* works is not complete. Precisions are required and it may be possible to optimize it. There are concurrent processes, and there is no details on the way following messages have to be treated: the *Chameleon* token itself, copies of the token, connection of a new site, disconnection of a site, an active site (*PCo* or *TPCo*) which becomes inactive (*SCo*), an *SCo* site which asks for becoming active, in addition, is it sure that the token is never lost ? When an *SCo* site becomes *PCo*, are there any token copies left ? Is there always at least one active site ? Are there any lost messages, somehow ? Does a site always know to which one transmit the messages ? Indeed, the real topology is on the token, whereas each site has its own copy of this topology, which is refreshed each time the token or a copy of the token is received. As a matter of fact, the copy of the topology on each site is different, and at a time, only one site knows the real one: the owner of the token. So, for example, what happens when an *SCo* site sends an *SCo2PCo*-request to a disconnected site ? Is it even possible ? On a first step, we specified the behavior of the protocol for 3 sites through a classical Petri net (no colour). The second step consisted in describing the *Chameleon* for 4 sites through a classical Petri net. Since it is difficult to read those Petri nets, finally, we transformed manually those Petri nets into one coloured Petri net. Since there is no tool that permits to validate the obtained coloured Petri net, we made manual verifications of this last Petri net [11]. Only this coloured Petri net is presented in this paper. A colour is given to each site, and the number of colours is not limited. Thus, This Petri net works whatever the number of sites is. To render it readable, we have chosen to represent it by parts, using 2 figures. Each figure deals with a particular behavior of the protocol. Some places, like *SCo, PCo, NC, Token, Signal* are drawn in different figures, but each place *PCo* refers to the same place whatever the figure is, as for the other places. In addition, we have assumed the fact that when a site becomes *Simple-Consumer*, then the next *Producer-Consumer* site considering the way the token is transmitted around the ring becomes its tutor.

4.3 Behavior of a Producer-Consumer and a Simple Consumer

The figure 1 describes the behavior of a *Producer-Consumer (PCo)* site. This Petri net considers the topology transmitted by the token. Indeed, each site has its own view of the topology and the reference topology is the one transmitted by the token. Thus, if a C_i-coloured mark is in the place *PCo*, that means that this site is *PCo* in the topology transmitted by the token, but this site may be *Simple-Consumer (SCo)* in the topology of another site. The topologies of each site are not represented in this model. The initial marking is not specified in the figure since it depends on the number of sites. Each site has a colour, and there is an order between those colours. This order corresponds to the way the token is transmitted around the logical ring of sites. Indeed, at initial time, there is a mark of each colour in the place *PCo*, and for each site *i*, there are *ITER_i* C_i-coloured marks in the place *Iterations*. The token is given to one site *i*, thus, there is one C_i-coloured mark in the place *Token* and one $Succ(C_i)$-coloured mark is placed in *Next*, since, a priori, the next site to which the token will be transmitted is $Succ(C_i)$-coloured one. In order to update an object, one site i has to be the owner of the token. Before determinating the next active site, all the signals received by the other sites have to be treated. Under those conditions, i determinates which site to send the active token, and the token copies. The active token is sent to the next active site before the inactive token copies to its tutored sites. The figure 2 shows the behavior of a *Simple Consumer (SCo)* site. In order to become active, one site i has to treat all the inactive token copies left before sending a signal to its tutor. If it is not the wright tutor, the signal is forwarded

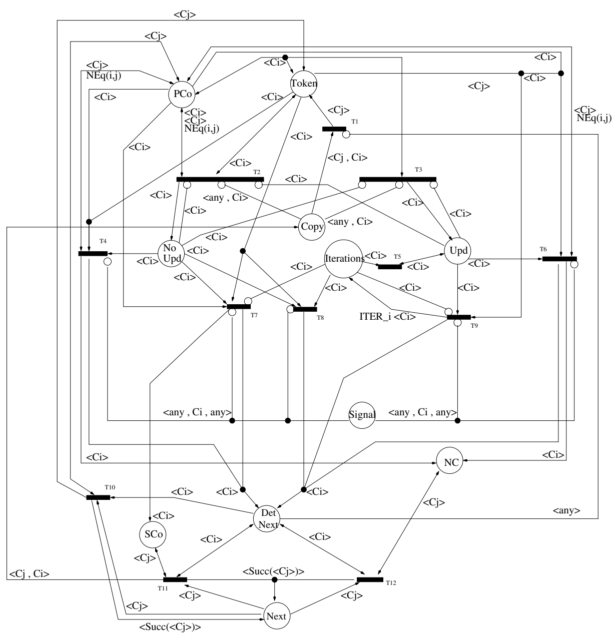


Fig. 1. Behavior of a PCo

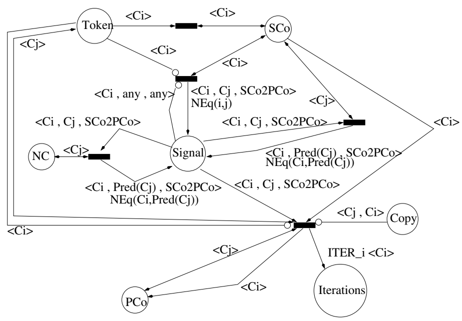


Fig. 2. Behavior of a *Simple Consumer*

to the next active site. i will become PCo the next time its tutor will be the owner of the active token. The detailed behaviors of an SCo and PCo sites are described in [11].

5 Model Analysis. Details Given to the First Specifications

In [11], we verified some properties of the obtained coloured-Petri net for 3 connected sites and 4 token revolutions without any updating before comming *SCo*. We drove different significant scenarii and drew significant parts of the *accessible marking graph*. Indeed, since this Petri net is coloured, unlimited and contains inhibitor arcs, it is neither possible to use a software to automatically find out its properties, nor possible to draw the whole *accessible marking graph*. Thus, this study has been made manually and we verified different properties according to significant initial markings (cooperative system states) and scenarii (site actions). Assuming there are one C_i-coloured mark $\forall i \in \{1; ...; n\}$ in the place *PCo*, one C_i-coloured mark $i \in \{1; ...; n\}$ in the place *Token* and one $Succ(C_i)$-coloured mark in the place *Next*, In [11], we have verified that the Petri net is alive. Indeed, it is possible to cross each transition, and there is no lock in our protocol. Thus, there is no way to go to an inconsistent state. In addition, there is always at least one mark in the place *PCo* which means there is always one active site at least. All sites can not become SCo or disconnected at the same time. We have verified also that (*PCo, SCo, NC*) constitutes an invariant marking for each colour. That means the state of each site is never lost and always either PCo, SCo, or NC. In addition, each site has always only one consistent state on the token. If we consider only the colours in the place *PCo* at any time, there is only one mark of those colours in the place *Token* or in the place *DetNext*. This property proves the token is never lost and there is only one active token. One of the active sites is either in critical section, or determining which site is the next one. In addition, there is always no more than one mark in one of the places *Upd* or *NoUpd*. That means there is no more than one site in critical section. The Petri net is also not limited because the number of marks in the place *Token* is not. Indeed, there are the active token and also all the inactive token copies sent to the *Simple-Consumer(s)* in this place. When a site is SCo, it is inactive, so it can consumes the inactive token copies at its own rythme. Thus, the number of copies is only limited by the available space left in this SCo. Nevertheless, this site may become active again. So, in order to keep a consistent view of the production area, it is necessary to concume the copies keeping the same order as the order in which they have been sent to it. For all those reasons, a timestamp must be added to the inactive token copies sent. Since the topology replicated on each site may be different from the others, inconsistencies may have occur. The reference is actually the topology transmitted by the *Chameleon* token. Moreover, asynchrionious signals have to be transmitted to sites, and the choice of the receiving site depends on sites replicated topologies. Our Petri net model validates the behavior of the connecting sites. Many specification points are clarified through this Petri net model. Indeed, we show the reasons why signals have to be treated before the active token transmission on the *PCo* sites in section 4.3. Since all signals modify the topology the token may be sent to the wrong site. In parallel, inactive token copies may also be sent to sites which do not need them. In addition, it is better to send the inactive token copies created during the last token arrival before entering in critical section again in order to minimize the differences between the production area and the production area replicated on SCo sites. As a matter of fact, inactive token copies must be consumed by SCo sites as soon as possible. In order to optimize the speed of the token revolution, it is better to transmit the active token to the next *PCo* site before transmitting the inactive token copies to the tutored sites. This model shows that one site can send no more than one signal to no more than one site and wait for the treatment to be complete, or to wait for a time-out achievement. In addition, an acknoledgment and a time-out mechanisms are required when a signal is sent. Indeed, if the receiver is in fact disconnected, the request may be lost, the receiver of a request must inform the sender if it is able to treat it. Indeed, if the receiver is SCo or if it is disconnected, the sender has to send the request again to another site. The *TeNeCi* platform will be based on our framework *CAliF* [2]. Over this platform, the *Pilgrim* and the *Optimistic Pilgrim* are the protocols that manage the concurrent accesses to the shared objects [12]. The results presented in this section prove that the protocol can be used for such a domain. Regarding this Petri net model, we also improved our communication protocol optimising the order of the tasks. Concretly, the active token is transmitted only to the active specialists connected to the platform, and the others receive inactive token copies. For example, when one specialist is drawing a circle over an X-Ray photography in a white board shared editor, we are certain that all the other specialists (active and inactive) will see this new circle. In addition, we have also proved that each specialist can modify this circle, even if he (or she) has not drawn anything for a long time. Furthermore, the way this circle is transmitted is optimized.

6 Conclusion and Further Works

We have defined a new protocol for concurrency management in collaborative work: the *Chameleon*. For a medical use, we had to prove its robustness. Coloured Petri Nets allow us to correct the *Chameleon* weaknesses we have discovered during this modelling. Indeed, when an inactive site sends a signal to its tutor in order to become active, this signal may not be sent to the right site, as the model has shown us. Thus, a time-out and an automatic forwarding mechanisms have been implemented in the first version of our communication protocol. In addition, the model proved that it is possible to send a lot of inactive token copies to an inactive site, so, a timestamp has been added to the inactive copies sent by the active sites to their tutors. Finally, when a site is seeking for the next active site in the shared topology, it is now not allowed to take into account the signals of its tutors, otherwise one site may be in an inconsistent state (both active and inactive). Thus, a lock mechanism during this step has also been implemented in the *Chameleon*. Due to those results, we have proved that our protocol is robust and can be used in medicine applications.

References

1. Garcia, E., Guyennet, H., Lapayre, J.C.: Group partitioning over corba for cooperative work. Journal of Cluster Computing, to appear (2005)
2. Garcia, E., Guyennet, H., Lapayre, J.C., Moulin, T.: Adapatative tele-application for remote neurology diagnosis. International Journal of Telemedicine and e-Health, to appear (2005)
3. Guan, S.U., Lim, S.S.: Modeling adapatable multimedia and self-modifying protocol execution. Future Generation Computer Systems **720** (2004) 123–143
4. Kristensen, L.M., Petrucci, L.: An approach to distributed state space exploration for coloured petri nets. Proc. 25th Int. Conf. Application and Theory of Petri Nets (ICATPN'2004). Lecture Notes in Computer Science, Springer **3099** (2004) 474–483
5. Park, G.L.: Performance evaluation of a list scheduling algorithm in distributed memory multiprocessor systems. Future Generation Computer Systems **20** (2004) 249–256
6. Schwarz, R., Mattern, F.: Detecting causal relationships in distributed computations: In search of the holy grail. Distributed Computing **7** (1994) 149–174
7. Peterson, J.L.: Petri net theory and the modelling of systems. Prentice Hall PTR, ISBN: 0136619835. (1981)
8. Koriem, S.M., Dabbous, T.E., El-Kilani, W.S.: A new petri net modeling technique for the performance analysis of discrete event dynamic systems. Journal of Systems and Software **72** (2004) 335–348
9. Jensen, K.: Colored petri nets and the invariant method. Theorical Computer Science **14** (1981) 317–336
10. Jensen, K.: An introduction to the practical use of coloured petri nets. W. Reisig and G. Rozenberg (eds.): Lectures on Petri Nets II: Applications. Lecture Notes in Computer Science **1492** (1998) 237–292
11. Garcia, E., Henriet, J., Lapayre, J.C.: Modeling the chameleon using a coloured-petri net : Proof and validation. LIFC Technical Report, num. RT2005-03, http://lifc.univ-fcomte.fr. (2005)
12. Garcia, E., Guyennet, H., Lapayre, J.C.: Study of an optimistic protocol for concurrency management in cscw. LNCS Proceedings of the 9th International Conference on CSCW in Design (CSCWD'05), IEEE, Coventry, UK, to appear (2005)

The Design of a Workflow-Centric, Context-Aware Framework to Support Heterogeneous Computing Environments in Collaboration*

Jinqiao Yu[1], Y.V. Ramana Reddy[2], Sentil Selliah[2], Vijayanand Bharadwaj[2], Sumitra Reddy[2], and Srinivas Kankanahalli[2]

[1] Department of Mathematics and Computer Science, Illinois Wesleyan University, Bloomington, IL 61761 USA
jyu@iwu.edu

[2] SIPLab, Concurrent Engineering Research Center, Lane Department of Computer Science and Electrical Engineering, West Virginia University,Morgantown, WV 26506 USA
{ramana.reddy, vijay.bharadwaj, sumitra.reddy, srinivas.kankanahalli}@mail.wvu.edu, selliah@csee.wvu.edu

Abstract. Networked based distributed collaboration has become pervasive and makes today's enterprises much more productive than ever before. The automation of the collaborative processes can be driven and managed by workflow rules. In addition, during recent years, ubiquitous computing and universal network connectivity have imposed another imperative requirement on collaboration systems. That is, remote participants with heterogeneous and dynamic connection and process capabilities should be able to join seamlessly in the collaboration at any time. Intelligent, context-aware, and workflow-centric collaboration is an essential prerequisite to fulfill that requirement. In this paper, we describe a generic framework, dubbed as EkSarva, which is capable of embedding workflow rules and context-awareness intelligence into collaborative sessions leading to increased opportunities for process automation

1 Introduction

Networked based collaboration in its most fundamental form consists of interaction and information generating and sharing among participants. Current groupware and collaboration tools mainly focus on facilitating information sharing. These tools typically bundle the numerous functionalities found in applications like email, document editors, calendars, and process management and provide a single interface for their usage. This tool-centric approach, we believe, only increases the usage complexity and is not adequate to capture the semantics of the collaboration operation. In addition, with the proliferation of mobile devices and ubiquitous computing, these traditional tools also are not flexible enough to support heterogeneous and dynamic usage patterns.

* This research is partially supported by the Bell Atlantic Endowment at West Virginia University. The views herein are those of the authors and do not necessarily reflect the views of the supporting agency.

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 22 – 29, 2005.

As an alternative, we are developing a generic collaboration framework where the workflow is embedded in the enactment system and connection-awareness is automatically maintained to support collaboration participants' diverse and dynamic requirements. In essence, we believe CSCW should be a design-oriented research area. The design of the above framework captures a deep understanding of the nature of collaboration work and its forms and practices.

The paper is organized as follows. We first describe the key design considerations of EkSarva. Then we review related work. Next, we explain the EkSarva framework and the formalisms it provides to model collaboration. Following this, we use an example to illustrate the usage of this framework. Finally, we conclude our paper with a brief discussion of our prototyping and future research extensions to improve it.

2 Key Design Consideration

2.1 Workflow Centric

In a society of knowledge workers, collaboration, in its most fundamental form, consists of generating information, sharing it with others in the same community. This sharing of information will result in further generation of new information which will trigger actions upon them. The underlying paradigm of this collaboration can be simply stated as "information sharing, when acted upon, results in a state-change of the project which triggers further information exchange-until the project goals are achieved – or we run out of time". This collaboration process is often referred as "Workflow". In the design of this EkSarva collaboration framework, we focus on the workflows driving the collaboration towards a common task.

2.2 Context Awareness

During recent years, ubiquitous computing and pervasive network connectivity have given rise to expectations of building a digital society, where remote participants with heterogeneous and dynamic connection and process capabilities can join seamlessly to accomplish a common task. This imposes a new requirement on today's collaboration framework: context-aware capabilities of adapting to heterogeneous computing requirements. The essence of context-awareness is invisibility; the changes of physical locations, process speed or other technical barriers should be transparent or invisible to collaboration participants. The collaborators should focus on the workflows embodying the collaboration process and moving on to achieve the final target instead of spending a significant amount of time and efforts in reconfiguring tools and adapting to new computing environments.

In the EkSarva framework, we also bring in context-awareness capabilities to fulfill this new requirement imposed by the emergence of a pervasive computing society. This context-awareness feature facilitates the movement of workflows driving collaboration towards its final targets without being interrupted to adapt to the changes of computing environments; the latter is made invisible to collaboration participants.

3 Related Work

Habanero [1] is a collaborative framework and an environment containing several applications. The Habanero environment consists of a client and server. The server is responsible for hosting and managing the sessions. The client interacts with the sessions using various applications. DISCIPLE [2, 3] is a framework for synchronous real-time collaboration. The main characteristics of DISCIPLE framework are a layered architecture, explicit knowledge-based support for software modules, and multimodal human/machine interaction. These frameworks provide the substrate for collaboration. They do not provide any mechanisms to embed workflow rules into the execution components. As a result human interaction is required, consistently, to progress a collaborative session. In [4, 5] Personal Agents (PA) are used to enhance collaboration. The agents are limited to knowledge management tasks in a collaboration environment. They do not participate in the enactment of the collaboration. The Chautauqua [6] consists of a workflow system at its base. The system also facilitates changes to the workflow during the enactment, thus providing a notion of dynamic workflow. Our system differs from the above projects in its ability to model collaboration with flexible and powerful formalisms. In addition, our system places workflow in the center and also provides adaptive services through context-awareness to meet the requirements of dynamic and heterogeneous collaboration computing environments.

4 Conceptualized Collaboration Framework

Modeling a collaboration session involves capturing the fundamental components involved in the collaboration sessions and the interaction among them. A collaboration framework cannot succeed without successfully modeling its targeted collaboration activities. The EkSarva collaboration framework conceptualizes the following entities involved in collaborations: Person(P), Project(P), Place(P), Situations(S), Signal(S) and Transcripts(T).A detailed description of this conceptualized framework can be found at [7].

5 EkSarva System Architecture

The architecture of EkSarva consists of two major components based on the functionalities of those subsystems[Fig 1].

5.1 Specification System

The specification subsystem is responsible for providing mechanisms to specify the framework component: Project, Person, Place, Situation, Signal and Smart Transcript (PPP/SST). This subsystem would be realized through the development of the Collaboration System Specification Shell (CSS) with a domain specific knowledge base. Through an interface, a Collaboration Initiator (CI, the person who has the authority to initiate collaboration) interacts with the shell to specify the above mentioned framework components. Further, the CI would provide additional information required for any customizations. The customizations are intended to create a collaboration environment that differs from the domain specific generic environments. In short, the above specification subsystem provides the initial configuration of the collaboration system.

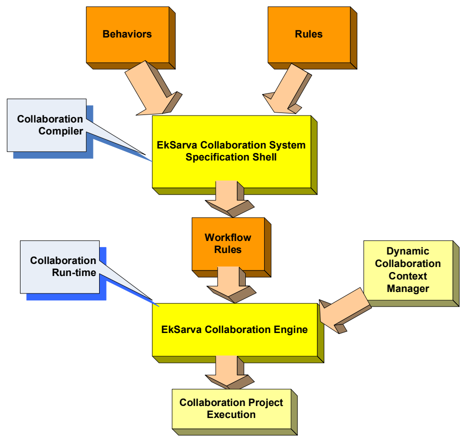


Fig. 1. EkSarva System Architecture

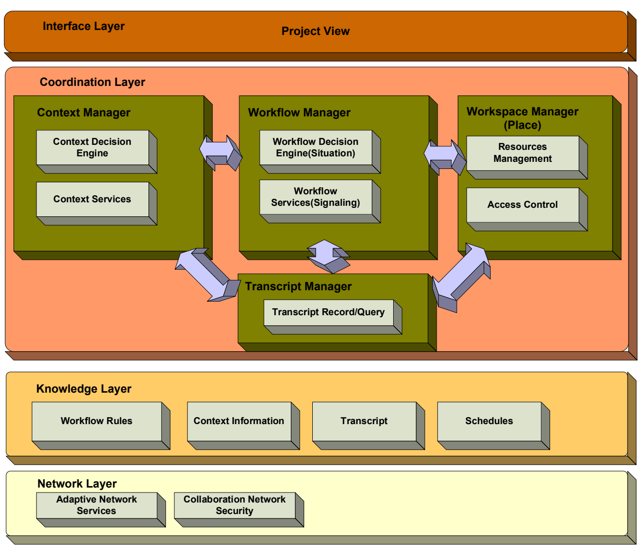


Fig. 2. The layered view of the framework

5.2 EkSarva Collaboration Engine

EkSarva Collaboration Engine is responsible for enacting the collaboration embedded in the workflow rules by providing coordination, synchronization and communication.

The collaboration engine processes the rules specified by the CSS and executes the workflow. Thus the two-stage process is analogous to a compiler of specification and a run-time engine for instructions generated. Mechanisms that are required to implement the above mentioned architectural components can be facilitated through a layered component model (Fig 2). Each layer corresponds to a software module providing the necessary services

5.3 Layered Structure of EkSarva Framework

EkSarva Framework can be structured in four layers: interface, coordination, knowledge and network. The interface layer, in addition to, providing a representational view of the underlying collaboration "Project", also enables interaction with the project; The coordination layer is the command center of a collaboration project; the knowledge layer contains the library of concepts that enables EkSarva to bring context awareness and intelligence to the collaboration process. The network layer provides network supporting services to the above layers. A more detailed description of the above layered structure can be found at [7]. But here we illustrate more detailed and revised functionalities provided by the coordination layer.

The coordination layer has the following main components (a) Workflow Manager, (b) Context Manger, (c) Workspace Manager, and (d) Transcript Manager.

a) Workflow manager is the controlling entity of the collaboration process. It is workflow manager that drives the process forward to different stages until the final goal is reached. The core of workflow manager is a workflow decision engine that makes decisions to control the flow of the collaboration process based on different collaboration situations.

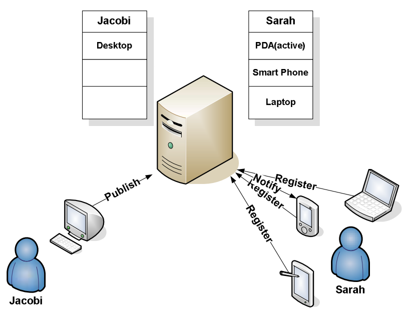


Fig. 3. Context Registration and Adaptive Notification

b) Context Manager is responsible for maintaining context-awareness through the entire collaboration process. The core of context manager is a context decision engine, which enables the framework to adapt to the changes of different collaboration computing environments, and these changes or heterogeneous computing environments are made invisible to collaboration participants by the services provided by it. Consequently, collaboration participants then can concentrate on the collaboration process instead of technical details.

After the EkSarva collaboration engine starts running and the collaboration process is initialized, collaboration participants can join the collaboration session through a registration process (Fig 3). During this registration process, the client tells the collaboration server what are his/her interests. Meanwhile, the client also registers his/her computing environment, including computation power, display size, connection type and speed, etc. The client can also register more than one device in the server. For example, Sarah (a client) can register a PDA, a smart phone and a laptop. All of the above information will be stored in a client profile associated with the client's logical ID in the collaboration process. The context manager manages a profile for each client and maintains the context-awareness through these profiles. During the collaboration process, suppose another client, Jacobi, starts a video conferencing. He can simply send the video to the server. Then the context-decision engine will send out this video stream to all subscribers (Sarach) and send it in a format which is suitable for each individual client based on his/her current computing environment. Sarah's heterogeneous and dynamic computing environment is then made invisible to Jacobi.

c) Workspace manager provides the necessary functionalities to store, retrieve and manage object that are constituents of a project. Policing of access control mechanisms of objects is also the responsibility of this manager.

d) Transcript manager is responsible for generating and maintaining "Smart Transcripts". Transcripts are used to "record" information about the activity of the collaboration, such as states, action items, important milestones, etc.

6 An Example Usage Scenario: Cyber Conference Room

To visualize how the EkSarva framework could be harnessed to facilitate collaboration, we illustrate a cyber conference room where geographically distributed conference attendees can participate. Within a company, several departments are holding conferences simultaneously. By signing up for departments, employees can become members of those respective conference collaborations (Fig 4). An employee views his or her workspace within a cyber conference room via the Project view. This view provides all relevant conference information, such as reports, data, deadlines, and so on. Submitting a report may be as simple as dragging and dropping the file from his/her personal view into the project view. The EkSarva personal space then submits the file to the cyber conference group space. The coordination layer of the projects makes this operation transparent to the user. This layer is responsible for interpreting the rules present in the Knowledge Layer, which specify how each action should be handled based on the corresponding workflow rules and collaborative computing environment context.

In order to support heterogeneous and dynamic collaboration environment, context manager maintains each client's context profiles collected at the registration process. Based on these client profiles, collaboration information and data such as multimedia presentation will be filtered or tuned to fit each client's capability. For instance, in the above example, John is on the way to his office and he only has a personal PDA connected to the Internet. Therefore, based on his computing environment context, only a small size multimedia presentation version will be sent to his PDA. After John arrives at his office, he switches from the PDA to his laptop. This change of computing environment is detected by the collaboration server. The follow-up media stream will be changed to higher quality to reflect this change. But all of these computing environment changes are made invisible to other meeting attendees by the services provided by the context manager in the server.

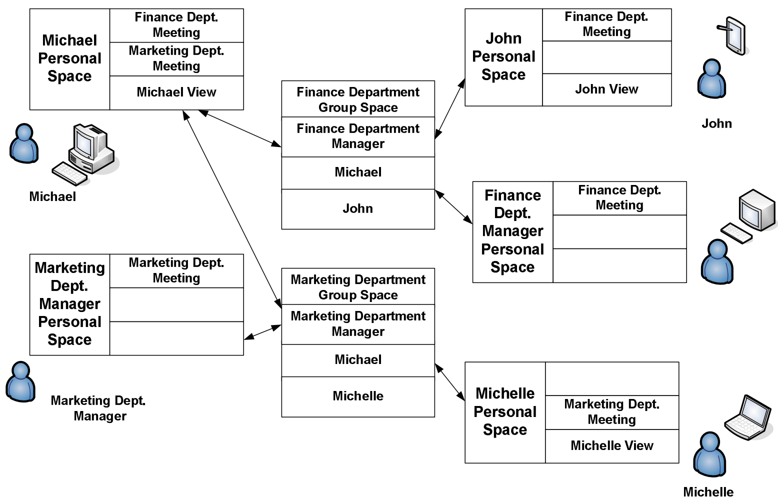


Fig. 4. Cyber Conference Room

7 Conclusion and Future Work

In this paper, we have presented a generic collaboration framework. Our main focus is the development of a flexible collaboration framework that is driven by intelligent workflow. In order to meet the heterogeneous and dynamic collaboration environment requirements, we incorporate context-awareness into the framework design. This also facilitates CSCW based collaboration in meeting the trends of ubiquitous computing. We are working on prototyping the above cyber conference room project to demonstrate the effectiveness of its modeling of collaboration processes and satisfaction of ubiquitous computing requirements. We are also exploring different mechanisms, including the client context profile maintenance and management, to better support the diverse client capability requirements.

References

1. A. Chabert, E. Grossman, L. Jackson, S. Pietrowicz and C. Seguin, "Java object-sharing in Habanero", Communications of the ACM, 1998, Vol. 41, No. 6, pp. 69-76.
2. A. Krebs, M. Ionescu, B. Dorohonceanu, I. Marsic, "The DISCIPLE System for Collaboration over the Heterogeneous Web", Proceedings of the Hawai'i International Conference on System Sciences, January 6-9, 2003.
3. W. Li, W. Wang, I. Marsic, "Collaboration transparency in the DISCIPLE framework", Proceedings of the international ACM SIGGROUP conference on Supportinggroup work, 1999, pp. 326-335.
4. F. Bergenti, M. Garijo, A. Poggi, M. Somacher, J. Velasco, "Enhancing Collaborative Work through Agents", VIII Convegno dell'Associazione Italiana perl'Intelligenza Artificiale, 2002.
5. F. Enembreck, J. Barthes, "Improving Computer Supported Cooperative Design With Personal Assistant Agents", Trans. Of the Society for Design and Process Science, 2003, Vol. 7, No. 2, pp. 3-19.
6. C. Ellis, C. Maltzahn, "Chautauqua: Merging Workflow and Groupware", In Proceedings of the HICSS'96 Conference, 1996.
7. Reddy, R., Selliah, S., Bharadwaj, V., Yu, J., Reddy, S. and Kankanahalli, "Eksarva: An Intelligent Collaboration Framework", in Proceedings of the IEEE International Conference 'Intelligent Systems', Varna, Bulgaria, June 22-24, 2004.

Do Tangible User Interfaces Impact Spatial Cognition in Collaborative Design?

Mary Lou Maher and Mi Jeong Kim

Key Centre of Design Computing and Cognition
University of Sydney, Australia 2006
mary@arch.usyd.edu.au, mkim9133@arch.usyd.edu.au

Absract. Developments in digital design workbenches that combine Augmented Reality (AR) systems and tangible user interfaces (TUIs) on a horizontal display surface provide a new kind of physical and digital environment for collaborative design. The combination of tangible interaction with AR display techniques change the dynamics of the collaboration and have an impact on the designers' perception of 3D models. We are studying the effects of TUIs on designers' spatial cognition and design communication in order to identify how such tangible systems can be used to provide better support for collaborative design. Specifically, we compared tangible user interfaces (TUIs) with graphical user interfaces (GUIs) in a collaborative design task with a focus on characterising the impact these user interfaces have on spatial cognition.

1 Introduction

Digital design workbenches, table-top Augmented Reality (AR) systems equipped with tangible user interfaces (TUIs), have been proposed as an alternative for the design review meetings since they allow designers to intuitively modify the spatial qualities of 3D designs and keep communication channels open by preserving traditional mechanisms such as verbal and non-verbal communication. This research focuses on the way the digital design workbench supports design thinking and design communication. Several researchers have proposed that such tangible interaction combined with AR display techniques might affect designers' cognition and communication (Tang, 1991; Bekker et al., 1995). However, they have not posed any evidence in a systematic way. To date, the central preoccupation of research on TUIs has been in a developmental direction, which usually describes the fundamental ideas behind their tangible systems, their prototype implementation, possible application areas and some initial usability results. We aim to obtain empirical evidence about the potential impact of TUIs by investigating if and how the tangible presence of the virtual objects on the design workbench affects designers' spatial cognition and design communication in collaborative design. This paper presents some preliminary results of a pilot study using protocol analysis, which considers what factors are associated with designers' spatial cognition in 3D modelling.

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 30 – 41, 2005.

2 Tangible Presence and Spatial Cognition

Since a human's cognitive ability is strongly bounded, to carry out complex reasoning without the aid of tools is very difficult. Norman (1991) defined tools as cognitive artefacts, which do not modify the computational power of the human's cognitive ability but instead modify the content of the knowledge involved in the elaboration process (Rizzo et al., 1995). That is, tools of thought – cognitive artefacts – are external aids that enhance cognitive abilities. Thus, we predict that designers' perception of spatial knowledge will be improved when using TUIs, and this may be due to the tangible presence of virtual objects.

2.1 Tangible User Interfaces and Physical Interaction

AR technology blends reality and virtuality to allow the seamless interaction between physical and digital worlds. Thus AR research has focused on the linkage between digital information and physical objects by superimposing, tagging and tracking objects. The term "augmented reality" is often used to refer to such TUIs in which 2D and 3D computer graphics are superimposed on physical objects. TUIs turn the physical objects into input and output devices for computer interfaces to restore the richness of the physical world in human-computer interaction. Arias et al. (1997) argued that new HCI approaches need to combine physical and digital environments to augment the weaknesses of the other by using the strengths of each since the "reflective conversations" are very different between the two environments. Digital design workbenches that combine AR technologies and TUIs provide a physical interaction by putting tangible 3D objects containing digital information into a single environment.

According to Wang et al. (2001), the strengths of physical interaction can be explained by two aspects. Firstly, physical interaction provides direct, naïve manipulability and intuitive understanding. It is very natural to pick up and place a physical object; certain characteristics such as size, weight, colour and shape can be used to communicate meaning. Secondly, physical interaction provides tactile interaction; the sense of touch provides an additional dimension of interaction. Seichter and Kvan (2004) posed the concept of "augmented affordance" to explain interactions using TUIs in AR systems. Norman (1988) regarded the study of affordances of objects, originated by Gibson (1979), as the start of a psychology of things since affordances result from the mental interpretation of things, based on our past knowledge and experience applied to our perception of the things. He proposed that "affordance" refers to the perceived and actual properties of the thing that determine just how the thing could possibly be used. According to Seichter and Kvan (2004), TUIs can be seen as offering a conduit between the real or perceived affordances implied by the physical properties of the interface tool and the affordances created by the digital behaviours in the virtualised interface.

2.2 Designers' Spatial Cognition and Spatial Representation

As a consequence of the diversity of approaches and related disciplines, there is little consistency in what is meant by the term "spatial" (Foreman & Gillett, 1997). We associate the designers' perception of the form and spatial relationships of the design components with the designers' spatial cognition. The meaning of 'space' to the designers is not an abstract of empty space, but rather of the identity and the relative locations of the objects in space. Space then is decomposed into particular objects and the spatial relationships among them viewed from a particular perspective. The spatial relationships may include functional reasoning since design is required to satisfy intended functions. In addition, it has been argued that touch is also a spatial modality, where the close linkage between motor and spatial processes has been emphasized. Kinaesthetic information through a haptic system provides us with the ability to construct a spatial map of objects that we touch (Loomis & Lederman, 1986). It is the movement of a hand repeatedly colliding with objects that comes to define extra-personal space for each individual, as a consequence of repeatedly experienced associations (Foreman & Gillett, 1997). Thus, the movement simulated by the mouse in desk-top systems lacks tactile and kinaesthetic feedback that normally accompanies movement.

Based on the assumption that people often use general purpose verbs and prepositions when the context is sufficiently clear to disambiguate them, we will investigate language spatial representation. Language draws on spatial cognition so that we can talk about what we perceive and it thereby provides a window on the nature of spatial cognition (Anibaldi & Nualláin, 1998). Jackendoff (1983) suggested a level of mental representation devoted to encoding the geometric properties of objects and the relationships among them in space, in order to express our spatial experience in talking about objects and talking about spatial relationships. Gesture is also recognized as a good vehicle for capturing visual and spatial information as it is associated with visuospatial content (Wagner et al., 2004). People produce some gestures along with their speech, and such speech-accompanying gestures are not just hand moving (Lavergne & Kimura, 1987). Speech and gesture are both characterizing the spatial relationships among entities, which are closely related to and may even be beneficial for cognitive processing (Goldin-Meadow, 2003). Wagner et al. (2004) found that the movement of hands facilitates recall of visuospatial items as well as verbal items.

2.3 Digital Design Workbenches

We reviewed various digital design workbenches: metaDESK, iNavigator, BUILD-IT, PSyBench, URP, MIXdesign and ARTHUR system. Ulmer and Ishii (1997) constructed the metaDESK system with a focus on physical interaction to manipulate the digital environment. Standard 2D GUI elements like windows, icons, and menus, are given a physical instantiation as wooden frames, 'phicons, and trays, respectively. iNavigator is a CAD platform for designers to navigate and construct 3D models, which consists of a vertical tablet device for displaying a dynamic building section view and a horizontal table surface for displaying the corresponding building plan geometry. From the user's perspective, the display tablet is served as "a cutting plane" (Lee et al., 2003). BUILD-IT developed by Fjeld et al. (1998) is a cooperative planning tool consisting of a table, bricks and a screen, which allows a group of designers, co-located around the table, to interact, by means of physical bricks, with models in a virtual 3D setting. A plan view of the scene is projected onto the table and a perspective view of the scene is projected on the wall. Brave et al. (1998) designed PSyBench and inTouch, employing telemanipulation technology to create the illusion of shared physical objects that distant users are interacting with. Although still in the early stage, it shows the potential of distributed tangible interfaces. URP developed by MIT media lab is a luminous tangible workbench for urban planning that integrates functions addressing a broad range of the field's concerns such as cast shadows, reflections and windflow into a single, physically based workbench setting. The URP system uses pre-existing building models as input to an urban planning system (Underkoffler & Ishii, 1999). MIXDesign allows architects to interact with a real scale model of the design by using a paddle in a normal working setting, and also presents an enhanced version of the scale model with 3D virtual objects registered to the real ones (Dias et al., 2002). ARTHUR system is an Augmented Round Table for architecture and urban planning, where virtual 3D objects are projected into the common working environment by semi-transparent stereoscopic head mounted display (HMDs). Placeholder objects (PHOs) and wand are used to control virtual objects (Granum et al., 2003).

These various configurations of digital workbenches, with and without augmented reality, show a trend in developing technology that supports designers in creating and interacting with digital models that go beyond the traditional human-computer interface of the keyboard, mouse, and vertical screen. The different configurations described above draw on specific intended uses to define the components and their configuration. Few of the publications about digital workbenches evaluate the new interface technology with respect to spatial cognition or improved understanding of the spatial relationships of the components of the digital model. In this paper, we consider the existing digital workbenches as defining a class of design environments that use TUIs to be a departure from the traditional GUIs that designers are currently using to create and interact with a digital design model. While TUIs and GUIs will continue to be alternative design environments for digital models, we focus on the differences between them in order to clarify the role and benefit of TUIs for designers.

3 Comparing GUI-Based with TUI-Based Collaboration

In devising an experiment that can highlight the expected improvement in spatial cognition while using TUIs, various scenarios have been considered: Face-to-face collaboration with physical models versus TUI-based collaboration with tangible digital models; Face-to-face collaboration with pen and paper versus GUI-based collaboration with mouse and keyboard; GUI-based collaboration with intangible digital models versus TUI-based collaboration with tangible digital models. We chose the third category for this research because it will enable us to verify if and in what way tangible interaction affect designers' spatial understanding of 3D models in computer-mediated collaborative design.

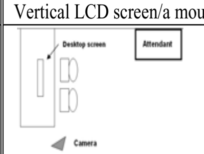
3.1 Design Collaboration in GUI vs. TUI Environments

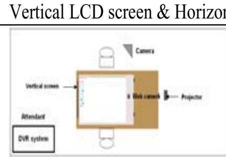
A GUI is a graphical user interface to a computer, which features a pointing device, icons and menus, and desktop on the display screen. Thus, we consider a 3D CAD system for the design environment in a GUI environment. Current CAD systems allow designers to explore design alternatives as modifiable models, but they produce indirect interaction with 3D models using a mouse. Despite the physical form, the mouse has no physical contextual awareness and lacks the efficiency of specialized tools. It is regarded as a time-multiplexed input device controlling different functions at different times (Fitzmaurice, 1996). The ability to use a single device for several tasks is a major benefit of the GUI, but given the nature of interaction where only one person can edit the model at a time, the GUI environment may change interactivity in collaborative design (Magerkurth, 2002). On the other hand, TUIs have been offering an alternative to standard computer user interfaces, where physical objects are representations and controls for digital information, thereby restoring some of the tangibility of physical models. We consider a digital design workbench for the design configuration in a TUI environment. The digital design workbench is specifically configured for 3D design and visualization (Daruwala, 2004), where designers can manipulate 3D virtual objects in a semi-immersive environment in real time and can be spatially aware of each other as well as the design. The direct hands-on style of interaction afforded by the physical objects provides designers concurrent access to multiple, specialized input devices. That is, they are "space-multiplexed" input devices, which can be attached to different functions, each independently accessible (Fitzmaurice, 1996). Using AR techniques, TUIs merge the display and task space, which may support shared understanding due to the interaction that draws out ideas in a conversational manner (Arias et al., 2000).

3.2 Experiment Set-Up

We chose ArchiCAD as an application for the GUI-based collaboration and a design workbench constructed by Daruwala and Maher (2004) for the TUI-based collaboration. Inspired by the fact that tabletops and walls are the main working surfaces during collaborative design sessions (Ma et al., 2003), they proposed a customized digital workbench that provides both horizontal and vertical display surfaces to facilitate multiple views of the 3D model. In addition they included a collection of 3D blocks with tracking markers in ARToolKit to offer a form of tactile influence on the design as handles to the virtual objects. The digital workbench environment is designed to demonstrate a range of novel input techniques for 3D design, so it is not engineered as a targeted end-user application.

Table 1. Two experiment settings

<table>
<tr><th></th><th>GUI-based collaboration</th><th>TUI-based Collaboration</th></tr>
<tr><td>Hardware</td><td>Desktop computer - Mouse & Keyboard</td><td>Digital design workbench - 3D blocks</td></tr>
<tr><td>Application</td><td>ArchiCAD</td><td>ARToolkit</td></tr>
<tr><td>Display/Task space</td><td>Vertical LCD screen/a mouse & keyboard</td><td>Vertical LCD screen & Horizontal table/Horizontal table</td></tr>
<tr><td>Settings</td><td>

</td><td>


</td></tr>
</table>

To simulate a situation similar to a design review meeting where designers review the current state of the 3D plan and modify it, we developed two renovation scenarios of the same degree of difficulty: the layout of a home office apartment and the layout of an interior design office. Designers are required to collaborate on renovating the studio to satisfy intended functions to a pre-defined set of specifications in design briefs. The designers did not have access to a pen device or to the 2D view in ArchiCAD to make the conditions similar to the TUI-based collaboration. A set of 3D objects were made available in the application's library in order for the designers to focus on design configuration rather than designing the furniture components during the design sessions. The working time in each design environment was limited to 20 minutes.

Table 2. Experiment design

	Experiment 1		**Experiment 2**		**Experiment 3**	
Sessions/Task	**1st GUI(TA)**	**2nd TUI (TB)**	**1st TUI (TA)**	**2nd GUI (TB)**	**1st TUI (TB)**	**2nd GUI (TA)**
Activity	Extending and renovating the studio by placing the furniture					
Subjects	A pair of 2nd year architecture students					
2D tool/ View	No pen was allowed to use / 3D view in ArchiCAD was required					
library	30 pieces of furniture were provided in order for designers to focus on design rather than furniture selection					

*Task A (TA): Home office apartment, Task B (TB): Interior design office

4 Segmentation and Cognitive Actions

We have done a pilot study that is an adaptation of protocol analysis, a methodology for studying problem solving processes that has been widely used in design cognition. Rather than ask the designers to think aloud, we recorded their conversation and gestures while they were collaborating on a predefined design task. The data collected for analysis includes verbal description of spatial knowledge and non-verbal data such as gestures. We are not trying to capture evidence of all of the designers' cognitive actions during the design process, but instead focus on the contents of what designers do, attend to, and say while designing, looking for their perception of discovering new spatial information and actions that create new functions in the design.

4.1 Segmentation

Segmentation is dividing the protocol into small units according to a rule. One way of segmentation is to divide protocols based on verbalization events such as pauses or syntactic markers for complete phrases and sentences (Ericsson and Simon, 1993). Another way is looking at the content of the protocol, and divide the protocols into small units along lines of designer's intentions (Suwa et al., 1998). We take the latter approach, thus a change in subjects' intention or the contents of their actions flagged the start of a new segment. Focusing on a specific part of a design, designers emphasized a different aspect of the design of that specific part, which showed what they actually were doing or attending to. Consequently, a single segment sometimes comprises one sentence, and sometimes several sentences.

Table 3. Segments of the protocols from one of the TUI sessions

Seg. N	Time	Transcript	Intentions
9	1:40	S: Cos I've seen in how in some studio apartments how they've got these bookshelves that partition the room ….and then just chuck some crap in there	S is retrieving previous knowledge on bookshelves that partition the room.
10	1:49	A: Okay. Um. Should we go about rearranging it? S: Alright, so…we're gonna move these, like, that way A: So, first, these are all living stuff but you're uh…	They are rearranging 3D blocks by moving all living stuff to other side.
11	2:04	S: How about if we partition in there and we… A: …Put some bookshelves like this. Uhh.. yeah… some bookshelves	A is putting bookshelves on the horizontal surface attending to partition

4.2 Cognitive Actions

For each segment, we classified designers' cognitive actions into four categories including visual and non-visual information based on Suwa's definition (1998): 3D modelling actions, perceptual actions, functional actions and design communication. The three categories correspond to the levels at which incoming information is thought to be processed in human cognition. 3D modelling actions correspond to sensory level, perceptual actions to perceptual, and functional actions to semantic. Another action category, the design communication, is included to reveal specific aspects of the collaborative interactions. We paid attention to the information of whether or not actions are new for each design action and the dependency on each other in order to measure the correlation between them and designers' spatial cognition. The first category, 3D modelling actions, refers to physical actions that are functionally processed with design thinking. This category includes the selection, placement and relocation of pieces of 3D blocks made by designers. The second category, perceptual actions, refers to actions of attending to visuo-spatial features of the artefact or spaces they are designing with. Perceptual actions are inherently dependent on physical actions. The attention to the feature or relations, creation of new space or relations, and unexpected discovery of objects or spaces that occurred were investigated as a measure of designers' perceptive abilities for spatial knowledge. The third category, functional actions, refers to actions of conceiving of non-visual information, but something with which the designers associate visual information. Whenever instances of functional actions were found it was necessary to interpret what perceptual actions or 3D modelling actions were dependent on them. In terms of functional actions, we focused on discovering 'Re-interpretation', which was applied to a segment when a designer defined a different function from previous one when s/he revisited. The fourth category, design communications, refers to actions of communicating with one another regarding design ideas and the completion of the design task itself.

5 Analysis

The following analysis is a preliminary interpretation of the data collected. We focussed on finding patterns of designers' behaviours and cognitive actions by interpreting the information shifts, specifically looking for significant differences in the data collected from the GUI sessions and the data collected in the TUI sessions.

5.1 Observation of Designers' Behaviours

Through the observation, we noticed that designers in the GUI sessions discussed ideas verbally and decided on a solution before performing 3D modelling actions whereas designers in the TUI sessions communicated design ideas by gesturing at and moving the objects visually and decided on the location of each piece of furniture as they were manipulating 3D blocks. That is, designers in the GUI sessions spent a lot of time explaining their ideas to one another verbally while designers in the TUI sessions used 3D blocks to test and visualize design ideas. In terms of collaborative interactions, the TUI environment enabled designers to collaborate on handling the 3D blocks more interactively by allowing concurrent access to the 3D blocks and to produce more revisited 3D modelling actions before producing the final outputs. These results may be caused by the different properties of the tools. Designers in the GUI environment shared a single mouse compared to multiple 3D blocks, thus one designer mainly manipulated the mouse and the other explained to his partner what they were focusing on. On the other hand, with the direct, naïve manipulability of physical objects and rapid visualization, designers in the TUI environment seemed to produce more multiple cognitive actions and completed the design tasks faster.

Fig. 1. GUI-based collaboration

Fig. 2. TUI-based collaboration

The above results motivated us to be interested in the revisited 3D modelling actions in terms of idea developments and designers' spatial cognition. We speculate that the revisited 3D modelling actions in the TUI sessions uncover information that is hidden or hard to compute mentally and suggest that this will play an important role in supporting designers' spatial cognition and producing a design solution.

5.2 Cognitive Actions in the GUI and TUI Sessions

Looking more in depth into the content of cognitive actions, we found different patterns of behaviour between the GUI and TUI sessions. In terms of perceiving the location of an object or space, designers in the GUI environment focused on the individual location itself while designers in the TUI environment attended more to spatial relations among objects or spaces. For example, when designers worked on the layout of a sink in the home office apartment, designers in the GUI session just clarified the location of the sink without noticing the problem in relation to the bedroom whereas designers in the TUI sessions found that the current location of the sink is wrong by perceiving the spatial relation with the bedroom.

Table 4. Perceptual actions on the location of a sink

Session	Transcript (GUI)	Session	Transcript (TUI)
GUI 2	B: Kitchen and dining area A: Yep B: Which she does not yet have… well **she has a sink [laugh] in her ba-bedroom,** and then living/meeting area A: Yep… and a working area	TUI 1	A: It shouldn't be near the bathroom or I mean, I think it shouldn't be **near the bedroom**, sorry. It shouldn't have **a kitchen sink**. S: Yeah that's what I was thinking. Why is **it next to the bed**? A: It's a bit odd, and it's also just….not…normal
		TUI 3	B: coz.. The sink sink, **sink dosen't need to be in the bedroom** A: needs to be in the kitchen B: yeah sink in the kitchen. sink over here for now A: mm hmm

When placing an object in the GUI environment, designers put it in a relevant area simply considering the function of the object, and sometimes they attended to the shape or style of the object itself. On the other hand, designers in the TUI environment created and attended to a new spatial relation created by placing an object, with respect to other objects or space around it. For example, regarding the placement of a new desk, designers in the GUI session emphasized the function of a desk for a computer programmer and placed it in the corner. However, designers in a TUI session considered two locations for the desk, in the corner or near the window, and then they decided to put it near the window because the designer can look out the window by creating a spatial relation between the desk and window.

Table 5. Cognitive actions on the placement of a desk

Session	Transcript (GUI)	Session	Transcript (TUI)
GUI 2	B: we need a work table A: desk and stuff B: she's **a prgrammer right? so she'd need** a, ummm A: that one's got a little computer thing on it B: hummm (as in yeah) A: that one? B: yeah A: and that can go in the **corner**….	TUI 3	B: we need a desk, first of all, for his um office area.. maybe one of this..is it like a desk? A: ok B: maybe in the **corner** there B: now we want the desk to go **near the windows** A: ok B: so **he can look out the window**

In comparison to the GUI system, the TUI system allowed designers to discover a space and features of an existing object unexpectedly when it is revisited, and which were never paid attention to. These perceptual actions were classified as unexpected discoveries, which are usually associated with ‘Re-interpretation’ actions. Re-interpretation and unexpected discoveries have been regarded as the driving force for exploration of new design ideas, thus we can hypothesize that the TUI system facilitates generating ideas by improving designers’ spatial cognition. However, it is not yet clear that the idea fluency is directly connected to the quality of the final output. The following are examples of unexpected discoveries and re-interpretation extracted from the verbal protocols of the TUI sessions.

Table 6. Unexpected discoveries and Re-interpretation in the TUI sessions

Transcript	Interpretation
B: **Just little bit layout is not very**…. A: You don't like how they have drawers in the middle? B: No, I mean A: You end up with **empty space** in the middle. how this sofa faces onto her	B felt that the layout is little bit strange and then A is discovering **an empty space** in the middle. **(Unexpected discovery)**
B: and ahh… the dining table can go… can go her… near the kitchen A: I'm thinking of having the dining.. The kitchen and dining area closer to.. B: Closer to the bed? A: Yeah. pit it closer to this section here B: That ahh **That might get crowded over there**. I could…move it around here?	B is discovering that **the area is getting crowded** (feature) by placing the dining table and sink near the bed **(Unexpected discovery)**
S: (Interrupt) ya know how they have those kitchens that are just two long rows? A: Oh yeh S: And then that would be like, become like the bar. **The breakfast bar**.	After discovered **the shape of kitchen** to be two long rows (Unexpected discovery), A is **defining a new function** of the cabinet in the kitchen. **(Re-interpretation)**

6 Conclusion

The results of the pilot study indicated that the GUI and TUI sessions produce different outcomes in terms of designers' cognitive actions. Compared to the GUI sessions, designers in the TUI sessions exhibited the following patterns of behaviour:

- performed multiple cognitive actions in a shorter time
- re-visited a previous design frequently while coordinating design ideas
- created and attended to spatial relations such as local and global relations
- discovered a space or feature of an existing object unexpectedly
- produced more re-interpretation actions

From the above results, we can assume that the TUI system is an effective design environment for the spatial layout design since it encourages designers to attend to or create spatial relations between artefacts or spaces. However, more protocols have to be analysed to reinforce these findings.

In the next phases, we will consider different experimental situations including a single designer using the think aloud method, we will develop a coding scheme derived from our observations in the pilot study and theories of spatial cognition, and will analyse the results using both qualitative and quantitative methods.

While there may be significant differences in spatial cognition in using TUIs, we do not think that digital workbenches will replace GUI systems. Rather, new developments in TUIs will provide alternative design environments that complement existing GUI systems. Knowledge of the implications of the differences in spatial cognition provide a basis for developing and implementing new design environments as well as provide guidelines for their most effective use.

References

Anibaldi, L. and Nualláin, S. (1998) A Computational multi-layered model for the interpretation of locative expressions. In Nualláin, S. (ed.) Spatial Cognition. Jonh Benjamins B.V. pp249-266.

Arias, E. G., Eden, H. and Fischer, G. (1997) Enhancing communication, facilitating shared understanding, and creating better artefacts by integrating physical and computational media for design. In Proceedings of the Conference on Designing Interactive Systems, ACM Press, New York, NY, 1-12.

Arias, E., Eden, H., Fischer, G., Gorman, A. and Scharff, E. (2000) Transcending the individual human mind – create shared understanding through collaborative design. ACM Transactions on Computer Human Interaction, 7 (1), pp 84-113.

ARToolKi. http://www.hitl.washington.edu/artoolkit/tutorials.htm

Bekker, M.M., Olson, J.S. and Olson, G.M. (1995) Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design. In Proceedings of the Symposium on Designing Interactive Systems (DIS)'95. pp 157-166.

Brave, S., Ishii, H. and Dahley, A. (1998) Tangible interface for remote collaboration and communication. In Proceedings of CHI'99. pp 394-401.

Daruwala, Y. (2004) 3DT: Tangible input techniques used for 3D design & visualization. Honors thesis, the University of Sydney.

Dias, J.M.S., Santos, P. and Diniz, N. (2002) Tangible interaction for conceptual architectural design. In Workshop of the First IEEE International Augmented Reality Toolkit 2002.

Ericsson, K. A. and Simon, H. A. (1993) A protocol analysis: verbal reports as data (revised edn). MIT Press, Cambridge MA

Fitzmaurice, G. (1996) Graspable User Interfaces. PhD thesis, the University of Toronto.

Fjeld, M., Bichsel, M. and Rauterberg, M. (1998) BUILD-IT: an intuitive design tool based on direct object manipulation. In Gesture and Sign Language in Human-Computer Interaction, Lecture Notes in Artificial Intelligence, Vol. 1371, Wachsmut and Fröhlich, Editors, Springer-Verlag, Berlin. pp 297-308.

Foreman, N. and Gillett, R. (Eds.) (1997) Handbook of spatial research paradigms and methodologies Vol. a. Hove, UK: Psychology Press.

Gibson (1979) The ecological approach to visual perception. Erlbaum Associates, New York.

Goldin-Meadow, S. (2003) Hearing gesture: How our hands help us think. Cambridge, MA: Harvard University Press.

Granum, E., Moeslund, T.B. and Störring, M. (2003) Facilitating the presence of users and 3D models by the augmented round table. In PRESENCE Conference, Aalborg, Denmark, October 2003.

Jackendoff, R. (1983) Semantics and Cognition. Cambridge, England: Cambridge University Press.

Lavergne, J. and Kimura, D. (1987) Hand movement asymmetry during speech: No effect of speaking topic. Neuropsychologia 25. pp. 689-693.

Lee, C.H., Ma, Y.P. and Jeng, T. (2003) A spatially-aware tangible user interface for computer-aided design. In Proceedings of the Conference on Human Factors in Computing Systems (CHI'03). pp 960-961.

Loomis, J.M. and Lederman, S.J. (1986) Tactual perception. In K.Boff, L. Kaufman and J. Tomas (eds.), Handbook Perception and Human Performance. New York: Wiley.

Magerkurth C. and Peter, T. (2002) Augmenting tabletop design for computer-supported cooperative work. In Workshop on Co-located Tabletop Collaboration: Technologies and Directions at CSCW'02, 2002.

Norman, D.A. (1988). The design of everyday things. New York: Basic Book.

Norman, D.A. (1991). Cognitive artefact. In J.M.Carroll (Ed.), Designing interaction. Cambridge, MA: Cambridge University Press.

Rizzo, A. and Marti, P. (1995) Distributed cognition and artefacts.
Seichter, H.and Kvan, T. (2004) Tangible interfaces in design computing, Virtual Environment 2.
Suwa, M., Purcell, T. and Gero, J. (1998) Macroscopic analysis of design processes based on a scheme for coding designer’s cognitive actions. Design Studies, 19(4). pp 455-483.
Tang, J.C. (1991) Findings from observational studies of collaborative work. In International Journal of Man-Machine Studies, 34. pp 143-160.
Ullmer, B. and Ishii, H. (1997) The metaDESK: Models and prototypes for tangible user interfaces. In Proceedings of User Interface Software and Technology (UIST)’ 97, 14-17 October, 1997. pp 14-21.
Underkoffler, J. and Ishii, H. (1999) Urp: a luminous-tangible workbench for urban planning and design. In Proceedings of the Conference on Human Factors in Computing Systems (CHI’99). pp 386-393.
Wang, Q., Li, C., Huang, X. and Tang, M. (2001) Tangible interface: integration of the real and virtual.
Wagner, S.M., Nusbaum, H. and Goldin-Meadow, S. (2004) Probing the mental representation of gesture: Is handwriting spatial? In the Journal of Memory and Language 50, pp 395-407.

Cooperation in Highly Distributed Design Processes: Observation of Design Sessions Dynamics

François Laborie[1], David Jacquemond[2], and Matthieu Echalier[1]

[1] EADS Corporate Research Center,
Centreda 1, Av. Didier Daurat, 31700 Blagnac, France
{francois.laborie, matthieu.echalier}@eads.net
http://www.eads.net

[2] Airbus France, 316 Route de Bayonne, 31060 Toulouse, France
david.jacquemond@airbus.com
http://www.airbus.com

Abstract. In the recent years, industrial design programs have shifted from a co-located, sequential process to a globally distributed, concurrent process. Therefore, communication and coordination within and among teams have become major challenges. Surprisingly, several studies demonstrate that collaborative tools exist but are very seldom deployed nor used. Indeed, co-located working sessions remain a critical cooperation milestone for otherwise remote teams. In order to better understand the dynamics of those meetings, and identify what makes them so critical for the design process, we analysed working sessions of aeronautical design teams. This article reports the results of our observations based on three criteria: information types, information flow and interactions performed.

Our study shows the critical importance of private communication channels, and investigates how they relate to the working session's process. It highlights the need of parallel visualization channels for cross-references, and analyses the use of interactions depending on ownership and context.

1 Introduction

Industrial design programmes for complex products, such as cars or aircrafts, are heavily relying on efficient collaboration among all actors involved in the project. Two major factors have been further increasing this dependency in the recent years.

On the one hand, design organisation has been increasingly challenged by the distribution of design teams and the distribution of programmes in general. This trend is general to most design industries. As reported by Simienuch and Sinclair in their study of automotive industry, percentage of an automobile being outsourced has risen from 5-15% in 1989 to 40-80% in 1997 [1]. Analogically in aeronautics, the average number of nationalities involved in a program has risen

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 42–51, 2005.

to 40 countries in the latest programs. As a consequence, one meeting out of three involves international partners. Global distribution of work is a reality.

On top of this increase of distributed teams, concurrent engineering organisation has been adopted as the norm for most large industrial design projects [5]. Concurrent engineering is based on [2]:

- simultaneous/parallel design activities with strong interdependencies between each parallel team,
- integration of all lifecycle constraints in the early phases of the project
- and development teams made of multidisciplinary backgrounds.

So concurrent engineering is yet another factor that reinforces the need for effective cooperation. Indeed, early involvement of all lifecycle constraints assumes, among others, a multidisciplinary cooperation between manufacturing, support and designers all along the project. Similarly, strong interdependences between teams distributed in several countries calls for frequent and efficient cooperation and coordination solutions.

Influenced by those two factors, design has shifted from a co-located sequential process to a globally distributed, concurrent process. Consequently, the existing very strong need for cooperation, who already occupied more than 50% of a designer's daily activity.[3], has now become a strategic element. It is critical that cooperation be efficiently supported among distributed muldisciplinary teams.

Different solutions exist to support remote cooperation, especially one-to-one collaboration. High-end solutions such as video-conferencing, data conferencing and application-sharing are now widely available on the market. However, we observed that those solutions are very seldom used for real-time cooperation and meetings. Teams still heavily rely on face-to-face co-located meetings and "low-cost" solutions such as email and telephone to cooperate. Those observations corroborate findings of other recent studies [3], [4], [5]. This lack of usage is usually said to be due to a lack of training and availability of "sophisticated" tools, and to the "natural" initial resistance of most users to new technologies. We argue in this paper that this lack of usage is also due to a lack of adequacy between existing collaboration tools' functionalities and actual needs.

In order to demonstrate that hypothesis and assess in what designer's needs are specific, we studied current work practices of aeronautical design teams during a set of design meetings called *working sessions*. This article presents and analyses the results of observations of those working sessions based on three criteria: information types, information flow and interactions performed. It helps understand what makes co-located design meetings so critical that no cooperation tools makes possible its occurrence in a distributed environment yet.

In a first section, the article justifies the choice of working sessions as a test-case in an industrial design context. A second section presents the methodology used to analyse the working sessions. The next two chapters then detail observation's results before comparing them to existing works in the following chapter. In a last section, the article eventually discusses perspectives for the introduction of relevant technological tools in the cooperation environment.

2 Working Sessions as a Use Case

Studies on environments supporting distributed design cooperation often focus on punctual projects limited in time and aiming at co-designing a whole product during meetings. Supporting environments must therefore facilitate actual design during the sessions. For instance, two remote student teams must cooperate during a limited number of meetings to design a complex mechanical sub-system in [6]. This type of configuration is not representative of large industrial programmes' reality.

Programmes' organisation follows a rigorous product breakdown-structure, so that each team works on a specific sub-component's design. Most design work is therefore accomplished within the co-located design teams. Meetings between distributed teams aim at coordinating the various sub-components coherency and interfaces. Very little if any design is performed during the meetings.

In those programmes, meetings are declined in several types of "reviews", for instance to validate the interfaces compatibility between sub-components, to verify the respect of initial constraints, to assess design maturity or to discuss planning and managements. Observations showed that most of those reviews are formal presentation meetings, leaving very little room for exchanges and debates beyond oral negotiation.

In parallel to this formal coordination process, slightly less formal meetings are organised to discuss more technical points. They involve experts and designers of remote teams for one to three days and give way to information exchanges and technical design solution's negotiation [7].

Those *working sessions* are by far the most interactive and complex of all design reviews for distributed team's we observed. The working sessions are the only ones gathering different technical teams for in-depth design discussions, involving sketches, technical drawings and open design discussions. This complexity has been the main reason for us to focus on working sessions. Analysis of working session's dynamics provides the most complete requirements in terms of information types, flow and interactions performed.

3 Methodology

The research methodology involved observing a selected number of working sessions over a fixed time period. These observations were organised to improve our understanding of the dynamics and specificities of working sessions gathering remote design teams.

Known as "structured observation", this methodology was made popular by Henry Mintzberg [8] in his landmark study of managerial work. Structured observation involves observing the designers as they participate to the working session. Each observed event, such as a verbal contact or a piece of shared document is coded by the researcher and categorized in pre-established records. In order to perform this study we adapted the structured observation methodology to the specificities of design meetings. A *chronology record* describes topics patterns, *documents record* describes each piece of document used, a *participants record* describes each participant, a *public topics* keeps track of each interventions during a topic discussion, documents used and interactions performed, a *sidebar discussions record* describes each private discussion and an *actions record* keeps track of all decisions made. Data are then synthesized and analysed.

More than 800 interventions have been recorded during four working sessions of one day each. The data were synthesised and then analysed at three different levels. The first level detailed the types of information manipulated, their format and origin. The second level involved analysing information flow, detailing where and how designers used the information. The third level analysed how designers interact with the information during a working session with respect to the previous two levels.

The analysis further differentiates the interventions performed depending on two different contexts. The first context is during a public discussion, involving the majority of participants around a given topic. The second context considers private discussions, also called sidebar discussions [9], when a few participants engage in a parallel discussion for a limited duration of time. The next two chapters detail the most significant analysis results for both contexts.

4 Public Discussions Dynamics

Public discussions are the most visible context where designers share and discuss information. Most remote collaboration tools focus exclusively on the public discussion as the only communication channel among distant users. However, each level of analysis draws key conclusions, which highlight frequent misconceptions or gaps between the reality and available tools.

4.1 Environment Layout

Working sessions usually involve five to seven disciplines from two or three nationalities for meetings lasting one to three days. Participation fluctuates between less than ten to around twenty participants. The latest being the maximum observed. Under those circumstances, the working session tends to be more formal, with less technical debates and more information sharing.

Our observations took place in one single room, all participants gathering around a long square table. Classical meeting room layout and tools were available: one video-projector, two whiteboards, a paper-board, a telephone and a fax. It was interesting to note that less than 50% of all participants brought laptops to the meetings, so that not all participants could connect to the available video-projector.

4.2 Information Formats

Our initial analyse was to evaluate the format of documents used during public conversations. It appears that designers tend to use less tangible documents, such as paper-based documents or whiteboard drawings, than a few years earlier. Observations made by Guthrie [3] three years ago showed that as few as 13% of documents used during aeronautical design meetings were electronic. Our observations registered that this figure has risen to 70%. This is mostly due to the availability of laptop computers and generalisation of office tools such as Microsoft® PowerPoint® for presentations.

However, with 30% of all documents used, tangible documents still play a significant role in public discussions. They are still privileged for fast and intuitive interactions, such as drawings and annotations and unplanned print-outs fetched from the suit-case at the last minute are still widely used during discussions.

Tangible documents would be hard to integrate to distributed sessions, however, the key ideas of intuitive interaction and unplanned, on the spot, use of documents may be considered.

While a lot of research effort is made to enable Digital Mock-Up (DMU) and Computer Aided Design (CAD) models to be used during distributed meetings, their actual use during working sessions was very seldom. Still, most presentations and documents contained snapshots and still images of the digital models. The images were used to highlight an innovative design solution or point out an existing clash but almost never required interactive models to be used.

Three possible reasons can explain that DMU and CAD applications are seldom used during working sessions. First, the lack of adequate hardware to run CAD software in meeting rooms. Second, the obligation for one the actors to "pilot" the CAD-station [10] and perform required interactions on the model. Third, because they are time consuming: both in set-up times and tedious manipulation delays while they are mainly used for snapshots

Analyse shows that at least two actors of each team brought documents to use in public discussions during working sessions. About 50% of the documents used had not been initially planned on the topic. The less formal the working session is the more unprepared documents are used.

Those points stress the need to add unprepared documents dynamically to an opened discussion. No matter who the actor is, all actors need to be able to access this functionality at any time.

4.3 Information Flow

Not surprisingly given the large proportion of digital documents, 67% of all documents are made public using the retro-projector. The table and whiteboard are also used on a lower percentage to display paper-documents and sketches. This observation only confirms that large displays are used to make information public.

A more surprising observation is the need for parallel visualisation of multiple documents. Observations showed that up to 26% of the documents used were displayed in parallel with other documents. In most cases two documents were used simultaneously, for instance a formal presentation of a topic was displayed on the projector while a designer showed related paper-based details of the same topic on the table. On a very small percentage of occasions (4%) up to three documents were used simultaneously, an electronic presentation, a whiteboard sketch and a paper document with cross references between all three documents. The need for parallel visualisation channels is usually based on a reference document, stating validated elements, and related documents presenting specific contributions. Parallel channels enable cross-references between documents, which both structure and helps to justify a presentation. It also enables to keep background information visible, which often facilitates shared understanding in a tongue all participants do not always master.

Collaborative tools almost never allow to simultaneously display different documents. This can be understandable for small display solutions in face to face meetings, but should be taken into account in large group meetings environments.

4.4 Interactions Performed

Observations showed that interactions have always been performed sequentially. Designers never simultaneously interacted with the same document. At some points, three experts have for instance been hotly discussing and drawing on a whiteboard, each expert with his own pen. But social rules made that if two experts were making a move to draw on the board, one was always stopping to prevent simultaneous interaction with the same support.

Therefore, there is no need to implement simultaneous interaction capabilities in meeting displays.

An analysis of interactions per displays and per documents format clearly shows that displaying documents, navigating within documents and pointing are the most performed of all interactions during a public discussion (85%). The smaller percentage of annotations, authoring and drawing is almost exclusively performed on tangible documents (95%).

In a significant percentage of cases (38%), several attendees interacted with the same document. This is particularly true for whiteboards, were in most cases two experts author, point and annotate the same document, but it was also observed for interactions with electronic documents. For example, one document was transferred to a laptop and displayed on the projector, the presenter points and makes references to his document while the laptop's owner navigates in the slides. During a discussion another expert asks for the mouse and circles a particular area of the public document to ask a question. In this case no less than three participants interacted with a single document. *There is a need for the environment to support different types of interactions on public documents. Any attendee should be able to interact with a public document from his seat.*

5 Private Discussions Dynamics

Frequently neglected in cooperative tools, private discussions, also called sidebar discussions, take an important part to the working sessions [11], [9].

5.1 Form and Objectives

Results showed that as much as 36 private discussions took place on average during working sessions. A large majority involved members of the same team or same nationality, but 22% of sidebar communications gathered actors from different nationalities.

Few actors usually participate to those discussions, 75% only involved two actors, and 19% involve three, so that 90% of them took place in the working session room.

There is a need to establish dynamically private communication channels among participants of different teams in parallel to the main discussion.

When analysing discussion's purpose, we found that the large majority of private discussions (72%) took place to extend and discuss a previous public topic. Designers usually engage in sidebar discussions to consolidate their confidence in an information discussed, or give and ask confirmation to their pairs. The remaining 28% of private discussions initiated a new discussion, for instance to prepare a new meeting on a specific topic.

A majority of private conversations use documents as a support to discussions (52%). As most private conversations have a direct link with a previous public topic, 20% of documents were previously displayed in the public space. Another 20% was authored directly during the conversation, all of them on tangible surfaces. After discussions, 17% of sidebar discussion's made their conclusions public, mainly orally. However, in 20% of cases documents used in the private space were also used to support their public conclusions.

A private discussion's environment should therefore permit to move seamlessly documents to and from the public space into the sidebar discussion space, or to author documents directly in the private space.

5.2 Display Surfaces and Interactions

Once on the private space, documents are predominantly visualised on usual private spaces: the table and laptops(resp. 53% and 39%). However, a few discussions took place around whiteboards, usually associated with public displays (9%). Mostly because the designers needed to sketch on existing drawings.

Again, there is a need to associate public displays, or their contents, to private discussions. Whether by offering the capability to switch between public and private mode or by providing equivalent functionalities in the private environments.

As for the public environment, a few occurrences of parallel use of documents during discussions have been observed. In most cases it associated electronic documents displayed on a laptop and technical sketches on paper notes.

Sidebar discussions may also require parallel visualisation of document.

6 Related Works

The need for parallel visualisation and seamless exchange and control of documents within a meeting environment have been studied in a few experiments of co-located ubicomp environments. For example the iRoom [12], CIFE [13] or iLand [14] associate three large displays and flexible control and document sharing capabilities between public and private spaces. However, those environments have been designed specifically for co-located meeting scenarios. Those works should be extended to the study of distributed scenarios.

Solutions to enable sidebar conversations have also already been proposed. Törlind et Larsson [11] tested an instant messaging solution and Marratech®[1] commercial solution offers private video channels. Those solutions only partially address the needs identified for private discussions as they are mainly limited to text or video. Moreover, those solutions have initially been designed to support one-to-one-to-one distributed users, not many-to-many collaboration, they therefore expect each user to collaborate through his personal computer not within a room or collective space.

7 Perspectives

By analysing working sessions dynamics, this article introduces a set of technological gaps to overcome in order to support distributed meetings. The next logical step is to design a cooperative environment that supports those functionalities. And to evaluate the impact and relevance of those new functionalities on distributed working sessions.

Two experiments have been conducted so far. The first is to introduce two digital visualisation channels in two remote environments. One is an interactive whiteboard, the other a classical projector. Designers are currently using the boards in local experiments. However, as they grow familiar with the functionalities, we observed an increase in annotations performed on digital boards and a few cases of parallel use of the two digital displays during meetings.

The second experiment was to introduce data-sharing capabilities to the environments. As users get used to the functionalities we observed during the latest co-located working sessions that designers used the data sharing functionality to display their private space on the public environment.

Next challenges will be to further evaluate the new solutions implemented during distributed meetings, and to integrate more intuitive documents flow and control capabilities together with sidebar discussions functionalities.

8 Conclusion

Through an analysis of aeronautical working sessions dynamics, this article investigated the patterns that cooperative tools should support for many-to-many cooperation scenarios. Based on semi-structured observations results, it focused on three levels of analysis to define the sessions dynamics: information types, information flow and interactions performed. We identified four major functionalities that current meeting environments fail to support effectively:

[1] Marratech: http://www.marratech.com .

- Unprepared documents are often dynamically integrated to the public discussion. The environment must facilitate on-the-spot information sharing.
- There is a strong need to share more than one document at the same time as multiple documents are often visualised in parallel, for cross-references.
- Multiple participants can sequentially interact with the same public document. Interaction solutions in the room should easily enable shared control of public documents.
- Participants often engage in private discussions. They transfer documents to and from the public space into the private discussion space, and author documents directly in the private space.

The authors are well aware of the limitations of these results who are confined to an observation of current patterns of collaboration in co-located design reviews. However, the observations stress major gaps between current practices and technological solutions available and call for further on-field studies of environments supporting those requirements distributed situations. It will then be interesting to evaluate the impact of those tools on distributed working sessions practices as new patterns may emerge.

References

1. Simienuch, C.E. and Sinclair, M.: Real time collaboration in design engineering: an expensive fantasy or affordable reality?, Behaviour and information technology, 18(5):361-371, (1999).
2. Prasad B.: Concurrent Engineering Fundamentals: Integrated Product and Process Organization. Prentice Hall, ISBN: 0131474634, December 1995.
3. Guthrie, C.: La Coexistence des Medias de Communication dans le Processus de Conception. Un projet de recherche-action dans une entreprise industrielle Ph.D. Dissertation, GREGOR Universit Paris1 Sorbonne, Fr (2004).
4. McGregor, S.P.: Towards a Prescriptive Design Process for Distributed Design Teams. Ph.D. Dissertation, DMEM University of Strathclyde, Glasgow, (2002).
5. Durstewitz, M., Kiefner, B., Kueke, R., Putkonen, H., Repo, P., Tuikka, T.: Virtual Collaboration Environment for Aircraft Design. In Proceedings of the Sixth International Conference on Information Visualisation (IV'02), London, UK. IEEE Press (2002), 502-507
6. Milne, A.: An Information-Theoretic Approach to the Study of Ubiquitous Computing Workspaces Supporting Geographically Distributed Engineering Design Teams as Group-Users. Ph.D. Dissertation. Stanford Univeristy, US (2005).
7. Détienne, F., Martin, G., Lavigne, E.: Negotiation in collaborative assessment of design solutions: an empirical study on a Concurrent Engineering process. Collaboration Technologies (2000)
8. Mintzberg, H: Structured observation as a method to study managerial work. Journal of Management Studies, vol. 7, no1, (1970), pp. 87-105
9. Chan, B: An experiment in supporting geographically distributed engineering design. Technical Report EDRC 12-69-95, Carnegie Mellon University (1995).
10. Lécaille P., Vink D.: Sociological and Ergonomic Aspects of Virtual World. The DMU-Vi Consortium (Digital Mock-Up Visualisation in Product Conception and Downstream Processes) Ref. DMU-Vi/1.5/3S-CRISTO/D1.5.4(2000)

11. Törlind, P., Larsson, A.: Support for Informal Communication in Distributed Engineering Design Teams. Proceedings of CIRP 2002, Hong-Kong (2002).
12. B. Johanson, A. Fox, T. Winograd.: The Interactive Workspaces Project: Experiences with Ubiquitous Computing Rooms. IEEE Pervasive Computing Magazine 1(2), April-June 2002
13. Fischer, M.; Hartmann, T.; Rank, E.; Neuberg, F.: Combining Different Project Modeling Approaches for Effective Support of Multi-Disciplinary Engineering Tasks. In Proceedings INCITE 2004, Langkawi, Malaysia (2004)
14. Straitz, N., Tandler, P., Müller-Tomfelde, C., Konomi, S.: "Roomware: Towards the next generation of human-computer interaction based on an integrated design of real and virtual worlds. J. Caroll (Eds), (2001) pp. 553-578

Designing Display-Rich Environments for Collaborative Scientific Visualization

Kyoung Shin Park[1] and Yongjoo Cho[2]

1 Digital Media Laboratory, Information and Communications University,
517-10 Dogok-dong, Gangnam-gu, Seoul 135-854, Korea
park@icu.ac.kr
2 Division of Media Technology, Sangmyung University,
7 Hongji-dong, Jongno-gu, Seoul 110-743, Korea
ycho@smu.ac.kr

Abstract. This paper presents a design study evaluating the shared workspace for Amplified Collaboration Environments (ACEs). ACE is a display-rich project room that enables distributed teams to work together in intensive collaboration campaigns. The study involved small groups of users performed a collaborative visualization and analysis task while varying the technology configuration for ACE's shared workspace. The result showed the participants benefited from the ability to see others' work, which helped enhance group awareness and performance.

1 Introduction

With the emergence of collaboration technologies, people can better communicate with one another and accomplish complex tasks while at distant locations. Major corporations launch global virtual teams to address multifaceted challenges of global competition, markets and coordination. Also, modern science and engineering research are often carried out by large, distributed and multidisciplinary research teams to produce significant discoveries by sharing computational resources, scientific instruments and massive data. The trend in the collaborative nature of scientific research and tremendous improvements of computer and network technologies suggest an opportunity to update and elaborate the original concept of scientific collaboration.

Amplified Collaboration Environment (ACE) is a display-rich project room that integrates ubiquitous tools and spaces to support a collaborative scientific investigation using advanced computation, collaboration and visualization technologies [5]. An ACE, such as the Continuum, adapts information to optimally display in a project room such that it allows distributed teams of experts to work together in intensive collaborative campaigns.

Fig. 1 shows the displays that comprise two fully constructed Continuum spaces at the Electronic Visualization Laboratory (EVL) at the University of Illinois at Chicago. The Continuum uses a variety of modular technologies: conferencing, content sharing, collaborative annotation, and wireless interaction. The Access Grid (AG) supports group-to-group collaboration in which a group of people at different locations can see and talk with one another simultaneously [2]. A passive stereoscopic display is used for 3D immersive data sharing. The high-resolution tiled displays

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 52 – 60, 2005.

Fig. 1. Two Continuum Amplified Collaboration Environments built at the Electronic Visualization Laboratory at the University of Illinois at Chicago

provide shared content views of text documents, web pages, spreadsheets, graphs and charts, and scientific visualizations. The collaborative annotation is supported by shared touch-screen whiteboard on which collaborators may jot down notes and sketch diagrams. Also, remote access interface is supported to encourage users to work on these displays collectively.

While collaborative visualization is a key component of data analysis process in scientific investigation, most widely available visualization tools are designed for single user models in which all users must cluster around a single workstation for collaboration. Over the past two decades, many synchronous collaboration tools have been developed by industry and research institutes, but most are designed for interaction among co-located users in a room (e.g., Colab [7]) or distributed users sitting at separated desktop computers (e.g., Commune [1]). While these tools are valuable, they are not capable of supporting the kind of interaction that occurs in real science campaigns because scientists need to collaboratively query, mine, view, and discuss visualizations of enormous data sets (in the order of terabytes) in real time. In this paper, we present an ACE's shared workspace model, targeted for accommodating interactions among distributed teams and sharing data intensive visualizations.

This paper presents the shared workspace models found in previous synchronous collaboration systems. It then describes a set of iterative design studies of distributed teams who used the Continuum spaces to perform a collaborative visualization and analysis task while varying the technology configurations. It will describe the changes of the Continuum technologies, discuss the findings, and compare them with previous shared workspace models.

2 Shared Workspace Models

In this section, we review some of more important shared workspace models found in previous collaboration systems. This review covers synchronous collaborative work of co-located or distributed group members sharing information over shared visual workspace.

- Collaborative work side-by-side model is a spontaneous co-located collaboration where individuals work side-by-side at adjacent terminals. Their activities can range from discussing their current work, comparing results, and sometimes leaning over to ask their neighbor for assistance and to monitor their neighbor's activities.
- Extreme programming model is a practice in which two programmers work side-by-side at one computer, continuously collaborating on the same design, algorithm, coding or testing [10].
- Shared surface model is where three to four people work around a large sheet of paper or a whiteboard with each individual having a pen for pointing or drawing. This model also includes Single Display Groupware (SDG), which offers a shared public display and multiple inputs from users for small co-located group work [8].
- The CoLab model is an electronic meeting room that has a large public display and individual workstations to give users the ability to work on private or public information artifacts [7].
- Presentation model is where all meeting participants share information artifacts presented on a large public display that are pushed from a speaker's private workspace (such as, a laptop computer).
- The war room model is a project room that contains whiteboards, flipcharts, tack boards, and walls stuck full of information that are visible and accessible to all members [6].
- Distributed shared surface model, such as Commune [1], is where all distributed participants see and share the same information and the interface of the work surface while allowing simultaneous access and providing methods to convey gestures to remote collaborators using tools like telepointers.
- Distributed CoLab model includes relaxed-WYSIWIS system (e.g. GroupSketchpad [3]) that allows distributed members to share, view and control information contents publicly or privately.
- Distributed presentation model, such as Forum [4], allows distributed audiences to view a video of the speaker as well as speaker's presentation slides. This model provides display pushing from the presenter's desktop screen to remote participants' desktops to share the information contents.
- Distributed meeting room model involves the coupling of several meeting rooms and remote offices as well as integrating them into virtual meetings. This model can be described as multi-users, multi-devices and room-to-room interaction.

3 The Design Study of Evaluating ACE's Shared Workspace

This section describes a set of iterative design studies conducted to examine the system configuration of ACE's shared workspace. It presents the design changes of the Continuum technologies and findings and lessons learned from the study.

Table 1. System configuration changes over iterative design studies

Study	Group	System Configuration			
		Conference	Tiled display	With personal display	With full-screen
Study 1	Group 1 Group 2	Full-AG & mini-AG	Distributed corkboard		
Study 2	Group 3 Group 4	Full-AG & Enhanced mini-AG*	Distributed corkboard	With close-up personal display*	
Study 3	Group 5 Group 6	Full-AG & Enhanced mini-AG	Distributed corkboard	With close-up personal display	With full-screen*
Study 4	Group 7 Group 8	Full-AG & Enhanced mini-AG	Presentation display*	With close-up personal display	With full-screen

3.1 Method

Table 1 shows the changes of group and system configuration over the four iterative design studies. Two groups of four students were assigned for each design study. For each study, the system configuration was varied, and the study evaluated mainly on the configuration of the tiled display. All students participated in two design studies and solved different problem sets with other group members. The groups were reorganized in the third or the fourth study.

Sixteen computer science graduate students volunteered as subjects in this study. All students had experience with computers and collaboration technologies such as email and instant messaging. Some students have used Microsoft's NetMeeting® or other online meeting room systems. None of them had prior experience with the XmdvTool [9], an information visualization tool, used as a part of this study. Most students had little to moderate experience with correlation statistics. All students expressed fairly high interests in collaborative work using the Continuum technologies.

The groups were given a pre-test survey (e.g. technology familiarity, comfort, interest, and domain knowledge) and received one-hour training about the Continuum's hardware and software technologies and some basic concepts of correlation statistics and multivariate data analysis required in their first exposure to the tasks. The group members were then separated in two Continuum spaces (two users in each room) and asked to perform a collaborative information visualization and analysis task. A post-test survey and interview was followed shortly after to give feedback about the usability of Continuum technologies.

In this task, the groups were asked to perform an exploratory data analysis on a multivariate dataset using the information visualization tool, called XmdvTool. The groups briefly scanned the text of the dataset to become familiar with the variables and to read the hypotheses. They then examined the Cereals dataset or the Boston housing dataset on the XmdvTool to answer questions and wrote up a report of the group findings on the whiteboard. The groups were asked to solve five focused questions where members would need to find evidence to verify or refute hypotheses. The groups were also asked to solve two trend questions where members would need to search for inquiry trends in the dataset.

All groups were recorded using video cameras. An evaluator in each room also took observational notes on group behaviors. All activities of group members on the computers were also captured into log data files. The work quality and group work process, such as number of problems solved correctly, members' participation, and group interaction was also measured. A post-test survey was collected to measure user satisfaction with the technology and the precision of the work process.

3.2 System Configuration

The first design study evaluated the distributed corkboard of the Continuum. The Continuum uses the tiled display as a large distributed corkboard where information is visible to all members and member's activities are available at a glance. As shown in Fig. 1, each Continuum space has 4-node tiled display (2 by 2 wall mounted layout), an Access Grid display (4 cameras and 2 microphones in the full-AG setting, and 1 camera and microphone in the mini-AG setting), and a shared touch-screen whiteboard. The SpaceGlider program allowedïusers toïnavigate their mouse pointers across the boundary of the tiled display screens. SpaceGlider is an intuitive approach to the problem of navigating and controlling multiple computers that may exist in physically separate locations [5]. The shared whiteboard between rooms were connected using the NetMeeting's shared whiteboard application.

The second design study evaluated the addition of personal displays that mirrored one of the tile screens. The personal displays (tablet PCs) were provided to support a close-up view. This configuration change was made to help user interaction with the wall mounted tiled display. The mini-AG was improved by adding another microphone and a camera with a magnifying filter to capture a wider close-up view of the collaborator. This is done to help casual interaction between distributed participants. SpaceGlider was used to connect the four tile screens and the whiteboard.

The third design study evaluated the scalable distributed corkboard tiled display, which was provided to support smoother transition between individual work and group work. It allowed users to view either the distributed corkboard or one screen maximized over the entire tiled display. The tiled display, by default, showed four individual screens (i.e., full-screen option is off) that had distinct background colors for easy identification of workspace. Users could turn on or off the full-screen option at any time to maximize the screen over the entire tiled display. The Switcher program was provided for users to access the tiled display and the whiteboard. Switcher allowed users to jump to any of the displays [5].

The fourth design study evaluated the presentation display to see how the user's awareness of their remote collaborators' work factored into their combined problem-solving ability. In this study, group members were assigned to their own tablet PCs that had distinct background colors, and one could choose to make his/her workspace appear on the tiled display for all members to see. This configuration allowed only one individual workspace (tablet PCs) to be visible on the public group display (the tiled display) at a time. Hence, it provided more private workspaces while it also allowed users to make their workspace visible to all group members. This time participants could switch only between their own workspace and the whiteboard on their tablet PC. Other settings were the same as the third study.

3.3 Observations

The distributed corkboard tiled display supported *high visibility of all group members' work activities*. It allowed members to casually glance over at others' work on their tile screens, which helped them maintain necessary group awareness between distributed members. The participants pointed out that they found useful information or answers, received new ideas and learned from other's activities, which saved them from having to ask questions. For those having difficulty, remote collaborators could immediately address their problems, as their screens were available for everyone's inspection. The distributed corkboard also helped the group members to *easily refer to the contents on the screens*. It is observed that the participants often asked the others to "look at" one of their screens to point out someone's finding on the screen during group discussion.

The presence of personal displays (i.e., tablet PCs) seemed to encourage the group members to *focus on their individual work*, but it also reduced the amount of casual glancing by the members. The members worked more on close-up personal displays and occasionally checked the distributed corkboard tiled display or the local collaborator's personal display to see others' activities. With the personal displays, the distributed corkboard was mainly used to mediate remote collaboration, such as assisting remote collaborators and facilitating group discussion. Interestingly, the groups indicated that they were still aware of others' activities even with reduced casual glancing.

The scalable distributed corkboard provided less public visibility, because the full-screen made others' workspaces private. It also helped users *read texts or graphs more easily* and supported somewhat *easy transition between individual work and group work*. The third study groups used this full-screen option to read trends in the graphs from one's finding during group discussion. This full-screen option was also useful for individual uses, such as to make a personal scatter plot graph bigger. However, it did not interfere with other members' work since they could still work on their personal displays. The participants also said they used the full-screen occasionally to grab other's attention to start a group discussion. While the scalable distributed corkboard reduced visibility to some extent, our participants were still aware of current contexts since the full-screen was used for short periods.

The presentation display with personal displays provided *more private visibility*. It allowed only one person's work (i.e., tablet PC) to be visible on the presentation display (i.e., tiled display) at a time. This presentation display disallowed casual glancing over other members' work. This less visibility resulted in reduced group performance and interaction due to the extra step required to show individual work on the public display. We called this the "show me" pattern – That is, members were asking for and offering to show his/her workspace in order to share information with other members. This was a source of delay when the group wanted to share information since group members had to explicitly ask to make information visible. More importantly, given the presentation display, the groups expressed the need to see others' work simultaneously for comparison purposes as they did with the distributed corkboard. They reported that *they only looked at the presentation display during group discussions*.

4 Discussion

The distributed corkboard provided a public display that allowed group members to interact with any tile screen and to see other's work at any time. This is similar to the shared surface model with the only difference being the group having high-resolution workspaces. The distributed corkboard with personal displays encouraged group members to focus on individual work on their personal displays while the distributed corkboard in users' periphery served as a group display for awareness and remote collaboration. This is similar to the CoLab model with the difference being the group being able to see other members' work. The scalable distributed corkboard with the personal displays is also similar to the CoLab model with the difference being the group having both the distributed corkboard for awareness and the full-screen for group focused work. The presentation display with the personal displays is modeled for the presentation model where individual work becomes visible on the presentation display by "display-pushing" from private personal displays. This, however, is a source for delayed information sharing among members in our intensive collaborative work.

A traditional model for collaboration is a group of users clustering around a single workstation, discussing their ideas and planning their next actions. The collaborative work side-by-side model encourages more individual work while it allows spontaneous group collaboration. The extreme programming model strictly forces a pair to work together using a single workstation thereby disallowing individual parallel work, but assisting the pair to establish greater shared understanding. The shared surface model provides a WYSIWIS group display with multiple input controls to allow members to work simultaneously on this shared workspace. The WYSIWIS provides a common view among members, which helps them build shared understanding during distributed collaborative work [1]. However, WYSIWIS is too restrictive and often results in members having to switch the shared workspace sequentially due to less information visible at one time.

The CoLab model is motivated to relax the over-strictness of WYSIWIS to support both group work and individual work [7]. However, the awareness problem is particularly severe in the distributed case due to non-visibility of individual work when members work simultaneously on different parts of the shared workspace [3]. In the presentation model, individual work is not visible until the individual is presenting it on a public display. Hence, this model is appropriate for group members working together on a single item of interest, such as a Power-Point presentation. The war room model is most apt at intensive collaborative work since it provides high visibility of all members' work and also allows members' spontaneous and simultaneous interaction with these artifacts. The ACE extends the shared workspace of the war room model to distributed collaboration. Our design study confirmed that the participants benefited from highly visible, high-resolution shared workspaces. This feature helped them work simultaneously while maintaining necessary group awareness.

5 Conclusion

Amplified Collaboration Environments are integrated ubiquitous tools and environments that allow distributed researchers to work together in intensive collaborative campaigns. The Continuum is an ACE specifically designed to support collaborative scientific investigation in which collaborators can visualize, analyze, and solve problems. Unlike other previous shared workspace models, ACE's aim is to support interactions among distributed teams with a lot of information artifacts over multiple input controls and display technologies.

This paper presented an exploratory design study conducted to evaluate ACE's shared workspace for enhancing distributed teams' collaborative work. The study asked small groups of users in two separate Continuum spaces to perform a collaborative visualization and analysis task under a variety of technology configurations. The study result confirmed the value of having all information artifacts visible to all group members during intensive collaboration. Such high visibility helped members to be aware of each other's activities and the problems others faced. When this visibility was reduced, members explicitly showed or asked to show information from their private personal display to the public group display. This research will be continued in the direction of improving and evaluating the high-resolution graphics and collaboration technologies of the Continuum to support real world distance collaboration.

Acknowledgement

We would like to thank the subjects participated in this design study. We would also like to greatly acknowledge the contribution of students and faculties at the Electronic Visualization Laboratory at the University of Illinois at Chicago, especially Jason Leigh and Andrew Johnson, for their assistance and extreme dedication to supporting the experiment process.

References

1. Bly, S. A. and Minneman, S. L. Commune: A Shared Drawing Surface. In Proceedings of Conference on Office Information Systems, Boston, Massachusetts, pages 184-192, 1990.
2. Childers, L., Disz, T., Olson, R., Papka, M. and Stevens, R.: Access Grids: Immersive Group-to-Group Collaborative Visualization. In Proceedings of IPT Workshop, 2000.
3. Gutwin, C. and Greenberg, S.: A Descriptive Framework of Workspace Awareness for Real-Time Groupware. Journal of Computer Supported Cooperative Work, 11(3-4):441-446, 2002.
4. Isaacs, E. A., Morris, T., and Rodriguez, T.K. A Forum for Supporting Interactive Presentations to Distributed Audiences. In Proceedings of Computer Supported Cooperative Work, pages 405-416, 1994.
5. Leigh, J., Johnson, A., Park, K., Nayak, A., Singh, R. and Chowdry, V.: Amplified Collaboration Environments. In Proceedings of VR Grid Workshop, Daejun, Korea, 77-85, 2002.

6. Olson, G. M. and Olson, J. S. Distance Matters. Human-computer Interaction 15(2 and 3):139-179, 2000.
7. Stefik, M., Bobrow, D. G., Foster, G., Lanning, S. and Tatar, D.: WYSIWIS Revised: Early Experience with Multiuser Interfaces. ACM Transactions on Office Information Systems, 5(2): 147-167, 1987.
8. Stewart, J., Bederson, B. B. and Druin, A. Single Display Groupware: A Model for Copresent Collaboration. In Proceedings of CHI'99, pages 286-293, 1999.
9. Ward, M. O., XmdvTool: http://davis.wpi.edu/xmdv/.
10. Williams, L. A. and Kessler, R. R. All I really need to know about pair programming I learned in kindergarten. Communications of the ACM. 43(5):108-114, 2000.

Visualization of Interactions Within a Project: The IVF Framework

Benoît Otjacques[1], Monique Noirhomme[2], and Fernand Feltz[1]

[1] Public Research Center – Gabriel Lippmann,
Information Systems Research Unit, 41, Rue du Brill,
L-4422 Belvaux, Luxembourg
{otjacque, feltz}@crpgl.lu
[2] University of Namur, Computer Science Institute,
21, Rue Grangagnage, B-5000 Namur, Belgique
monique.noirhomme@info.fundp.ac.be

Abstract. Almost all projects can be considered as cooperative undertakings. Their strategic management as well as the daily operations causes numerous interactions to occur, either among persons or among persons and resources. These interactions have been studied from various viewpoints but few researchers have focused on their visualization. The graphical representation of the cooperation is however a powerful tool to help the project participants to get a correct understanding of the situation. This paper proposes thus a structuring framework (IVF - Interaction Visualization Framework) of the visualization techniques used to display such interactions. Three basic axes of classification are used to structure the study. Which objects are visualized? Why are they visualized? How are they visualized? For each axis, several properties have been identified and the admitted values have been specified. This work can be considered as a first step towards a structured view of the 'visualization of cooperation' domain.

1 Introduction

Almost all projects can be considered as cooperative undertakings. In fact, considering that, in most of the cases, several persons join their efforts towards the project common goal, the three basic functions of cooperative situations, namely communication, co-production and sharing, and coordination [8], are always present to a certain extent. In order to produce the promised deliverables, numerous interactions occur all along the project life cycle, either among persons or among persons and material or immaterial objects. These interactions can really be considered as the visible face of the cooperation. It is thus not very surprising that many researchers have studied them in different contexts.

For instance, in the construction industry, the interactions among the participants of a project have been studied from very various viewpoints such as trust [9], distribution of communication media [5], modes of information dissemination [10], influence of communication media on design performance [6], or distributed cognition [12].

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 61 – 70, 2005.

Nevertheless, notably fewer researchers have focused their work on the specific challenges associated with the visualization of these interactions, or in other words, on the visualization of the cooperation. This viewpoint is, however, really worth being investigated. Indeed, human beings know for centuries how powerful may be a graphical representation of a problem for helping to solve it, which is usually summarized by the famous citation '*A picture is worth a thousand words*'. This paper aims thus to reflect on a global framework that would be able to characterize the visualization techniques aiming to represent some interactions such as those arising during a project.

The paper is organized as follows. First, the motivation of the research is explained. Second, some definitions are proposed in order to precisely position the authors' work. The framework used for the analysis of the visualization techniques is then described in details. The next section highlights the potential of application of the framework with an example. The paper ends with a conclusion and some reflections on further works.

2 Motivation of the Research

The intended result of this work is thus to propose a taxonomy which could be used as a basic tool by researchers studying the visualization of interactions. Indeed, to the limit of our knowledge, no taxonomy especially dedicated to this kind of visualization techniques has already been proposed. Such a framework would, however, be clearly of prime interest as the following reflections point it out.

Herbert Simon stated some decades ago that '*the first step to understanding any set of phenomena is to develop a taxonomy*' [14]. Several researchers have also drawn the attention on the reasons why taxonomies, classifications or frameworks (these terms are often used with quite similar meanings) are so crucial in any science. In this paper, we limit ourselves to some specific authors who belong to the information visualization domain in order to demonstrate that the arguments apply in our context but the reader must be aware that numerous other researchers have discussed this point.

According to Lohse and his colleagues [7], '*Classification lies at the heart of every scientific field. Classifications structure domains of systematic inquiry and provide concepts for developing theories to identify anomalies and to predict future research needs.*' Wehrend and Lewis [16] explain that, with a common conceptual framework, '*workers in any area can place their techniques, so that abstract similarities among problems in different application areas can be identified, and new techniques can be readily transported across application lines*'. Gruia-Catalin and Cox [4] mention that a taxonomy is a '*vehicle through which we carry out a systematic review of current systems, techniques, trends, and ideas …*'.

In any scientific domain, a taxonomy is necessary for several reasons, among which the most important are probably:

- allowing a systematic and rigorous review of current techniques and ideas;
- positioning specific works in the whole research field;
- clarifying the concepts of the field, which is a required step before elaborating a new theory;

- identifying similarities and differences among the research findings;
- pointing out new research directions;
- highlighting lacks in research proposals.

This list is certainly not exhaustive but it is believed to include the most significant elements and it offers enough good reasons to reflect on a new taxonomy.

3 Definitions

Before all, it may be useful to remind the scope of the different scientific domains dealing with graphics and data. Card, Mackinlay and Shneiderman [3], who are among the most renowned authors in this field, define '*visualization*' as '*the use of computer-supported, interactive, visual representations of data to amplify cognition*'. They also propose a list of working definitions aiming to clarify the relationships among the concepts related to information visualization. The most useful of them in the context of this paper are the following. '*Scientific visualization*' concerns '*the use of interactive visual representations of scientific data, typically physically based, to amplify cognition*'. '*Information visualization*' deals with '*the use of interactive visual representations of abstract, non physically based data to amplify cognition*'.

The research work described in this paper can now be better positioned: it relates to the information visualization field and does not encompass the scientific visualization domain. In other words, the visualization of physical artifacts (e.g. a combustion engine, a chemical molecule, or a power plant) is not considered in our reflection. Instead, we focus on the graphical representation of abstract objects, such as meetings, e-mails or access to shared documents.

After having specified to which scientific discipline our work belongs, the research field needs to be accurately delimited. The research aims to study the visualizations techniques used to display the interactions occurring during a project, i.e. the interactions linked to a professional context. Therefore, the concept of interactions, which is a central element of our research work, has also to be defined. In this paper, an '*interaction*' is defined as '*any kind of communication or exchange among two or more persons, or between one or many persons and some resources used in the project*'. Typical examples of such interactions are phone calls between an architect and a contractor in a construction project, e-mail exchanges between a project manager and the project members, face-to-face meetings between the project manager and the client, successive operations on a document located on a web site, or planning modification requests by some project participants.

4 Interaction Visualization Framework - IVF

4.1 General Approach

A global framework, called IVF (*Interaction Visualization Framework*), has been designed to underpin the analysis of the visualization techniques aiming to display some interactions (called *InterVis* techniques in this paper). It relies on a very rational approach. Any visualization technique can be qualified by an object to be visualized

(*What?*), a goal for which the object is visualized (*Why?*) and a technique with which the object is visualized (*How?*). This generic framework has been instantiated in our research context and has been applied to distinguish the *InterVis* techniques.

In each of the three above-mentioned aspects, a set of properties has been identified and the values allowed for each property have been specified. Most of the time, the values are not exclusive, which means that a given *InterVis* technique can simultaneously take several values.

Some of the properties (e.g. *Retinal Variables*, *Entity Covering Level*) are derived from previous taxonomies in the information visualization domain [1], [2], [13], [15]. Other properties (e.g. *Interaction Media*, *Integration*) do not belong to these previous taxonomies. The structuring work described in the paper is thus original in two ways. One one hand, it identifies some new properties. One the other hand, it offers a specific collection of criteria especially dedicated to the classification of *InterVis* techniques.

The following sections describe the properties associated to the three basic viewpoints: *What? Why? How?*

4.2 Objects to Be Visualized: *What?*

First we have to discuss the objects to be visualized. In accordance with our previous definition, two kinds of objects are concerned: the interacting entities (persons and / or resources) and the interactions themselves. Very logically, the properties used to describe the object to be displayed are thus grouped in two distinct sets.

The first set of properties qualifies the entities in interactions.

- Two kinds of entities are distinguished: '*persons*' and '*resources*'. Obviously, in this approach, the persons are not considered as resources, which are limited to material (e.g. device) or immaterial (e.g. document) objects. Two values are allowed for the '***Interacting Entities***' property, namely '*person-to-person*' and '*person-to-resource*'.
- The '***Entity Visualized Objects***' property allows specifying whether the visualization technique displays entities and / or attributes of these entities. It may take two values: '*entities*' or '*entity attributes*'. For instance, in this context, a project participant is considered as an entity while the role of this participant in the project (e.g. manager) is an entity attribute.
- The '***Entity Temporal Aspects***' property specifies whether some time-related aspects (e.g. history) of the entities are displayed. Due to the very specific nature of time, a dedicated property has been added while time-related aspects might have been considered conceptually as attributes of the entities. This property admits two values: '*time sensitive*' and '*not time sensitive*'. For instance, if the visualization shows only the presence of interacting persons, this property takes the '*not time sensitive*' value, but if it also displays the moment when the persons joined the project the '*time sensitive*' value is set.
- The next property, called '***Entity Covering Level***', expresses the level of covering of the data space. It can take three values: '*one entity*' (i.e. visualization of one specific object), '*some entities*' (i.e. a specific subset of all objects) and '*all entities*' (i.e. all objects). The example of project participants can still be used to

illustrate this property. If one specific participant is displayed, the property takes the value '*one entity*'. If the visualization concerns the subgroup of all participants employed by a given company, the property takes the value '*some entities*'. Finally, if all participants of the project are showed, the property is set to '*all entities*'.

- The '***Entity Granularity***' property refers to the level of data aggregation allowed by the visualization technique. If only individual entities can be displayed, the property value is set to '*individual entities*' but if some groups of entities can be visualized as such, the property gets the '*aggregated entities*' value. For instance, if the visualization technique only allows displaying documents as separate graphical objects, the '*Entity Granularity*' property gets the '*individual entities*' value. If it allows showing a group of documents as a single graphical object, the property is set to '*aggregated entities*'.

The second set of properties relates to the interactions themselves.

- The '***Interaction Visualized Objects***' property allows specifying whether the visualization technique displays interactions and / or attributes of these interactions. It may take two values: '*interactions*' or '*interaction attributes*'. For instance, in this context, an e-mail is considered as a basic object while the 'priority' of this e-mail is an object attribute.
- The '***Interaction Temporal Aspects***' property specifies whether some time-related aspects (e.g. history) of the interactions are displayed. Due to the very specific nature of time, this dedicated property has been included but, conceptually, time-related aspects might have been considered as attributes of the interactions. This property admits two values: '*time sensitive*' and '*not time sensitive*'. For instance, if the visualization shows the moment when each interaction occurred, it gets the '*time sensitive*' value but if it displays only that some interactions occurred, the '*time not sensitive*' value is set.
- The second property expresses the level of covering of the data space. The '***Interaction Covering Level***' property can take three values: '*one interaction*' (i.e. visualization of one specific object), '*some interactions*' (i.e. a specific subset of all objects) and '*all interactions*' (i.e. all objects). The example of e-mail exchanges can still be used to illustrate this property. If the technique permits to show a specific e-mail, the property takes the value '*one interaction*'. If the visualization concerns all the e-mails sent by a specific participant, the property takes the value '*some interactions*'. Finally, if all e-mails exchanged during the project can be displayed, the property is set to '*all interactions*'.
- The '***Interaction Granularity***' property refers to the level of data aggregation allowed by the visualization technique. If only individual interactions can be displayed, the property value is set to '*individual interactions*' but if some sets of interactions can be visualized, the property gets the '*aggregated interactions*' value. For instance, if the visualization technique visualizes each individual e-mail as a single graphical object, it gets the '*individual interactions*' value. If it allows representing all the e-mails exchanged between two project participants with a unique graphical object, the property is set to '*aggregated interactions*'..

- The '***Interaction Arity***' property is defined by the number of entities that take part in the interaction. Two values are allowed for the interaction arity: '*binary*' (i.e. interaction between two entities) and '*multiple*' (i.e. interaction between more than two entities).
- The '***Interaction Centricity***' property specifies whether the interactions that are visualized are relative to a given central person ('*egocentric*') or concerns all interacting people without favoring a specific individual ('*sociocentric*'). For instance, an e-mail client is typically '*egocentric*' as it shows all interactions between the connected user and the other persons.
- The '***Interaction Media***' property refers to the tool(s) supporting the interactions. Some possible values are '*face-to-face meetings*', '*e-mail communications*', '*phone calls*', '*video-conferencing*', '*web site access*', …

4.3 Goal of the Visualization: *Why?*

The goal of the visualization is the second aspect that we use to classify the visualizations of interactions. This aspect concerns the context of use for which the visualization is designed. Indeed, it is crucial to elicit the reasons why some interactions are displayed. Two properties are associated with this viewpoint.

- The '***Visualization User Role***' property specify whether the technique is dedicated to be used by an actor which takes part in some of the interactions ('*participating actor*') or by someone which is external to the interactions ('*external observer*').
- The '***Visualization Purpose***' property indicates the essential purpose of the visualization. Two basic goals are identified: '*interaction support*' and '*interaction analysis*'. The first value concerns the cases when the basic purpose of the visualization is to support the user to interact with other entities and the second value applies when the visualization is used to analyze the interactions between some entities.

4.4 Visualization Techniques: *How?*

The last classification axis concerns the visualization techniques themselves. Intentionally, no property is defined to specify what might be called the 'visualization type' (e.g. *pie chart*, *tree* or *parallel coordinates*). In fact, such a property would not be consistent with the framework philosophy that aims to characterize the *InterVis* techniques in a way as generic and lasting as possible.

- The '***Display Space Dimensions***' property is used to indicate how many space dimensions are used to represent the data graphically. Three values are allowed: '*1-Dimension*', '*2-Dimensions*' and '*3-Dimensions*'. This property is defined as the number of coordinates necessary to position any data in the visualization display space. It may be useful to point out that it does not refer to the number of dimensions of the data. In Bertin's terminology [1], this property refers to 'plan variables'. It does not concern the retinal variables such as color or size. For instance, a timeline is '*1-Dimension*', a planar tree is '*2-Dimensional*' and a virtual reality representation is '*3-Dimensional*'.

- The '***Entity Graphical Mapping***' property indicates which kind of graphical object is used to represent the entities in interaction. The possible values are '*point object*', '*linear object*', '*surface object*' and '*volume object*'. For instance, in usual 'node-links' graphs, this property takes the value '*point object*' because the entities are displayed as point-like nodes.
- The '***Interaction Graphical Mapping***' property indicates which kind of graphical object is used to represent the interactions among the entities. The possible values are '*point object*', '*linear object*', '*surface object*' and '*volume object*'. For instance, in usual 'node-links' graphs, this property takes the value '*linear object*' because entities are displayed as links.
- The '***Retinal Variables***' property specifies which retinal variables are used to represent data values in the visualization technique: '*color*', '*grey level*', '*shape*', '*orientation*', '*size*', '*grain*', and '*texture*'. For instance, in a node-link graph, the frequency of the interactions can be represented by coloring the links ('*color*') or increasing the link width ('*size*').
- The '***Display Space Boundaries***' property is used to specify whether the graphical representation uses a fixed ('*fixed*') or variable ('*variable*') display space, depending on the size of the data set. For instance, this property is set to '*fixed*' for a treemap and to '*variable*' for a classic node-link graph.
- The '***Geometrical Distortion***' property specifies whether the visualization technique shows distorted views of the data ('*distorted*' vs. '*not distorted*'). In a distorted view, the values of the data are not all equally mapped with their geometrical representation within the display space. The fisheye lens and the 'Perspective Wall' are some examples of distorted techniques.
- The '***Abstraction Level***' property indicates whether the graphical elements used in the representation are purely abstract shapes ('*abstract*'), remind some real objects via metaphors ('*metaphorical*') or mimic reality ('*pseudo-real*'). To illustrate these values, representing e-mails by arcs takes the '*abstract*' value; using expressive faces to display e-mail priority takes the '*metaphorical*' value; and representing meetings within a virtual 3-D world takes the '*pseudo-real*' value.
- The '***Animation***' property indicates whether some movement or animation is included in the graphical representation of the interactions. Two values are allowed: '*animated*' or '*static*'. In order to avoid confusion, it must be mentioned that this property does not refer to the dynamics of the interactions themselves.
- The '***Interactivity***' property refers to the presence or absence of means for the user to interact with the graphical representation. A visualization technique will be called '*interactive*' if the user can dynamically modify the selection of objects to be displayed, the displayed properties of these objects, or a parameter of the graphical output used to display them. If the user cannot modify any aspect of the data representation, the visualization technique is called '*not interactive*'.
- The '***Integration***' property explains whether the visualization technique is integrated with the interaction tool ('*integrated*') or not ('*independent*'). For instance, an e-mail visualization module can be integrated within an e-mail client ('*integrated*') or it can be used to display an e-mail log file issued from an external application ('*independent*').

5 Application of the IVF Framework: An Example

This section illustrates how the IVF framework is currently used to support the design of new visualization techniques. The e-mail conversation is one of the very usual ways to cooperate and several researchers study how to enhance the graphical representation of these interactions. In this context, the authors have recently developed a technique called *Mat'Graph* [11] (cf. Fig. 1). It may be noted that the IVF framework was not used in the design process because it has been developed afterwards. Nevertheless, the next developments of this technique, which are currently under way, clearly benefit of the IVF framework.

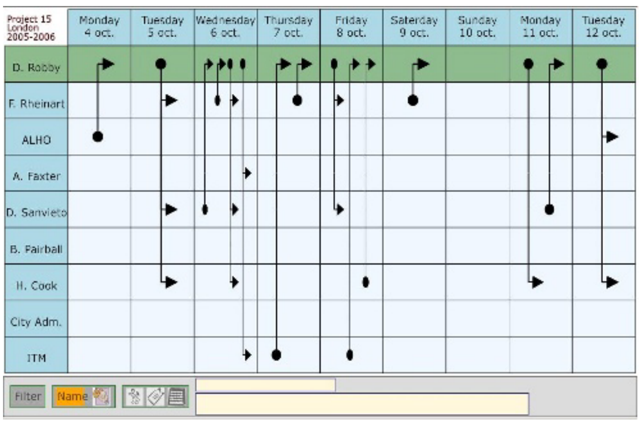


Fig. 1. *Mat'Graph* Visualization of e-mails

Indeed, the design process has been modified and the new procedure is described hereafter. First, the IVF framework is used to qualify the visualization to be improved - namely *Mat'Graph* in the example - as well as other potentially appropriate visualization proposals. Second, the values given to the IVF properties in the case of *Mat'Graph* are systematically scrutinized and are compared with those encountered in the other existing techniques. Third, the modifications that are supposed to be the most meaningful are identified. Finally, a new visualization technique (*Mat'Graph2*) is developed as an evolution of the former one. This design methodology is applied in an iterative manner in order to progressively refine the visualization technique as the two following examples illustrate it.

The first example relates to the representation of some interaction properties. The *Mat'Graph* visualization was given the '*color*' value for the '***Retinal Variables***' property. Having characterized other techniques, it appeared that it could sometimes be useful to use the '*shape*' of some glyphs (i.e. another value cumulatively allowed for the '***Retinal Variable***' property) to convey information about a property of the messages. *Mat'Graph* was thus adapted and it was decided to visualize the priority of the e-mails by the shape and the color of the glyph at the beginning of the arrows associated to them. The meaningless black circle in *Mat'Graph* was thus replaced by a red square when the message priority is high (cf. Fig. 2). A second occurrence of the refining process concerned the modification of the value given to *Mat'Graph* to the '***Interaction Media***' property. Instead of displaying only the e-mail-based interactions, it was decided to visualize also the '*face-to-face*' meetings. Some vertical yellow lines in the matrix cells of the participants are used for this purpose (cf. Fig 2.).

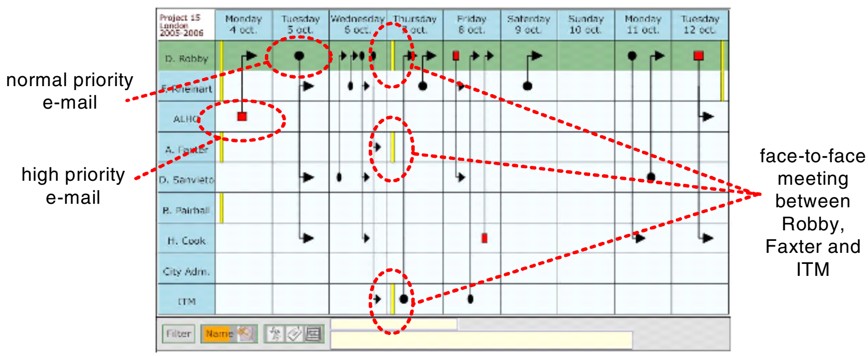


Fig. 2. *Mat'Graph2*, an IVF-based evolution of *Mat'Graph*

Those successive refinements of the *Mat'Graph* visualization technique could probably also be achieved by a less formalized process. Nevertheless, the IVF framework certainly eases the work of identifying the parameters of the design space where some improvements are possible.

6 Conclusion

The proposed framework is a first attempt to classify the visualization techniques aiming to display the interactions occurring during a project, which are seen as the visible face of the underlying cooperation. Several properties have been identified to characterize the *InterVis* techniques according to what is visualized, the purpose for which it is visualized and the techniques used to visualize it. Further works related to the IVF framework will explore three main directions. First, some complementary reflections are needed to assess whether some new classification criteria should be added. Similarly, the range of possible values for some properties should probably be more refined. The framework will also be confronted to a large set of *InterVis* techniques in order to evaluate its completeness. Finally, as it has been illustrated, the framework is currently used to support the design of new *InterVis* techniques.

References

1. Bertin J.: Sémiologie graphique, 1998 reprint of the 2nd édition of 1973, Editions des hautes Etudes en Sciences Sociales. 1998.
2. Card S. and Mackinlay J.: The Structure of the Information Visualization Design Space, in Proceedings of the IEEE Symposium on Visualization (InfoVis' 97), 18-25 October 1997, Phoenix, Arizona, USA, pp. 92-100.
3. Card S., Mackinlay J. and Shneiderman B.: Readings in Information Visualization, Using Vision to Think. Morgan Kaufmann Publishers, Academic Press, San Francisco, 1999. pp 6-7.
4. Gruia-Catalin R. and Cox K. C.: Program Visualization : The Art of Mapping Programs to Pictures, in Proceedings of the 14th International Conference on Software Engineering, 1992, Melbourne, Australia, pp. 412-420.
5. Howard R. and Petersen E.: Monitoring communications in partnering projects, in Electronic Journal of Information Technology in Construction, Vol. 6, 2001, http://www.itcon.org/2001/1, accessed 18 March 2005.
6. Kvan T. and Gao S.: Frames, Knowledge and Media, An investigative Study of Frame Systems within Computer and Paper Supported Collaborative Design Process, in Proceedings of the 22th Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 2004), September 2004, Copenhague, Denmark, pp. 410-417.
7. Lohse G. L., Biolsi K., Walker N. and Rueter H. H.: A Classification of Visual Representations, in Communications of the ACM, Vol. 37, N° 12, December 1994, pp. 36-49.
8. Lonchamps J.: Le travail coopératif et ses technologies, Hermes Science Publications, Paris, 2003, p. 27-28.
9. Mohamed S.: Determinants of Trust and Relationship Commitment in Supply Chain Partnerships, in Proceedings of 2nd International Conference on Innovation in Architecture, Engineering and Construction, June 2003, Loughborough, UK, pp. 617-625.
10. Otjacques B., Post P. and Feltz F.: Management of Information Flows during Construction Projects, in Proceedings of the 20th CIB W78 International Conference on Information Technology for Construction, April 2003, Auckland, New Zealand, pp. 278-285.
11. Otjacques B., Feltz F., Halin G. and Bignon J.-C.: Mat'Graph: Transformation matricielle de graphe pour visualiser des échanges électroniques, to appear in Proceedings of the 17th French-speaking Conference on Human-Computer Interaction, September 2005, Toulouse, France, ACM Press.
12. Perry M.J.: Distributed cognition and computer supported collaborative design: the organization of work in construction engineering, PhD Thesis, Brunel University, UK, 1997.
13. Roth S. F. and Mattis J.: Data Characterization for Intelligent Graphics Presentation, in Proceedings of Conference on Human Factors in Computing Systems (CHI'90), April 1990, Seattle, Washington, USA, pp. 193-200.
14. Simon H.: The Sciences of the Artificial, MIT Press, 1969.
15. Tweedie L.: Characterizing Interactive Externalizations, in Proceedings of Conference on Human Factors in Computing Systems (CHI'97), March 1997, Atlanta, Georgia, USA, pp. 375-382.
16. Wehrend S. and Lewis C.: A Problem-oriented Classification of Visualization Techniques, in Proceedings of the 1st IEEE Conference on Visualization, October 1990, San Francisco, California, USA, pp. 139-143.

An Interaction Interface for Multiple Agents on Shared 3D Display

Jeong-dan Choi[1] and Chi-Jeong Hwang[2]

[1] Telematics Solution Research Team, Telematics and RFID Research Department,
Electronic and Telecommunications Research Institute,
161 Gajeong-dong, Yuseong-gu, Dajeon, 305-350, Korea
jdchoi@etri.re.kr

[2] Image Processing Laboratory, Computer Engineering Department,
Chungnam National University,
220 Gung-dong, Yuseong-gu, Daejeon, 305-764, Korea
cjhwang@ipl.cnu.ac.kr

Abstract. In the multi-user VR game, wide stereoscopic display is an important component that supports immersion and perception to the gamer. Because the space in front of the display allows multi-participants to collaborate and guarantees view with the entire field of view for each user. This paper describes a method to integrate user interfaces for interactions among users wear VR devices on wide stereoscopic display. The key technologies of our proposed system include the problems of coordination with various devices to treat user events in real-time, synchronization of PC clusters which renders the VR contents, and seamless display of multi-channel display surfaces. To accomplish this, we propose a VR game display platform composed of multi-projectors, stereoscopic see-through head mounted device, and gesture recognition system for collaborative interactions. Our preliminary results showed that the proposed strategy well suits high resolution application at real-time frame rate for VR game.

1 Introduction

Many conventional VR arcade games have typically used monitors as their display devices. While monitors were cost-effective size but easy to use, they failed to convey a feeling of immersion to their users. We believe that multi-user VR game display requires a wide screen model which multi-user could collaborate in immersive environments. Nowadays there are many applications for wide screen using project-based multi-channel stereoscopic display, such as CAVE® [1], immersive desks [2], and their derivatives. These are often used as superior tools for data visualization and analysis of these very large amounts of data because of the benefit of wide FOV (Field Of View) [6]. So, in this research area, it is only necessary that simple communication device such as keyboard or mouse input command for navigation or passive manipulation of virtual environments. Moreover multiple users can use the same input command device as their interface to interact or collaborate, so that their command is

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 71 – 78, 2005.

controlled by remote distributed machine on via network. So it is allowed that multiple users use same device. But, if they share device in the same spot area such as wide screen, then respective multiple events must be treated by one single control machine. In this respect, we consider the issues of interface metaphor that multiple VR devices can serve the serve the purpose of interaction very well. As well as, there are many growths in the user interface. These interfaces must have a high degree of clarity and allow the interaction in the form of human-like behaviour through computer-aided tools. It is important to consider the goals of these entities, their possible actions, and the information available to them. One user's action provides pre-described strategies to the different players. We are interested in the construction of 3D stereo display allows the real-time interaction not only between a user and another but also between user and 3D object onto that.

In this paper, we propose a VR game display interfaces which delivers a sense of presence in the game environments through the collaboration and communication to users. Our proposed display platform ensures that coordinated interfaces achieve different from typical networked VR system preserving overall system performance. And also we introduce a preliminary example of VR safari game which is based on a projection-based multiple windows rendering system on passive 3D stereo display, synchronization and intensity-blending of multi-channel system for wide field of view on different PCs.

2 Our Proposed System Components

This section describes in detail our proposed VR Game system and contents, with some specific game devices. Our system overview is graphically shown in Fig. 1.

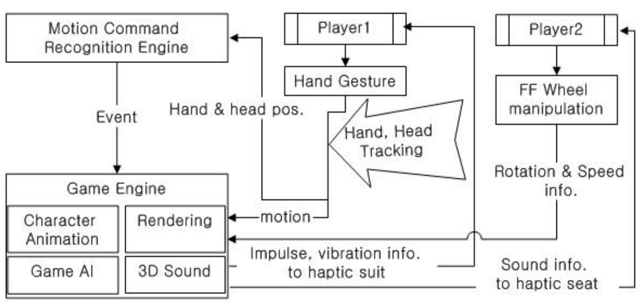


Fig. 1. Our proposed system overview

Target contents is a virtual world of real jungle hunting adventure and on-line role-playing game (RPG) supported by more than 2 persons. Main system supports surrounded display with PC clusters [3][4]. And it drives various VR interfaces with DirectX[9] APIs that is composed of DirectInput for haptic jacket, seat and force feedback wheel, DirectMusic for 3D stereo sound, DirectPlay for network services and Direct3D for rendering of X file format. One user plays game performing gestures with hand glove. The hand gesture is tracked by magnetic sensor, which is attached hand glove. And then the hand's orientation and positions are transmitted to Motion Recognition Engine. The other user drives the jungle off-road with force feedback steering wheel. The extracted game event is transferred to Rendering and Animation Engine. And the motion of actor is generated with virtual actor on screen in real-time. Player 1 wears SEE through HMD and player2 wears polarized glasses. Both players enjoy the jungle hunter with wide stereo screen. So, our topics of several 3D display interfaces and real-time interactions [8][11] are discussed in the order written.

3 Considerations of Interactive 3D Display Interface

Our system extends the concepts of guaranteed entertainment and predictable performance into multi-agent domains such as cooperating teams of VR Game. Agents will communicate to achieve overall team goals. While executing their allocated tasks, agents will respond to ongoing events in real-time. And triggering dynamic action tailored to the current context.

We describe our efforts in developing a 3D display for the improvement of collaborative abilities and maximization of entertainment. This section covers user interface flow, which includes input and output device information. And we describe some requirements for coordination of input and output devices.

3.1 Proposed Metaphor for Game Interface

We basically use a spatial common display for share communication interface. First, it is required that there be adequate wide for multi-agents. Second, multi-agents must wear some device to pass their intents through the display. And last we explore the notion of a agent for assisting users communication.

To solve these problems, we use a projector-based display system. This 3D stereo system includes LCD projectors with polarized filter and PC clustered rendering systems. The PC rendering system is clustered for synchronization. This clustered system has been developed last year [7]. And users incorporate 3D wide visualization capabilities with stereoscopic see-through HMD and polarized glasses.

Fig2 shows our event flows and interfaces. The beginning of game is started from the tracked information of director's hand. The tracking system traces the latest position and orientation of the hand gesture. The motion recognition engine interprets the motion as predefined game command. The motion is the metaphor for game events and communication with other collaborator. Game events include cast spells upon a game objects, change of attack plan, and movement sign and so on. Director's motion rendered into display is the method for game mission. So, the other collaborators understand the intention and effect. And also the motion's result is converted for output device into strength of sound and digitalized information for haptic suit and seat.

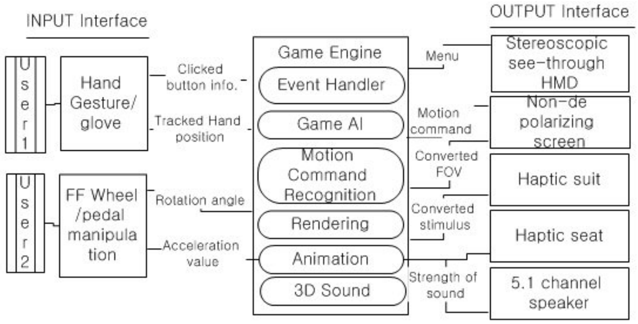


Fig. 2. An diagram of event flows and used interfaces

3.2 Shared Large-Sized Display Using Multi-channel

PCs, consisted of the Multi-Channel PC Cluster, are connected to their corresponding projectors. These projectors display their own portions of game display, and, thus, when they display simultaneously, a user can enjoy an immersive game display with a wide FOV. However, since the display surface for the projectors are shaped arbitrarily, the user must endure watching the distorted game display affected by the shape of the display surface. Furthermore, if there are cases where a display area for a projector is overlapped with another, these will cause an irregular game display in terms of its intensity. These problems can be fixed by using structured light technique and intensity blending. In addition, a laser tracker is used to track user's position to render the game content on the user's viewing point.

Geometry calibration. There are multiple projectors and screens which must be edge-aligned. We use a set of a reference plane; arbitrary display surface, CCD camera and projector pairs, and a laser tracker (see Fig. 3). Each camera and projector pair is used to extract a shape of its corresponding portion of display surface based on a Structured Light Technique. The same projector is also used to project a game content on the display surface.

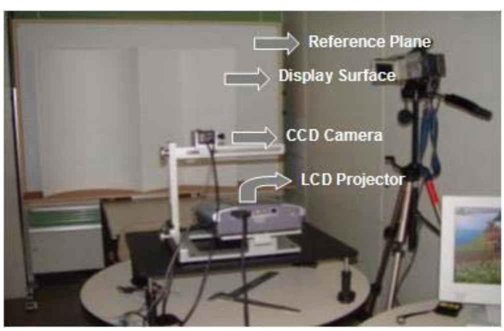


Fig. 3. The composed system for geometry calibration

Projector is used to display an interference pattern on its display area and camera is used to capture the pattern on the area [10]. The Structured Light Technique is a well-known two-wavelength phase shifting interferometer. The basic concept of phase-shifting method is to extract heights of the object surface from an interference pattern created by a phase-shift of interferometer. 4-frame phase-shifting method is normally used, and, from this method, we can acquire four interference patterns by shifting a phase of $\pi/2$ four times. The formula expresses the intensity of acquired interference pattern at (x, y).

$$I_p(x, y) = I_0(x, y)\{1 + \gamma(x, y)\cos[\phi(x, y) + \Delta]\} \tag{1}$$

$I_0(x, y)$ is an average moiré intensity at (x, y), $\gamma(x, y)$ is a standardized visibility of moiré interference pattern, Δ is a degree of phase-shifting, and $\phi(x, y)$ is a height of the surface at (x, y). By the 4 frame phase-shifting method, we can get following four different formulas from following formula.

$$\begin{aligned} I_1(x, y) &= I_0(x, y)\{1 + \gamma(x, y)\cos[\phi(x, y)]\}, \\ I_2(x, y) &= I_0(x, y)\{1 + \gamma(x, y)\cos[\phi(x, y) + \pi/2]\}, \\ I_3(x, y) &= I_0(x, y)\{1 + \gamma(x, y)\cos[\phi(x, y) + \pi]\}, \\ I_4(x, y) &= I_0(x, y)\{1 + \gamma(x, y)\cos[\phi(x, y) + 3\pi/2]\} \end{aligned} \tag{2}$$

Therefore, we can acquire the height map of the object surface from $\phi(x, y)$. However, since these heights are calculated from arctangents, there can be ambiguities of height in every 2π. And then two-pass rendering technique is added. (see Fig. 4).

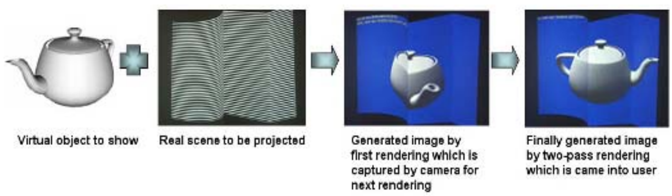


Fig. 4. Object rendering onto arbitrary display surface

Intensity blending. This multi-projector display does not rely on an assumption that each projector is dedicated to its unique part of the display surface. Thus, there may be an overlapped display surface area by projectors. When this overlap occurs, the intensity of the overlapped area becomes the sum of intensities of the projectors involved with the area. Thus, this situation produces an irregularity of brightness to the multi-projection display. Our method is to assign each vertex of display surfaces in overlapped area different alpha value depending on the distance to its nearest edge of the display surface. The vertices of an overlapped area in a display surface are same with its corresponding ones in the other display surfaces. The alpha value is expressed as (3).

$$\alpha_i(x, y) = \frac{d_i(x, y)}{\sum_i d_i(x, y)} \tag{3}$$

$\alpha_i(x, y)$ is an alpha value of a vertex at (x, y) which is in the display surface i. $d_i(x, y)$ is the distance between the vertex and its nearest edge of the display surface i. Fig. 5 shows a display without the intensity blending and one with it. Cleary one can see the effect of the intensity blending from the overlapping edge.

Fig. 5. The result scenes of intensity blending

3.3 Private Display for Collaboration and Personal Works

Apart from completeness usability, immersion and entertainment are the key quality aspects for VR game system. There are many situations in which we would like to interact with private objects in the surrounding public contents. An Augmented Reality can make this possible by presenting private information that enriches the public worlds. But there are only a few applications in augmented reality gaming and significant potential in this area. And also see-through HMD is used to overlay private information on the contents displaying on the wide screen. We use HMD to determine which weapon is selectable in nearby-space and which object is targeted in far-space. The interface presents two windows to the operator. The first window carries private information. We currently use a navigation and selection weapon menu. And the second window displays surround contents formed through the lenses.

3.4 User Interactions with Real-Time Virtual Character Animations

A user is displayed as a virtual character on shared display. We use a skinned mesh for the character that is stored as an X file, and it is rendered using the Direct3D. For the user's immersion into the game, whenever the user makes a command by gesture, movement of the arm of the virtual character is generated in real-time according to the sensed hand position of the user. For this purpose we implemented a real-time inverse kinematics (IK) routine [5].

Two kinds of IK solver were implemented. One is an analytic solution for 2-link IK problem and the other is for multi-links by using the inverse rate control. We assume the links are arms. Our 2-step IK solution for the 2-link case consists of finding the incremental transformations for the elbow and the shoulder. If an original transformation matrix is R and a new needed incremental transformation is ΔR, then a new transformation matrix becomes ΔR·R. The detailed 2-step algorithm is as follows. First, we calculate the desired elbow angle $\theta'2$ from L1, L2, the goal position, and the base joint position. L1 and L2 are the length of the upper arm and the lower arm, respectively. If we designate L3 as the distance from the base to the goal, θ′2 is given by

$$\theta_2^{'} = \cos^{-1}\left(\frac{L_1^2 + L_2^2 - L_3^2}{2L_1L_2}\right) \tag{4}$$

Δθ2 is (θ′2 - θ2), and the incremental transformation for the elbow, ΔR2, is a rotation by Δθ2 with respect to the axis of rotation that is perpendicular to the plane formed from the three joints. The second step is positioning the end-effecter to the goal position by adjusting the shoulder joint. This is similar to the first step and Δθ1 is calculated. The axis of rotation for ΔR1 is the line perpendicular to the plane formed from the base, the goal position, and the intermediate end-effecter position in the first step. Without additional calculations for exploiting the redundancy in the elbow joint, our algorithm produces a fast and natural-looking solution given an initial natural configuration. Initial configuration is restored after each execution of the algorithm to avoid losing the original posture and drifting with an unfeasible posture.

For the multi-link IK we applied the inverse rate control as in. This method uses the Jacobian matrix and numerical integrations. The process can be abstracted by following equations.

$$\boldsymbol{x} = f(\boldsymbol{\theta}),\ \Delta\boldsymbol{x} = J\Delta\boldsymbol{\theta} \tag{5}$$

$$\Delta\boldsymbol{\theta} = J^{+}\Delta\boldsymbol{x},\ J^{+} = J^{\mathrm{T}}(JJ^{\mathrm{T}})\text{-}1 \tag{6}$$

Where J is a Jacobian matrix and J+ is a pseudo inverse of J. In order to obtain a real-time result, the multi-link IK routine was formulated to solve an incremental problem i.e. a final configuration becomes the initial configuration of the IK problem at the next frame.

4 Experimental Results and Conclusions

We described the design and implementation of a real-time, collaborative, entertainment VR interfaces on 3D display system. We developed projection-based 3D stereo system, rendering using DirectX, intensity blending, geometry calibration of screen and user interactions using real-time virtual character animations.

We evaluate three major components: interactivesness, completeness of game mission, and realtime performance. First two of them are tested through the depth and size correction for stereo-scopic displays. The effective use of stereoscopic display systems requires the perception in stereopsis especially to represent the physical objects as the realistic virtual scene. And we measure the frame rate to achieve real-time performance(fig.6). The contents consist of objects which are non-playerable and playerable with average 30000 triangles and simulated by a physics engine. The results shows that the frame rates are 60 billion polygons per second and uniformable. So our proposed system is resonalble to be used VR game platform.

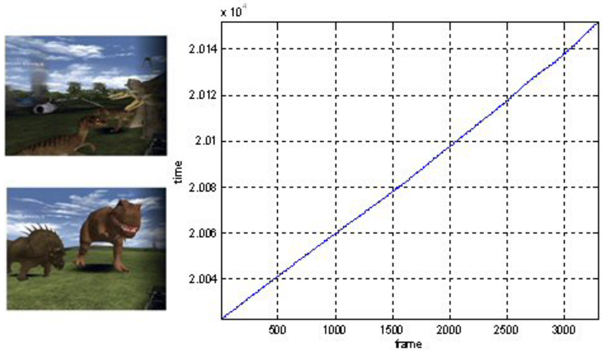


Fig. 6. Frame rates evaluation

Finally, our current research includes some futurer works on the wide interactive display such as the automatic calibraion of multiple projectors and color.

References

1. Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., and Hart, J. C., : The CAVE:audio visual experience automatic virtual environment. Communications of the ACM, Vol. 35, No. 6, (1992) 64-72
2. Funkhouser, F., and Li, K., : Large-Format Displays. IEEE CG&A, Vol. 20, No. 4, (2000) 20-21
3. Brian, W., Constantine, P., Vasily, L, : Scalable Rendering on PC Clusters. IEEE CG&A (2001) 62-70
4. Raskar, R. : Display Grid: Ad-hoc Clusters of Autonomous Projectors. SID04 (2004)
5. Tolani, D., Goswami, A. and Badler, N. I.. : Real-time Kinematics techniques for anthropomorphic limbs. Graphical Models. Vol. 62, No. 5. (2000) 353-388
6. Funkhouser, F. and Li, K.: Large-Format Displays. IEEE CG&A Vol. 20, No. 4. (2000) 20-21
7. Jeong-Dan, C., Ki-Jong B. et al. : Synchronization method for Real Time Surround Display using Clustered Systems. ACM Multimedia. (2002) 259-262
8. Choi K. H. : A Probabilistic Networks for Facial Feature Verification. ETRI Journal, Vol. 25, no.2, (2003) 140-143
9. Microsoft Corporation. DirectX 8.1 programming reference
10. Ming H. : Surface Effects on Alignment of Graphic and Real Objects in a Stereoscopic Augmented Reality Environment. CHI (2001) 441-412
11. Pattie M., Bruce B., Trevor D., and Alex P. : The Alive System: Wireless, Full-body Interaction with Autonomous Agents. ACM Multimedia Systems, No. 5. (1997) 105-112

Concurrency Control and Locking in Knowledge Base for Rapid Product Development

Dieter Roller, Marian Mešina, and Constanza Lampasona

Institute of Computer-aided Product Development Systems,
University of Stuttgart, Universitätsstr. 38, D- 70569 Stuttgart
{dieter.roller, marian.mesina,
constanza.lampasona}@informatik.uni-stuttgart.de

Abstract. Rapid Product Development (RPD) is a product-development technique that includes experts collaborating in a multi-disciplinary environment. Products consist of components. A component can be used in various products. These can be represented by a network structure (semantic network) consisting of vertices (objects) and edges (relations). Modifying a vertex can change other vertices within the network. In the Active Semantic Network (ASN) these modifications are propagated automatically. The RPD-Information is stored in the ASN knowledge base. Changes of referenced objects should be automatically taken into account within the ASN-database. Reciprocally, changes of an object attribute should lead to modifications in the referenced files. For the synchronization, object-lock mechanisms and access-control to the knowledge base are necessary. The concept presented here allows locking objects and also individual attributes. Object status and attributes allow regulating the access and the necessary steps for consistency re-establishment after system crash.

Keywords: Active semantic networks, locking, concurrent engineering, rapid prototype development (RPD), Knowledge basis consistence.

1 Introduction

This contribution was accomplished in the context of the SFB-374 research project, whose activities are focused on the development of methods and tools to reduce iteration cycles in product development. Active Semantic Networks (ASN) were developed for the improvement of product development practices [01]. One very important task was setting up a knowledge base containing and administrating cross-linking information and all development-relevant data.

The knowledge base most represents the current state of several parallel-running development-processes. Which objects can be processed in parallel must be clearly regulated. Finding and using synchronization techniques belong to the most important tasks.

2 ASN Knowledge Base

The ASN-knowledge-base consists of the ASN-database and a collection of files referenced in the database. Some ASN objects can be fully described in the database

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 79 – 85, 2005.

(e.g. a working group), others only partially by their attributes (e.g. a construction unit); their remaining information is contained in referenced files (e.g. CAD models).

The ASN-database is based on the URDM (Unified Relational Data Model [02], [03], [04] and [05]), a simple database design, which allows representing very complex databases with only a few tables without losing information.

The basic idea is having object-specific tables (RDBS) or classes (OODBS) with different data field/attributes and converting them into a new table/class named Attributes. This new table/class has at least three fields/attributes: Object_ID, Attribute and Attribute_Value. Other field included is Status(from…to) to establish the validity period.

All objects are collected in the table/class Objects, which at least has the attributes: Object_ID, Object_Type and Status(from…to).

The relations between objects in URDM are record sets of a table named Net that has the fields: Object1, Relation_Type, Object2 and Remark.

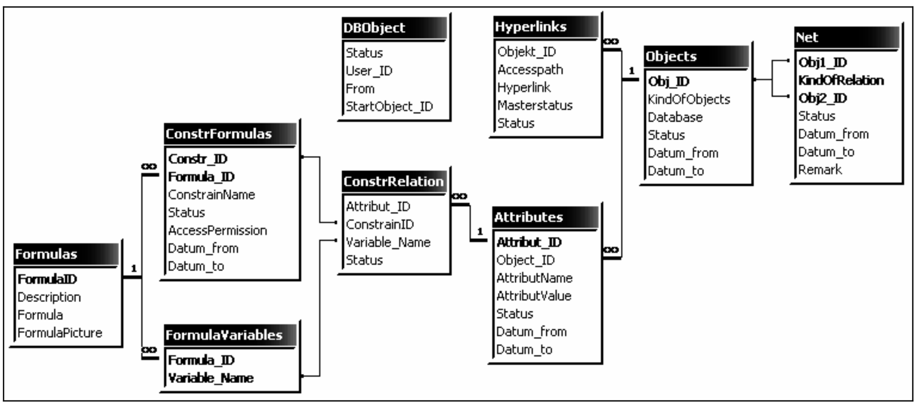


Fig. 1. Simplified presentation of the Unified Data Model

The URDM-design is presented in figure 1. The tables Objects, Attributes and Net would be sufficient for using the database.

The table Hyperlinks was added in order to gather the places where the referenced objects are stored. All copies of the referenced objects are included in the Hyperlinks table, but only one can have the Master status. The relation “A is a version of B” may be added using a name convention. Naturally, all referenced objects and their copies are saved where they are needed the most.

3 Transaction Security

The problem of transaction security is very complex and in this contribution only some important aspects are considered. For the transaction and data security a conception was elaborated, which differentiates between consistency states and access methods to the knowledge base.

3.1 Access Methods

Read Access 1: The read data (e.g. project costs) should stay unchanged some period of time. While performing this reading, all related dependent objects should be locked.

Read Access 2: This read mode corresponds to the usual Read (just reading attribute values, without durable consequences). With the read mode 2, objects dependent on the read object should not be locked.

Change Access: Object attributes or hyperlinks assigned to it are changed. All dependent objects must be locked.

3.2 Knowledge Base Consistency States

Consistency state 1: Data in the database is complete and consistent. Object attributes contained in the database fulfill all constraints conditions and correspond to the attribute values in the referenced files. Also the attributes values in referenced files correspond to the values in the database.

Consistency state 2: The ASN knowledge base is not in consistency state 1. However, the database contains the necessary information, in order to cause automatically or manually the consistency state 1 if necessary. Precisely here the status attributes play a central role.

Example: The attribute values of some objects were changed. The calculation of the attributes dependent on constraints can be accomplished e.g., by the solution of equation systems. The solution could take relatively much time and be seen as a long transaction. In order to reach consistency state 2, in a first step objects and object-attributes status are set, which show, which input sizes will be dealt with. Setting status attributes can be seen as a short transaction. A system crash during the solution of equations does not mean an information loss, since the status attributes are only put back after the complete execution of the computation procedures. The status attributes set before the system crash show which changes must be still accomplished. In normal operation, the knowledge base is nearly always in consistency state 2 and only in the "quiet phase", after the execution of all transactions, in consistency state 1. The ASN remains also after the system crash in the consistency state 2

Consistency state 3: The ASN knowledge base is neither in consistency 1 nor in consistency 2. However, the ASN users can bring the knowledge base into one of them. Consistency state 3 is a legitimate but unwanted state. Examples: Some referenced objects were deleted or changed without declaring it in the status attributes. Objects in the database were erroneously deleted and referenced files or dependent objects exist further. In case of a total loss of the ASN database it should be possible to continue working with different applications.

The transition from consistency state 3 to 1 or 2 must be accomplished partially manually by the system administrator. It is important that, during the ASN operation, consistency state 2 can be achieved through short transactions. The transitions can be activated with the help of the implementation of ECA (Event, Condition, Action) rules directly or with coordination agents.

3.3 Undo Function

The tables Objects, Attributes and Relations have defined validity period attributes. It is possible to achieve old database states using this attributes. This is an important aspect of the security philosophy.

3.4 Referenced Extern Data and Applications

Changes of referenced external files, e.g. CAD models, are accomplished by external applications (Catia, MS Excel). Such changes can be regarded a long transactions of the knowledge base, but not as transactions of the database. They do not fulfill according to ACID the atomicity criterion. Processed intermediate stages of work can be stored as versions. For each version of an external file an object with suitable attributes in the database can be provided and assigned. User-specific programs must be developed for the automatic adjustment of attribute values in the database and the external referenced files.

4 Lock Mechanisms and Attribute Status

In earlier ASN versions the locks could be set only on objects. In the current version it is possible locking several attributes and their status attributes.

4.1 Short and Long Transactions

Different changes in the ASN knowledge base take different times. Transactions can be roughly divided in 4 categories.

T1. Set/change of attribute values. Acceptance - attribute values are known before the transaction. It concerns very short transactions in the database, where ACID (Atomicity, Consistency, Isolation and Durability) conditions are always fulfilled.

T2. Determination of object dependence, marking of selected objects and attributes by setting status values. Determinations are based on the Constraints tables or on table Net. These transactions are also relatively short and the computing times lie in the seconds range.

T3. Dissolution of constraints. Constraints are mathematically very different and can lead generally complex nonlinear sets of equations and/or inequation systems. This can lead to computing times within the range of milliseconds up to minutes.

T4. Into this category belong the manual treatments of objects with many attributes, treatments of referenced files, and manual synchronization of object attribute values in the ASN database and in the referenced files. These transactions are often very long, the treatment of a complex CAD model can take up days, weeks or even months.

Since the marking procedures of the objects belong into category T2 and take relatively short time, they should initiated only by one user at a time. The marking accesses are regulated by the table DBObject. The status value busy means that a marking procedure was started. The remaining data-fields show who started the marking procedure and when and to which object the marking procedure was started. The status value free means that a new marking procedure can be started.

4.2 Status-Attribute Values

All objects on which object A depends are called "predecessors of the object A". All objects whose attribute values are dependent on object A are called "successors of the object A". Cyclic dependence (closed dependence loops) are possible.

Case 1: determination of status attribute values with read access 1 - reading of all attributed of object A is intended:
0 - Initial state, no lock;
1 - Object is read, all predecessors must be marked;
2 - Object is read, all predecessors are marked;
3 - Reading, the object A is being read;
4 - Object was read, predecessors must be unmarked.
With object state 1-4 all attributes of predecessors are automatically locked.

Case 2: determination of status attribute values with read access 1 - reading of only certain attributes of object A is intended:
0 - Initial state, no lock;
151 - Object is read, all predecessors and their relevant attributes must be marked;
152 - Object is read, all predecessors and their attributes are marked;
153 - Reading;
154 - Object was read, predecessors and relevant attributes can be unmarked.

Status values (marking) of selected attributes:

11 - Selected attribute is read;
12 - Objects and attributes, of which the selected attribute is dependent, are marked
13 - Reading, the value of the selected attribute is being read
14 - Value of selected attribute was read, predecessors' attributes can be unmarked.

The use of the status attribute values described in the case 2 increases the possibility for parallel treatment of ASN objects because not whole objects but only certain attributes become locked. However, in order to be able to use the attribute locks, dependences between individual object attribute values must be created. If this information is not available, it can only be worked with locks as those described in case 1. With the read access 2 no object or attribute locks must be set.

The cases 3 and 4 refer to change accesses in the ASN knowledge base.

Case 3: determination of status attribute values with change-accesses to object A. These change-accesses concern all attributes of object A and all attributes of dependent successors:
0 - Initial state, no locks;
61 - Object is being changed; all successors and referenced objects must be marked;
62 - Object is being changed, all successors are marked;
63 - Recording procedure, it is being worked on object A;
64 - Object has been changed, all successors must be updated
65 - All successors were updated, successors must be unmarked.

With status 61-64 all attributes znd referenced files are automatically locked.

Case 4: determination of status attribute values with change-accesses to object A That change-accesses concerns only certain attributes of Object A and attributes of successors:
0 - Initial state, no locks;
161 - Object is being changed; all successors and their relevant attributes must be marked;
162 - Object is being changed; all successors and their attributes are marked;
163 - Recording procedure;
164 - Object has been changed; all successors and referenced objects must be updated;
165 - Successors must be unmarked.

Status values (mark) of selected attributes:

21 - Attribute is being changed;
23 - Attribute is in treatment;
24 - Attribute value was changed.

The use of the status attribute values described in case 4 increases the possibility for parallel treatment of ASN objects because not whole objects but only certain object attributes become locked. However, in order to be able to use the attribute locks, dependences between individual object attribute values must be created. If this information is not available, it has to be worked must only with locks as those described in case 3.

Status Values in the Table Hyperlinks
0 - Initial state;
31 - Object to that the hyperlink file pertains is being changed; the hyperlinked file must be treated later;
32 - Object to that the hyperlink file pertains was changed; the hyperlinked file must be treated. When the treatment of the hyperlink file is done, the status attribute value is set to 0 again;
33 - Hyperlinked file is being changed; object corresponding to the hyperlink file and further dependent objects must be marked;
34 - Hyperlinked file is being changed; object corresponding to the hyperlink file and further dependent objects are marked;
35 - Hyperlink file was changed. The object corresponding to the hyperlink file and further dependent objects must be updated.

5 Summary

In this contribution examples were shown, on how the status attribute values of the objects and attributes can be set in order to reach consistency 2 in the knowledge base at any time. Even if the system would crash, status information is present for reaching the complete consistency of the knowledge base. The status values of an object and its predecessor (read-access 1) or its successors (change-access) change coordinately.

The coordinated changes of status value can be realized by ECA rules or by coordination agents. The system for the definition of the status values suggested here, can be still refined and further developed. The status values of objects have first soft locks. They warn the potential user, which would access information, which is would soon changed or that change access may be incompatible with other user's accesses. Nevertheless, if an access is to take place, arrangements between the users of the knowledge base must be first met.

The use of the status attributes offers support during the objective definition and locks when needed but using only as few as possible.

References

[1] Roller D.; Bihler; Eck O.: "ASN: Active, distributed Knowledge Base for Rapid Prototyping". In: Roller, D. (ed): Proceedings of 30th ISATA, Volume "Rapid Prototyping in the Automotive Industries & Laser Applications in the Automotive Industries", Automotive Automation Ltd., Croydon, England, 1997, pp. 253-262

[2] Brückner A.: US Patent 6.208.992 B1. March 27th, 2001.

[3] Schempp T.: ISIS - Informationssystem für historische Statistik, Universität Zürich 1998. Online: http://www.hist.unizh.ch/gs+edv/ss1998/databank/isis/isdmo3.html

[4] Mešina M.: ASN und die relationalen Datenbanksysteme (Realisierungsaspekte). Bericht Nr. 2002/03, Juli 2002. Institut für Informatik, Fakultät Informatik, Universität Stuttgart

[5] Roller D.; Mešina M.:Unified Relational Datamodel and Rapid Prototyp Development. In: Proceedings of the 14th International Conference "Process' Control 03", June. 8-11, 2003, Strbské Pleso, High Tatras, Slovakia,.ISBN 80-227-1902-1, pp. 63

[6] Mešina M., Roller D., Lampasona C.: Unified Relational Data Model and Data Retrieval for Rapid Prototype Development, TCME 2004 Lausanne, Schweiz, April 13.-17.

Collaborative Solution for Cooperation, Coordination and Knowledge Management in the Ceramic Tile Design Chain

Carlos Vila, Fernando Romero, Vanesa Galmés, and Ma Jesús Agost

School of Technology and Experimental Sciences,
Department of Technology, Universitat Jaume I,
Avd. Vicente Sos Baynat s/n, 12071, Castellón, Spain
vila@tec.uji.es
http://www.cinei.uji.es

Abstract. This paper presents a solution for cooperation between companies that belong to a supply chain in the ceramic tile sector, where competitiveness requires constant innovation in product design. Since it is becoming ever more frequent for these companies to design products and work collaboratively within the framework of the Extended Enterprise, a need arises for a methodology for the successful implementation of collaborative practices. The main contribution of this research work lies in the definition of a collaborative infrastructure for cooperative design that enhances interaction and knowledge management between different firms using information technology tools and, particularly, PLM solutions. The methodology used to achieve an efficient implementation is based on process modeling. Design processes are not always structured; there are continuous changes due to engineering modifications, and this fact makes modeling harder. Consequently, the definition of collaborative processes will allow a deeper understanding of product development activities.

Keywords: Virtual Enterprise, Collaborative Design, Product Lifecycle Management Solutions, Ceramic Tile Cluster.

1 Introduction

If the Extended Enterprise is to be efficient, it must have a good supply chain management, which requires improvements in the planning and management of interrelated complex systems in order to increase productivity.

Although the concept of supply chain was developed a little more than two decades ago, we could say that researchers are recently paying attention to solve the problems of virtual organizations [1]. This involves a terminology that was incorporated in order to define an advanced logistic technique that focused on the integral control of company flows [2].

Over the last few years, the level of integration actually involved in the definition of Supply Chain Management is being reviewed and improved. Nowadays, other concepts such as Synchronized Supply Chain, Supply Chain e-collaboration or Extended Concurrent Engineering (3D-CE; Product/Process/Supply Chain) have been incorporated into the cultural heritage of this area to greatly enhance these limitations.

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 86 – 93, 2005.

In this new era of globalization it is becoming more frequent for companies to design products and to work collaboratively within the framework of the Extended Enterprise (EE), since this enables them to reduce the number of operations and processes that add no value and thus achieve better and more robust products through supply chain management.

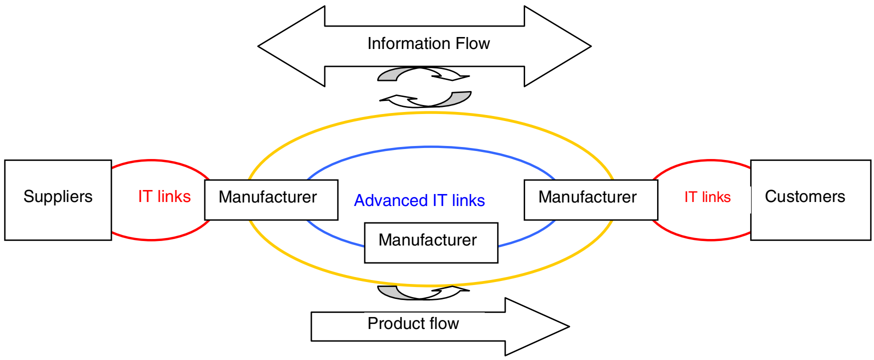


Fig. 1. Extended Enterprise information flow [3]

As we can see, Information and Communication Technologies (ICT) have allowed the integration of several activities within the company and with the enhanced capabilities of the Internet, the initial ideas of extended and virtual organizations can finally be deployed (Fig. 1).

However, from the technological perspective, it cannot be said that the solution could only come from the implementation of computer support tools. It is clearly recognized that Information and Communication Technologies, and more particularly those based on the web, can be of great help in integration and collaboration and, moreover, that the concept of collaborative engineering and cooperation is closely linked to the use of web services (e-collaboration). Yet, the real situation of many implementations reminds us that integration and collaboration cannot be achieved between the different agents involved by simply adopting the required technology.

Therefore, there is a need not only for a web-based tool for collaboration (CSCW or PLM) but also a framework that allows for the integration and coordination of product life cycle processes and the exchange of information between agents (expert applications or individuals).

2 Collaboration Framework for Supply Chain Models

Although the idea of the Extended Enterprise linked to the supply chain is relatively new, a small number of firms have begun to implement its principles. Competitive pressures, an instinct for survival, and the realization that their traditional business model was likely to fail, typically have driven these firms to adopt this new organizational model. For example, the Airbus consortium of four of Europe's aerospace companies joined efforts and fortunes to create an airplane, the Airbus, which was a response to Europe's shrinking share of the commercial airline business. Attempts are now being made to implement this initiative in other sectors that seek collaboration with new technologies in their product development process.

We could say that the Extended Enterprise is the entire set of collaborating enterprises, both upstream and downstream, from raw material to end-use consumers that work together to bring values to the market. The advantages of the Extended Enterprise derive from an enterprise's ability to quickly utilize the entire network of suppliers, vendors, buyers, and customers. The flows of information that lie at the core of the coordination and collaboration among network members not only link disparate information sources, they also provide an opportunity to build knowledge-based tools. Enterprises engage in longer term partnering relationships built around mutual goals and accompanied by a very rich and deep exchange of information.

However, we can find several network topologies that can configure an extended Enterprise or combination of them (Fig 2).

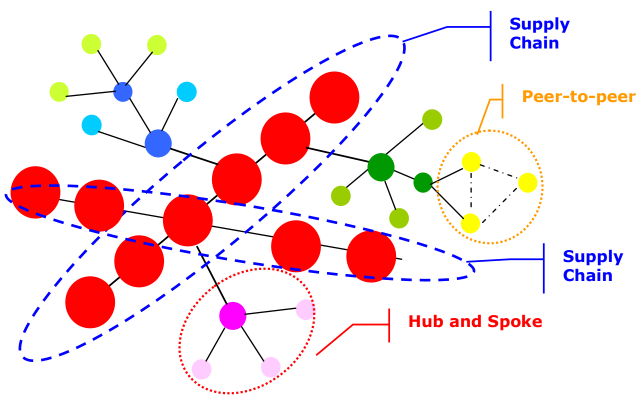


Fig. 2. Different types of configuration for an Extended Enterprise

Success or failure is now a function of the collective performance of the enterprise and not individual actions. Companies must participate in effective chains, linked by trust and a belief that one partner will not act opportunistically at the other's expense.

But the success of e-collaboration, mainly in the product development process, is not possible merely by investing in technological solutions. Suitable implementation of these solutions is also required and the enterprise must be correctly prepared for this new project; such preparation will be the first step of the implementation stage. First, in this preparatory stage, it will be necessary to have a global knowledge of the situation of the enterprise, as well as how and to what extent it will be affected.

The Collaborative Engineering (CoE) concept confirms the same principles as Concurrent Engineering but on a wider scope. The concurrent development of products and their processes remains the core of the philosophy. However, new concepts such as coordinated decision-making (related to product and project) and information technology have been added.

In this way, the core of Collaborative Engineering is the workgroup, usually geographically dispersed, working in an Extended/Virtual Enterprise context. As in

Concurrent Engineering, a workgroup implies people from different fields, departments and even enterprises working together on the development of the product and throughout its life cycle. Concurrent Engineering was not focused on how this collaboration and cooperation among the team members takes place. However, Collaborative Engineering does focus on this issue. The new enterprise models of Extended and Virtual Enterprise imply a new way of working, where the team members are geographically dispersed and collaboration must fit the inter-enterprise contracts [4].

This implies the need for a cooperation framework for Information and Communication Technologies which can support not only cooperation and integration among the members but also good process and project management and knowledge management [5]. With this aim, a research project is being developed in order to apply these ideas regarding the design process within a ceramic tile supply chain [6], where virtual enterprises can be created supported by a cluster with a collaborative infrastructure.

3 Research Approach

Spain is the second most important ceramic tile manufacturer and exporter in the world. Within the ceramic tile sector 75% of the supply chain's stakeholders are gathered in the area of Castellón, which accounts for 90% of the Spanish production. This success is, in part, due to the informal relations among the different members of the supply chain. These relations constitute the beginning of a structured collaboration called "*collaborative ceramic cluster*".

From this predominant position, the Castellón ceramic tile sector will have to face new challenges in the new Product Introduction and Development environment. These companies are aware of the difficulties; such as the high rate of references that forces the sector to manage more than 6.000 references at the same time, and the launch of new products comes up against problems since companies are bound by fair timing.

For these reasons it is extremely important to invest in new product development and innovation processes to improve the quality of ceramic tiles. Moreover, ceramic tile enterprises with a high capacity for organizational learning and design management will be the most competitive and profitable enterprises.

To help enterprises to face these challenges, the *Industrial Enterprise Innovation Center* (CINEI) is developing a research project (CE-Tile) that promotes business collaboration and resolves information and activity management problems in new product design and introduction processes.

Therefore, the ultimate aim of the project is to solve problems in the management of the information associated to the development and introduction processes of ceramic products by creating a process model using Information and Communication Technologies, together with the Internet, to set up a network that enables companies to collaborate and to share knowledge.

The experience of the sector in the implementation of ERP systems (for shipping, invoicing and production) using applications for large business management solutions, like SAP, has made us see the need to have some models adapted to the typology and specific characteristics of the products and processes of the sector.

One specific objective that stands out is the development of a pilot project, which is based on a commercial software solution (SAP PLM), and allows the integration of diverse sub-systems such as creative and graphic design, product development and so on, carried out by the enterprise or by others in the supply chain.

The pilot project is being developed supported by a methodology that reproduces life-cycle phases of the Extended Enterprise with ICT tools to build up the Virtual Enterprise (VE). This VE is created with the objective of designing and developing a floor tile series/model. The enterprises that belong to the sector supply chain and are involved in the development of this model form the VE [7]. With the aim of improving cooperation, concurrency, information and knowledge exchange and reducing time to market, the pilot project is developed in four phases:

1. Creation of the basic infrastructure of the collaborative cluster and the foundations of the VE.
2. Creation of the VE and configuration of the operating infrastructure.
3. Execution of the experience of series/model design and development.
4. Analysis of the results from the pilot project. Implementation planning in manufacturing.

The solution reached, and the experience gained during the first two phases, have shown the need for an implementation team consisting not only of the supply chain firms themselves and the project leader (CINEI) but also the support of a consulting enterprise and a software developer company. The real problems arose when finding a general framework for collaboration. Therefore, a general solution was developed in order to make it feasible to extend it to other supply chains.

4 Findings

The first studies on product development processes have illustrated that, traditionally, there is a certain amount of collaboration among all the actors in the design chain. This collaboration is interdisciplinary and takes place among commercial technicians from the enamels and colorings industries, graphic designers, production and product development technicians from the ceramic tile industries, as well as mold manufacturers. Although this collaboration has not been accomplished to any large extent, it is a consequence of the competitiveness that exists within the sector, which is expected to be really fierce. Moreover, it has also been detected that capital assets are sometimes crossed between companies involved in the logistic chain and, more particularly, in the design chain.

Therefore, for the first phase, a basic infrastructure of the collaborative cluster and a foundation for the VE was set up and a model of the current state of product design and development process was produced, so that a diagnosis could help to propose the final collaborative solution.

After the analysis, a descriptive model of cooperation was developed with IDEF0 methodology in order to describe the activities that transmit and transform the information (Fig. 3). It shows how people should act or work together for the particular purpose of designing ceramic tiles and to help other partners willingly when it is requested. It also recommends the collaborative project management activities that coordinate timetable and product related information so decisions can be more efficient. This model goes beyond the idea of a process model in that it includes strategic level and medium level processes and reflects cooperation and coordination activities.

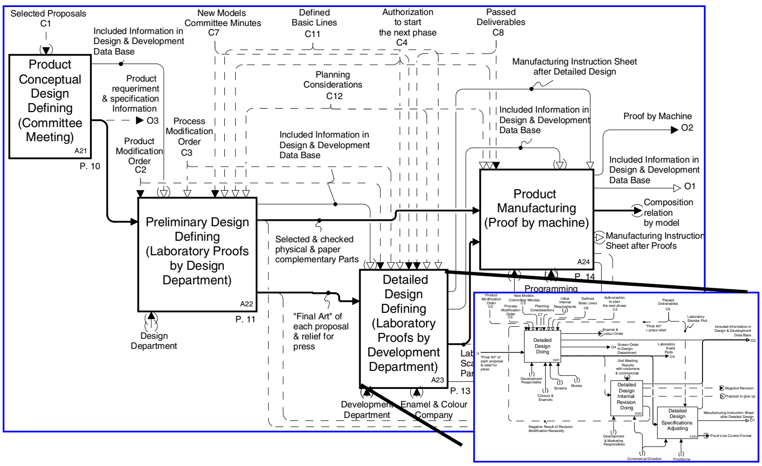


Fig. 3. Process model for Collaborative Design

A second model (operational level) was carried out to set up the different interactions among the members of the VE collaborative team. The model helped to clarify the workflow process as well as the communication channels to transmit information among the different technical activities using an ICT tool (Fig. 4).

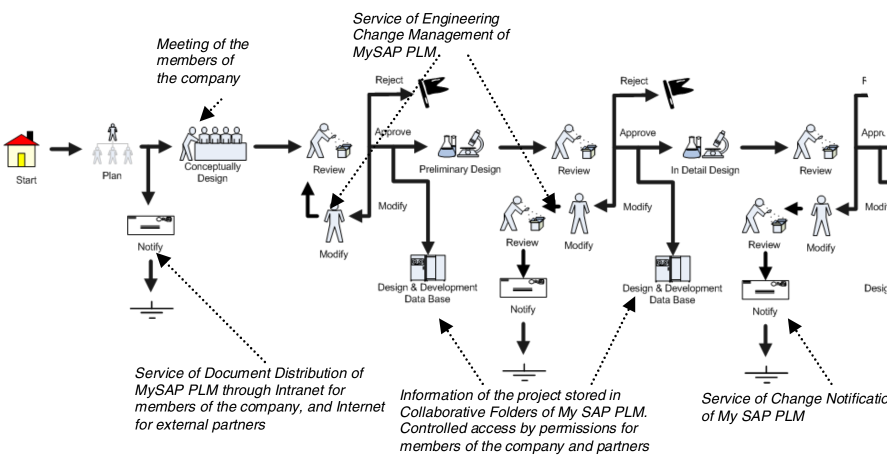


Fig. 4. Event Flow Model

Finally, a third model was created in order to capture the knowledge generated during the collaborative design process (Fig. 5). It helped in the selection of processes and functionalities of the PLM tool and the customization required by the models.

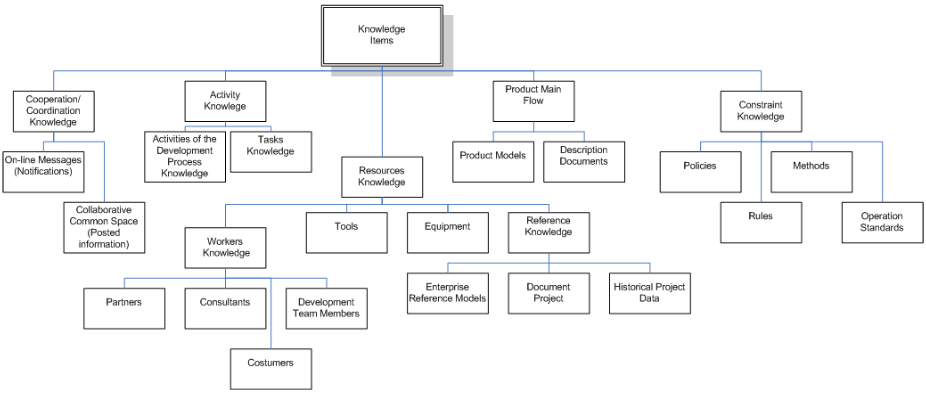


Fig. 5. Knowledge Model

During the second phase the operating infrastructure for the pilot project was determined and the business processes that had to be improved were identified. A number of supply chain companies were selected for the pilot project (third phase) and the bases of the VE agreements were established. As a result, the configuration of the operative infrastructure of the Design Project allowed the definition of the Ceramic Collaborative Cluster (Fig. 6).

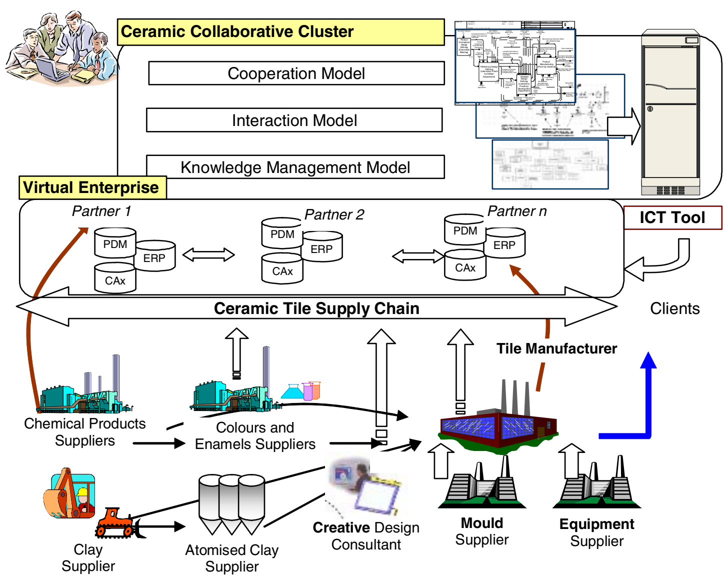


Fig. 6. Collaborative solution and infrastructures for the Ceramic Tile VE

5 Conclusions

The main contribution of this research work lies in the definition of the minimum descriptive models required to build up a collaboration infrastructure. These models determine the needs and requirements that must be demanded for the implementation of collaborative solutions with ICT tools – in this particular experience with a Product Lifecycle Management application. The development of these descriptive models is important to researchers because, on the one hand, the literature on this subject is scarce and, on the other, they provide us with a deeper understanding of the activities associated to product development by defining the collaboration processes (coordination, cooperation and knowledge management) that take place.

Acknowledgement

CE-TILE project, "Collaborative Product Engineering Development of an Information Model and an Implementation Methodology for the Supply Chain in the Ceramic Tile Sector", is partially sponsored by the Spanish National Program for Design and Industrial Production (DPI2002-02141). We are also grateful to the following companies that belong to the ceramic supply chain Tau, Macer, Torrecid and Kersa, and also for the support of SAP and CAPGEMINI Consulting to this research project.

References

1. Stephens, S.: Supply Chain Operations Reference Model version 5.0: A new tool to improve supply chain efficiency and achieve best practice. Information Systems Frontiers, Vol. 3, 2001, p. 471-476.
2. Lummus, R. R. & Vokurka, R. J.: Defining supply chain management: a historical perspective and practical guidelines. Industrial Management & Data Systems, Vol. 99, 1999, p. 11-17.
3. Davidow, W. H., and M. S. Malone, 1992. The Virtual Corporation, HarperCollins, New York.
4. Perrin, O., Godart, C. (2004). A model to support collaborative work in virtual enterprises. Data & Knowledge Engineering 50, 63-86.
5. Ho, Cheng-Ter; Chen, Yuh-Min; Chen, Yuh-Jen; Wang, Chin-Bin (2004). Developing a distributed knowledge model for knowledge management in collaborative development and implementation of an enterprise system. Robotics and Computer-Integrated Manufacturing 20 439-456.
6. Muckstadt, J. A., Murray, D. H., Rappold, J. A., & Collins, D. E.: Guidelines for collaborative supply chain system design and operation. Information Systems Frontiers, Vol. 3, 2001, p. 427-453.
7. Camarinha-Matos, L.M., Afsarmanesh, H., Rabelo, J.R. (2003). Infrastructure developments for agile virtual enterprises. Int. J. Computer Integrated Manufacturing, vol. 16, nº 4-5, 235-254.

GCE: Knowledge Management Applied in a Design Reengineering Process

Jonice Oliveira[1], Jano Moreira de Souza[1,2], Mauricio Lima[1] , and Ricardo Farias[1]

[1] COPPE/UFRJ - Computer Science Department - Graduate School and Engineering Research, Federal University of Rio de Janeiro, Brazil
[2] IM/UFRJ - Institute of Mathematics – Federal University of Rio de Janeiro
{jonice, jano, mauriciolima}@cos.ufrj.br
rfarias@lcg.ufrj.br

Abstract. In order to reach competitive advantage, manufacturing firms need not only to coordinate business and cross-border activities, but must also integrate and improve internal processes, coordinate access to internal and external knowledge, exploit and develop their employees' capabilities. In a manufacturing context, there are several levels (equipments ↔ processes ↔ products ↔ market) which are related, and the competitive advantage and success in the market depends on the synergy of these levels. In all phases, we have people and specific information associated, and the handling of specific information by people, in a context, together with a usage experience, comprises knowledge creation. In this work, we describe a case at a Brazilian Oil Company. This company was losing knowledge on its first level of the manufacturing chain. In this paper, we will describe this problem, as well as the procedures to solve this problem, and the Knowledge Management environment created to support the process.

1 Introduction

In their search for competitive advantage, manufacturing firms need not only to coordinate business and cross-boarder activities, but must also integrate and improve internal process, coordinate access to internal and external knowledge, exploit and develop their employees' capabilities. These internal processes are totally related to design operations.

Penrose, in [1], argued that a firm is not a homogeneous unit, but rather a "collection of resources", in which the heterogeneity of the services contained in its resources makes a fundamental difference to the strategic goal of the firm. As shown in Fig. 1, in a manufacturing context, there are several related levels, and the competitive advantage and success in the market depends on the synergy of these levels. In all phases, we have people and specific information associated, and the handling of specific information by people, in a context, together with a usage experience, comprises knowledge creation.

At a first level, we have the equipment used by operators and handled by technicians. Equipment is an important part of the manufacturing process because it comprises the basis for the transformation process and product creation. Their non-operation or sub-operation compromises the entire process.

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 94 – 102, 2005.

At a second level, we have the processes, which can conceive, design or handle the placing in the market of products. Specific activities of a process need to be executed by people with particular expertise or competence.

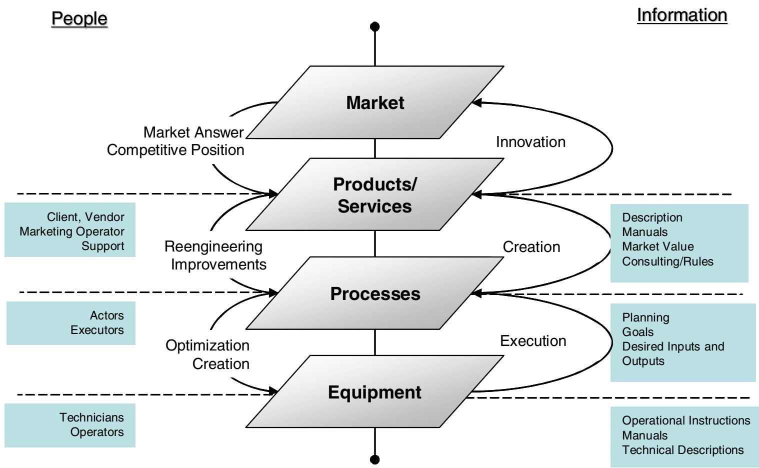


Fig. 1. Related levels and their relationships in a manufacturing chain

The result of a process is a product (third level). This product, or the services related to it (as support, consulting or technical support), will be launched onto the market and at the last level lies the market position, as the consumers' acceptance or rejection of the new product and the competitive reactions. The latter level purports wide horizons and can be considered as strategic. There are linkages between the four levels, as operational decisions trigger improvement activities, and improved activities augment the strategic decision space.

In all stages, we have knowledge creation (which should be managed) and a good performance of a stage brings forth significant contributions to the other stages. Looking at the first level: **equipment**. Knowledge about equipment allows for the choice of the right equipment to be used in a production process. The appointment of persons who can handle equipment with success guarantees fast execution. In general, operating standards - described in manuals, technical descriptions and operational instructions - communicate sets of codified rules to the shop-floor personnel and are used to guarantee the control of the manufacturing process.

Processes are a set of activities executed by people with different competences. In a manufacturing process, a team is usually multidisciplinary, and the process model and the activity specifications comprise a kind of knowledge. The documentation and codification process is a facilitated, but participative, process which generates knowledge by extracting tacit knowledge from those involved [2]. Depending on the human resource practices, this process may be iterative, continuously constructing new knowledge, permitting, as an example, that well-trained and experienced operators may collaboratively find better ways of doing the work. These constitute improvement activities (in the **processes** stage), which usually result in updating the related documentation, or solely in establishing new practices. In addition, improvement activities may be more formally organized as planned meetings with structured agendas which use data from performance measurement systems and personal experiences to improve operations and their associated documentation. Well defined and properly executed processes allow **products** to reach the **market** faster, as well as the creation of better and innovative products.

Creativity and innovation of products based on knowledge is the core of competition in global manufacture industry at the beginning of the 21st century [3]. Clients' opinions, competitive positions and product information itself in the form of drawings, recipes, nominal capacities, and others, is translated into process data under the form of a reengineering process and improvements. Consequently, this forces the creation of new equipment or improvement of the equipment used to lead faster and better processes.

It has been shown that the knowledge management practices within the process development function and across the product-process interface significantly influence the performance of the process development process [4]. The entire chain (equipment ↔ processes ↔ products ↔ market) has to work as a gear: in unison and continuously.

To exemplify these questions, we describe in this paper a case at a Brazilian Oil Company. This company was losing knowledge on its first level of the manufacturing chain: it was "forgetting" about its equipment. This then reflected on the entire production. Firstly, in section 2, we will describe the problem. Next, we describe the procedures to solve this problem, in section 3, as well as, the Knowledge Management environment created to support the process (section 4). Finally, in section 5, we comment on the conclusions and our future works.

2 The Problem

The problem faced by the cited enterprise is typically a problem of "knowledge loss". This company owns several pieces of equipment from finding and extracting petroleum to the manufacturing of byproducts. At the refinery, several pieces of equipment are used but, in due time, a great deal technical information about these was lost. Knowledge about handling the equipment, its types, applications, purposes and operation, was only in the minds of more experienced employees. That is, if a more experienced employee, holder of critical knowledge about important equipment, is dismissed, retired or transferred to another section, he/she takes this knowledge with him/her. That is, the company loses its knowledge.

Another problem, which we deem of greater complexity, is that the company does not have the ability of knowing what it actually knows, in other words, the competences mastered by organization, thereby bringing forth greater difficulties to discover specialists in different kinds of equipment. The attempts of solutions are described in section 3 and section 4.

3 The Equipment Reengineering Process

This process purports some steps, as follows:

1. A person responsible for the sector defines a set of more important equipment, which will be documented.
2. A piece of equipment from this set is chosen.
3. An operator or technician, who knows about the equipment and its functions, is sought within the institution.
4. The equipment is photographed.
   4.1. If possible, the equipment is opened to be photographed inside.
   4.2. If not, the equipment is photographed only from the outside.
5. The operator or technician, who was chosen in step 3, is interviewed about the equipment. All information is documented.
6. Information about the equipment is searched in external sources (books, magazines, internet, etc.)
7. A designer analyzes the photos and all information obtained about the equipment and drafts some sketches, as shown in Fig 3.
8. A number of experts in graphical computing, with the sketches and all information about the piece of equipment, perform the 3D modeling of the equipment, as shown in Fig 2.
9. A simulation of the equipment in operation/running is made, and recorded in a video archive format.
10. A number of experts validate the simulation.
    10.1. If not correct, return to step 8.
    10.2. If it is correct, all information (information obtained, photos, sketches and simulation) are available to any employee, in an organized and categorized way, in organization and the process returns to step 2.

As the company does not have a specific way to document and search by competences, finding a person with a specific expertise is heavy work. The search is based on a sequence of interviews, attempting to identify a certain person to participate in the process.

The idea is to use these simulations as sources of learning, permitting that a inexperienced workers learn as an equipment works.

Fig. 2. A part of Reducer Pump (P-6401) modeling

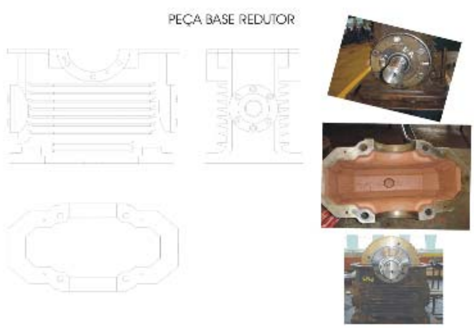


Fig. 3. Reducer Pump (P-6401) sketches and photos

4 Knowledge Management Environment

An environment was envisioned to provide all support to the Equipment Reengineering Process, as well as to provide all tools to help putting in place of knowledge in the company. This environment is titled GCE, the acronym of "Gestão de Conhecimento em Equipamentos" (Equipment Knowledge Management). The GCE architecture, as shown in Fig.4, is detailed in the following sections.

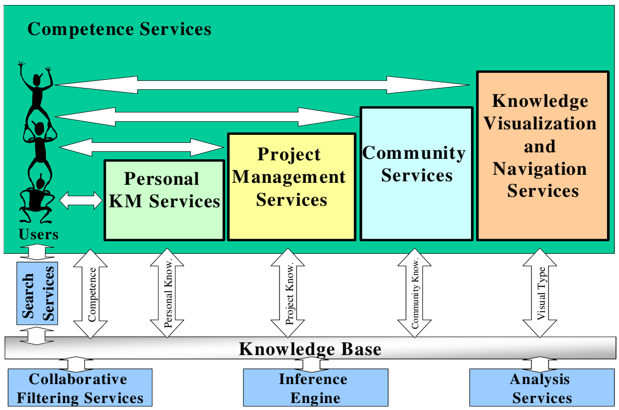


Fig. 4. GCE Architecture

4.1 Personal KM Services

This module is responsible for providing functions for an employee to manage his/her personal knowledge, as well as information about him/her. This module provides services such as:

- Curriculum Vitae – The curriculum vitae is one way of keeping information about a person. This information comprises the name, personal information such as address, e-mail, home page and phone, academic background, professional activities, language skills, scientific production, advisory and prices. In addition to this information, the user should state in what his/her competences and his/her degrees of expertise are.

- Personal Blog – it acts as a tool to provide personal knowledge management. and reflects some topics about the user's profile. In the GCE, by combining the output of several chosen weblogs, you obtain a tailor-made publication and contact with people with the same interest as yours. Moreover, the evolution of employee's interest and involvement about a topic is registered, as a chronological record of his/her thoughts, knowledge, references and other notes that could otherwise be lost or disorganized

- Mental Maps - In the GCE, the user can construct mental maps to define concepts, help in brainstorm sessions and simplify the job of discussion graphically. Users can display some concepts as public, so that anyone can see the concept, its relationship with other public concepts, and its definition.

4.2 Project Management Services

This module is responsible for providing services to: i) define and execute a project by a workflow tool, ii) store the knowledge created during project execution, and iii) permit knowledge reuse.

- Define and Execute a Project – The project manager or responsible employee creates the project model, with the activity sequence and the needs for the execution of each activity: competences necessary for execution, inputs, outputs and tools. The project-model creation is defined in a workflow graphic mode, and process execution control is performed by a workflow engine.

- Knowledge Storage during the Project – During activity execution, the involved parties can add all kinds of explicit knowledge to this activity, such as reports, documents, manuals, comments, suggestions, better practices, mistakes and ideas. Then all steps in the activity execution can be documented, and in the future we can track all kinds of created knowledge and the context in which it happens, an important factor for an inexperienced employee to learn about the process in a domain. The process model itself is a kind of knowledge which should be stored for further access.

- Knowledge Reuse – Previous process-models and the knowledge described in their instance activities can be reused in full or in part by other users.

4.3 Community Services

One of the main focci in the GCE is virtual-community creation, created by groups of people with a common interest, who could exchange information and work together. This module displays a number of tools to improve the interaction between people in a community, as such survey, forum, news, scheduled e-meetings, public library and events. Also, all conversation in an e-meeting is stored and categorized in one or more Knowledge Areas.

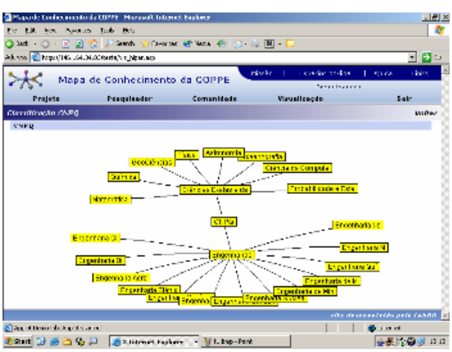


Fig. 5. Hyperbolic Tree with a Knowledge Areas Classification

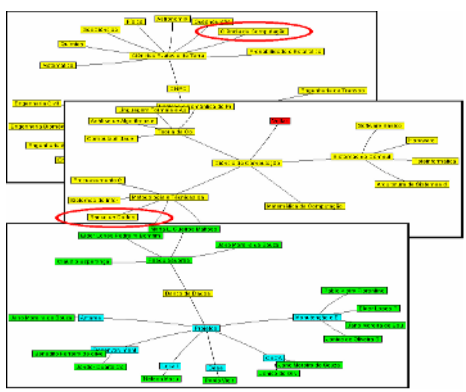


Fig. 6. Navigation in Hyperbolic Tree

4.4 Knowledge Visualization and Navigation Services

It is composed by:

- Hyperbolic Tree –In the GCE, as shown in Fig. 5 and Fig. 6, the hyperbolic tree is used to visualize the Equipment Ontology. The user can navigate by this ontology and by all information related to a knowledge area, as projects (in green), competences, people (in blue) and communities.

- Conceptual Project Map – permits a tree-like visualization of a specific project. Thus the user can see all related information of a project, such as materials, contributions, participants, collaborators, managers, knowledge areas and competences which are pre-requisites and the workflow model. This map permits searching over it, and opening a matching node.

4.5 Competence Services

Currently, the main problem in an organization is the inability to discover what it knows, what are its abilities are, what its employees and experts know and in which knowledge areas they work. A way of reasoning the competences of a person is by analyzing the documents that this person creates, edits and manipulates, and the frequency of these operations. These documents can be texts (as publications or notes), mental maps, project definitions, e-mails, blogs, and others. The services provided are:

- SMiner – Fundamentally, the SMiner function is to mine competencies based on texts/publications. In our approach, we use the Stemming technique to measure the relevance of a term by removing suffixes in an automatic operation. Terms are related to a person's competence, and these competences and the expertise degree (using the measure of relevance of a term) are stored in the database. Finally, after the mining, it is possible to check the mapped abilities in a report provided by the application.

- Competence Searcher - Moreover, as important as identifying competences is searching for them and inferring the most similar competences when we do not find them. Thus, in our approach, the competences are searched following this order of priority: **i)Declared competences** – the competences that the person thinks she/he has. These abilities are recorded by the person herself and saved; **ii)Project competences** – correspond to the competences found in the employees' developed projects. We assume that if a person has worked in a project and executed an activity which requires some competence to be executed, then, this person has this competence; **iii)Extracted competences** – recovered from published text mining by Sminer, described in the previous section; **iv)Community competences** – collected from the communities in which the user participates or contributes. It means that issues discussed in communities can be understood by their members.

The result of this search is an ordered list with these criteria.

- Weakness and Strengths - In this work, we use the mapped competences of an institution to measure the weakness and strengths. Knowing weak and strong points, that is, knowledge areas where the institution has a good representativeness and competences that should be developed further, respectively, the institution can be better positioned, developing planning and strategies to continue or to improve its current position, enabling continuous knowledge dissemination, and increasing internal interaction and collaboration.

- People Recommendation for a Project and Community Recommendation. The results of the competence search can be used as a support tool in decision-making when project managers are choosing employees. Also, after the employee's competence identification, our approach does the following: i) if there are no any communities about a topic, the environment searches for a number of people with similar interests and proposes new community creation or ii) suggests existing communities that match in the profile of the future member. Depending on the topic, a person can belong to more than one community.

4.6 Collaborative Filtering Services

In this service, documents, people, mental models, community and process models are recommended to a new user based on the stated preferences of other, similar users and communities. We use a Collaborative Filtering Spider which collects, from the Web, lists of semantically related documents and assembled in communities.

4.7 Inference Engine

Reasoning about profiling is made in the “Competence Services”. This module is responsible for extending these functionalities in order to infer cases or solutions (Case Based Reasoning), and reusable content.

4.8 Analysis Services

Some tools are used for observing knowledge evolution, the evolution of communities and comparative analysis by employees, departments and organizations.

All GCE modules are implemented and in use, with the exception of the Collaborative Filtering Services, Inference Engine, Analyses Services modules, which are in the development phase.

5 Conclusion and Future Works

Our work is based on an actual equipment knowledge loss problem in a manufacturing company, a Brazilian Oil Company. Initially, we discuss about a common manufacturing process and the knowledge exchange in it, we describe the problem and the solution. This environment, called GCE, envisions to aim not only the Equipment Reengineering Process, but also stimulates collaboration among people, knowledge flow in the organization, knowledge creation and dissemination.

One future work comprises the use of details of the employees’ courses done in the organization, as well as his/hers certifications, to identify competences. Currently, competences are inferred using declared positions, information about projects, publications and communities.

Presently, the GCE is a centralized-base environment. As future work, we envision extending it to a distributed scenario. Moreover, we will measure how much the GCE is aiding people and the organization to manage their knowledge and disseminate it. Currently, all the results are subjective.

References

[1] Penrose, E.T., The theory of the growth of the firm. Wiley: New York, 1959.
[2] Benezech, Daniele, Lambert, Gilles, Lanoux, Blandine, and Loos-Baroin, Jocelyne (2001): Completion of Knowledge Codification: an Illustration through the ISO 9000 Standards Implementation Process, Research Policy, 30, 1395-1407.
[3] Yunhe Pan, Shouqian Sun, Enwei Bao, "Developing Situation and Tend of CAD Technology ", Journal of Computer Aided Design & Computer Graphics. , 1999, 11(3), pp. 248-252
[4] Pisano, Gary P. (1997): The Development Factory: Unlocking the Potential of Process Innovation, Harvard Business School Press, Boston, MA.

A Constraint Maintenance Strategy and Applications in Real-Time Collaborative Environments

Kai Lin, David Chen, Chengzheng Sun, and Geoff Dromey

School of Information and Communication Technology,
Griffith University, Brisbane, QLD 4111, Australia
Kai.Lin@student.griffith.edu.au
{D.Chen, C.Sun, G.Dromey}@griffith.edu.au

Abstract. A constraint expresses a relationship among one or more variables. Constraints are very useful in the development of collaborative applications, such as collaborative CAD and CASE systems, but satisfying constraints in the presence of concurrency in collaborative systems is difficult. In this article, we discuss the issues and techniques in maintaining constraints in collaborative systems. In particular, we also proposed a novel priority strategy that is able to maintain both constraints and system consistency in the face of concurrent operations. The strategy is able to resolve constraint violations in multi-constraint systems and is independent of the execution orders of concurrent operations. To illustrate the applicability of the proposed priority strategy, the applications of the approach in various collaborative systems are discussed in detail.

1 Introduction

A constraint expresses a relationship among one or more variables. Interactive applications, such as CAD and CASE systems, may specify many constraints to confine the relationships or states of constrained objects.

Maintaining constraints automatically is advantageous in collaborative systems. A task demanding people to work collaboratively is often complex and may contain many constraints. Thus, it is very practical and powerful for collaborative systems to maintain constraints automatically on behalf of users. For example, when people work collaboratively to design a project using Java Class notation, many conflicts may arise if a system only relies on individuals to maintain Java single inheritance constraint.

Constraints are very useful in handling complex tasks in collaboration scenarios, but satisfying constraints in the presence of concurrency in collaborative systems is difficult. Concurrent operations may result in some constraints becoming difficult to satisfy even though they may be maintained easily in single user environments. For example, it is hard for us to satisfy the constraint defining "$X=Y+Z$", when three users concurrently change the values of X, Y and Z. In addition, interferences among constraints may be very intricate and difficult to coordinate in collaborative systems.

Much work has been done on the maintenance of constraints in single user interactive applications [2], [3], [5], [13]. However, maintenance of constraints in concurrent environments has many new features which cannot be handled by single user strategies, such as ensuring both constraint satisfaction and system consistency, handling the constraint violations generated by concurrent operations, etc.

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 103 – 110, 2005.

In this paper, we discuss the issue of maintaining constraints in collaborative environments and propose a novel strategy to achieve both constraint satisfaction and system consistency in real-time collaborative systems. The proposed strategy is independent of the execution orders of concurrent operations. In addition, the applications of the proposed strategy in various collaborative systems are discussed in detail.

The rest of this article is organized as follows. In section 2 we discuss problems of constraint maintenance in real-time collaborative systems and propose a novel strategy that is able to handle constraint violations and maintain system consistency in collaborative environments. Section 3 introduces the applications of the proposed strategy. Comparison with related work is introduced in section 4 and the major contributions and future work of our research are summarized in the last section.

2 A Priority Strategy of Maintaining Constraints in Collaborative Systems

Collaborative systems are groupware applications to support people working together in groups, such as electronic conferencing/meeting, collaborative CAD and CASE [10]. Constraints are very useful in collaborative systems, which can confine and coordinate concurrent operations.

To meet the requirement of high responsiveness in the Internet environment, replicated architecture is widely adopted in collaborative systems. Shared documents are replicated at the local storage of each collaborating site, so that operations can be performed at local sites immediately and then propagated to remote sites [1], [10]. However, maintaining both consistency and constraints in collaborative systems adopting replicated architecture is difficult, which is illustrated in the following two scenarios:

Scenario 1. A horizontal-line constraint, C_1, restricts *left-endpoint.y=right-endpoint.y* of any horizontal-line. Two users concurrently move both endpoints of a horizontal-line to different vertical positions.

Scenario 2. There is a constraint C_2 that confines objects *A* and *B* should not overlap with each other. On the initial document state, *A* is at position P_a and *B* is at P_b. Two users concurrently move *A* and *B* to the same position P_c from different sites.

There is a contradiction between satisfying constraints and retaining operations' display effects in both scenarios. The problem is caused by concurrent operations competing to satisfy the same constraint in different ways. To characterize these operations, we define competing operations group of a constraint.

Definition 1. A Competing Operations Group of constraint *C*, denoted by COG_C, is a set of user operations, $\{O_1, O_2, \ldots O_m\}$, such that:

(1) For any $O_i \in COG_C$, there is $O_j \in COG_C$ while O_i and O_j are concurrent,
(2) The executions of all the operations in COG_C will result in a constraint violation of *C*, which cannot be restored if all these operations retain their display effects,
(3) For any $O_i \in COG_C$, the executions of all the operations in $COG_C - O_i$ will not generate the condition described in (2).

In multi-constraint systems, concurrent operations may violate unspecific constraints, as illustrated in the following scenario:

Scenario 3. A collaborative system enforces two constraints, C_1 and C_2, which define "*X.color=Y.color*" and "*Y.color=Z.color*" (Here, notation *X.color* represents the color of object *X*). Two users concurrently generate operations O_1 and O_2 that change *X* and *Z* to different colors respectively.

In scenario 3, the concurrent executions of O_1 and O_2 form competing operations group of neither C_1 nor C_2, because C_1 can be satisfied while both O_1 and O_2 retaining their display effects, if we do not enforce C_2. For the same reason, C_2 can be satisfied by un-enforcing C_1. However, if both users' operations retain their display effects, C_1 and C_2 cannot be satisfied at the same time. In this scenario, the concurrent executions of the two operations form a competing operations group of an unspecific constraint of the system. To characterize these operations, we define competing operations group of a system.

Definition 2. A Competing Operations Group of system *S*, denoted by COG_S, is a set of user operations, $\{O_1, O_2, \ldots O_m\}$, such that:

(1) For any $O_i \in COG_S$, there is $O_j \in COG_S$ while O_i and O_j are concurrent,
(2) The executions of all the operations in COG_S will result in that all the constraints enforced in system *S* previously cannot be satisfied at the same time if all these operations retain their display effects,
(3) For any $O_i \in COG_S$, the executions of all the operations in $COG_S - O_i$ will not generate the condition described in (2).

According to the definitions 1 and 2, if constraint *C* is enforced in system *S*, a competing operations group of *C*, COG_C, must be a competing operations group of *S*, COG_S.

If a competing operations group of system *S*, COG_S, is generated, to maintain all the constraints enforced in *S*, one operation in COG_S will lose its display effect. If different operations lose their display effects at different sites, divergence occurs. To solve this problem, a priority strategy is adopted. If there is a conflict in retaining all operations' display effects in collaborative systems with constraints, the strategy masks operations' effects according to their priorities. For instance, in scenario 1, if the operation changing *left-endpoint.y* of a horizontal-line has a higher priority than the other operation, its display effect will be retained while the display effect of the other operation will be masked at each site.

In this paper, the timestamp of an operation denotes its priority, the bigger the timestamp the higher the priority. For any two operations O_a and O_b, the timestamp of O_a is bigger than the timestamp of O_b, if and only if O_b total ordering precedes O_a [10].

If the execution of operation *O* generates a set of competing operations groups of system *S*, $COGS_S = \{COG_1, COG_2, \ldots COG_n\}$, $1 \leq n$, *O* must be contained in each $COG_i \in COGS_S$, $1 \leq i \leq n$. If *O* is the operation with the lowest priority in a $COG_i \in COGS_S$, $1 \leq i \leq n$, *O* will be masked. As a result, no other operation should be masked in any $COG_j \in COGS_S$, $1 \leq j \leq n$, $j \neq i$, because *O* is also an operation in COG_j and is masked.

Masking an operation, *O*, may cause some masked operations that have lower priorities than *O* restore their display effect (i.e. be unmasked). For instance, in scenario 2, two users concurrently execute O_1 and O_2 that move objects *A* and *B* to P_c respectively. If O_2 is masked to satisfy the overlapping constraint, when O_1 is masked by other operations, O_2 may recover its display effect. Therefore, if an operation, *O*, is masked to satisfy the constraints of a system, all the masked operations that have lower priorities than O should be checked. If their executions on the new document state can ensure the satisfactions of these constraints, they should be unmasked. On the other hand, unmasking an operation O_j may cause other operations that have lower priorities than O_j be masked.

Based on the above discussion, we propose a priority strategy as follows:

For any operation O that is ready for execution in a multi-constraint system S:

(1) If the execution of O will not generate competing operations group of S, O can be executed directly at the site,

(2) If the execution of O will generate a set of competing operations groups of S, $COGS_S = \{COG_1, COG_2, \ldots COG_n\}$, $1 \le n$, and O is the operation with the lowest priority in a $COG_i \in COGS_S$, $1 \le i \le n$, O will be masked at the site,

(3) If the execution of O will generate a set of competing operations groups of S, $COGS_S = \{COG_1, COG_2, \ldots COG_n\}$, $1 \le n$, and O is not the operation that has the lowest priority in any $COG_i \in COGS_S$, $1 \le i \le n$, to satisfy all the constraints in S, O should be executed after the operations with the lowest priorities in each $COG_i \in COGS_S$, $1 \le i \le n$, are masked.

(4) Suppose O_i is the operation that has the highest priority amongst all the masked operations in each $COG_i \in COGS_S$, $1 \le i \le n$. After the execution of O, all the masked operations that have lower priorities than O_i should be checked in descending order of their priorities. If their executions on the new document state can ensure the satisfactions of the constraints enforced in S, they will be unmasked.

The above priority strategy can maintain both constraints and consistency in collaborative systems, which is independent of the execution orders of concurrent operations. It can be adopted in many collaborative applications, including CAD, CASE, spreadsheets, graphical interface toolkits, simulation systems, etc.

3 The Applications of the Priority Strategy

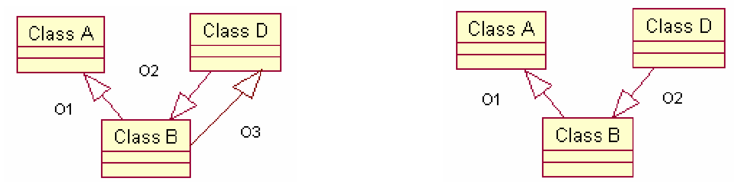
To illustrate the applicability of the proposed priority strategy, in this section we discuss the applications of the approach in various collaborative systems.

3.1 Maintaining Constraints in Real-Time Collaborative CASE Systems

Collaborative CASE systems support people working together on large projects. For this many users need to work on shared documents to get the maximum benefits in terms of productivity gains. Real-time collaborative CASE systems allow a group of users to view and edit the same documents at the same time from different sites to analyze requirements and create designs. They provide several benefits, such as easy interaction among designers, sharing document easily and collaboratively integrating requirements into a design, etc.

In Object-oriented CASE systems, constraints are widely adopted to confine the relations of different Classes. For example, each Java Class should satisfy single inheritance constraint and Class hierarchy is always acyclic. Therefore, the proposed priority strategy can be applied to maintain constraints in collaborative CASE systems, as illustrated in the following scenario:

Scenario 4. A collaborative CASE contains three Java Classes, *A*, *B* and *D*, on the initial document state. Three users concurrently generate operations. O_1 denotes that *B* extends *A*. O_2 requires that *D* inherits *B* and O_3 represents *B* extends *D*, as shown in Fig. 1 (a). We use a notation *O.priority* to represent the priority of operation *O*. In this scenario, $O_1.priority > O_3.priority > O_2.priority$.



(a) (b)

Fig. 1. Three concurrent operations violate Java single inheritance constraint and form cyclic Class hierarchy in (a). After O_3 is masked, the Class hierarchy of the three Java Classes is shown in (b).

In scenario 4, the executions of the three concurrent operations violate Java single inheritance constraint and form cyclic Class hierarchy, because the executions of O_1 and O_3 cause *B* extend two Classes, and O_2, O_3 form a circular inheritance. In this scenario, the maintenances of the two constraints interfere with each other. According to the relative priorities of the three operations, O_2 should be masked to satisfy acyclic Class hierarchy constraint and O_3 should be masked to maintain Java single inheritance constraint. However, if O_3 is masked to enforce Java single inheritance constraint, there is not cyclic Class hierarchy anymore, so that O_2 can retain its display effect, as shown in Fig. 1 (b).

The proposed priority strategy can handle the interferences among constraints' maintenances in collaborative environments, which is independent of the execution orders of concurrent operations. For example, in scenario 4, if O_2 is masked previously to ensure acyclic Class hierarchy, when O_3 is masked to maintain Java single inheritance constraint, O_2 will be unmasked. On the other hand, if O_3 is masked before the execution of O_2, the executions of O_1 and O_2 will not generate any competing operations group of the system, so that both of them can retain their display effects. Under both conditions, the final results are identical with the one shown in Fig.1 (b). The proposed strategy can be adopted in collaborative CASE systems to maintain constraints.

3.2 Maintaining Constraints in Collaborative CAD Systems

Constraints have been used in a number of ways in CAD systems, such as to denote the special requirements of geometric design, to maintain consistency among multiple views on data, etc.

The proposed priority strategy can be adopted to maintain constraints in collaborative CAD systems, which is demonstrated by the following example:

In CAD system *S*, the radiuses of three circles, *A*, *B*, *D*, representing the wheels of a tricycle, must be the same. Thus, two constraints, C_1 and C_2, should be maintained, which define "*A.radius=B.raduis*" and "*B.radius=C.raduis* " respectively. Three users concurrently generate operations O_1, O_2 and O_3 from different collaborating sites, which change *A.radius*, *B.raduis* and *C.raduis* respectively. O_1*.priority*> O_2*.priority*>O_3*.priority*.

Suppose at a site, the three operations are executed in the order O_1, O_2, O_3:

(1) O_1 will not generate COG_S of the system on the initial document, so that it can be executed,
(2) When O_2 arrives, it will be masked, because O_1 and O_2 form a competing operations group of C_1 and O_1*.priority*>O_2*.priority*,
(3) When O_3 is ready for execution, it will be masked too, because O_1 and O_3 generate a competing operations group of *S* and O_1*.priority*>O_3*.priority*.

Thus, after the executions of the three operations, both O_2 and O_3 are masked while O_1 has display effect.

At another site, the three operations are executed in the order O_3, O_2, O_1:

(1) The execution of O_3 on the initial document state will not form competing operations group of *S*. Thus, it can be executed,
(2) When O_2 is ready for execution at the site, O_3 will be masked, because O_2 and O_3 form a competing operations group of C_2 and O_2*.priority*>O_3*.priority*,
(3) When O_1 is ready for execution, O_2 will be masked, because O_1 and O_2 form a competing operations group of C_1 and O_1*.priority*>O_2*.priority*,
(4) After O_2 is masked, O_3 will be checked. Because O_1 and O_3 form a competing operations group of *S* and O_1*.priority*>O_3*.priority*, O_3 cannot recover its effect.

After the three operations are executed in different orders at different sites, each site achieves the same final document state. The proposed priority strategy can be adopted to resolve constraint violations in multi-constraint collaborative CAD systems.

3.3 Maintaining Constraints in Other Collaborative Applications

The proposed priority strategy is a generic solution to handle constraint violations in multi-constraint collaborative systems. It can be applied in many kinds of collaborative applications. For instance, it may be used by a spreadsheet system to maintain the constraints confining the relationship between different cells, a simulation system representing current and voltage relationship of a complex circuit to confine that concurrent operations always satisfy Ohm's law, a graphic editing system to coordinate the concurrent operations that update graphic objects, such as coordinating the operations that concurrently change the *left, right* and *width* of a rectangle, etc.

The proposed strategy masks operations' display effects according to their priorities in case there is a conflict in retaining all operations' display effects in collaborative systems with constraints. The approach is independent of systems and constraints. However, how to detect competing operations group and mask operations is constraint and application dependent.

4 Related Work

There is a large body of research related to constraint maintenance in single user environments [2], [3], [5], [13]. However, maintaining constraints in concurrent environments has many new features that cannot be handled by the strategies adopted in single user environments.

CAB [7] and SAMS [9] are related to constraint control in collaborative environments. CAB presents an active rule based approach to modeling user-defined semantic relationships in collaborative applications and explores a demonstrational approach for end-user customization of collaboration tools to support the definition of those relationships. However, just as its author stated that many complications of maintaining constraints in collaborative environments, such as how to handle constraint violations and coordinate interferences among constraints, are not investigated in CAB.

The intention of SAMS is to achieve semantic consistency by integrating semantic constraints to the operational transformation approach. SAMS is based on XML resources. Its application in other environments has yet to be investigated. Moreover, SAMS uses loose constraint satisfaction strategy that allows constraints be violated.

By comparing with the above approaches, we proposed a priority strategy that is able to maintain both constraints and system consistency in the face of concurrent operations in collaborative environments. The strategy is independent of the execution orders of concurrent operations and can resolve constraint violations in many kinds of multi-constraint collaborative applications.

5 Conclusions and Future Work

Constraints are very useful in collaborative systems, which can confine and coordinate concurrent operations. However, maintaining constraints in real-time collaborative systems is a challenging task. The difficulties are caused by concurrent operations that violate constraints. Being able to solve this problem is crucial in the development of collaborative CAD and CASE applications.

In this paper, we proposed a novel priority strategy to solve this problem. This solution ensures constraint satisfaction while maintaining consistency. Our solution allows concurrent operations be executed in any order. It can be adopted to handle constraint violations in many kinds of multi-constraint collaborative applications. The applicability of the proposed strategy is discussed in detail.

We are currently investigating the scenario that user operations manipulating the constrained variables are concurrent with constraints' additions and deletions. If constraints can be added to or deleted from a system when users are manipulating the constrained variables, maintaining consistency will become very difficult. How to solve this problem is currently being investigated and will be reported in our subsequent publications.

References

1. Begole James et al.: Resource Sharing for Replicated Synchronous Groupware, IEEE/ACM Transactions on Networking, Vol. 9, No.6, Dec. (2001) 833-843.
2. Borning, A. and Duisberg, R.: Constraint-Based Tools for Building User Interfaces, ACM Transactions on Graphics, Vol.5, No.4, Oct. (1986) 345-374
3. Borning, A. et al.: Constraint Hierarchies, Lisp and Symbolic Computation, Vol. 5 No. 3, Sept. (1992) 223-270.
4. Dourish, P.: Consistency Guarantees: Exploiting Application Semantics for Consistency Management in a Collaborative Toolkit, In Proceedings of the ACM Conference on Computer Supported Cooperative Work, ACM, New York (1996) 268-277
5. Freeman-Benson, B. et al.: An Incremental Constraint Solver, Communications of the ACM, Jan. (1990) 54-63
6. Ignat, C.-L, Norrie, M.C.: Grouping in Collaborative Graphical Editors, ACM Conference on Computer-Supported Cooperative Work, Chicago, USA, Nov. 6-10, (2004) 447-456
7. Li, D. and Patrao, J.: Demonstrational Customization of a Shared Whiteboard to Support User-Defined Semantic Relationships among Objects, ACM GROUP'01, Boulder, Colorado, USA, Sept. 30-Oct. 3, 2 (2001) 97-106
8. Sannella, M. et al.: Multi-way versus One-way Constraints in User Interfaces: Experience with the DeltaBlue Algorithm, Software-Practice and Experience, Vol. 23(5) (1993) 529-566
9. Skaf-Molli Hala, Molli Pascal and Ostér Gerald: Semantic Consistency for Collaborative Systems, the Fifth International Workshop on Collaborative Editing Systems Hosted by the 8th European Conference of Computer-Supported Cooperative Work, Helsinki, Sept.15, (2003)
10. Sun, C. et al.: Achieving Convergence, Causality-Preservation, and Intention-Preservation in Real-Time Cooperative Editing Systems, ACM Transactions on Computer-Human Interaction, 5(1), Mar. (1998) 63-108
11. Sun, C. and Chen, D.: Consistency Maintenance in Real-Time Collaborative Graphics Editing Systems, ACM Transactions on Computer-Human Interaction, Vol. 9, No.1, Mar. (2002) 1-41
12. Sun, D. et al.: Operational Transformation for Collaborative Word Processing, ACM Conference on CSCW, Chicago, USA, Nov. 6-10, (2004)
13. Zanden, B.: An Incremental Algorithm for Satisfying Hierarchies of Multi-way Dataflow Constraints, ACM Transaction on Programming Languages and Systems, Vol.18, No.1, Jan. (1996) 30-72

Design of a Cooperative Working Environment for Mobile Devices*

M. Mascaró, B. Estrany, J. Laredo, E. Sidaner, and Y. Luo

Universitat de les Illes Balears, Departament de Matemàtiques i Informàtica,
07122 Palma de Mallorca, Spain
{mascport, tomeu.estrany, dmilyu0}@uib.es

Abstract. GEMMA is a cooperative control system developed to manage and coordinate a fleet of mobile resources (boats, planes, employers, etc) dedicated to maintain clean coastal and beach waters in the Balearic Islands. All the process is supervised by a control center in several workstations. In addition, a cooperative working environment was designed and developed for a group of mobile devices. In this article we present a project extension that involves the control tasks of mobile computers. Using portable computers, PDA and mobile phones, controllers can develop their control tasks supervising the coastal state and be online with all the system and resources in the Gemma project. All the communication processes are established by TCP-IP connections via GPRS/UMTS.

1 Introduction

The Balearic Islands have become one of the most important vacation resorts in Europe in the Mediterranean Sea area. The Balearic Government in collaboration with the University of Balearic Islands, has been designed and developed a system to control and maintain the quality of their coastal area, "The GEMMA System". GEMMA is an interactive and cooperative real time IT system to collect environmental marine waste.

GEMMA project offers an integral solution to implement a control and management system for mobile resources (ships, planes, employers, etc) to reach our environment maintenance goal.

During the year of 2005 a new step has been made. A cooperative working environment was designed and developed to adapt GEMMA for a group of mobile computers (PDA, mobile phones and notebooks) by GPRS/UMTS communication support. Controllers, mechanics and others employers can monitor the actual situation and work online in real time no matter where they are, without any equipment more than a mobile computer with the GPRS/UMTS support. In addition, portable devices can update all the information by themselves in real time in the display and show to the final user the real time situation. Even for the PDA users, we provide the real aspect orthographic photographic information on the PDA's. This is supported by the Orthographic Photographs Web Map Server in our control center.

* This project is an extension of the GEMMA project, supported for the Government of the Balearic Islands.

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 111 – 118, 2005.

2 General Cooperation and Coordination Support

The implemented system allows cooperative work of 3 controllers, 12 coastal boats, 27 beach boats, 1 mechanic and 1 plane. The controllers can manage the efforts of all the mobile resources using GPRS and phone call connections. In all time controllers can verify the state and locations of all mobile resources. Simultaneously mobile resources are connected to each controller to inform them of the possible actual incidences and their response.

2.1 The Cooperation Support Between the Control Center, the Controllers and Other Mobile Resources

All the events are placed into the system in real time. The following explains the coordination of the tasks of the cooperative work in the GEMMA system.

Boats and planes situation: One MOVILCOM[1] device is installed on each boat. To maintain the always updated information in the system and considering the average speed of the boats, each MOVILCOM sends its position to the SCDT Server in the control center every 60 seconds. The SCDT Server updates all the ship positions in the Data Base Server immediately when it gets any position change event. On the other hand, each controller's GUI application (GEMMA GUI) makes an update every 15 seconds of the information displayed on the monitor according to the information stored in the Database Server. The real time situation of the mobile resources in the system is delayed in the worst case only 15 seconds. Therefore, the real time state of the elements in the system is guaranteed (see figure 1).

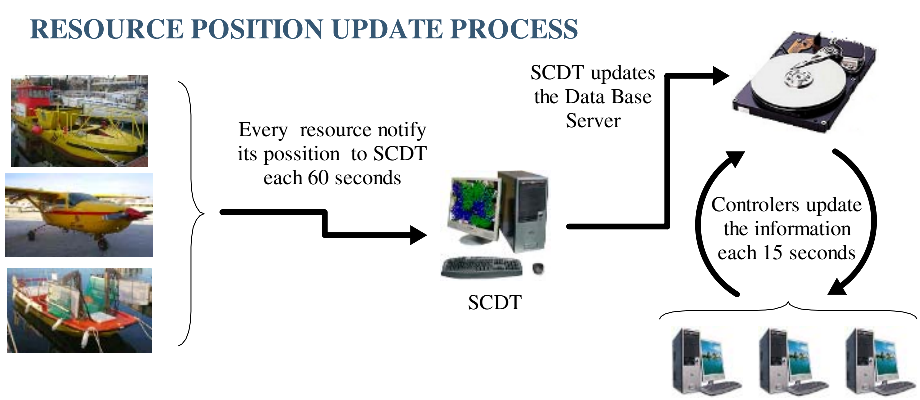


Fig. 1. Resource position update process

[1] MOVILCOM: Mobile communication device of the ships. This device can send the GPS position and several events to the Control Center. It can be make call phones. All the communications take place by GPRS/UMTS using voice or a TCP-IP channel.

Control Center sends an event to some ships: When any of the GEMMA GUI applications wants to send an event to one mobile resource, it can send a message to the SCDT[2] Server. When the SCDT receives a message from a controller, a TCP-IP communication is established with the MOVILCOM of the target mobile resource (see figure 2). Controllers can send several events to the mobiles. The most relevant events are:

- Reprogram MOVILCOM sentences.
- Text messaging.
- Ask for a phone call conversation.
- Establish a work route.
- Ask for the state of the MOVILCOM.
- Informs about some incidence (waste, emergency, weather, etc).

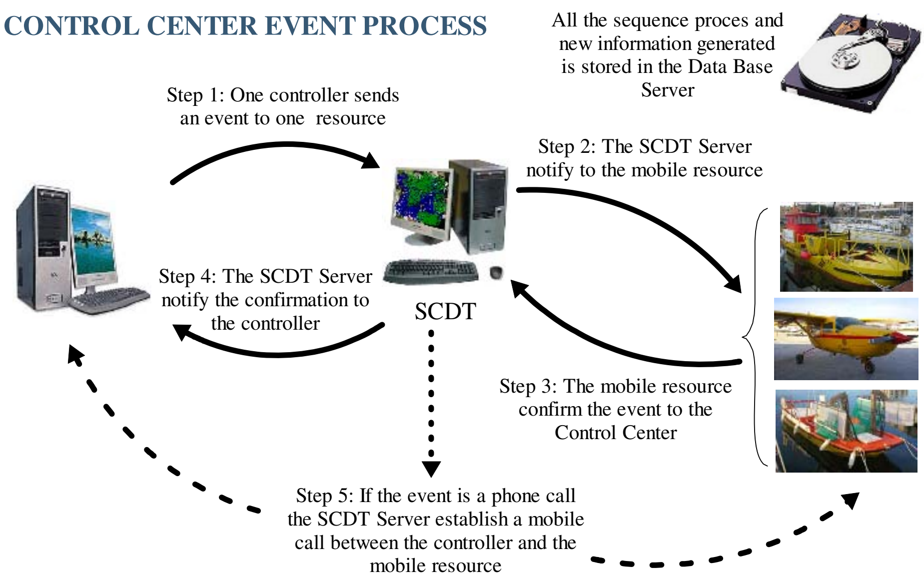


Fig. 2. The steps of how a controller sends an event to one mobile resource

When some MOVILCOM receives an event message, it starts the correspondent task and sends a confirmation event to the Control Center (SCDT). While the SCDT doesn't receive confirmation of any message during the security interval of time, it tries to send the message to the MOVILCOM another time, until it gets an answer from the target MOVILCOM. After that, the SCDT Server confirms the action to the controller.

[2] SCDT: Communication Server: controls all the communication messages and phone calls that take place in the GEMMA System. All the communications takes places by GPRS/UMTS devices.

In the case that the controller sends phone call event to some ships, when the SCDT receives the confirmation of the mobile resource, it starts to manage a conversation process between the two entities. The conversation stops when one of the members of the communication, controller or marine (or pilot), hangs up the phone device.

One Ship sends an event to the Control Center: It's the same process as before but in the reverse direction.

Fly ships can send geographical referenced images to the control center: Fly mobile resources has been equipped by a laptop and a geographical reference camera. It can take pictures and associate the GPS position to them. The images in EXIF[3] format are posted by an SCP action (Secure Copy) to the Control Center Server. When GEMMA server receives an EXIF image, it extracts the geographical reference information, and creates a new incidence at this geographical point in the Database Server. After notifying to the SCDT this new incidence, the new incidence appears immediately in all the GEMMA GUI applications. The controllers can then study the picture of the new incidence in their display if necessary.

3 Heavy Cooperative Action Control

The communication manager in the SCDT controls all the processes in a cooperative mode. Several controllers can interact with all the entities (boats, plane and other controllers) in GEMMA system in real time. Without cooperative action control, undesired events can appear. The most important undesired events and some policies are explained below.

3.1 Several Commands Given to the Same Mobile Resources

All the events in the cooperative GEMMA system can appear any time. How to guarantee there is no conflicted commands given to the same mobile resource? In GEMMA the answer is simple and clear – there is a central control. For example, when two commands are send to the same mobile resource at the same time from different controllers, two message events arrives to the SCDT. The Communication Server (SCDT) puts them in a priority FIFO list. The message list in SCDT has a FIFO policy. However, some critical events can be put at the head of the list immediately (SOS, important event). Messages are sent to the ships, by turn, using the process exposed in section 2.1.

3.2 No More Than One Controller Can Modify the Same Data at the Same Time

All the commands applied to the Server Database are made by a java function. This java function accesses the Data Base via servlets[4] using the "synchronized" java

[3] EXIF: Image format that can take associated to the JPEG image some GWS84 information (lat/Lon).

[4] Servlet: Java application located in a Java Server Pages Server. Java Server Pages is a Web Server extension.

feature. This feature guaranties the exclusive access to the Data Base. If some database modification petitions arrive at the same time, they are managed in a FIFO policy. This process is entirely transparent to the application which modifies the database.

3.3 The Information That all the Unit Receives is Up-To-Date

All the information changes in a several parallel processes in real time. How to guarantee the information that all the unit receives is up-to-date? The requirement of the system guarantees a delay within 15 seconds in the updated information process. That's the maximum delay that can occur in the system as we have explained in section 2.1.

4 Light Cooperation Control

A reduced small version of the GEMMA GUI (little GEMMA GUI) has been developed to enable a pocket PC access to the information of the GEMMA System.

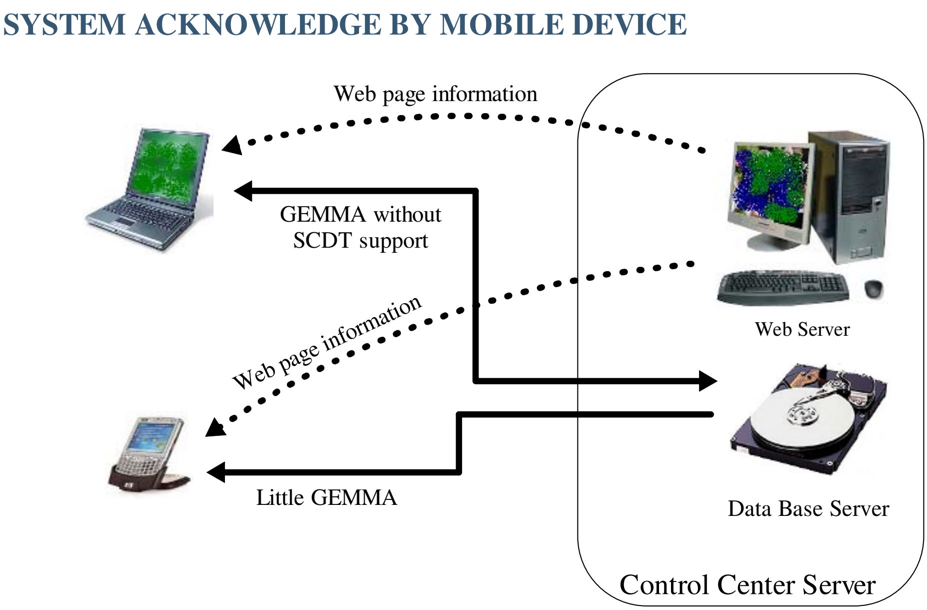


Fig. 3. Access to the GEMMA System information by a mobile device

4.1 The GEMMA Laptop Version

Controllers can run the GEMMA application on a laptop or a workstation outside the Control Center connected to the internet by a mobile Ethernet card. In this mode all the GEMMA application capabilities are available except the communication with the SCDT features (See figure 3).

4.2 Special Treatment in Programming on PDA and Mobile Phone

To develop the reduced version of GEMMA for Pocket PC we adapt the java application to the requirements of the JVM[5] for the mobile devices. Little GEMMA requires a JEODE JVM installed on pocket PC. This JVM can manage a Java program compatible with the JDK 1.3 version, but can't support the SWING Java API. To solve this problem, special treatment has been taken. We modified the GUI features of the application to adapt them to another GUI library (AWT). In addition, the Little GEMMA version can interact with all the information and can display it but it can't modify any data (see figure 4). Unlimited number of such pocket PC's can use the GEMMA system.

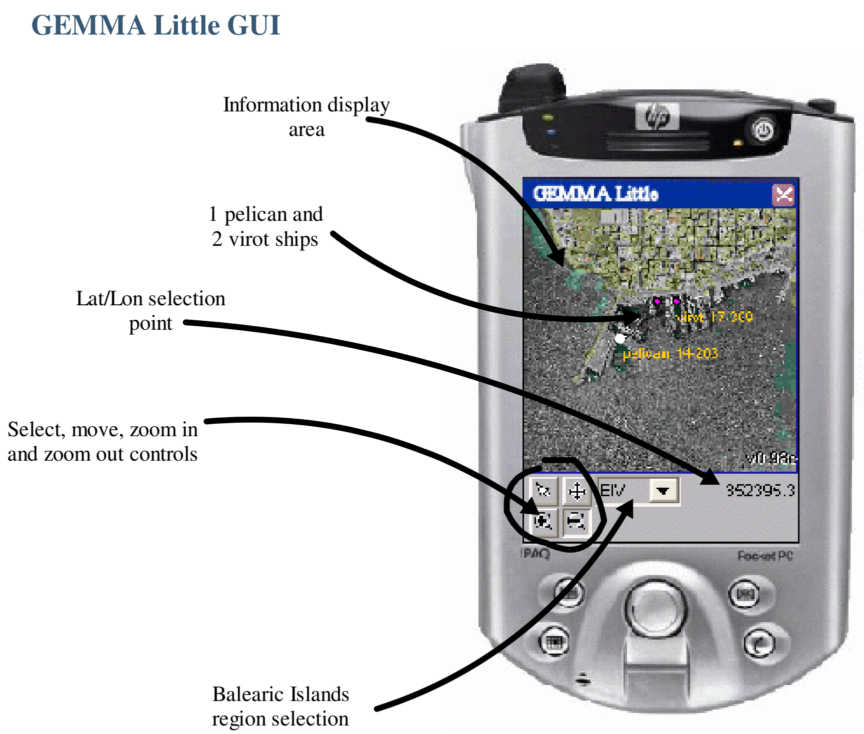


Fig. 4. Gemma reduced version for mobile devices

5 Rendering all the Information in the Graphical User Interfaces

The main Control Center application was developed entirely in Java. This feature makes possible the use of GEMMA in any platform with J2SDK installed. From a

[5] JVM: Java Virtual Machina.

free Java application to render static GIS information application, the “Alov Map” [1], a new interactive render system GEMMA GUI, was developed. GEMMA GUI is able to render GIS information and refresh some contents when the system requires it. In this way, the system can show the real positions of the entities, boats, planes and incidences. It can offer the controller extra information about them. In addition, the system is capable of establishing a phone call or sending a message selecting any item on the screen [5] [6].

6 Group Communication Support

The system offers several concurrent tools to manage the communication between the controllers, boats and the rest of entities. It supports the communication between mobile resources and control center by GPRS, GSM and SMS.

6.1 Communication Manager (SCDT)

The communication manager manages all the communications between the Control Center and the rest of entities which can interact with the system – the cleaning boats, police, observer planes, coast guard, city councils, etc. The communication system is the module that was not developed using free software tools in order to satisfy the requirements of the communication supplier company. The communication manager in the SCDT was developed in Delphi under windows platform. It offers several concurrent tools to manage the communication between the controllers, boats and the rest of entities. By using the GPRS/GSM technology the system manages text message, data files, and phones calls to implement direct communication between the entities in the process. All the communication process is completely integrated in the graphical user interface, automatic and transparent to the final users.

SCDT can create point to point TCP-IP connections with all the MOVILCOM – the communication device of the mobile resources using a private virtual network. It can make several actions: to establish a communication between two entities, to program the mobile devices tasks remotely, to send SMS messages, periodical event notification tasks, etc.

7 Results

The GEMMA System is full operability since 06/01/2004. The response and results of the Cooperative GEMMA Control System are satisfactory.

To see the actual state of the GEMMA project and a real time information about the system functionality visit http://80.39.95.27/gemmaweb/index.html. You can see complete 2004 collected marine waste statistics in the Menorca Island in figure 5.

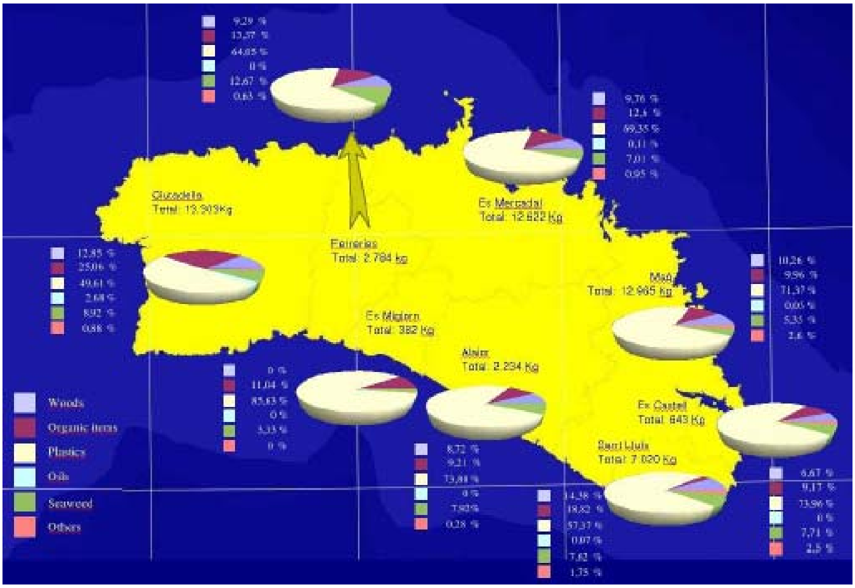


Fig. 5. Statistics of collected marine waste in year 2004 (Menorca Island)

References

1. Alov Map. http://www.alov.org/. University of Sydney, 1999-2004.
2. MITOrthoServer. http://ortho.mit.edu/mitorthoserver. Massachusetts Institute of Technology, 2001.
3. OpenGIS Consortium. http://www.opengis.org/. 2004.
4. Fernández P., Bejar R., et alt. Web mapping interoperatibility in practice, a Java approach guided by the OpenGIS Web Map Server Interface Specification. University of Zaragoza. European Project TIC98-0587, 2000.
5. Estrany B., Mascaró M., Bibiloni A., Serra G., Sideaner E., Laredo J., Capo B., Alcover B., A Cooperative Control System to collect environmental marine waste. University of the Balearic Islands (Spain). CDVE2004, Palma de Mallorca, España. 2004
6. B. Estrany, M. Mascaró Portells, J. M. Aguiló, L. Arqueros and Y. Luo. An integral IT system for large scale coastal environment control. Proceedings of the Fourth IEEE International Workshop on Soft Computing as Transdisciplinary Science and Technology (WSTST'05), pp.755-764, Takeshi Furuhashi, Mario Koppen, Azuma Ohuchi, Yukio Ohsawa (Eds.), Springer-Verlag, 2005..

Mobility in Environmental Planning: An Integrated Multi-agent Approach*

D. Borri, D. Camarda, and A. De Liddo

Dipartimento di Architettura e Urbanistica,
Politecnico di Bari, via Orabona 4, 70125 Bari, Italy
Tel. +080.5963347, Fax +080.5963348
d.camarda@poliba.it

Abstract. Mobility infrastructure planning is an increasingly crucial aspect of environmental planning, essential to boost regional economies and social relations, as well as critical for environmental impacts involved.

Structuring inherently complex issues and problems is a major challenge of mobility planning. Today, therefore, a major issue is the setting up of system architectures that take into account the impacts of the mobility system on environmental and social quality.

A multi-agent approach is seen as useful to let local stakeholders interact, meanwhile learning from each others' attitudes and cognitions. The present paper deals with the building up of an IT-based multi-agent platform in public interaction forums, particularly envisioning alternative mobility scenarios in the plan of the Italian province of Foggia.

1 Introduction

Multiple source knowledge acquisition (MSKA) is increasingly used in current strategic regional planning for problem setting and for acquiring distributed knowledge concerning social visions and aspirations and in general commonsense baseline information for the preparation of strategic plans and policies [4]. In contrast with MSKA spreading fortunes, only modest progress seem to have occurred, in the last years, on using MSKA for specific problem solving which traditionally needs highly focused technical expertises [11].

In regional and urban planning, analysis and design of mobility and transportation issues and systems are a well consolidate engineering work: they have typically developed according to a formal model approach which does not give much room to ill-structured knowledge and statements [8]. Only very recently attention for subjective behaviour has been paid, optimization techniques in the field have incorporated subjective functions [1], and in general mobility issues have been interpreted considering non-material factors as relevant as the material ones in explaining the structure of observed phenomena [13].

* The present study was carried out by the authors as a joint research work. Nonetheless, chapter 1 was written by D.Borri, chapters 2, 4 were written by D.Camarda, chapter 3 were written by A. De Liddo. The authors are particularly grateful to G. Caratù, L. Grassini, M. Patano, for their important contribution to the discussion ground of the present paper.

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 119 – 129, 2005.

Problem solving in current systems engineering remains anchored to routine procedures which are based in general on a rational knowledge-in-action sequence (even if now not necessarily linear as in the past) of the following type: analysis>defining alternative solutions>choicing the best or a satisfying possible solution>using this solution for tentative problem solving>controlling performance in solving the problem>modifying the route if needed>etc (see the *Graph plan* family, e.g., [18]).

It is quite clear that all of these model steps relate to a well defined technical expertise and tool capacity even if the structural and organizational (Zeleny) limitations affecting that rationale have been definitely stated since many years (the works by Simon, Lindblom).

The use of MSKA in regional strategic planning in general implies knowledge engineers who define possible contents and aims of knowledge interaction, select participants in knowledge interaction, set and manage socio-physical arenas for knowledge interaction, and facilitate and record the knowledge play [5]: participants – stakeholders – in knowledge interaction are variously selected (in general aiming at representing knowledge senders and receivers extracted by the society at large involved by strategic planning), their discourse being of course strongly influenced – substantially and procedurally – by the cognitive stories and styles characterizing each of them, the (structural and circumstantial) characteristics of the knowledge experiments set up by the knowledge engineers, and last but not least ill-structured causalities.

Depending on its typical features, MSKA addresses a type of knowledge – a social knowledge – which is strongly different from the technical knowledge which derives from systematic and in-depth work on phenomena and related problems and conform to knowledge standards and conventional rationality [7][14][15][12]. Social knowledge from MSKA have been recently more credited than the technical one when facing environmental problems which are strongly influenced by social knowledge and action.

Mobility an transportation issues and problems in regional strategic planning mostly deal with the search for optimal or at least satisfying movement of persons and goods within social communities and economic markets. Time and safety reliability of transport multimodal corridors – which relate to the physical characteristics of the corridors – are analyzed on the basis of technical standards which progressively enter the normative and legal domain, meanwhile social events and valuations and aspirations influence the technical standards as they solicit them to take account of their values and needs.

The technical analysis of transport nets is for this reason of geometrical (physical) and social nature. This comes from technical observation of net form and from social needs expressed by local/non-local users and enters the circuit of political debates and programs.

The present case study of the use of MSKA concerns the evaluation of social compliance and in general knowledge robustness of technical choices regarding mobility and transportation issues in regional planning for the area of Foggia, north of Puglia. After the present introduction, a synthetic outline of the multi-agent system architecture will be drawn out, discussing the mechanics of the whole interaction process during the environmental plan set-up. Chapter 3 deals with the evolution of the substantial contents of the interaction, specifically focusing on mobility issues within the basic environmental planning process as emerged from the contributions of expert and non-expert knowledge agents. Brief comments on outputs achieved and research perspectives end up the paper.

2 Exploring the Process Architecture

The forum experience occurred in the province of Foggia, in Apulia region (Southern Italy), where the provincial administrative board set up and started a participatory process to draw the environmental structure plan of its territory in 2002.

A first draft plan prepared by the planning office showed a set of six future visions of the province, basing on experts' studies and researches, as canvases to draw the final plan on. Such visions ranged from environmental to managerial issues, and were represented by thematic geographical and conceptual maps[1].

The Department of planning of Bari Polytechnic (DAU) was then expected to prepare a planning process architecture to support the sharing, enrichment and evaluation of this set of visions. The general aim was to let decision-makers issue the final plan with related policy program, with a deeper and more aware knowledge of issues and perspectives.

The involvement process was organized into 20 interactive meetings with the community, through an iterated process of modification of expert visions. Face-to-face interactions were paralleled by a web-based virtual forum, mainly aiming at further enriching and exploring issues. This process layout was drawn out and organized according to, and as an evolution of, research activities carried out by DAU in the last few years [4] (figure 1).

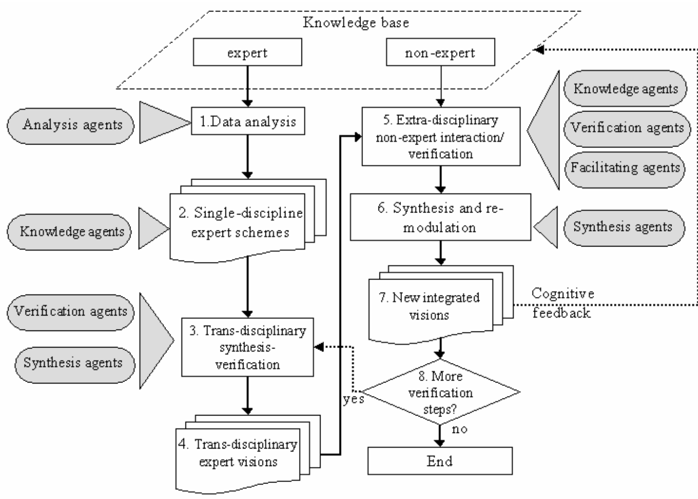


Fig. 1. The process architecture

As said, a multi-agent structure performing different activities was involved, with specific roles in expediting stages and fostering intermediate outputs. Expert visions were the result of the work of expert agents in a preparatory phase, that took place before the starting off of the forum process. The initial analytical work (step 1) was

[1] They were: (i) an ecological network, (ii) the development of infrastructures, (iii) strategies of external linkages, (iv) the small town network, (v) history and culture, (vi) the open territory.

performed by data-mining experts and researchers of the planning group, who worked on existing data-sets, maps and other informational media to structure and make the initial knowledge base available for subsequent experts' actions. Basically, these experts acted as substantial knowledge agents, coming from different domains, to draw out sectoral commentaries (step 2) as bases for final expert visions.

After the work of domain experts -knowledge agents-, all sectoral schemes needed to be matched, synthesized and transformed into coherent strategic visions (step 3). Again, researchers and staff from the planning group acted as verification agents, comparing all branch contributions with the aim of matching trans-disciplinary issues, eliminating redundancies, suggesting amendments and integration. Also, they tried to group issues, topics, problems under aggregate tags, so acting as synthesizing agents aiming at providing the essential framework to draw out trans-sectoral visions. This outcome of strategic expert visions was finally achieved (step 4), providing the basic representation for the subsequent enrichment of visions with non-expert contributions. The first stage of the non-expert enrichment of visions (step 5) represents the actual forum-based interaction, where stakeholders involved acted as knowledge agents, giving their substantial cognitive contribution to the process. Languages, forms of knowledge representation, behavioural inputs are commonly reported as being nor formal, neither standard when dealing with non-expert interactions, like in this case [10]. There was the need of building up an evolving platform able to allow the exchange of the different knowledge structures lying in each stakeholder, expressed in textual, verbal, graphical, gestural forms, with real-time feedback. Therefore, the process was supported by an architecture of inter-operating media tools, hybridized by human control. In this delicate context, DAU knowledge engineers acted as verification and facilitation agents during forum sessions, being particularly careful in avoiding filtering and interference in coordinating the interaction process.

After the forum phase, the subsequent stage (step 6) was carried out by DAU and planning office experts, acting as synthesizing agents, somehow similarly to step 3. They grouped new or amended issues under aggregate tags, trying to preserve concept linkages and languages as intrinsic guidelines and minimize external interpretations of data. This delicate step, although generally reported as being semantically risky [2][17], was reputed necessary in order to have an usable input for the next steps. An improvement of this approach was studied by the group in a preliminary stage, trying to involve speech and gestural recognition software tools to minimize human interpretation, basing on previous field studies[2] [4][19]. Due to the limited time available, the attempt could not result in an actual architectural improvement, but represents an intriguing research follow-up. Meanwhile, a record of original oral and graphical representations of concepts expressed by stakeholders has been kept for any possible future references.

Outputs of the previous step acted as framework for the drawing out of new strategic visions, purposely amended with the non-expert knowledge (step 7). As a cognitive by-product of the interaction, these knowledge addictions become part of the whole knowledge base, enriching and expanding its contents. Subsequently, after passing the

[2] See the experiences carried out at Carnegie Mellon under the Camera Assisted Meeting Event Observer (CAMEO) project, funded by the US Defence Agency (www.cs.cmu.edu/~cameo).

if-then node for possible re-verifications by domain experts (step 8), visions were ready to be used as strategies of the provincial plan, and the process mainstream ends up.

With reference to available literature, real processes of collective participation in environmental planning often report mixed successes, particularly in terms of actually raising the level of knowledge and information gathered, as compared to the standard knowledge coming from experts and formal databases. In real contexts, stakeholders' participation can be fickle, variable, scarce, due to a number of reasons ranging from unwillingness of wasting time to inadequate promotion and information [3][10]. Such problems add up to the increasingly known problems of handling informal, non-standardized forms of knowledge in non-expert stakeholders [17]. Up to date, several efforts have been devoted to cope with the latter, and even DAU research activities is largely oriented to that aim [5][10]. However, the former issue has been substantially ignored as an exploration frontier, rather considering it as a given constraint.

Despite this current trend, this real planning process proved to be fairly articulated and useful to explore some terms of variations of stakeholder participation. Therefore, as an advancement of previous works, it was tried to draw out possible qualitative relations that can describe the multi-dimensional "shrinking" or "enlarging" of this imaginary cognitive sphere. This attempt was carried out, in order to explore if and to what extent can local administrators be supported in their decision and policy actions by -numerically and qualitatively- variable multi-agencies in governance processes.

3 Mobility Issues and the Environmental Planning Process

Foggia province areas have different environmental, cultural and development characteristics: the mountain zone in named Sub-Appennino, the costal zone is named Gargano and the flat zone is named Tavoliere (figure 2).

Such geo-morphologic variety is the mirror of many differences in the transport system and particularly in the road network. A relatively efficient network between the towns of the Tavoliere contrasts with a scarce connection in the Sub-Appennino and Gargano areas, both in terms of amount and quality of transport (low level of service and insufficient state of maintenance of roads). Objective of the observation of data experiences, collected during the computers-based forum, is to check if a symmetry exists between quality of expert knowledge (collected in the mobility study of experts) and common knowledge emerged during the interaction meeting, trying to show and explain possible symmetries or asymmetries.

The visions are built and represented by cognitive maps, that is a cognitive negotiation approach already used from our research group in previous cognitive experiments [5]. Cognitive maps can be bottom-up, or as real-time synthesis carried out by synthesizing agents (knowledge engineers) [6][9][16].

In the present case, the second mode was chosen, in order to represent a large amount of expert knowledge in clear and simple concepts to submit to a large community of non-experts. Therefore, a knowledge engineer mediation allowed a real-time advancement of the six visions submitted, with results immediately transferable into the ongoing planning process.

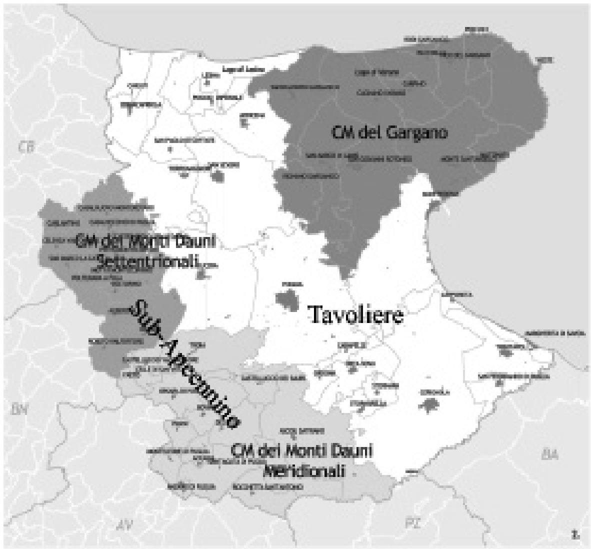


Fig. 2. Foggia province boundaries and partitions

3.1 Expert and Forum-Based Visions

Map in fig.3 shows the expert vision of main problems/solutions for mobility system.

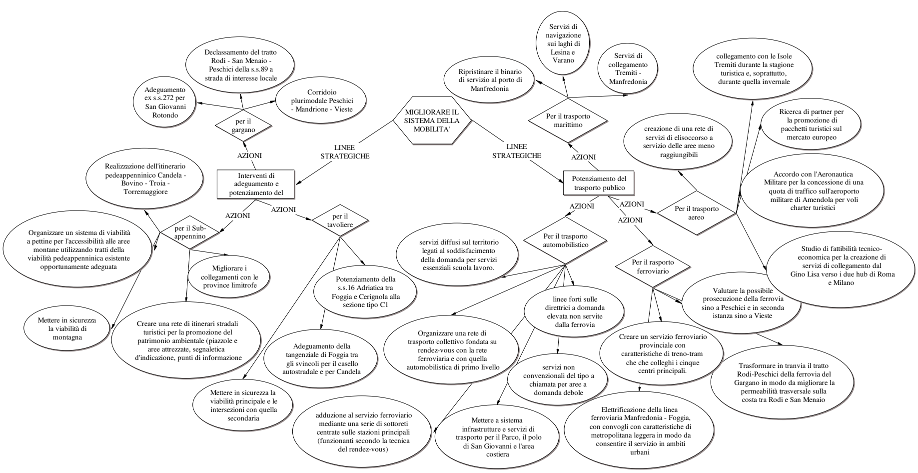


Fig. 3. Cognitive map of the mobility issues in the expert vision

The vision is organized into two strategic lines (rectangular boxes) aimed at improving the mobility system, i.e., new and upgraded roads, and development of public transport.

The first one strategic line is specified in actions for territorial areas, following the geographic partition previously described (Gargano, Subappennino, Tavoliere) (rhomboid boxes). The second strategic line concerns the development of public transport and it is described for each transport mode (air, sea, railway and road transport) (rhomboid boxes). Strategies and specific actions are summarized in the following table (figure 4).

First Strategy	Actions			
	Sub-Appennino	Tavoliere		Gargano
new infrastructures of adjustment and development of the road network	1)Actions to enhance safety on the mountain road system 2)improvement of connections between the towns of the province 3)tourist itinerary for the promotion of environmental and storic patrimony	1)improvement 2)adjustment and 3)Actions to enhance safety in the intersections between main and secondary roads		1)corridor connecting the railway line to the coastal road network 2)adjustment of the connection road to the religious Sanctuary of San Giovanni Rotondo
Second Strategy	Actions			
	Air transport	Sea transport	Railway transport	Road transport
development of public transport	1)improvement and activation of existing airports 2)network services of helicopter-aid for the less reachable areas	1)improvement of Manfredonia harbour 2)navigation service for main lagos of the province	1)railway line extension along the coast line 2)tram service that connects the five main towns of the province 3)light	1)interconnections between the railway network and the road network of first level (following the rendez-vous technique) 2)transport services for the Natural Park in the Gargano area 3)call-services for the areas with a low transport demand

Fig. 4. Synthesis of the mobility issues in the expert vision

As said before, during interactive meetings the six strategic visions have been represented in cognitive maps and for validation and enrichment, with a process supported by cooperative computer-based visualization. Starting from the observations and proposals collected during the meetings we have extrapolated an ex-post cognitive map that describes the community vision of the mobility system, similar to the one shown in figure 3, as base for our considerations.

The procedure of construction of the map, performed by synthesizing agents, is articulated as follows. First, isolation of the observations and proposals about the improvement of mobility system, then synthesis of the observation contents in concepts (nodes) and classification of the concepts by strategies (strategic lines) and specific project indications and proposal (sets of action). Subsequently, actions are subdivided for each geographic area and transportation system. Finally, concepts are organized in one general map, following the criteria used for expert visions, i.e., following progressive levels of deepening (objective – strategic lines - actions).

The procedure results show that communities have characterized three strategies of improvement of the mobility system: the development of public transport, the refitting, adjustment and development of road network infrastructures and improvement and rationalization of the existing networks of transport.

The first strategy was detailed very little, whereas the second one indicates infrastructure actions for geographic area. It can be observed that the first two strategies coincide with expert ones, while the third strategy is not only introduced as a new proposal but it is also in contrast with the second one in terms of perspective.

The spirit at the base of this strategy is that development is not necessarily linked to new infrastructure equipment, but indeed some experiences seem to show that infrastructures can constitute barriers to the expansion (fig. 5).

First Strategy	Actions			
	Sub-Appennino	Tavoliere		Gargano
development of public transport	1)terminal bus service 2)improvement of the connections between towns of of the province			1)improvement of the connections between towns of of the province expecially in the winter season
Second Strategy	**Actions**			
	Sub-Appennino	Tavoliere		Gargano
new infrastructures of adjustment and development of the road network	1)development of an intermodal system road-railway-sea	1)connections between road network of first and second level 2)improvement of the intermodal poles of exchange between the main centers of the province		1)construction of Garganica highway 2)development of harbour infrastructures: detailed lists of actions for various coastal towns and villages 3)improvement of the politics of mobility towards the overseas lands
Third Strategy	**Actions**			
	Air transport	Sea transport	Railway transport	Road transport
improvement and rationalization of the existing transport networks rather than new infrastructures	1)improvement of existing airports rather than construction of a new airport	1)improvemnt of existing harbour infrastructures rather than building of new ports for private use	1)No to the proposal of extension of the coastal railway 2)No to the realization of one metropolitan light	1) widening of existing road rather than building of new roads 2)implementation of safety actions and maintenance of the road network 3)valorization of "tratturi"(old country road) and light infrastructures

Fig. 5. Synthesis of the mobility issues in the Forum-based vision

3.3 Comparative Insights

The expert and forum-based visions have been compared with two different levels of observation: an upper-order one concerning strategic lines and an inferior-order one concerning project proposal and action lines. At level of strategic lines, proposals and suggestions are related to political choices, particularly in mobility, but they also reflect, unavoidably, more width perspectives of social and economic development of the province. That is to say that different strategic lines reflect different ways to wish and imagine the future of the territory.

At this level the two visions turn out to be asymmetric, being symmetry a quite good overlapping of knowledge content in terms of quality and quantity between expert and common visions. The two visions are asymmetric at level of strategic lines because the forum-based vision comprises the strategic lines of the expert vision but enriches them with an alternative (and in part opposing) strategy. The new added strategy is founded on the community needs more than on the total effectiveness of the system. It aims at improving and rationalizing the existing network more than at creating new ones, coherently with the protection of the naturalness of the places. Point of strength of the not-experts knowledge was the ability to reflect on new and original strategic lines that reveal vocations and objectives of the installed communities. This information and proposals could act as suggestions in order to address the technical studies to the search of alternative solutions that take into account the vocations of the territory.

At level of actions the asymmetry is even more obvious. The expert vision defines more organic and specific actions, and they result also more coherent regarding the delineated total picture of the overall vision. It is obvious in this case that the vision has been written up from a homogenous team of specialists with the same objectives and the same cognitive baggage. Only in one case the forum-based vision proposes specific and detailed actions, i.e., the case of sea transport topic and harbour infrastructures, suggesting that this topic was neglected during the specialists' analysis. This case may show that the community itself have a good argumentative and deepening aptitude also on specific project topics.

Going back to the involvement of the installed communities in the environmental planning process (analysed here in relation to the transports topic) and to the prime objectives of appraisal and enrichment of the expert visions, the process has produced positive results. The non-expert knowledge negotiation has produced concepts, reflections and proposals that have enriched the expert vision not only in possible strategic lines but also in specific actions (e.g., the development of sea transport). Apart from the enrichment ability of this phase of consultation, it has also shown a sure validation ability, that is the ability to enhance some issues not sufficiently dealt with in expert visions -issues of primary interest for people who live the territory.

4 Comments and Final Remarks

The present paper has dealt with the setting up and implementation of a participatory MSKA system architecture to support environmental planning. Although increasingly diffused in general planning, modest advancements and case studies are reported for specific problem solving in environmental planning -a domain intrinsically complex and difficult to be managed. Following an experimentation thread started years ago in the Polytechnic of Bari [5], this time the background was selected in a real and complete forum-based process, with a particular emphasis on mobility infrastructure planning and management, as increasingly crucial aspects of environmental planning. Through the building up of an ad-hoc multi-agent architecture in public interaction forums, mobility scenarios in large participatory sessions were envisioned, compared and fine-tuned, in the real context of the Italian province of Foggia.

The carrying out of the interactive sessions aimed at enriching a layout of expert visions with the contribution of non-expert, commonsense knowledge. The involvement of a multi-agent architecture was justified by the quest for grasping/exchanging the multifaceted knowledge involved and the different operational abilities required to perform those tasks. Under this view, the level of knowledge exchanged, as well as a deeper comprehension of the mechanics of related cognitive processes were main objectives of the research, aiming at more effectively supporting public decision-making. From a substantial standpoint, the forum-based interaction attained some important vision results, showing both strengths and weaknesses. A strength of the not-experts knowledge contribution was the suggestion of new and original strategic lines, more coherent with local needs. A character of weakness was the lack of contextual coherence, articulation and deepening of issues, due to the presence of heterogeneous competences and languages, as opposed to the intrinsic linearity and richness of the starting expert visions. For this reason, external synthesizing agents need to formally adapt contribution to fit the planning purpose and effectively support subsequent policy actions.

In terms of efficacy of system architecture organization to support aware decision-making, the experimentation can be considered as fairly successful. The multi-agent approach allowed an easier and more complex knowledge exchange, consistent with the complex nature of environmental issues and stakeholders' behaviours. Conceptual mapping was an helpful tool in this sense, both in the generation/transmission of complex ideas and in supporting the work of the various process agents involved. Unfortunately, it was not possible to perform a bottom-up use of concept mapping through an unfiltered web-based interaction among participants, that could have allowed a direct comparison with previous works [5]. However, this can represent an intriguing action to be carried out in future researches. As a general remark, being mobility issues more particularly sensible to technical and expert visions than other environmental and socio-economic aspects of planning, it should be interesting to check if the above considerations are replicable in other issues as well. This may be a theme of reflection to be further analyzed in the future, also basing on the outcomes of the overall interaction process in the same planning case study.

References

1. Avineri, E., Prashker, J.N.: Sensitivity to Uncertainty: Need for Paradigm Shift. Proceedings of the 82nd Annual Meeting, Transportation Research B. Washington D.C. (CDROM) (2003).
2. Bloom, P., Peterson, M.A., Nadel, L., Garret, M.F. (eds.): Language and Space (Language, Speech, and Communication). Cambridge, MIT Press (1999).
3. Blowers, A.: Power Participation and Partnership: The Limit of Cooperative Environmental Management. In Glasbergen, P. (ed.): Co-operative Environmental Governance. Kluwer Academic Press, Dordrecht (1998) 229-250.
4. Borri, D., Camarda, D.: Dealing with Multi-Agents in Environmental Planning: A Scenario-Building Approach. Studies in Regional and Urban Planning 10 (2004) 89-98.
5. Borri, D., Camarda, D., De Liddo, A.: Envisioning Environmental Futures: Multi-agent Knowledge Generation, Frame Problem, Cognitive Mapping, Lecture Notes in Computer Science, 3190 (2004) 230-237.
6. Bougon, M.G.: Uncovering Cognitive Maps. The Self-Q Technique. In: Morgan G. (ed.): Beyond Method: Strategies for Social Research. Sage, Beverly Hills (1983) 173-188.
7. Bourdieu, P.: The Logic of Practice. Stanford, Stanford University Press (1992).
8. Cascetta, E., Cantarella, C.: A Day-to-Day and within Day Dynamic Stochastic Assignment Model. Transportation. Research A 25 (1991) 277–291.

9. Eden, C.: Using Cognitive Mapping for Strategic Options Development and Analysis (SODA). In Rosenhead, J. (ed.): Rational Analysis for a Problematic World: Problem Structuring Methods for Complexity, Uncertainty and Conflict. Chichester, Wiley (1989) 21-42.
10. Khakee, A., Barbanente, A., Camarda, D., Puglisi, M.: With or Without: Comparative Study of Preparing Scenarios for Izmir with Computer-Based and Traditional Brainstorming. Journal of Futures Studies 6(4) (2002) 45-63.
11. Lei Zeng, M., Mai Chan, L.: Trends and Issues in Establishing Interoperability among Knowledge Organization Systems. Journal of the American Society for Information Science and Technology 55(5) (2004) 377-696.
12. Maciocco, G., Tagliagambe, S.: La Città Possibile. Territorialità e Comunicazione nel Progetto Urbano. Bari, Dedalo (1997).
13. Mokhtarian, P. L., Salomon, I.: Modeling the Desire to Telecommute: The Importance of Attitudinal Factors in Behavioral Models. Transportation Research A 31(1) (1997) 35-50.
14. Scandurra, E.: L'Ambiente dell'Uomo. Milano, Etas (1994).
15. Stengers, I.: Power and Invention: Situating Science. University of Minnesota Press (1997).
16. Tegarden, D.P., Sheetz, S.D.: Group Cognitive Mapping: A Methodology and System for Capturing and Evaluating Managerial and Organizational Cognition, Omega 31 (2003) 113-125.
17. Warren, T., Gibson, E.: The Influence of Referential Processing on Sentence Complexity. Cognition 85 (2002) 79-112.
18. Weld, D.S., Anderson, C.R., Smith, D.E.: Extending GRAPHPLAN to Handle Uncertainty & Sensing Actions. In Proceedings of the 15th National Conference of the American Association on Artificial Intelligence (AAAI-98). AAAI Press (1998) 897-904.
19. Wesson, R.M.; Corradini, A. Cohen, P.R.: A Map-based System Using Speech and 3D Gestures for Pervasive Computing. Presented at the 4th IEEE International Conference on Multimodal Interfaces, October 14-16, Pittsburgh (2002).

Research on Mobile Agent Based Information Content-Sharing in Peer to Peer System

Shaozi Li[1,2], Changle Zhou[2], and Huowang Chen[1]

[1] School of Computer, National University of Defence Technology, Hunan, 410073, China
[2] Computer Science Department, Xiamen University, Xiamen, 361005, China
{szlig, dozero}@xmu.edu.cn

Abstract. This paper presents our recent research work on mobile-agent based information content sharing in peer to peer system. Our work focuses on how to share the text content and recommend information based on mobile agent. The prototype uses terminology derived from the ant colony metaphor, including nest - a peer entity, and ant – an autonomous agent. Two mobile agents are proposed to implement the content-based text retrieval and information recommendation effectively. These agents are then evaluated through implementations and analysis. The simulated results demonstrate the advantages of our approach in the P2P system, and its potential applications in complex electric commerce and distributed information sharing.

1 Introduction

A peer to peer (P2P) system is an aggregation of autonomous computers, which communicate and coordinate their actions through message transmissions over the computer network [1]. The motivation for a P2P system is resource sharing. It connects hundreds of thousands of users into a vast organization in which users share various resources, such as files, processing cycles, and storage space. There are many existing P2P applications, e.g. Napster [2], kaZaA [3], and eMule [4] that coordinate huge internet-scale files for sharing. With their unprecedented resource sharing power, P2P systems are gaining more and more attention and becoming increasingly popular throughout the world.

Research on P2P systems concentrates on how to locate resources in a distributed way and improve their efficiency. Existing applications, such as file-sharing, only provide basic and simple services which cannot satisfy the demand. The reason is that users need a variety of application services for different objectives. For example, we use a P2P file-sharing application to download a file by using the file name hash. Usually, we only know the file's name, not its content. Therefore, we may be deceived by fake files that have incorrect names. If we dispatch a program which analyzes the file at the remote peer before downloading it, we can reduce the network load and obtain the right file.

Agent-based P2P systems incorporate mobile agents to develop customized services which share remote resources according to users' requirements. Users can extend the functions of a remote peer by dispatching a mobile agent to it. Once the mo-

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 130 – 139, 2005.

bile agent arrives, it will provide a specific customized service on the remote peer for its creator. Mobile agents are active objects which have autonomous behavior, executing states, and network locations. The mobile agent technology is a simple and flexible way for users to develop customized services on remote peers.

In this paper, we propose the prototype of information retrieval and recommendation based on the mobile agent in P2P system. The remainder of this paper is organized as follows. Section 2 presents the application scenario of our prototype and the components in this prototype, including the nest and ant. Section 3 and 4 describe the algorithms used in two ants. Section 5 is the simulated result of information content retrieval based on the lexical-chain. Finally, Section 6 gives the conclusion and future work.

2 The Prototype's Model

The prototype uses terminology derived from the ant colony metaphor. It is composed of interconnected nests. Each nest is a peer entity sharing its storing information content. Nests handle requests originated from local users, by generating one or more *ants* – autonomous agents, which travel across the nest network to try to satisfy the request. Ants can observe their environment, perform simple local computations, and act based on these observations. The actions of an ant may modify the environment, as well as the ant's location within the environment. In the prototype, emergent behavior manifests itself as swarm intelligence whereby the collection of simple ants of limited individual capabilities achieves "intelligent" collective behavior [5]. According to the user's requests, there are two types of nets in our prototype. One is called SearchAnt, which searches the information to meet the user's requests. The other one is called RecomAnt, which recommends the user's information to other user who need it.

Before describing the nest and ant, we first give the description about application of our prototype.

2.1 Application Scenario of Our Prototype

Information sharing among people exists everywhere. In our daily life, we need to exchange information with other people who have similar interests. For example, Michael is a graduate student of computer science. He plans to choose natural language processing as his thesis topic. Therefore, he needs to collect up-to-date information in this area. On the other hand, John is a professor whose research interest is also natural language processing. Now, he has published several papers and wishes people who have the same interest can read them as quickly as possible.

To meet the requirements of two people, we can use a P2P file-sharing application to download a file by using the DHT (file name hash). However, if we only know the file's name, not its content, we may find that the content of this downloaded file is not interested. This will generate a significant amount of unnecessary traffic due to little utilization of the retrieval data.

As a new paradigm, mobile agent technology is expected to reduce network traffic and enhance the efficiency of information retrieval by moving code to remote peer, filtering them locally in parallel, and returning with a comparatively small result. For example, Michael can creates one or more mobile agents (SearchAnts) to find the peers that have the information about natural language processing and want to share them to other people. In the same way, John also creates one or more other mobile agents (RecomAnts) to find peers that are interested in his research. If they find the objects, the connection will be built directly between the creator and the object. Then the sharing information will be downloaded from the provider.

2.2 The Nest

In Figure 1, we illustrate the architecture of a nest which is composed of three logical modules: communication, ant scheduler and information sharing service.

The communication layer is responsible for the discovery of new nests, ~~for~~ management of network topology and movement of ant between nests. In the network, each node has a unique identifier. In order to communicate with a remote node, its identifier must be known. The set of nests known to a node are called its neighbors.

The ant scheduler module multiplexes the nest computation resource among visiting ants. It is also responsible for enforcing nest security by providing a "sandbox" for ants, which can limit the resources available to ants and prohibit ants from performing potentially dangerous actions.

The information sharing service is responsible for the interaction between the user and other nest, which includes building the connection with object nest, downloading the sharing information and creating the ant according to the user's request.

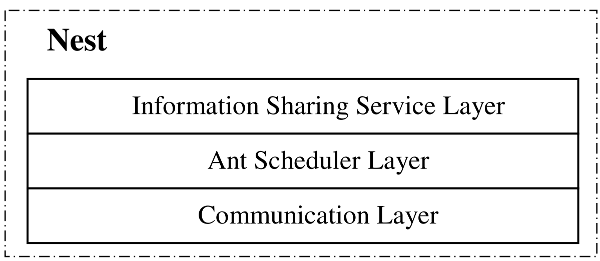


Fig. 1. The architecture of nest

2.3 Ants

Ants are generated by nests in response to user requests; each ant tries to satisfy the request. An ant will move from nest to nest until it fulfills its task. And then (if the task requires this) it may return back to the originated nest. Ants that cannot satisfy their task within a *time-to-live* (TTL) value are terminated. During the moving, the ant carries its state, which may contain the request, results or other ant specific data. The ant algorithm may be transmitted together with the ant state, if the destination nest does not know it; appropriate code transfer mechanisms are used to avoid downloading the same algorithm more than once, and update it when a new version of the same algorithm is available.

Ants do not communicate directly with each other; instead, they communicate indirectly by leaving information related to the service. This form of indirect communication, also used by real ants, is known as stigmergy [5].

In our prototype system, the behavior of an ant is determined by its objective. If the objective of ant is to search the information that it's nest needs, it will calculate the similarity between the texts that is shared in the nest and user. In Section 3, we describe the process in detail. If the objective of ant is to recommend the information, another algorithm will be used, which is described in section 4.

3 The Algorithm of SearchAnt Based on Lexical Chain

A lexical chain [6] is a series of words constituted by some adjacent words under the same topic. These words appear simultaneously in the same lexical environment, as they are expressing the same thing. Lexical chain is now used in information extraction, information retrieval, checking unsuitable words in a text, analyzing and computing texts' similarity, constructing automatic links for super texts, segmenting texts, words disambiguating, etc [7, 8]. To find the information content that user needs, SearchAnt is dispatched out. It analyzes the texts and calculates the similarity with user's interests using the algorithm based on lexical chain. The following is the detail of this algorithm.

3.1 The Way to Form Lexical Chain

Before building lexical chain to express a text, we should firstly consider how to choose the words, in other words, which words are suitable to be candidate words. After analyzing the text, we delete all empty words: pronouns, modal verb, prepositions or adverbs with a subordinate clause and articles. For those words that appear frequently, such as good, do, taking, etc., we also put them into the table that includes the unused words. The remained words are all candidates.

The next problem is the word's relationship. According to it, we build lexical chain in dictionary. We use WordNet (G. A. Miller and R. Beckwith etc., Cognitive Science Laboratory, Princeton University, USA) as the source knowledge. WordNet is a machine dictionary based on psycholinguistic principles. It uses our familiar spellings to express morphology and the synonym set Synsets (a list of synonyms which can be substituted by each other according to certain context) to express the meaning of a word. So far, WordNet embodies about 95600 lemmas, including 51500 words and 44100 compound words. They are organized roughly 70100 acceptations or synonym sets, describing relationships such as hyponymy, synonymous, antonymous, meronym /holonym, etc.

Here we consider 3 relationships: extra-strong, strong, and medium-strong.

a) Extra-strong:
   It means the repetition of two words, neglecting their distance.
b) Strong:
   It can be considered in 3 situations (set a window between two words, usually 7 sentences):

b1) Both two appear in the same synonym set, e.g. human and person appear in the same set {person, individual, someone, man, mortal, human, soul}

b2) There is some kind of semantic relationship between some certain sets in each synonym set of the two words, e.g. one synonym set of precursor {predecessor, precursor, antecedent} has a antonymous relationship with successor's synonym set{successor}. We call this relationship horizontal connecting.

b3) If one is a phrase or compound word, and some words in this phrase or compound word appear in other synonym sets, then we don't consider what kind of relationship this containment is, e.g., private_school. The synonym set {private_school} is included by the synonym set {school}. Although there is a relation of hyponymy, we consider school and private_school as strong.

c) Medium-strong:

There is a path connecting two words (define the distance between two words as a window, usually 3 sentences) and we limit the path length between 2 and 5. Figure 2 shows the medium-strong path between apple and carrot.

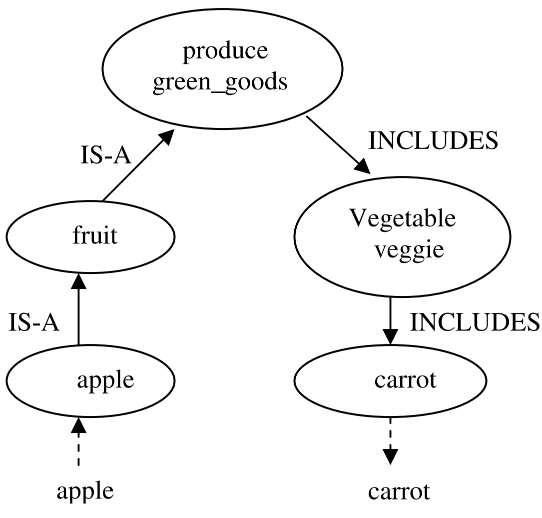


Fig. 2. The medium-strong path between apple and carrot

According to relationship between key words, the formula for computing a weight value is as follows.

$$Weight_{ij} = \begin{cases} 3C & \text{Extra-strong} \\ 2C & \text{Strong} \\ C - \| path_{ij} \| - \beta *(\#turn) & \text{Medium-strong} \\ 0 & \text{Others} \end{cases} \quad (1)$$

In the formula (1), C, β are constants, $path_{ij}$ is the semantic distance between two words , #turn means how many times the directions change in the path.

The basic idea is as follows: For each word W_i in the text, if W_i is a candidate word and there is some relationship between W_i and the lexical-chain (the interval between them cannot be too big), then we calculate the power value between W_i and the lexical-chain and append W_i into the lexical-chain.

3.2 Expressing Text

Traditional information filtering technique primarily filters out relevant texts by key words searching and statistics. It just respectively computes the frequency of each key word in a text, which neglects the relationship between key words and the topic. This kind of technique has many defects. In the following, we analyze three main factors that caused these defects and give the corresponding solutions.

3.2.1 Influence of the Useless Words

That means the characteristics that appear in the all categories can not represent a category. Some of these words belong to the set which is composed of the unused word, and the others carry little information. So we will delete them.

3.2.2 Influence of the Relationship Between Words

It can be divided into two instances: one is due to synonyms (e.g. Chinese word "计算机" and "电脑"), while the other is due to certain semantic relationship between words. In the medical category, for example, certain relationship exists in "doctors", "nurse", "hospital", "sickbed", "operating room ", "diagnosis", "infection", "state of an illness", "antibody", etc. The existence of one characteristic can replace others in some sense. The frequency that each characteristic appears respectively may be little or overlaid by some irrelevant words of high frequency. When the above distance formula does not take account of such influence, it will also lead to inaccuracy in distance computing. For this reason, we should take the semantic relation between words account. If they express the same topic, their semantic distance in a dictionary is short, and they can be put together automatically during the process of text analyzing and considered comprehensively while computing the similar degree.

E.g.: Information of some words from a text as follows: {{information: 3, technique: 1, Bayesian-technique: 1, datum: 2, model: 1, area: 1} {computer: 4}}, the figure after each word means the times this word appeared in the text.

If we only consider the frequency of each word, computer's frequency is highest. Yet we find other words are related strongly in semantics and can complement each other, so each one's importance is elevated in this way.

3.2.3 Influence of Inequal Status Between Words

Although the importance of each key word to the topic can be shown by the times of appearance, it is not enough. Usually, we needn't read the whole passage, but we can find out the topic accurately from the title or the first paragraph. This means there are some characteristic words supporting certain topic strongly (decision characteristics). Their existences decide the topic to a great extent. However, in a vector space model, this kind of decision may be submerged by the influence of numerous non-decision characteristics. In that case, we bring forward the concept of characteristic district.

Text characteristic district is a district that can show a text's topic, including headlines, abstracts, key words and references. Since not all texts contain the abstract, key words or references, we let these structure units alternative. Based on the sample statistics, some Chinese researchers found that the coherent degree of natural science thesis' headlines and contents is 98% in local Chinese periodicals, and the degree in news texts is 91%. Almost every article contains a headline. For this reason, a headline is one main text characteristic.

Clue words are those summarizing or generalizing words, such as "anyway", "in a word", "to sum up", etc. We will enhance the importance of the words included in characteristic districts, and intensify the power of the words after clue words to enhance their importance.

According to the above discussion, we will use lexical chain to express user templates and unknown texts. Firstly, we analyze texts handed over by user and build lexical chain to express them. Then we build user template which automatically learns during filtering to express users' interest and meet their interest better. For those unknown texts, we adopt the same method to build lexical chain to express their meaning.

3.3 Analyzing Text

However, not all words in a text can be used to build lexical chain. Only key words that can express the meaning of a text most clearly will be selected. The methods for expressing texts are as follows:

Preprocessing Text: Withdraw etyma and recognize phrases;

Part-of-Speech Tagging: Part-of-Speech Tagging for words in the text;

Extracting Key Words: Get rid of the following words in texts: articles (e.g. a, an, the), preposition or adverb with a subordinate clause (e.g. to, of, in), modal verb (e.g. would, must), and conjunction (e.g. and) etc.. Let's make a definition of W(s,w,c), in which w means a word, s means the word's sequence in a text, and c means the part of speech. For example, (12, think, verb) means the 12th word in W is 'think' and its part of speech is verb. We also can evaluate different powers for different parts of speech to show their importance, of which nouns are usually most important. For those words appeared in headlines, the first or last paragraph, or the beginning or end of a paragraph, we can also enhance their power. In addition, we can set a threshold value to get rid of those words with the frequency below it.

Expressing texts with lexical chain: Now, we get series of words. After lexical chain built automatically, we will get the text's lexical-chain expression.

3.4 Filtering Text

So far, texts and users' interest are expressed with the lexical chain. The relevant degree between texts and users' interest can be estimated by the cosine value of the formula :

$$\cos(a) = \frac{\sum_{ij} V_i * T_j}{\sqrt{\sum_i V_i^2 + \sum_i T_i^2}} . \tag{2}$$

Here, $V=\{V_1,V_2,\ldots,V_n\}$ is a vector expression of a text's lexical chain, $T=\{T_1,T_2,\ldots,T_n\}$ is a vector expression of lexical chain of users' interest. The less α is, the closer these texts are related to users' interest.

Among all texts, we can make an order of relevant degree to feed back users according to this value, or we can set a valve value k. If the relevant degree between texts and users' interest is above k, it means that these texts can meet users' interest. We'll return the ordered texts to users, and get rid of the remaining texts which under the value. We can also take users' feedback into account. If almost every text we filtered out is considered interesting to users, we reduce k. Otherwise, we increase it.

4 The Algorithm of RecomAnt

When the peer wants to recommend some information to others, RecomAnt will be created and dispatched out. It uses the information recommendation algorithm to find the destination peer. There are many different techniques to implement the recommender systems [9]. So far, collaborative filtering is the most successful recommender system technology. The main idea of collaborative filtering is to recommend new documents to the users based on the similarity of their tastes and others'.

In general, collaborative filtering is a three stage process of finding similar users (neighbors), computing predicated ratings, and applying the predictions as recommendations to the user [10].

To generate predictions for a user, the system first identifies his "neighbors", whose interests are highly correlated to the user's. There are several possible options for the correlation coefficient, and the most common one is the constrained Pearson correlation:

$$\text{correl}(u_i, u_j) = \frac{(u_i - z)^T (u_j - z)}{|u_i - z||u_j - z|} , \tag{3}$$

where u_i and u_j are two users' ratings vectors, and z is the neutral rating which is subtracted from the vector. For simplicity, we assume that z is subtracted off in the original ratings matrix, although in practice z must be taken into account when presenting a recommendation to the user. Now, equation (3) becomes (4) as follows:

$$\text{correl}(u_i, u_j) = \frac{u_i^T u_j}{|u_i||u_j|} , \tag{4}$$

which is also known as the cosine similarity measure in information retrieval.

In the second stage**,** the system generates predictions for the user by computing a weighted average of each neighbor's rating scaled by their correlation value,

$$r_{i,j} = \frac{1}{|N(u_i)|} \sum_{u_k \in N(u_i)} correl\ (u_i, u_k) * r_{k,j} \quad . \tag{5}$$

Using this formula, we can do prediction for users' rating. With more rating information, we may make a better recommendation for users in the future.

The third stage is the application of the prediction, which involves adding z back to the prediction and presents the prediction to the user as a recommendation if it is high enough. If the user follows some of the recommendations and provides feedback to correct the predictions, the system learns the user's tastes and his relationship to the community over time.

5 Experimental Results and Analysis of the Content-Based Filtering Model

Two main criterions for evaluating the information retrieval systems are precision and recall rate. The efficiency of a text retrieval system is generally described by average precision. And the visual explanation is the area of the precision/recall curve line.

This paper uses the medical corpus OHSUMED on TREC-9,which is a subset of MEDLINE in the famous National Library of Medicine and constituted by the medical literatures from 1988 to 1991, including 348，566 texts from 270 medical periodicals, with a content of 400MB. In addition, literatures in 1987 are used as training corpus, while literatures between 1988 and 1991 as testing corpus [11]. The results of our comparative experiments are shown in Table 1.

In traditional vector space models, each key word is considered respectively. So we can only use the times that a key word appeared, but not make full use of the information how words are correlated in a text. When the lexical chain is used, the precision will be improved.

Table 1. The experimental results

	The traditional keyword based system	The lexical chain based system	Difference
Average precision	38.53%	47.46%	8.93%

6 Conclusion and Future Work

This paper presents the result of our recent research work on mobile-agent based information content sharing in P2P system. Our work focuses on how to share the text content and recommend information based on mobile agent. The prototype uses terminology derived from the ant colony metaphor, including nest – a peer entity, and ant – an autonomous agent. Two mobile agents are proposed to implement the content-based text retrieval and information recommendation effectively. The simulated results show some advantages of the proposed approach based on lexical chain and collaborative filtering algorithm in P2P system, and its potential applications in complex electric commerce and distributed information sharing. In the future, we will implement this prototype system with JXTA and use it in some applications.

Acknowledgement

This project is supported by The Natural Science Fund, Science & Technology of Fujian Province (Project Number: A0310009, 2001J005), Academician Fund of Xiamen University, P.R. China. Many thanks to Dr. Qingyang Hong, Miss Yang Cao, Dazhen Lin and Mr. Feng Guo, for their contributions to this paper.

References

1. George Coulouris, Jean Dollimore, and Tim Kindberg. Distributed systems: Concepts design. Addison-Wesley, third edition, 2001.
2. Napster Inc. The napster homepage. Retrieval 2001 from http://www.napster.com/
3. KaZaA. Kazaa file sharing network. Retrieval 2002 from http://www.kazaa.com/
4. eMule. Emule-project.net. Retrieval 2002 from http://www.emule-project.net/
5. E. Bonabeau, M. Dorigo, and G. Theraulaz. *Swarm Intelligence:From Natural to Artificial Systems*. Oxford University Press, 1999.
6. J. Morris and G. Hirst. Lexical Cohesion Computed by Thesaural relations as an Indicator of the Structure of Text**.** Computational Linguistics, 17(1):21-48, 1991.
7. G. Silber and K. McCoy. Efficiently Computed Lexical Chains as an Intermediate Representation for Automatic Text Summarization. Computational Linguistics, 29(1), 2003.
8. H. Gregory Silber and Kathleen McCoy. Efficient Text Summarization Using Lexical Chains In Proceedings of the 13th International Conference on Intelligent User Interfaces, IUI2000, pp.252-255, New Orleans, January, 2000.
9. Terveen L. and Hill W. Human Computer Collaboration in Recommender Systems. In J. Carroll (Ed.), Human Computer Interaction in the New Millenium, Addison-Wesley, New York, 2001.
10. Ian Soboroff , Charles Nicholas. Collaborative Filtering and the Generalized Vector Space Model. In: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, p.351-353, Athens, Greece, July, 2000.
11. S. Roberton,D.Hull, The TREC-9 Filtering Track Final Report, Proceeding of the Ninth Text Retrieval Conference (TREC-9), Feb. 2001.

Environmentally-Aware Security Enforcement (EASE) for Cooperative Design and Engineering*

Larry Korba, Yuefei Xu, Ronggong Song, and George Yee

Institute for Information Technology, National Research Council of Canada,
Building M-50, Montreal Road, Ottawa, Ontario
{Larry.Korba, Yuefei.Xu, Ronggong.Song, George.Yee}@nrc-cnrc.gc.ca
http://www.iit-iti.nrc-cnrc.gc.ca

Abstract. In cooperative design and engineering, work context has dramatic implications when it comes to building understanding and teams, and getting the work done. From an information security perspective, the context of operation for software or hardware can drive the expectations for security, and may indeed determine the levels of security policy that would be required for collaborative operation. There is a place in this domain for proactive, context-based security implementation. This paper describes a context-centric security enforcement system intended for cooperative design and engineering environments that we call: Environment-Aware Security Enforcement (EASE). We describe the application of this approach to an existing client-server, e-manufacturing application.

1 Introduction

In computer supported cooperative work environments work context can have a dramatic bearing upon the effectiveness of a collaborating team. Yet building effective cross-organizational collaborative environments is a big challenge, especially from the security perspective [1]. Context may be used to tailor information delivery and sharing based upon location, available resources, perceived activity needs, and expertise requirements. To support this idea, there have been a number of recent developments in assessing and using context in collaborative work environments.

When considering it from a security standpoint, context may be used to determine where, when, how and for how long individuals and organizations may share information with each other. The greatest benefit of establishing context-aware security mechanisms is to enable enforcement of security for mission-critical distributed applications in conformance with the security expectations of all collaborators in all contexts of their work. Ideally the collaborative environment would take into account the networks, computer operating environment, the tasks at hand, and other factors for all collaborators when determining if and how they may be allowed to work together.

In this work we present a model, design and prototype implementation for a system that uses a variety of contextual information to enforce whether or not collaborative environments may be usable. This builds upon research we have presented else-

* National Research Council Paper Number 48255.

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 140 – 148, 2005.

where [2]. The type of contextual information we use is typical of the information used when building security policies for organizations. The design provides a means for setting and enforcing security policy for the operating contexts of collaborators. The system assures that security requirements for all clients are properly maintained for secure collaborative operations. Within the system, policies may be set for different aspects of a user's context, including: computer or software platform, network connections, locations (physical context), social and work behaviors (cultural and social contexts), and the nature of the information being shared or built (information context). As extensions to this work, historical contexts relating the nature of activities over a period of time may also be used as a trigger for precautions or allowances in collaborative environments. In the current implementation, a variety of software-based sensors are used to determine the context of the user. Security agents running on client computers are responsible for controlling and monitoring these sensors. The security agents also communicate with a policy agent during collaborative interactions to enforce the security policies. Depending on the policies and the client operating contexts, the security agent may prevent or proactively enable computer activities locally, or deploy services in support of secure collaborative operation. An additional challenge associated with this work is that we want to secure operations in a pre-existing collaborative software (an e-manufacturing application). Our approach is to provide security enforcement while minimizing the impact upon the existing legacy e-manufacturing application.

This paper is organized in the following way. Section 2 describes the problem we are addressing with this work. With a description of the target domain in place, we describe our approach in general, detailing the architecture, and implementation in Section 3. In Section 4 we describe relevant previous research related to this area. Our Discussion and Conclusions section follows in Section 5

2 Problem Statement

There have been a great many technologies developed to help groups of people collaborate more effectively. Networked computers form the basic substrate upon which different collaborative technologies have been built. Computer Supported Collaborative Work (CSCW) research involves investigation and development of approaches that make collaboration between users using this substrate more effective. Advances of CSCW have been applied to cooperative design and engineering.

Indeed, computer networks have become prevalent in all organizations. While organizations have been able to gain advantages in efficiencies and their work through their use, inter-networked computer systems also present a risk to the operation of organizations. In terms of cooperative design and engineering, a key concern is the assurance that proprietary information about the intellectual property owned by the organization or information about the company operations is available only to authorized individuals. Within an intranet environment, access privileges may be adequately controlled. Interconnecting intranets over the Internet to allow different organizations in different locations to collaborate, as would be the case for cross-organization collaboration, and/or design/production outsourcing, creates a liability in terms of the potential for unauthorized access to information, computers or devices on the company intranet.

Internet-based manufacturing involves sharing intellectual property in the form of detailed engineering and manufacturing information as well as competitive information in the form of order and costing details. The bottom-line here is that for general acceptance of an Internet-based cooperative design and engineering approach, the secrecy of the proprietary or competitive information must be maintained.

In addition to maintaining secrecy, Internet-based manufacturing must accommodate confidentiality of the organizations involved in the manufacturing process. Gathering and processing information about the activities of individuals or groups while managing or operating processes or machinery via computer networks can provide considerable detail concerning the ways in which the individuals interact as well as process-related information. In a highly competitive manufacturing environment, information about internal organizational operations must only be shared on a "need–to-know" basis. This work addresses the following questions:

- While the organizations involved in the collaborative work may have written policies describing how all participants are expected to behave, how will those policies be enforced?
- In a client-server context, can security policies apply security constraints on an end user's operating environment and based on the context of the end user's operations?
- Is it possible to add security enforcement to existing or legacy applications easily?

The next section describes our approach for answering the above questions.

3 Our Approach

Our approach is to apply policy-based security enforcement for a client-server application. Policies are created and managed centrally at the server. There are sensors at the client that measure whether or not each central security policy is maintained in compliance on the client platform. Enforcement and feedback to the user through the client application is done via the server. We detail first the design for EASE as applied to a client-server application, followed by a description of the implementation.

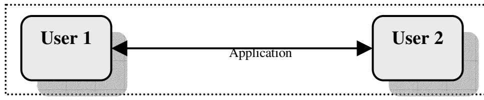
3.1 Design

In order to understand the design approach we have taken for EASE, we first describe the overall concept, then provide an example in the form of a specific implementation targeting an e-manufacturing application.

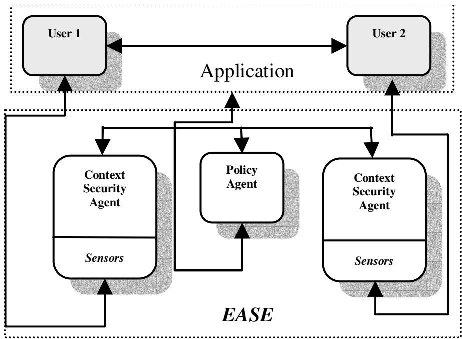
Figure 1 illustrates a system before and after being deployed with EASE. Illustrated in Figure 1(a) is the system to be protected. It may be a client-server or distributed application. The application may be designed for any particular purpose and may or may not have security functions built into its design. Indeed, while the system may include, for instance, methods to authenticate users, traditionally there are few methods available to enforce security compliance at the client end.

Figure 1(b) illustrates the application in operation with EASE in place. A policy agent, located at the server site, interacts with a security agent located on each platform where the application may be operated remotely. The policy server interprets an electronic version of the security policy for the applications that may run on the computer platform. The security agent and policy agent communicate with each other via a dedicated, secure communication channel. The policy agent communicates to the application in order to share information through messages and to enable or disable access to the application for any user. In order for the application to operate from any computer platform, the security agent must be present and must be in communication with the policy agent. Each security agent has a set of security (software) “sensors” that monitor different aspects of operation context for the application on its computer platform. Examples of the security policies that sensors and controlling software may monitor and enforce include the following:

- Ensure the operation environment is appropriate for the application. For instance for a Java application, it is important to ensure that the Java runtime version installed is adequate for the application to run correctly. In addition, to prevent security flaws and exploits, the system must ensure that the version of Java installed has not been tampered with.
- In order to protect distribution of copyrighted material or trade secrets, the system must ensure that when the application is running there are no other applications operating that may be used to garner inappropriate access to intellectual property or the application itself. These applications may include disassemblers, reverse compilers, screen snapshot software, file or system activity monitoring software, etc.
- Ensure that the computer platform running the application remotely is safe from viruses and Trojans. As with the above, this would amount to an enforcement of the security policy before the application is allowed to run.
- Ensure that only certain computers run the application. There may be other measures that can restrict access to the application. The environment’s security policy may restrict computers based upon IP address, MAC address, or computer hardware signatures.
- Ensure that the computer upon which the application runs remotely always has a security device installed. (The device may be a smart card, USB security device, other security dongle (wired or wireless). Security enforcement runs in a separate execution thread from the application, regularly.
- The policy agent may add further functions for improving authentication and authorization for the application. For instance, it may enforce password changes, or role-based authentication and authorization. The advantage of adding this function would centralize security administration for the application.
- Another item that may be monitored and reacted upon is an analysis of the situational behavior of all participants. If a user is considered to be behaving in an anomalous way, the application may disconnect the user. In other words, users will remain connected as long as they behave in the fashion expected by the administrators of the collaborative software system.



a) The application environment with two users, 1 and 2 connected through the application



b) The application environment with two users employing EASE.

Fig. 1. This diagram illustrates how EASE may be applied to an existing application

3.2 Implementation

To demonstrate EASE, we have created a software prototype. This section describes the implementation of the prototype.

This work builds upon the research and prototype development described in [3]. The web-based shop-floor monitoring and control program is called Wise-ShopFloor and is further described in [4]. A much simplified block diagram for the Wise-ShopFloor is shown in Figure 2. Web clients may access a variety of shop floor

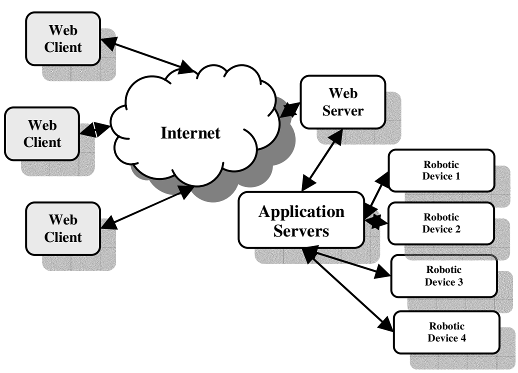


Fig. 2. Three-tiered architecture for e-manufacturing (Wise-Shopfloor) before the addition of EASE

equipment (e.g. robots, milling machines) via a web interface. Users are provided with three dimensional visualizations of the operations of machinery for control or operation purposes. The system uses a client-server architecture. The clients use a web browser. Java 3D provides the 3 dimensional visualizations of all machine operations. Java servlets form an application server used to access the various different machine tools. Before EASE was applied to the prototype, the primary security feature of the Wise-ShopFloor application was based on simple password authentication access control. The objective of our work was to add monitoring of the contexts of application execution and using the monitoring results to control context-aware access to the application.

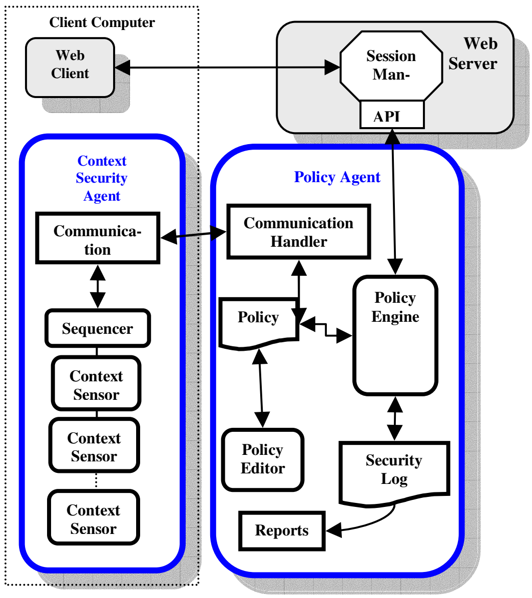


Fig. 3. This diagram illustrates the components comprising the system we build to provide Environment-Aware Security Enforcement (EASE)

Figure 3 illustrates our approach for adding EASE to the Wise-ShopFloor application. The Session Manager is the key place connecting EASE with the application. The interface consists of a simple application programmer interface (API). Figure 4 illustrates some of the components of the Policy Agent (PA) and Context Security Agents (CSA). The PA interprets policy, communicates with all context CSAs (one for every computer connected remotely to the application). Communication handlers at both the PA and CSA exchange commands, alarms and policy information. Depending on the nature of the policy, it is either interpreted at the PA where upon, the PA sends a message to the CSA to perform a policy compliance check, or it is sent to the CSA if a policy is to be scheduled regularly. In the latter case, the CSA schedules the appropriate sensor to perform the test periodically, signaling the PA when the test fails. When a sensor test fails, the PA sends a message to the session manager to deal with the consequence of the failure and ultimately provide feedback to the user. A policy statement in its simplest form contains the name of the sensor test, the frequency of the test, the consequence of its failure and the message that is sent to the Web Client for display as a pop-up window indicating a security failure. The failure may result in a warning to the user, or a message indicating that the application has disconnected.

The PA also houses a policy editor for creating and changing policies. Each rule is named and applied to operate with a single machine connected to the Wise-ShopFloor Application. Rules may also be applied to users. The rule is applied to one of the two tasks associated with most machines: Monitor or Control. The action associated with the rule may be either to permit or deny the selected task on the machine. As many sensors as required may be added to the rule. For each sensor, the administrator selects the sensor test to be applied, the expected response for the test, and the message to be sent to the application user if the test fails.

4 Previous Work

Two general research areas attributable to EASE are policy-based management (EASE uses policy rules to manage its operation) and pervasive/ubiquitous computing security (EASE enforces security based upon user and computer platform contexts).

McDaniel and Prakash have described an architecture for security policy enforcement [5]. Named "Antigone", the architecture offers a modular approach for adding security event detection modules. The system uses a transport layer mechanism and security-related events for handling by the detection modules. The authors describe optimization methods to reduce overheads in the architecture. Antigone is intended to be built around applications. With Antigone, the intended target for the security enforcement is appropriate behavior of the application as opposed to EASE where the target for enforcement is beyond the application extending to the context of work station operation and user behavior.

An example of an attempt to provide security enforcement for a computer platform at least at the level of the file system is given by Wolthusen [6]. The system described in this paper provides mandatory access control, encryption and auditing of file activity on an individual file basis for a distributed system. The system that the author has developed provides a holistic approach for handling file activity for the Windows NT file system. Ostensibly, the technology could be applied as a file context sensor for EASE, taking advantage of our effective management interface. While this approach would provide little or no advantage for the Wise-ShopFloor application, it would be advantages when EASE is adapted for distributed applications that have local file management requirements.

Schneider describes a practical way to enforce security policies by monitoring system and application processes by automata for safety-critical systems [7]. Each automaton is intended to deal with a security policy. The author describes and defines security policies as being "specified by giving a predicate on sets of executions. A target *S* *satisfies* security policy *P* if and only if $P(\sum_S)$ equals *true*." This definition applies well in EASE as well, since the decomposition of what might be complex policies into the conjunction of separate mechanisms used to enforce each of the component parts holds for our work. In addition, EASE targets enforcing security policies for distributed applications, whether they are safety critical or not. Moreover, EASE is not concerned with the policies for the application, per se, but rather those security policies applied to the execution environments for applications. The context sensors in EASE are automata-like in operation.

Covington et al. describe a context-aware security architecture (CASA) for emerging applications [8]. Similar to our work, CASA employs different sensors that monitor resources, systems and physical sensors to measure different contexts for participants in the application. CASA also uses context and object management layers as well as services for object and environment roles activation that influence an authorization service. The overall objective of this work is to provide more adaptive security services. With EAVE, we present and demonstrate a straight-forward approach for adding and managing security for a distributed, shared workspace using context and an agent-based methodology.

5 Discussion and Conclusions

There has been considerable research and development of policy management systems for network management and security management. However, there has been little work in the area of enforcement of security policies. The work that has been done in this area often involves incorporating security enforcement into the systems early on; at the beginning of the design process. In this work we take a new approach we call: Environment-Aware Security Enforcement (EASE): a fusion of work in policy-based management and enforcement and the context-based security research in pervasive computing. EASE offers a means of providing security policy enforcement for legacy applications. As was shown with our target Wise-ShopFloor application, the interface with the existing application was a simple API added to the legacy system to provide control over a user's authorization to use the application and messaging through the application for user feedback. The result nicely integrates the security enforcement functionality of EASE with the original application functionality.

EASE extends the security functionality of the target application to which it is applied by providing context-based security enforcement features. A variety of logical sensors may be applied to different aspects of a user's operating environment. Tests and enforcement measures may be taken to assure that the user is maintaining certain security requirements for operating an application even if the location of the user is outside the organizational physical and virtual boundaries.

The current proof-of-concept prototype demonstrates a limited number of sensors in operation. These sensors currently include: Java virtual machine integrity, operating system version, MAC and IP addresses. Future work will involve developing a more comprehensive set of sensors, including ones to determine whether or not an individual user is doing inappropriate or unexpected activities based upon system call information and social network analysis [9].

Acknowledgements

The authors acknowledge the National Research Council of Canada for its support of this work through their collaborative grant program with the National Science Council of Taiwan (2002-2005). We gratefully acknowledge the contribution of our collaborators in the NRC Integrated Manufacturing Technologies Institute for constructing the hooks in their prototype software required for our implementation. We also acknowledge the assistance of Myroslav Palenychka in the development of the prototype during his Jan.-Apr., 2005 work term within the Information Security Group.

References

1. Fuchs, L, Geyr, W., Richter, H., Poltrock, S. Frauenhofer, T., Daijavad, S. Enabling intercompany team Collaboration. Enabling Technologies: Infrastructure for Collaborative Enterprises, 2001. Proceeding tenth IEEE International Workshop on WET ICE 2001, Cambridge, June 20-22, 2001, pp. 374-379.
2. Xu, Y., Song, R., Korba, L., Wang, L., Shen, W., Lang, S., Distributed Device Networks with Security Constraints, IEEE Transactions on Industrial Informatics, 2005 (Accepted).
3. Shen, W., Lang, S., Korba, L., Wang, L., and Wong, B., 2000, "Reference Architecture for Internet Based Intelligent Shop Floors," Proceedings of the SPIE International Conference on Network Intelligence: Internet-Based Manufacturing, Vol. 4208, pp. 63-72.
4. Wang, L., Song, H. Cunningham, A. Development of Wise-ShopFloor for Web-based Monitoring and Control, Technical Report from the Integrated Manufacturing Technology Institute, National Research Council of Canada, IMTI-TR-023 (2004/02) January, 2004 (41 pages).
5. McDaniel, P., Prakash, A. A flexible architecture for security policy enforcement, Proceedings of the DARPA Information Survivability Conference and Exposition, Washington, DC, April, 2003, pp. 234-239, vol. 2.
6. Wolthusen, S.D. Security Policy Enforcement at the file system level in the Windows NT operating system family, Proceedings of the 17th Annual Computer Security Applications Conference, New Orleans, Dec 10-14, 2001.
7. Schneider, F.B. Enforceable Security Policies, ACM Transactions on Information and System Security (TISSEC), Vol. 3, Iss. 1, February, 2000, pp. 30-50.
8. Covington, M.J., Fogla, P., Zhan, Z., Ahamad, M. A context-aware security architecture for emerging applications, Proc. of the 18th annual Computer Security Applications Conference (ACSAC'02), Las Vegas, Dec. 9-13, 2003, pp. 249-258.
9. Temdee, P., Korba, L. Of Networks, Interactions and Agents: An Approach for Social Network Analysis, Proc. 6th International Conference on CSCW in Design, London, Ontario, July 12-14, 2001.

Grid Workflow Scheduling Based on Task and Data Locations*

Yujin Wu[1], Ning Gu[1], Yuwei Zong[2], Zhigang Ding[2], and Shaohua Zhang[2], Quan Zhang[2]

[1] Department of Computing and Information Technology, Fudan University, 200433, Shanghai, China.
{yujinwu, ninggu}@fudan.edu.cn
[2] Shanghai Development Center of Computer Software Tchnology, 201114, Shanghai, China.

Abstract. Grid workflow systems provide mechanisms to execute complex tasks which consist of related sub tasks. Due to the intensive computing and data transferring in Grid workflows, the locations of tasks and data have great impact to the execution performance of Grid workflows. In this paper, we present a novel approach to search for optimal Grid workflow scheduling effectively. We model workflow execution with fetching input data and running tasks, and present a optimized scheduling searching algorithm based on simulated annealing, which can find neighborhood scheduling fast. The experimental results show that our approach is effective and scalable.

1 Introduction

With the development of Grid computing technologies, it is possible to share resources and solve complex problems among dynamic, multi-institutional virtual organizations coordinately [1]. Grid workflow systems are Grid services responsible for constructing, verifying, execution and monitoring distributed workflows in Grid. Due to the large quantity of Grid resources with different speed, memory and network bandwidth, it is very difficult to schedule workflow tasks on Grid. Two factors must be considered in Grid workflow scheduling:

Task Location: workflow tasks can be run on different resources. Generally resources with higher performance run tasks faster. However, there may be more tasks assigned on fast resources, so later assigned tasks need to wait until previous tasks complete. A "fast and idle" resource may be more suitable for the task.

Data location: workflow tasks usually need some input data. If a task need to use data in other Grid resources, the data must be transferred to where task will be executed in. Consequently, if the location of data is "closer" to the task, the task can be started and completed earlier. In data-intensive workflows, data location has great effect on workflow's execution performance.

---

* Project Supported by the National Natural Science Foundation of China（Grant No.: 60473124）, 863 Hi-Tech Research and Development Program of China (NO: 2004AA1Z2390) and SEC E-Institute: Shanghai High Institutions Grid.

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 149 – 156, 2005.

In this paper, we present a novel approach to address these two factors in Grid workflow scheduling problem. We model a Grid task with two phases: fetching input data and executing the task. Then WETC algorithm is presented to compute a workflow's execution time under a given scheduling. To reach the short execution time, SASS and EWETC algorithms are presented to search for optimal scheduling in Grid with simulated annealing approach. Finally, we experimentally evaluate the model and the algorithms presented.

The rest of this paper is organized as follows. In Section 2, related work is discussed. The workflow scheduling model is presented in Section 3, and scheduling search process is described in Section 4. Experiment evaluations are presented in Section 5. Finally Section 6 contains our conclusions and directions for future work.

2 Related Work

Scheduling mechanisms are quite different in various Grid workflow systems. ICENI[4] predicts each task's running time on a specific resource, computes total cost of workflow, and searches for scheduling with lowest cost. However, it does not consider the data location and frequent data transfer within tasks. GridFlow[2] model workflow scheduling in two levels, and simulates workflow execution like a "just-in-time" way, so it can not provide the global optimal execution performance. Pegasus[3] map workflow tasks to resources before execution, but performance issue is not considered. In some other systems, tasks are scheduled to "best idle" resources just-in-time, which tries to reach the single task's best performance. This scheduling mechanism is simple and fast, but in most cases it is not optimal for the global workflow.

3 Grid Workflow Scheduling Model

In this section, we first present the model of Grid and workflow. Based on the model the WETC algorithm is presented to compute workflow's execution time under a specific scheduling.

3.1 Grid Workflow Model

We use DAG as our process model, which is used in most Grid workflow systems. First we define files used in workflows. We use F to denote all files in a workflow. Let $F=\{f_1,f_2,\ldots..f_m\}$ and let $|f_i|$ denote the size of file f_i. Next we define the Grid task.

Definition 1. A Grid task T_i is a tuple (I_i,C_i,O_i) where I_i is the set of input files, O_i is the set of output files, and C_i is the computing task T_i need to perform. We use T denotes all tasks in Grid. Not losing generality, we assume tasks in T are ordered.

Definition 2. A Grid workflow is a directed acyclic graph W = (V, E). V is the set of vertexes and E is the set of edges.

Each vertex in V is a Grid task T_i, so V is also the set of tasks in workflow W. The directed edge $(T_i , T_k) \in E$ means the start time of T_k must be equal to or later than the end time of T_i. We say T_i is a *previous task* of T_k and T_k is a *subsequence task* of

T_i. Let prev(T_i) = { previous tasks of T_i in W} and subs(T_i) = { subsequence tasks of T_i in W}. Obviously, only when all tasks in prev(T_i) finished, T_i can start, which we say T_i is *runnable*.

Definition 3. A Grid resource R_i is a binary tuple (P_i, B_i), where P_i is the computing power of R_i, and B_i is the set of Grid tasks that can be run in R_i.

A task running on a Grid resource has two phases: first, getting all required input files from the nearest resources where contain these files; second, performing the computing and write output files to disks in current Grid resource.

Definition 4. A Grid is a directed connected graph GR = (R, BW). R is the set of vertexes and BW is the set of edges in GR. Let R = {R_1,R_2,…R_u}. BW is the set of edges.

Each vertex in R is a Grid resource, so R is also the set of Grid resources in Grid GR. To each directed edge (R_i , R_j) $\in$ BW, we use bw(R_i , R_j) to represent the data transferring speed from R_i to R_j. Not losing generality, we set bw(R_i , R_i) = ∞. Now the model of Grid workflow has been defined. Moreover, we need to formally define Grid workflow scheduling.

Definition 5. Let p is a mapping from T to R denoted as *p: T->R*. We say p is a scheduling S of Grid workflow W, if for each t$\in$T and p(t) = (P_i,B_i), t$\in B_i$.

The Grid workflow model defined above explicitly regards locations of files and tasks as factors in Grid workflow execution and scheduling. Based on the model, the execution of a task in workflow is not only related to its previous tasks, but also related to tasks and resources providing files it used.

3.2 Workflow Execution Time Computing

Given a specific scheduling S of workflow W, we need to compute the execution time of W under S. First, we compute the execution time of a single task. It is the sum of transferring time of input files and the task's computing time, which can be described by follow formula:

$$runtime(T_i, R_j) = \sum_{f \in I_i} \left(\min \left\{ \frac{|f|}{bw(R_k, R_j)} \middle| f \ exists \ \ in R_k \right\} \right) + \frac{C_i}{P_j}$$

In order to describe the execution process of task T_i, we use three parameters *at(T_i), st(T_i), et(T_i)* which denote the time when T_i is runnable, the time when T_i starts to execute, and the time when T_i is completed. The runnable time at(T_i) is also the time when Grid workflow system assigns T_i to a Grid resource.The following definition formally describes the assigned order of all tasks in the workflow.

Definition 6. *A binary relation R_T* on workflow task set V is defined as follows: for each t_1, $t_2 \in$V, if at(t_1) < at(t_2), or at(t_1)=at(t_2) and order(t_1)<order(t_2), then (t_1, t_2) $\in R_T$, and it is donated as $t_1 < t_2$, otherwise (t_1, t_2) $\notin R_T$.

To describe tasks' enabling time in Grid resources, we use parameter *ct(R_i)* to denote the time when resource R_i has executed some tasks and wait for execute next task. The computing procedure is outlined in Algorithm 1.

The basic idea to compute workflow execution time is to compute every task by the order in R_T. In WETC, WaitingList stores the tasks waiting to be computed currently. WETC selects the minimum task in WaitingList to compute, and add its subsequence runnable tasks into WaitingList. The computing is repeated until WaitingList is empty. Tasks' maximum endtime is the workflow's execution time.

Algorithm 1. WETC(W,S)
Input: Workflow W, Scheduling S
Output: the workflow execution time

```
for each Ri,  let ct(Ri) = 0;
WaitingList =φ,
put all runnable tasks into WaitingList;
while ( WaitingList not empty) do
   select task t∈WaitingList, for all t1 ∈ WaitingList − {t}, t < t1;
   remove t from WaitingList;
   st(t) = max( at(t), ct(p(t));
   et(t) = st(t) + runtime(t , p(t));
   ct(p(t)) = et(t);
   for each t1 ∈ subs(t)  do
         at(t1) = max(at(t1), et(t));
         if  t1 is runnable then add t1 to WaitingList;
end while;
return max{et(t)}.
```

The following theorem shows the time complexity of WETC algorithm.

Theorem 1. Suppose each task uses no more than K files, the time complexity of WETC is O(nlogn +nuK + m).

4 Scheduling Search

The Grid workflow scheduling problem is a combinational optimization problem. Due to the large problem space, it is very difficult to find an optimal scheduling quickly. To reduce the searching time, we need to reduce the number of tested schedulings, and moreover, to reduce the time spent on workflow execution time computing under each scheduling.

4.1 Scheduling Search Based on Simulated Annealing

Simulated Annealing[5] is a generalization of the Monte Carlo method used for optimization of multi-variable problems. It is a heuristic neighborhood searching approach to find optimal solutions by the similarity of physics annealing process and solving process of combinational optimization problem. Here we apply simulated annealing to workflow scheduling search problem.

4.1.1 SASS Algorithm

In order to apply simulated annealing to Grid workflow scheduling search, we need to formally define neighborhood structure in scheduling space.

Definition 7. For Scheduling S_1 and S_2, S_2 is a *k- neighborhood scheduling* of S_2 and vice versa if there are only k tasks whose assigned resources are different in S_1 and S_2.

Algorithm 2. SASS(W)

```
Input: Grid workflow W
Output: Optimal scheduling S
generate random initial scheduling S;
CurrentExecutionTime = WETC(W,S), k = 0;
repeat
    for i = 0 to trialSteps do   /* a MAPKOB chain */
        generate 1-neighborhood scheduling NS;
        newExecutionTime = WETC(W , NS);
        if Metropolis_test() then
            S = NS;   /*  NS is accepted  */
            CurrentExecutionTime = newRuntime;
    k = k +1;
    compute TPk;
until no new scheduling accepted;
return the best scheduling found.
```

Algorithm 3. Metropolis Test Algorithm

```
Input: newExecutionTime ,CurrentExecutionTime, temperature TPk
Output: whether NS can be accepted
if (newExecutionTime < CurrentExecutionTime) then  return true;
  else
    deltaTime = newExecutionTime–CurrentExecutionTime;
    if exp( –deltaTime / TPk ) > random[0,1)  then  return true;
        else  return false.
```

The scheduling search is described as follows. First, we randomly generate a scheduling S. Then, scheduling NS, which is a 1-neighborhood scheduling of S, is generated. If the execution time under NS is less than that under S, then NS is accepted. Otherwise, NS is accepted by probability $e^{-dT/TP}$. Here dT is the difference of workflow execution time between S and NS, and TP is the control parameter of simulated annealing, which is called annealing temperature. dT and TP decide the probability by which a worse scheduling is accepted, as shown in Algorithm 3. The generation is repeated, and when a MAPKOB chain of generations are processed, TP is decreased, and therefore, a worse scheduling is accepted by lower probability. Finally, when no new scheduling is accepted in a full MAPKOB chain, which implies finding new better solution is nearly impossible, the searching procedure ends and the best scheduling the algorithm has ever generated is returned as the optimal solution. SASS algorithm is outlined as follows.

4.1.2 Annealing Parameter

Two parameters – annealing temperature TP_k and length of MAPKOB chain trialSteps impact the performance of SASS. Research on simulated annealing[5] shows that the value of TP_0 should be large enough to make initial acceptance ratio

$\chi_0 = e^{-dT/TP} \approx 1$. We set dT = workflow_runtime(W,S_0) / 5, and $\chi_0 = 0.9$ here. Then we have $TP_0 \approx 2$ * workflow_runtime(W,S_0)

Moreover, the attenuation of TP_k should be small. We set the attenuation function to be $TP_k = 0.9 * TP_{k-1}$ and trialSteps = n.

4.1.3 Time Complexity Analysis

Section 3 shows that the time complexity of algorithm WETC is O(nlogn+m). Algorithm SASS computes WETC for W*n times, which W is the number of MAPKOB chain computed. So the time complexity of SASS is $O(W(n^2\log n+mn))$.

4.2 Neighborhood Execution Time Computing Optimization

Simulated annealing can reduce the workflow execution time computing greatly. In order to do further optimization, we need to reduce time computing under each generated scheduling. We find that the computing process under two neighborhood schedulings S_1 and S_2 have much relativity. Using the relativity, we can optimize the computing on S_2 based on the computing process of S_1.

Theorem 2. Suppose scheduling S_2 is a 1-neighborhood scheduling of S_1, and the resource task t assigned in S_1 and S_2 are different, then in SASS, the executions under S_1 and S_2 before t in ScheduledList are identical.

Definition 8. A *snapshot* of scheduling S on task t denoted as *SNAPSHOT(S,t)* is the set {WaitingList, {at}, {st}, {et},{ct}} before the execution of task t is computed.

Algorithm 4. EWETC(W , LS, S)
Input: workflow W, last computed scheduling LS, current scheduling S
Output: the workflow execution time

```
suppose the resources that task t  is assigned to are different in LS and S;
restore  SNAPSHOT(S,t'),  t' is the nearest snap point before t;
while ( WaitingList not empty) do
      find minimum task t and computes t in the same way as WETC;
      if  the position of t is multiple of n/K  then
            save SNAPSHOT(S, t);   //  the position is a snap point
return max{et(t)}
```

Algorithm EWETC adds snapshot operation compared with WETC. EWETC makes a snapshot every n/K tasks, so K snapshots are saved in the computing of scheduling S. When beginning computing the 1-neighborhood scheduling NS, if S and NS assign task t to different resources, EWETC restores the nearest snapshot SNAPSHOT(S,t') before t and computes execution time of other tasks. By theorem 2 we have SNAPSHOT (S',t') = SNAPSHOT (S,t'), so the computing of S' can start from t'.

5 Experiment Evaluation

We do scheduling simulation on workflows whose number of tasks ranges from 20 to 1000 and compare the quality and execution time of the generated schedulings. The number of simulated Grid resources is from 6 to 50 and the bandwidth of network connection between resources is from 10Mbit/s to 200Mbit/s. The simulation program is written in Java and run on a PC with P4 1.5G, 256MB memory and Linux. Results are shown in Table 1 and Figure 1.

Table 1. Scheduling Searching Time (ms)

Task	20	50	100	500	1000
Resource	**6**	**8**	**10**	**30**	**50**
SASS	81	681	4,120	231,020	1,593,572
SASS +EWETC	75	582	3,419	150,609	1,069,219
Random-K	109	832	4,087	198,732	1,328,703

From Table 1, we can draw the following conclusions. With respect to algorithm running time, basic SASS algorithm is close to Random-k. Due to the use of snapshot, the execution time of the SASS + EWETC exceeds that of basic SASS when number of tasks is small. But with the number of tasks increasing, the efficiency benefit of snapshot begins to take effect. Finally, the execution time of SASS+EWETC is about 2/3 of that of basic SASS.

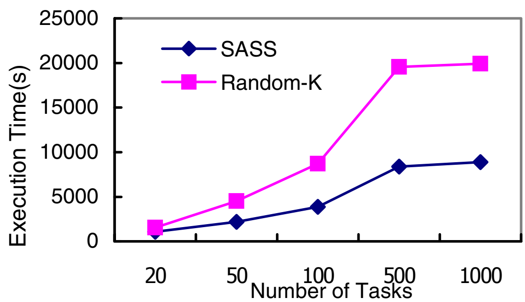


Fig. 1. Execution Time(s)

Considering the results of the algorithms shown in Figure 1, comparing SASS with Random-k, the result of the former is far better than the latter. Especially when the number of tasks increases, the advantage is more significant: the execution time of the scheduling generated by SASS is less than half of that of Random-k. This shows that SASS is quite effective.

We also implement SASS and EWETC in GFlow[6,7], a Grid workflow management system prototype. Some applications like distributed sorting running on GFlow work well under SASS and EWETC algorithm.

6 Conclusions and Future Work

The scheduling of Grid workflow system is essential to the workflow execution performance, and locations of data and tasks have much impact on workflow execution performance. This paper presents a workflow scheduling approach which looks data and tasks location information as important factors in workflow scheduling, and present optimized scheduling algorithm based on simulated annealing to reduce scheduling searching time. Simulation experiments show that the approach produces both good execution performance and scheduling results. In our future work, we are going to add more complex organizational form of Grid resources and Grid expense model into Grid workflow scheduling to further improve Grid workflow scheduling performance.

References

1. I. Foster and C. Kesselman, The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, 1999.
2. J. Cao, et al. GridFlow:Workflow Management for Grid Computing. In Proc. of the 3rd IEEE/ACM International Symposium on Cluster Computing and the Grid, pp. 198-205, 2003.
3. Ewa Deelman, James Blythe, et al. Pegasus: Mapping Scientific Workflows onto the Grid. Across Grids Conference 2004, Nicosia, Cyprus.
4. Laurie Young, Stephen McGough, et al. Scheduling Architecture and Algorithms within the ICENI Grid Middleware. In Proc. Of UK e-Science All Hands Meeting 2003,pp 5-12.
5. S. Kirkpatrick, C.D. Gellat and M.P. Vecchi. Optimization by Simulated Annealing. Science 220(1983),pp 671-680.
6. Shaohua Zhang, Yujin Wu, et al. Grid Workflow based on Performance Evaluation. In Proc. of the 2nd International Workshop on Grid and Cooperative Computing (GCC 2003).
7. Shaohua Zhang, Yujin Wu, et al. Adaptive Grid Workflow Scheduling Algorithm. Lecture Notes in Computer Science, Vol. 3252. Springer-Verlag, Berlin Heidelberg, New York(2004): 140-147.

The Construction Management Cooperated with Clients Using a Parametric Information Design Method

Koichi Aritomi[1], Ryosuke Shibasaki[2], and Nobuyoshi Yabuki[3]

[1] Senior Researcher, National Institute for Land and Infrastructure Management (NILIM)
Ministry of Land, Infrastructure and Transport, Japan
aritomi-k86qi@nilim.go.jp
http://www.nilim.go.jp/lab/qbg/eindex.htm
[2] Professor, Center for Spatial Information Science, University of Tokyo, Japan
shiba@skl.iis.u-tokyo.ac.jp
http://shiba.iis.u-tokyo.ac.jp/index-j.html
[3] Associate Professor, Department of Civil Engineering and Architecture,
Muroran Institute of Technology, Muroran, Hokkaido, Japan
yabuki@news3.ce.muroran-it.ac.jp
http://yb3.ce.muroran-it.ac.jp

Abstract. This paper presents a new construction management cooperated with clients using a parametric information design method for road construction. This method aims for re-designing and for managing to achieve the conformance of the completed form of roads to planned design with survey equipment. By using this method, we design a road with an information model that has alignments, profile and cross sections. The model can be changed according to the site condition in collaboration with the planning designer of the client. This method is able to manage the conformance of completed form accurately because the information model is precise. Furthermore, the file size is so small that the survey equipment can store the data, and it is easy to re-design the plan. Therefore, we can use the survey equipment easily for the conformance management of completed form accurately and to re-design a plan by using CAD quickly. The management by this method has been found to be efficient and accurate.

1 Introduction

Conditions at construction sites change frequently. Thus, the client makes changes of the previous design, negotiating with the contractors cooperatively. Then, the client confirms if the contractor has achieved the good conformance of the completed form with the design by checking many reports containing documents, photos and results of survey. Also the contractor has to aim for the completion by some negotiation and cooperation with the client. Otherwise the client cannot pay to the contractor.

However, the current re-design method is not efficient. Most contractors in Japan use the drawings that are modified manually in a traditional 2-D way because most clients in Japan have obliged the contractors to design with 2-D drawings. Then, the contractors have to check if the completed formation on the site conforms to the requirement of the client agreement by measurement manually [1].

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 157 – 165, 2005.

The parametric information design method of road design can re-design a previous design easily according to the site condition by using CAD. This information model embedded in the survey equipment can check conformance of the completed form of road construction efficiently and precisely so that we can re-design quickly and control conformances rationally [2].

We propose the new management method using the information model and show that it is efficient and accurate in this paper.

2 How to Manage Construction Cooperatively with Clients

2.1 What Is the Construction Management Cooperated with a Designer of the Client?

In the road construction, the first design is changed frequently according to the site condition because it is quite difficult to plan a terrain formation in advance precisely. Also, the planning to design detail terrain in advance precisely is an expensive task.

The contractor has to abide by the agreement with the client. If a contract between client and the contractor has some inconsistency with existing site conditions, the contractor has to report the fact that some inconsistency exists in the contract.

The client only orders the agreement to the contractor. If the client obtains such a fact report, he or she can change the agreement when he confirms the report is true.

In the current method, we have to make sure the report is true by measuring existing formation, make some drawings, and write many reports. It takes a long time to write the reports. According to the new terrain information, the designer of the client has to keep consistent with the existing site condition by re-designing the previous plans. The plans have many drawings including ground plans, profile sections, and cross sections. Sometimes, the contractor does the job that the designer of the client should do. This process is the construction management cooperated with a designer of the client traditionally.

Using the parametric information design method, this cooperation with the client is much easier. The construction management cooperated with a designer improves process of agreement between the client and the contractor.

2.2 Why Do We Have to Manage Conformance of Completion at the Construction Site?

One of the most important things for road construction management is conformance testing. The conformance testing aims to confirm if the completed objects of construction have achieved their purposes of the planned design. If the purpose was achieved, the client will pay to the contractor for completion work. The conformance test with the client is important in Japan.

The first example of conformance of completion is as-built management which confirms formation while a contractor is constructing a road. And the second example is that the client checks conformation of completed formation that the contractor made. The conformance management of formation is to confirm if the formation of road construction has matched with planned design. The contractor aims to get an acceptance of contract completion with the client by constructing roads completely. If the formation of road construction is not matched with the planned design, the client would not pay to the contractor.

The client usually has a standard for judgment to confirm completion. In Japan, an as-built management standard has been created by the client. The client can make sure if the formation differences of road construction to planned design is matched within the standard. The standard supports the judgment for reasonably close conformance of the completion. Therefore, the contractor could manage formation of road construction reasonably in the acceptable conformance by the client.

2.3 How to Use the Parametric Information Design Method.

The parametric design model can contain elements of 3-D composition. The elements are the alignment, the profile, and the cross sections. This model can compare the shape difference with planned design formation anywhere.

Once we get any formation of the road construction, we can compare any difference of the shape with planned parametric design (Fig.1.).

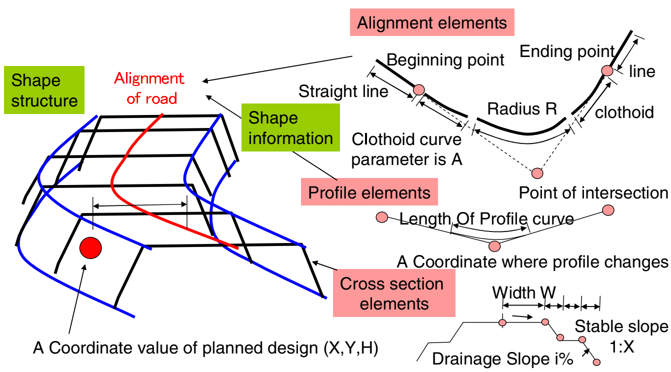


Fig. 1. Primitive Road Structure

2.4 The System Composition

The system composition we developed is shown in Fig. 2 and 3. The system uses information that has instruction, form work, expanded LandXML, and other files. The common module holds functions that are for file input/output, operation, graphical user interface. The common module can communicate with CAD system on the market by plug-in software or by API of the CAD system.

The designer of the client is able to use the system for parametric design. Once the designer creates information of parametric design, the worker in the construction phase can apply the information of parametric design to construction management. For example, the worker can make an instruction for stake out anywhere by using the parametric information. If the site condition changed, the client can re-design the previous drawings quickly. And the manager of the contractor can make sure information of working formation. The client can check the quality of the contractor's management with the agreement specifications by using information of parametric design. At the end of the construction, the contractor can make completion drawings (Fig.3).

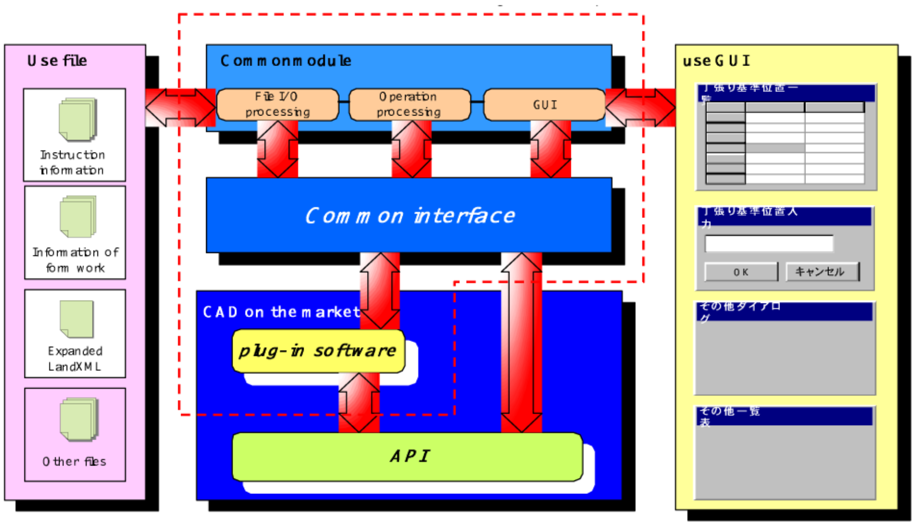


Fig. 2. Composition of experimental application

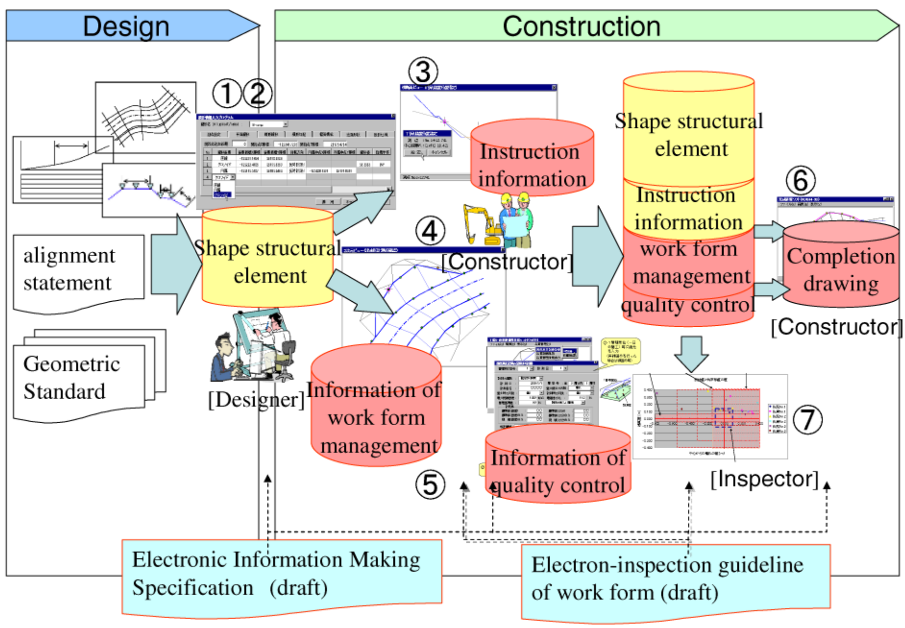


Fig. 3. Outline of experimental application

3 The Details of a Parametric Information Design Method

3.1 What Is Parametric Information Design?

Parametric design is a generic name for design by the variable operation. When this variable is manipulated, we can create the form of the purpose.

A parametric information design of this paper is a method to design information used in civil engineering. We proposed a primitive road information model for civil engineering. The primitive road information model is flexible and light. Nevertheless, the model has completely 3 dimensional forms of a road. We can re-design a road with this model as a parametric design method. We call this model "skeleton".

3.2 How to Design a Road with the Parametric Information Design Model

We made primitive road information with an expanded LandXML schema. LandXML schema is an international information model proposed by LandXML.org [3] and is a de facto standard as a terrain model. LandXML has many elements, e.g., units, coordinate system, Alignments, surfaces, etc. These elements are used to define terrain, road formations for civil engineering. We expanded this LandXML schema by adding four elements. They are "specifications", "CrossSections", "construct information", and "inspections". The "specifications" element is a set of specifications for construction requirements. The "CrossSections" element is basic information of design.

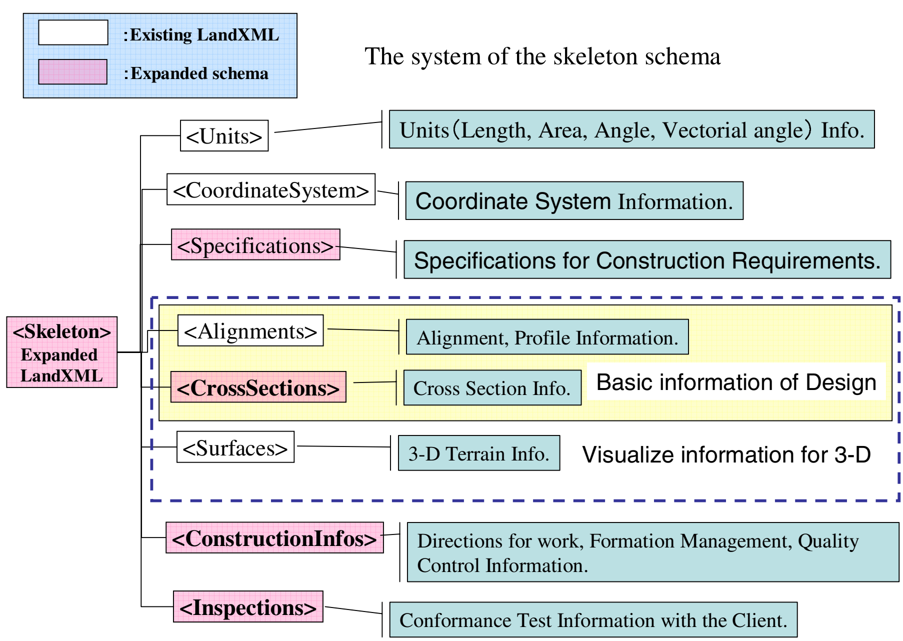


Fig. 4. The system of a skeleton schema

The "construction info" element is information to make at the construction stage. The "inspections" element is information inspected by contractor or client. These elements are shown in Fig.4.

Particularly, the "CrossSections" element is unique. The original LandXML has a cross section element, which is in "alignment" element. The element of original LandXML is for drawing cross sections with each station number, which usually has span of 20 meters. We call this way definition of section (where we call this type "pattern A" for shorten). The expanded "CrossSections" element is similar to "GradeModel" element of original LandXML. The "GradeModel" has elements composed of the order of the row and the each element of shapes. We call this way definition of element (where we call this type "pattern B" for shorten). These elements define parameters that are width, height, and inclination that change with each length (Fig.5). The current CAD system like Civil3D (AutoCAD) cannot show 3D models with this "GradeModel". We made an application system that can show 3D models with the "CrossSections" referring to "GradeModel".

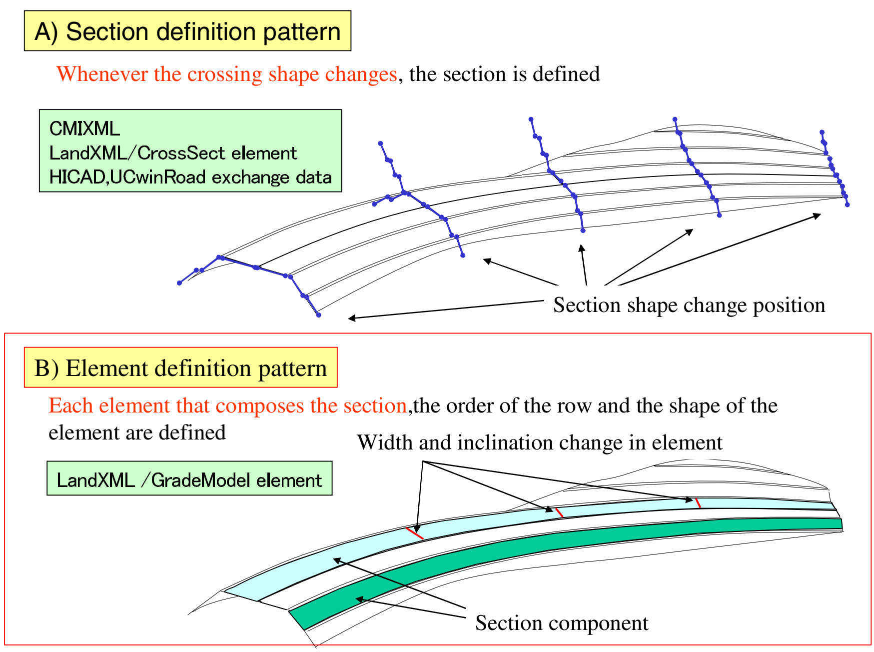


Fig. 5. The pattern of the "CrossSections" element of a skeleton schema

3.3 The Definition of the Parametric Model with the Expanded LandXML Schema

A sample of "CrossSections" element is shown in Fig.6. This "CrossSecctions" element is a type of pattern B. This "CrossSections" element has some elements, i.e., "CrossSection", "CrossSlope", "SlopeChange". The "CrossSection" element refers to an alignment of which name is “B ramp”. The “CrossSlope” element has some attributes, which are names of “LeftLane”, slope type, default slope value of this slope. The “SlopeChange” element has some attributes. The "staStart" attribute is a start length of the point in which a slope starts. The "staEnd" attribute is a end length of the point in which the slope ends. The “startSlope” and “endSlope” elements are attributes of the slope.

```
<CrossSections>
<CrossSection alignmentRef="BRamp">
<CrossSlope name="LeftLane" slopeType="%" slope="-2.0">
          <SlopeChange staStart="64.111" staEnd="120.445" startSlope="-2.0" endSlope="
8.0"/>
          <SlopeChange staStart="120.445" staEnd="150.416" startSlope="-8.0"
endSlope="-8.0"/>
          <SlopeChange staStart="150.416" staEnd="206.749" startSlope="-8.0"
endSlope="-2.0"/>
</CrossSlope>
```

Fig. 6. A sample instance of definition “CrossSections”

4 Application of the Prototype System

4.1 Preparation of Information for Agreement with Using the Parametric Information Model

We developed an experimental system for creating drawings, using the parametric information model. We used Civil3D 2005 of Autodesk. We added some modules to Civil3D 2005 as plug-in. Then, we can show drawings of the parametric information model (Fig.7).

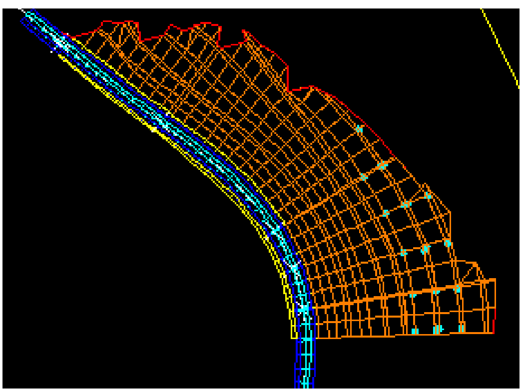


Fig. 7. A sample drawing rendered with parametric information model

4.2 How to Use the Prototype System

The site workers can check anywhere whether or not the formation is in good conformance with, using the parametric information. The process is as the following. (1) We install survey equipment that has parametric information at the site. (2) We calculate the point of the equipment. (3) At any cross section, we can set a target for the point that is compared with the planned form. (4) We see a target point at the site. (5).

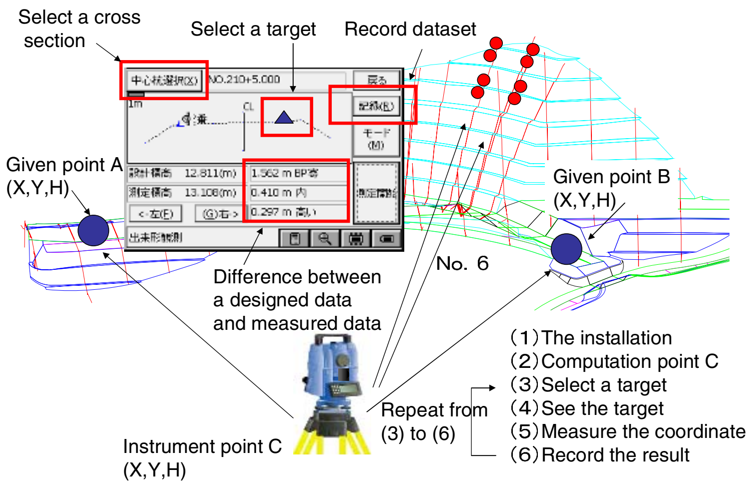


Fig. 8. Process of checking form at the site

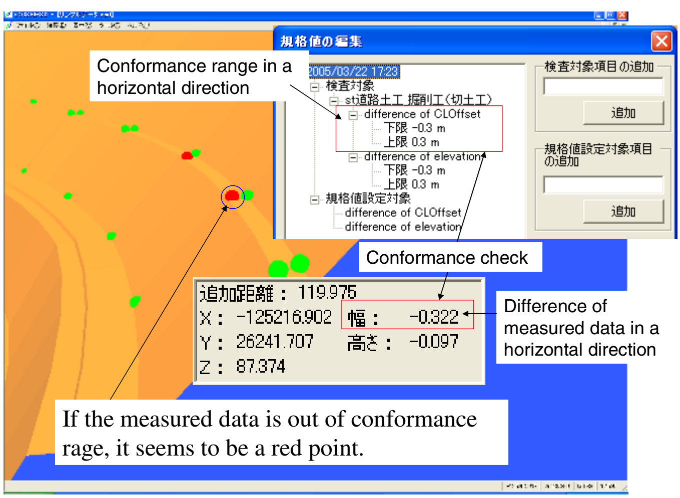


Fig. 9. A sample checking form on the experiment system

We measure a coordinate of the form. (6) We record the result of measurement (Fig.8). We repeat the process from (3) to (6) until all check points are measured. This prototype system was developed by expanding FC-1000 of TOPCON Corp.

By using the data given by the surveyor, the site manager can check the conformance of the site formation with the planned design (Fig.8). If the data is in good conformance, the check passes. If the data is not in good conformance, the check does not pass. We can define the range of conformance for reasonably close conformance. This prototype system was developed by expanding KOISHI-3D of KOISHI Corp.

5 Conclusions

In this paper, we showed that the cooperative construction management with designers using the parametric design model was useful, efficient and precise.

The parametric design model is able to contain the primitive road elements of 3-D composition. The elements are the alignment, the profile, and the cross sections.

We made the system that uses this information model containing instruction, form work, expanded LandXML, and other files. The designer of the client is able to use the system for parametric design.

Once the designer creates information of parametric design, the worker in the construction phase can apply the information of parametric design to construction management, e.g., ‘finishing stake’ instruction, form management, change of previous design, inspection, etc.

Then, the site workers can check anywhere whether or not the formation is in good conformance with, using the parametric information. Finally, the site manager can check the conformance of the site formation with the planned design.

Acknowledgement

We appreciate the cooperation of Dr. Katsumi Uesaka (NILIM), Mr. Nathan Crews (LandXML.org), Autodesk Inc., TOPCON Corp., KOISHI Corp., KOKUSAI KOGYO Co. Ltd..

References

1. Koichi Aritomi, Kensuke Matsuoka, Katsumi Uesaka, Tadashi Okutani: Proposal of an as-built management technology by three-dimensional design information, Journal of Construction Management Research, Japan Society of Civil Engineers, Vol.11, pp.81～90, 2004.
2. Koji Makanae: Application and Development of the Highway Design-Information-Model Utilizing the Parametric Alignment Model, Journal of Civil Engineering Information Processing System, Vol.13, pp.205-210, 2004.
3. LandXML.org: http://www.landxml.org

Theoretical Foundations for Knowledge-Based Conceptual Design of Complex Systems

Heriberto Maury Ramírez[1] and Carles Riba Romeva[2]

[1] Mechanical Engineering Department, Universidad del Norte, A. A. 1569,
Km. 5 Carretera a Puerto Colombia, Barranquilla - Colombia
hmaury@uninorte.edu.co
http://www.uninorte.edu.co
2 Mechanical Engineering Department, Universidad Politécnica de Cataluña, Av. Diagonal 647, 08028 Barcelona, Spain
riba@cdei.upc.edu

Abstract. The theoretical foundations that were developed and validated in a research work in order to the conceptual design systematizing of complex systems or processes, are presented in this paper. From the earliest design stages, two approaches of concurrent engineering, the Initial Specification and Conceptual Design, were integrated, the design for functionality (DFF) and design for environment (DFE). The obtained results are significant and novelty, because they determined the formal and heuristic required knowledge to the conceptual design systematizing. On the other hand, this work defined the nature and architecture of the required prototype software. The proposed foundations were validated on the conceptual design cases of the bulk solids handling and primary processing continuous system due to the obtained solutions with prescribed methodology were satisfactory, it was verified that the development foundations are rights.

1 Introduction

In this paper are presented the research work results, whose goals were to develop the knowledge engineering and define the characteristics of a software to assist the stage of the conceptual design of complex systems, integrating two concurrent approaches, DFF (design for functionality) and DFE (design for environment) from the earliest design stages, in other words, the initial specification and conceptual design.

In order to make possible a knowledge based system or knowledge based hybrid system was necessary to support on process design models such as: French [1], Pahl & Beitz [2], VDI-2221 [3]. However, these models must be adapted to apply them integrating different concurrent approaches in the conceptual design stage, because, they were developed before the concurrent engineering emerging.

Some of the first works with the referenced purpose were the Chacrabartti work [4] and Cabarrocas work, Chacrabartti developed methods and a software prototype to assist the conceptual design of systems to changing forces and movements, starting from a functional analysis of system to design. In this same direction, Cabarrocas ([5], [6], [7] y [8]) contributed to the theoretical foundations to conceptual design of

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 166 – 185, 2005.

drive systems with multiple operation modes. In our revision, there are other works in the conceptual design area with a DFF perspective, noting their references [9] y [10].

To reach this general purpose was necessary to employ the methodological stages that will be commented. In the first place, an overview of the state of art and the analysis of many real design cases were realized to know the different models of design process that were previously developed, and also it was important to determine their limitations when were applied them on real design cases with a concurrent approach.

In this way, the opportunities to reach the concurrence, as well as the design concepts and their relations were identified in the stage of the bibliographical exploration and the analysis of real cases. After of this, in the second stage was called stage of theoretical foundation formulation, with aim to solve the limitations of previous models, it was proposed one modified model for conceptual design, the methods for each one its stages, the total required knowledge including the previous concepts, the new design concepts as its heuristic components and their interrelations. Then, to verify the validity of proposed foundations, it was compared the goodness of the obtained solutions using these new theoretical foundations with the solutions obtained with traditional methods. And finally, the architecture and the required characteristics of a prototype computing system were proposed too.

For making easier the understanding of the developed contributions in this work, this paper has been divided in the following sections: the first section is about Composition, Interrelations and Nature of Required Knowledge. The second section is described an Integral Model for Conceptual Design "IMCD", in third section are exposed the methods for each one stages of this model, and the last section, the obtained results in validation stage are shown.

2 General Composition and Nature of Requiered Knowledge for Conceptual Design

The obtained results in this paper demonstrate that the required knowledge basis in a system of assistance to designer in systematic execution of the conceptual design stage must have a formal component and an heuristic component. The Table 1 presents the detailed composition of required knowledge basis.

The formal component must be composed by a new model or modified model of conceptual design process that allows the integrating of different concurrent engineering approaches, as well as the required methods to develop each design stage and a set of concepts related to the following areas: Engineering Design, Concurrent Approaches to integrate and particular concepts of application area.

On the other hand, the heuristic component must be integrated by rules, strategies and criterions that the designers obtain from their experiences aim to reduce the size of searching space and make easier the generation of solutions to the problem. Also, the component heuristic should allow to estimate the values of specifications to define in the initial specification stage. Additionally has been defined some elements as chain types and the types of flow branches that allow to simplify the design problem by its division.

Table 1. Nature and Composition Required Knowledge to Conceptual Design

Stage of Application	Of General Application in Engineering Design		Of Specific Application in Handling Continuous Systems	
	Formal: Models, methods and concepts	**Heuristic: Strategies, criteria and rules**	**Formal**	**Heuristic**
All Design Process	Design Process, Integral Model, Engineering Design Knowledge Basic Components: Formal (Models, Methods, Concepts,…) and Heuristic (Strategies, Criteria and Rules).	Rules to Selection the Methods by each design stage		
Initial Specification Stage	MEPEIS Method, Sufficient Initial Specification, Insufficient and Hiper-specification, Basics Components of Design Problem, DFX Requirement Categories, Redundant Requirements, Knowledge Components Relations, Flow Nodes	Strategies: Categorizing, Checking List Use and Depuration	Requirements List by Category, Flow Type: Continuous, Intermittent and in Storage	
Deducted Specification and Functional Construction	MEDGESF Method, Global Function, Basic Function, Derived Specification, Flow Chain, Flow Branch, Initial Basic Function Groups, Initials, Intermediate, Finals, Joint and Divergence Basic Function Groups, Flow Chain Types, Flow Branch Types, System Functional Structure's Matrix	Rules to doing Deducted Specification and Insert the Basic Function Groups	Typical Global Function List (12), Basic Function List (21): Initial Functions, Intermediate Functions, Final, Joint and Divergence Functions, Derivate Specification Matrix (System, Chains and Branches)	Matrix of Insertion Rules to: Initial Functions, Intermediate, Finals, Joint and Divergence Basic Function Groups
Synthesis of Solutions	MESISOLC Method, Global Solution, Elemental Solution, Global Conceptual Specification, Elemental Conceptual Specification, Heuristic Reduction, Heuristic Based Models to Conceptual Specification	Reduction Heuristic Criteria (DFX), Reduction Rules, Elemental Conceptual Specification Rules	Global Conceptual Specification Matrix and Elemental Conceptual Specification Matrix	Matrices de Criterios y de Reglas de Acotación, Matrices de Reglas para Especificación Conceptual
Solution's Evaluation	MEPEIA Method, Phase's Model, Emissions, Environment Impact Potential Forms, Heuristic Based Model to Estimate Emission Levels, absolute and relative potential impact indirect index	Estimation Emission Levels Rules	Environment Emission Evaluation Matrix, Emission Forms Basis: Consume Resources, Emissions and wastes	Emission Forms Levels Estimation Rule Matrix

And finally, it were defined also, the heuristic knowledge to realize the derivate specification, the functional construction, the synthesis of solutions and their environmental evaluation in the conceptual stage, which is very useful if it consider that the level of information available in this stage is very low.

Not only in this paper were defined the required knowledge components of this basis, but the necessary relations between its components. It is to say, it was define that kind of system structure should be similar to a Knowledge Based Hybrid System, which is an important contribution, because this does possible the actualization capabilities of Knowledge Basis, as well as the learning and explication capabilities, this last opinion is shared by Dixon and Finger ([11], ([12], [13] y [14]), who think that to reach these capabilities it is important has a clearly definition of the system architecture, as well as of the relations and structure required in the knowledge basis.

3 Integral Model for Conceptual Design "Imcd" and the Methods for Its Stages

In this search work with the purpose of integrate two concurrent approaches, it was modified the design process model developed by Pahl and Beitz [2], because the same one was proposed before the formal emerging of concurrent engineering. The proposed model was developed by identifying the limitations of the Pahl & Beitz's Model to integrate a concurrent engineering vision when were applied to the conceptual design cases of complex facilities for handling and continuous processing of bulk solids materials, considering in simultaneous the Design for Functionality and Design for Environment (Figure 1).

To overcome the exposed difficulties, the design researchers had developed many works to improve or modify the traditional design methodologies, we think that to implement with successful the concurrent engineering in the conceptual design a design meta-model should be obtained, as an result of the synthesis from multiple research works on specific areas.

In general, the Pahl and Beitz model consist of five stages, look it in the figure 1, these stages are: Task Clarification, Conceptual Design, Preliminary Sketch Development, Definitive Sketch Development and Detail Design.

Particularly, the Pahl and Beitz's Model describes better the product planning process and how you find a product idea, but it does not present how must be done the task clarification, because it presents the questions that the task clarification must answer, but it does not say how should you find the correct answers to these questions. Because Pahl and Beitz only describes the characteristics of a good specification, it is necessary to complement this work, with the development of new methods or stages that allow to reach it.

While Pahl and Beitz defined six steps for the conceptual design of products, in this work it is consider that in order to the conceptual design of complex systems or facilities is necessary to introduce justly behind a new stage called the Derivate Specification. The derivate specification allow to identify and characterize the independent flow groups inside the system with their inter-relations by means of a comparative analysis of the characteristics between the inlet and outlet flows, such information it must be integrated in clarification or Initial Specification.

Finally, this last analysis will allow to deduce new specifications called derivate specifications about the flow groups or independent flow chains that it are required to the system fulfils with its objectives . This characterization of the independent flow groups it will help in the decision about which functions must be inserted in the functional structure.

In order to complement the Pahl and Beitz's model, specially in the task clarification and conceptual design stages, it has proposed introduce the Derivate Specification Stage, to make easier the functional construction by means of a clearly definition of the flow relation between the inlets and outlets to the system. The previous goal is possible identifying the independent flow groups, that it has been called Flow Chains in this research work.

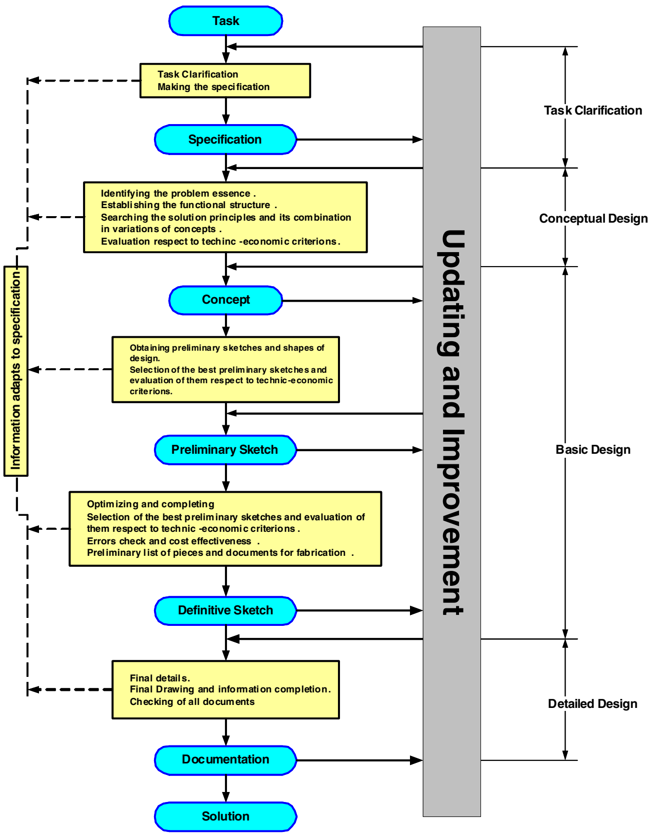


Fig. 1. Design Process Model of Pahl and Beitz

This flow chains would are consist of flow branches, as much the Flow Chains as the Flow Branches will must be characterized by means of new specifications deducted from the founded differences between the characteristics the inlet and outlet flows.

This a new stage has been be added to the design process model, with a new method that it has been developed for it, the obtained results with this allow to identify which rules of construction functional must be applied. In other words, the deducted specifications are the conditions that allow using the construction functional rules.

They had been defined and they must be integrated to the heuristic component of knowledge Basis. Thus, The Derivate Specification Stage is very useful, because it allow to simplify the design problem by its division in flow chains.

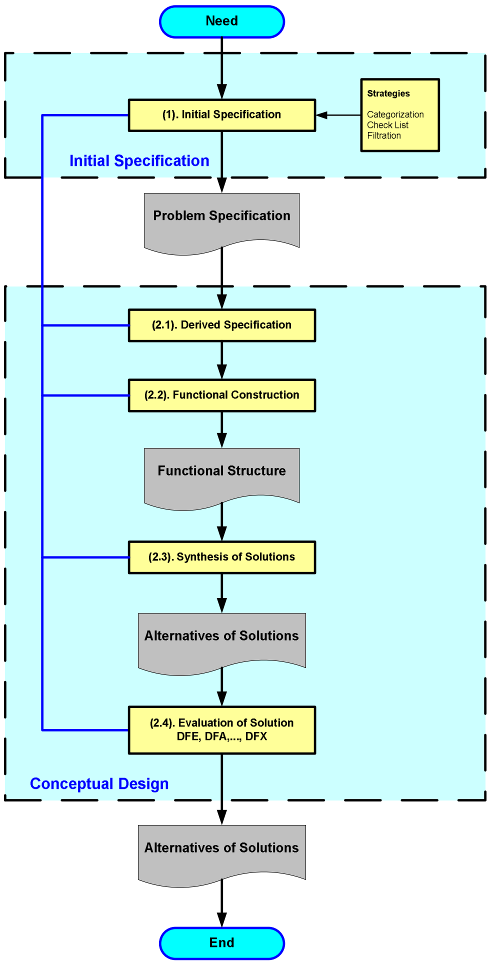


Fig. 2. Integral Model for Conceptual Design (IMCD)

Although the introduction of the Derivate Specification Stage to modify the model of Pahl and Beitz was validated on the conceptual design of the bulk solid materials handling systems and primary continuous processing systems, we consider that they could be applied it in any engineering area to obtain conventional or novelty solutions. The proposed modified model (IMCD MODEL) is shown at figure 2.

It is very import note that also, in the catalog conceptual design case a computing tool could be applied; for this purpose a the IMCD model and with the defined knowledge basis must be developed. It is believed that software this nature almost that should automatically generate the solutions; while in the generation of novelty solutions, the software behavior should be as an advisor through all the design process.

Additionally, in this research were proposed and were validated four methods in order to realize each one of the IMCD stages. These methods are:

- Sufficient Initial Specification Method (MEPEIS)
- Method to Generate the derivate specification and the functional structures (MEDGESF)
- Solution Conceptual Synthesis Method (MESISOLC)
- Method to evaluate the potential agents that affect the environment (MEPIEA)

Now in the following subsections, each method to stages of the IMCD model will be presented with detailing.

3.1 Sufficient Initial Specification Method (Mepeis)

This method as the developed knowledge basis to Initial Specification Stage allow to transform the problem design from a vague information state related to the user's need to an sufficient information state about the task clarification. This is possible, by means of the combination of three strategies: The Categorizing, The Check List Using and the Depuration of Requirements.

These strategies allow the obtaining a state of sufficient clarification about Design Initial Specifications; and also to verify the independence of the functional requirements that it has been fixed. In this way, the minimum information axiom and the independence axiom proposed by Suh N. [15] are satisfied. Now, it will expose the purpose of each one of the strategies. The graphical meaning of MEPEIS Method is shown in Figure 3.

The categorizing strategy allows to identify the components of design problem. It is to say, the same system, the environment and the inlet and outlet flows; as also it allows to define the requirement categories related to each concurrent approaches that it would wish to integrate in the process. After that, it has been defined these categories, the designer must analysis many design cases and extract to typical requirements for each category, these set of requirements must be attached to a non-purified check list: The non-purified check list employed should be used in future design cases to get the completeness of initial specification.

However up to this point, the independence of requirements there has not been verified and for this reason, a depuration strategy must be used. In the first place, this strategy would be useful to list the influential variables for each requirement and to identify sets of requirements that depend on the same group of variables. These sets have been called sets of requirements potentially redundant. These potentially redundant sets must be evaluated by designer, and finally, he must decide if they should be eliminated or not. Therefore, this last strategy aims the depuration of requirement sets that will be used in the initial specification.

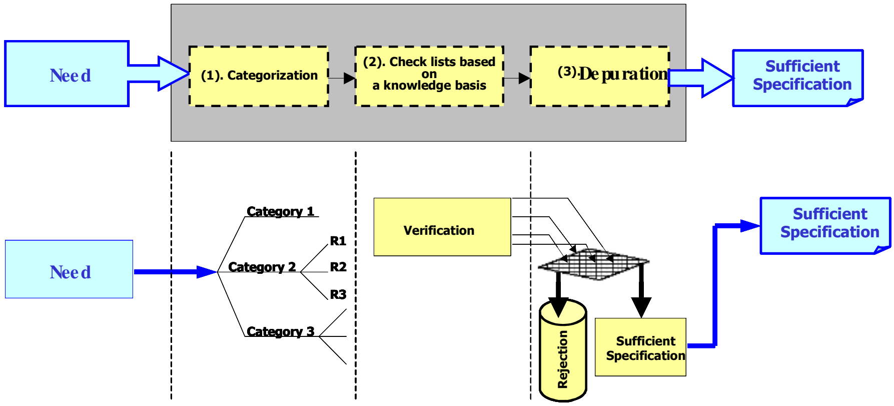


Fig. 3. Diagram of Sufficient Initial Specification Method (MEPEIS)

3.2 Method for Generating of Derivate Specification and Functional Structure (Medgesf)

In the same way with Method for Derivate Specifications and Generating of Functional Structures and with the developed knowledge for this stage, it is able to transform the information state about the design problem from information related with task clarification to information about of the required functional structure for the system that it must be designed.

This Method allows us to obtain the Derivate Specification and to realize the construction of functional Structure, which are activities includes into the IMCD model. In the first place, The MEDGESF method allows us to divide the systems into independent flow groups, which had been called Flow Chains.

After, these chains also are divided in flow branches and characterized by means of the derivate features deducted from initial clarification using rules include in knowledge basis. The derivate features are the starting point to employ another set of rules, which are within the knowledge basis, and that it has been defined as set rules for the insertion of functions.

The rules to integration of functions allow us to identify which functions must be introduced or applied. A novel concept that was denominated "Basic Function Groups" was developed in this work, to avoid the problems in the definition of the precedence relations between the functions that it must introduced (see this method in figure 4).

A group of functions is a set of basic actions, which are located in a habitual place and they have a typical aim within a flow branch. In this work were defined the different types of function groups that must integrate any kind of flow branches. These flow chains, flow branches and functional groups are an initial step in the definition of the required architecture of the system to design it

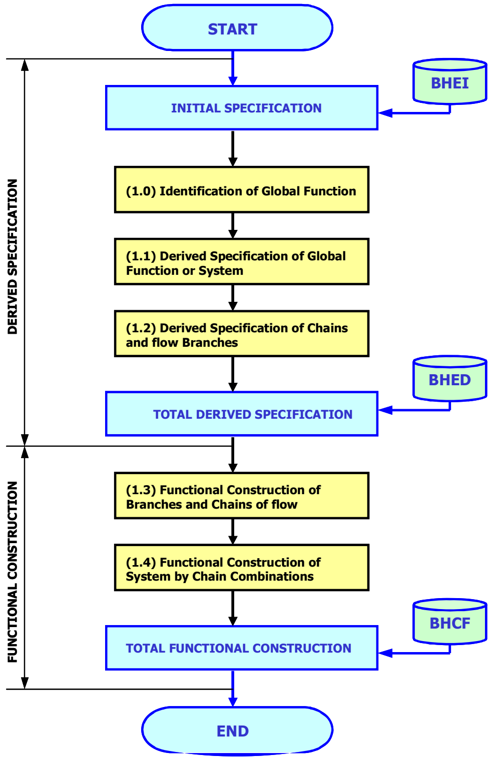


Fig. 4. Method to derived specifications and functional structure generation (MEDGESF)

3.3 Method for Conceptual Synthesis of Solution Alternatives (Mesisolc)

By other hand, The MESISOLC method allows to synthesize the alternatives of solution starting from functional structure, since it overcomes the problem of excessive size of the solution field by applying of a strategy of heuristic narrowing. This strategy should be employed including different approaches (DFX), and it was demonstrated in this research, that it has effectiveness higher than 99,99 percent for narrowing of the theoretical field of solutions. The strategy is one of the most important contributions of this research. The MESISOLC method joined to the developed knowledge basis allow to the transformation the state of design problem from a functional structure to a set of alternative solutions with reasonable size. (Figure 5).

In spite of the low level of available information in the conceptual design phase, the MESISOLC employs too a strategy of heuristic of modeling to estimate the specification values or hoped levels of performance for each one of the alternatives. For this objective are used different rules about distinct approaches (DFX) to include. The estimation of performance levels is supported on heuristic modeling, which become an important tool when it is desired to evaluate each one of the solutions with respect to a performance pattern predefined by the initial specification, such as it occurs in the conceptual design stage.

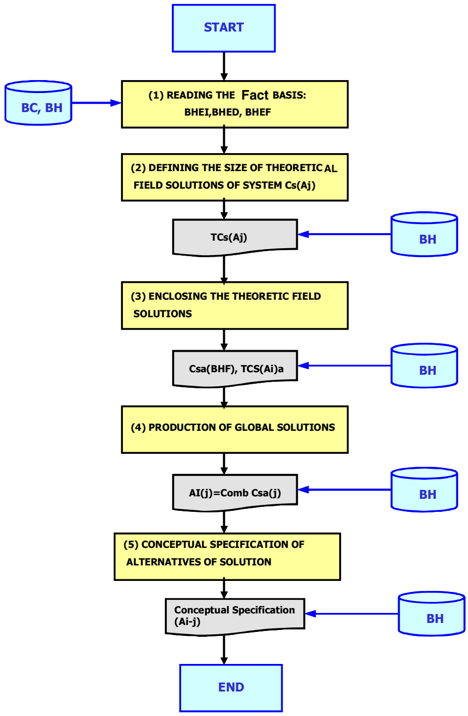


Fig. 5. Method to Conceptual Synthesis of Solution Alternatives MESISOLC

3.4 Method to Evaluate the Potential Environmental Impact (Mepeia)

Finally, it is possible by the using of the MEPEIA method and the developed knowledge basis to estimate the resources and the emissions of material and energy in all the lifecycle of the each alternative. In the same way, through of a collection of relatives parameters and criterions are ordered the alternatives in function of its environmental goodness, which is determined by the resources and emissions associated to the lifecycle phases of the each one of these alternatives, with independence of the specific answer of the environment on the same ones.

However, this method allows to consider the relative importance of these aggression forms in according to the particular circumstances of an environment or surrounding [16] y [17].

In the MEPEIA method are combined two strategies. The first one, it is the use of an model of phases and the second, is the employment of the weighed criterions, which allows the environmental evaluation in the lifecycle, integrating simultaneously different ways of environmental aggressions, with distinct relative weights in according to the type of surroundings where is want to install the system in process of design, Figure 6.

The model of phase allows us to analyze the interaction between the system and its surroundings throughout its lifecycle. While the weighed criterions leads to the obtaining a global relative comparison, which is very important in the decision taking with a high level of objectivity, in spite of the poor level of the available information in the conceptual design stage.

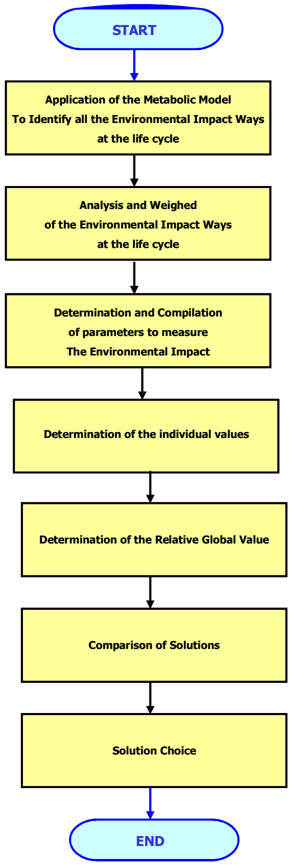


Fig. 6. Method to Evaluate the Potential Environmental Goodness MEPEIA

4 Validation of Model (Imcd) and the Knowledge Bases

To apply the model and developed methods on specific problems in the field of the continuous systems to handling and the primary processing of bulk solid materials, it is necessary to introduce a particular knowledge related to this field, as well as, it is required a knowledge of general level that must be included in the knowledge basis.

In this research, the two kinds knowledge were developed. It is think that the developed model (IMCD) and its methods are of general applicability, while, to they can de used in specific fields must be developed the required particular knowledge. Two cases of the employed in validation are described in the attachments of this paper.

4.1 Case 1

It will starting with the most simple of studied cases, an Mining Industry requires to built a harbor dispatch system that it is able to handle coal to a rate of 800 m3/h from rear dumps of 20 m3 to ships of 15000 m3 of capacity. For reasons of low depth of waters, the ships should be anchorage to 100 meters of the coast (see Figure 7). Apply the developed methodology to obtain the solution alternatives.

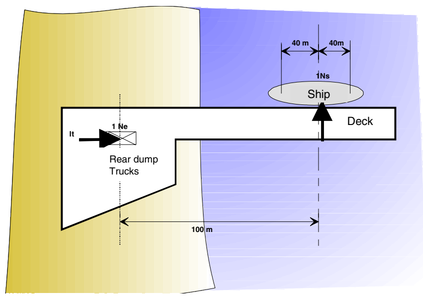


Fig. 7. Case 1, General View

After the stages of initial specification, derived specification and functional construction had been developed, it was required to obtain the functional structure of system, which is shown in Figure 8.

In this case, the obtained functional structure has eight functions: Reception, Conduction, feeding, Conduction, Transport, Ship loading and Conduction. If it is desired to obtain the theoretical field of solution alternatives from them, like a systematic combination of individual solution concepts for each function, then the size of theoretical field of solutions will be explosive, like is shown in the two equations, after of figure 8.

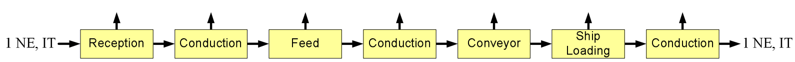


Fig. 8. Functional Structure of Harbor Dispatch System

$$TCS(1 - Ccrp - 1Rpes) = \prod_{p=1}^{pmáx=8} TCS(1 - Ccrp - 1Rpes - pfbi) = 2\times3\times7\times3\times11\times3\times8\times3 = 99792 \quad (1)$$

The expression (1) shows the theoretical size of the solution field that it was obtained by means of systematic combination. While the expression (2) allow to calculate the size of solution field after to apply the narrowing heuristic strategy.

$$TCSa(1 - Ccrp - 1Rpes) = \prod_{p=1}^{8} TCSa(1 - Ccrp - 1Rpes - pTCSa(fbi)) = 1x1x3x1x2x1x2x1 = 12 \quad (2)$$

Table 2. Solution Alternatives to a Coal Harbor Dispatch System[1]

Number	Position								ALT. ID.
	1	2	3	4	5	6	7	8	
1	R(TPP/L)	C(RE)	ED(C)	C(INT)	T(CS)	C(INT)	CM(MBR)	C(T)	Alt(Ai-1)
2	R(TPP/L)	C(RE)	ED(C)	C(INT)	T(L)	C(INT)	CM(MBR)	C(T)	Alt(Ai-2)
3	R(TPP/L)	C(RE)	ED(C)	C(INT)	T(CS)	C(INT)	CM(MBRRe)	C(T)	Alt(Ai-3)
4	R(TPP/L)	C(RE)	ED(C)	C(INT)	T(L)	C(INT)	CM(MBRRe)	C(T)	Alt(Ai-4)
5	R(TPP/L)	C(RE)	ED(VCB)	C(INT)	T(CS)	C(INT)	CM(MBR)	C(T)	Alt(Ai-5)
6	R(TPP/L)	C(RE)	ED(VCB)	C(INT)	T(L)	C(INT)	CM(MBR)	C(T)	Alt(Ai-6)
7	R(TPP/L)	C(RE)	ED(VCB)	C(INT)	T(CS)	C(INT)	CM(MBRRe)	C(T)	Alt(Ai-7)
8	R(TPP/L)	C(RE)	ED(VCB)	C(INT)	T(L)	C(INT)	CM(MBRRe)	C(T)	Alt(Ai-8)
9	R(TPP/L)	C(RE)	ED(P)	C(INT)	T(CS)	C(INT)	CM(MBR)	C(T)	Alt(Ai-9)
10	R(TPP/L)	C(RE)	ED(P)	C(INT)	T(L)	C(INT)	CM(MBR)	C(T)	Alt(Ai-10)
11	R(TPP/L)	C(RE)	ED(P)	C(INT)	T(CS)	C(INT)	CM(MBRRe)	C(T)	Alt(Ai-11)
12	R(TPP/L)	C(RE)	ED(P)	C(INT)	T(L)	C(INT)	CM(MBRRe)	C(T)	Alt(Ai-12)
Alternatives	Reception	Conduction	Extraction	Conduction	Trans port	Conduction	Ship loading	Conduction	
Basic Functions									

This situation could be solved using a the heuristic strategy of narrowing, it was developed in this research. Through this strategy was obtained effectiveness higher than 99,99 % in the reduction of the size of theoretical solution field, using criterions and narrowing rules, only of functional and environmental nature, that were defined for each one of the solution sets for the basic functions.

When are contrasted the sizes of the theoretical field with narrowed field, it is possible to affirm that this strategy reduces significantly the size of the narrowed solution field.

The size of theoretical field defines a great variety of possibilities, but at the same time this fact is a weakness, because is practically impossible their total evaluation. For this reason, it was demonstrated that a heuristic narrowing strategy like the developed, it is powerful means to transform the theoretical field, of excessive size, to a narrowed field with reasonable size.

To generate systematically this narrowed field of solutions, it was reasonable to employ a similar process to the morphological charts. In Table 2 each row correspond a general solution alternative. For example, the first row correspond to the Alternative No. 1, which in this case, will be integrated of a receiver or pyramidal hopper (R(TPP/L)) to receive and contain the flow comes from the rear dump trucks, leaving it ready to be conduct by skirt board (C(RE)) and extracted from there by a belt feeder or extractor (ED(C)), after this would be conduced by an intermediate chute C(INT) up to a standard conveyor belt T(CS), that would transport it to a mobile mechanical ship-loader CM(MBR), which finally, would give it to the ship-vessels through a final conduct or telescopic chute C(T).

In spite of the poor level of detail that exists in the conceptual phase, the conceptual specification and the environmental evaluation of the twelve alternatives formulated must be obtained, for this aim the MEPEIA Method was used, the results can be

[1] In order to the meanings of abbreviations see No.1 appendix.

observed in the Table 3. In the weighed environmental evaluation that was obtained, it was included all the stages of the lifecycle of system and their ways of potential aggression.

Table 3. Global Evaluation of Relative Environmental Goodness in the Life Cycle for the narrowed solution field

FIA(l)	Material Resources		Energetic Resources		Material Wastes		Wastes of Energy		Solid Emissions		Noise Emissions		Total
ID. Alt	Irel(Ai-j)	W(l)	Irel(Ai-j)	W(l)	Irel(Ai-j)	W(l)	Irel(Ai-j)	W(l)	Irel(Ai-j)	W(l)	Irel(Ai-j)	W(l)	Sum(Irel*W(l))
(Ai-1)	0.38	0.14	0.26	0.2	0.27	0.12	0.27	0.17	0.33	0.2	0.33	0.17	0.27431
(Ai-2)	0.57	0.14	0.32	0.2	0.31	0.12	0.32	0.17	0.33	0.2	0.33	0.17	0.35801
…	…		…				…				…	…	…
(Ai-12)	0.5	0.14	0.41	0.2	0.42	0.12	0.41	0.17	0.33	0.2	0.33	0.17	0.36768
Irel*W(l)	0.14		0.2		0.12		0.17		0.2		0.17		1.0

According to the results shown in the last column of Table 3, it is easy to visualize that the best alternative with respect to the environment goodness is the number No. 1 (Ai-1), considering the weigh established for the distinct criterions, because it had the lowest value to the relative aggression factor.

4.2 Case 2

Now, it will expose some results obtained in other practical case, which is more complex than the past. In this case, it is desired to obtain the conceptual alternatives for a crushing system preparing of concrete aggregates, in which the material without treatment is transported from the quarry by rear dump trucks. The system must produce aggregates with specified sizes for the concrete production. The required sizes ands their flows can be view in the Figure 9.

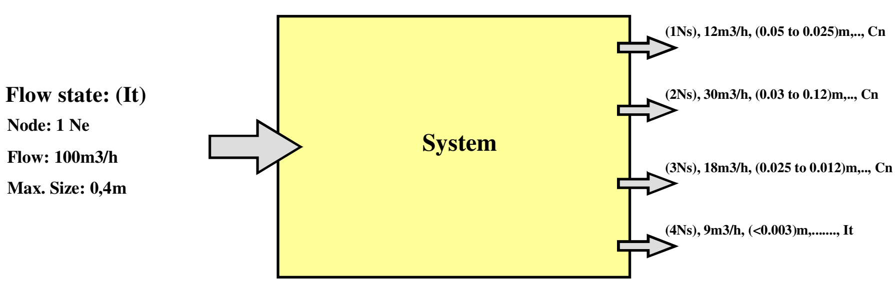


Fig. 9. Black box of a Crushing and Classification

Also, after the specification stage and derived specification were concluded, it was obtained for this system the functional structure (see figure 10). This is very more complex than the last, because it had four material outlets. From this block structure was obtained a theoretical field of 6,04*104 alternatives through a systematic combination of concepts for each one of functions. Evidently, it was impossible to evaluate it totally with traditional tools this theoretical field by its excessive size.

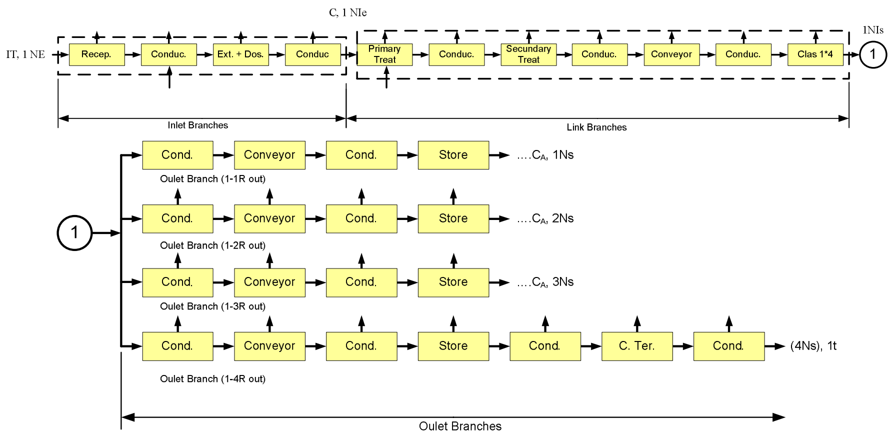


Fig. 10. Functional Structure of an Aggregates Production Plant

Table 4. Comparative Matrix of Weighed Potential Impact of different alternatives in the lifecycle

	Forms of Environmental Aggression during the Lifecycle FIA(l)												
	Material Resources		Power Resources		Material Wastes pre-operation		Power Wastes		Particulate Material Emissions in Operation		Noise Emissions in operation		
Id. Alt.	Itrel	W(l)	$Itrel_l$	$W_{(l)}$	$Itrel_l$	W(l)	$Itrel_l$	W(l)	$Itrel_l$	W(l)	$Itrel_l$	$W_{(l)}$	Igrel(Ai-j)
(Ai-1)	0.469		0.53		0.48		0.578		0.53		0.497		0.53661
…...	…...	0.14	….	0.2	….	0.12	…	0.17	…	0.2	…	0.1 7	…
(Ai-192)	0.51		0.47		0.52		0.422		0.477		0.503		0.46339
TOTAL	0.14		0.2		0.12		0.17		0.2		0.17		1.0

The application of a narrowing heuristic strategy with functional and environmental criterions reduced the largely size of the solution field from 6,04*104 to 3888 alternatives, which allowed to demonstrate that the effectiveness of the landmark strategy is superior to 99,999 %. Also it is obvious that if this strategy was applied including other criterions (DFX), easily the size of the solution field obtained it could be bellow to 200.

In this case, it was applied additional narrowing criterions, such as reliability and initial cost; the narrowed solution field was reduced from 3888 to 192. The executed environmental evaluation for the field of feasible solutions can be appreciated on Table 4. It was omitted for reasons of available space, in this document, the evaluation results of all the feasible alternatives.

In general, if it is considered the goodness of the obtained results for all studied cases and special for the two presented cases, it had demonstrated that the application of model IMCD to real design problems is effective and convenient, because it allows to the integration concurrent engineering approaches without misfortunes, approaches such as Design for Functionality (DFF) and Design for Environment (DFE) were integrated.

5 Conclusions

In order to facilitate the understanding of the conclusions, it will be presented them from two points of view. In the first place, it is consider the novelty of the obtained results and it is considering the future research works that could be developed.

Importance and Novelty of the Contributions

Such as it appears in three related sections, although this research is in the same line of the works of Chacrabarti and Cabarrocas ([18], [19], [20], [8]) functional (DFF) is newness, but the obtained results demonstrate its effectiveness and complementariness with relation to these previous works, because it does possible to develop the conceptual design of complex systems or macro-systems integrating the approaches of DFF and Design for environment (DFE). Therefore, it allows to the generation of solutions integrated by systems in themselves (machines), instead of simple concepts; in agreement with the bibliographical revisions realized, this it would be the first work of this nature.

The contributions of this research are not only the development of a model modified for conceptual design MIDC, but the methods and the formal and heuristic knowledge for the systematic execution of the conceptual design integrating the approaches of design DFF and DFE. In other words, also in the research was developed totally the required knowledge engineering starting from the particular study of design of the continuous systems to handling and primary processing of bulk solid materials

Perhaps in this research according to Riba C. [21] the most important concept within the developed concepts is of the Derived Specification. Because with the traditional models was not possible the systematization of the functional construction from the initial specification, by the lack of a stage where were identified and characterized the independent groups of relations between inlet and outlet flows of the system to design. The value of this concept is that by a side, it simplifies the design problem by its division in flow chains and flow branches. And by the other hand, it directs us towards the functional construction by means of the characterization of the chains and the branches, which allows us to identify the rules for insertion of the basic groups of functions (GFB) and where should be them introduced.

The novelty and importance of the developed methods can be synthesized like follow:

- *Method for Sufficient Initial Specification*. Unlike Kusiak [22], [23], [24], y Suh [15] whom described the characteristics of a suitable specification, in this research was developed the MEPEIS method, which combines three strategies for get it. They are Categorization, Check Lists Using and Depuration

Strategy. The useful of this method had been demonstrated by two reasons: First, because it allows us to transform the design problem from the state of vague clarification about the need to a state of adequate clarification, through of a sufficient number of requirements. And second place, by the integration in the conceptual design of the DFF and DFE approaches from this early stage.

- *Method for Generating of Derived Specification and the Functional Structure MEDGESF.* It is possible conclude that with the MEDGESF method, in which are combined a heuristic and systematic approaches to execute a preparatory step to the functional structuring called Derived Specification, it is easier to the transformation of the design problem from initial level of requirements to the level of functions with their relations. It is possible to reach it with a coherent combination of "algorithms" and rules, which simplify it or to divide the problem in chains and branches, and they allow their characterization too. With it, it is possible to identify so much the global functions as the insertion rules of the groups of basic functions, that progressively they will be employed to obtain the functional construction of the system.
- *Method for the generation of Conceptual Solutions MESISOLC.* Also, it was demonstrated that with this method of systematic and heuristic nature (MESISOLC), it is possible to transform the design problem from functions level to a field of solutions with reasonable number of feasible alternatives, which is one of the main objectives of the conceptual design phase.

  Definitely, the success of this method was verified by the high level of effectiveness of the strategy of heuristic landmark that it was introduced (> 99.999 %), which has allowed surpassing the disadvantage of an explosive size of the theoretical field of solutions that it is obtained when are employed of the traditional systematic methods of synthesis. In addition, the bibliographical exploration had ratified that this contribution is novel in relation with the previous works, since they had not approached explicitly this problematic.
- *Method for Environmental Impact Evaluation, MEPEIA.* It was developed a method whose newness consists of the employment of a metabolic model and weighed criterions. This combination allow us to realize an evaluation considering multiple forms of aggression with dissimilar relative weigh throughout the lifecycle of each one of the generated alternatives.

The first strategy allows the analysis of the interaction between the system and its surroundings during the phases of its lifecycle, with the purpose of identifying the aggression ways. The second one, it allows us to reach a relative global comparison. In the early phases of design, this last one strategy allows to a multi-criteria evaluation instead of an ordinal mono-criteria evaluation derived from absolute values employed by Graedell [25].

Future Works

The following complementary research works can be visualized from the obtained contributions and considering the premise that a general design methodology will be progressively obtained by means of multiple works of application on specific areas:

- It is necessary, to develop completely an improved computer tool to assist this process with the purpose to increase the performance level and to include the explication and learning mechanisms, which are required in this kind of systems.
- It is necessary too, the development of the foundations for integration in a SIADC system of 3-D schematization (Layouts) of the solutions, considering an adequate management of space and physical constraints. The simultaneous considering of spatial restrains and the functional design, it will lead us to the insertion of basic functions to evade the physical obstacles. Because functional changes will reflect on the solution composition, it is opportune the considering simultaneous of the spatial limitations and the functional design, especially in the early phases of design. With this, it will be possible to reduce significantly the costs and the development time. However, it is an ambitious goal regarding that the geometrical and dimensional characteristics of the commercial equipments are very different.

References

1. French M. J.: Engineering Design, The Conceptual Stage. London, Heneiman. (1981)
2. Pahl G., Beitz W. (Wallace K.: editor).: Engineering Design. A Systematic Approach. Springer-Verlag (1984).
3. VDI 2221, Sheet 1.: Systematic Approach to Design of Technical Systems and Products. VDI-Verlag (1987). pp-1-32.
4. Chacrabarti, A. & Bligh, P.: An Approach to Functional Synthesis of Solutions in Mechanical Conceptual Design. Part I. Research in Engineering Design, n6, (1994). pp. 127-141.
5. Cabarrocas J., y Riba C. Contribuciones a la síntesis funcional de sistemas mecánicos en la Fase de Diseño Conceptual. Actas del XIII Congreso Nacional de Ingeniería de Proyectos, (1997). Vol. 2, pp. 734–741.
6. Cabarrocas J., y Riba C.: Evaluación de los modelos de representación de la estructura funcional de los sistemas mecánicos para su modularización. Bilbao, Anales de Ingeniería Mecánica, XII Congreso Nacional de Ingeniería mecánica de España (1997).
7. Cabarrocas J., y Riba C. Caracterización de los modos de operación de las máquinas como base para la estructuración de sus sistemas de accionamiento. Terrassa, Anales de Ingeniería Mecánica, XII Congreso Nacional de Ingeniería mecánica de España. Año 12, Vol. 1, (1998). pp. 232 – 237.
8. Cabarrocas Joan.: Disseny Conceptual basat en la síntesi funcional de sistemes d´accionament amb multiples modes d´operació. Barcelona, Universidad Politécnica de Cataluña (1999).
9. Chew M., Shen S., Issa G. & Issa G.: "Kinematics structural synthesis of mechanims using knowledge – based systems". ASME; Vol. 117, (1995). pp. 96 – 103.
10. Chiredast M., Gea H. C., Kikuchi N.: Structural configuration examples of an integrated optimal design process. ASME, Vol. 116, (1996). pp. 997 – 1004.
11. Dixon, J. R.: On research towards a scientific theory of engineering design. Artificial Intelligence for Engineering Design Analysis and manufacturing. Vol. 1. No. 3, (1987).
12. Dixon J. R.: Knowledge-Based Systems for Design. ASME, Vol. 117, (1995). pp. 11 – 16.

13. Finger S. y Dixon J. R.: A review of research in mechanical engineering design. Part I. Research in Engineering Design, Vol. 1, No. 1, (1989). pp. 51 –67
14. Finger S. y Dixon J. R.: A review of research in mechanical engineering design. Part II. Research in Engineering Design, Vol. 1, n 1, (1989). pp. 121-137.
15. Suh N. P.: Principles of Design. Oxford United Press, (1990).
16. Maury H., Cabarrocas J. y Riba C.: Aplicación de las técnicas de Ingeniería Concurrente al Diseño Mecánico. Terrassa, Anales de Ingeniería Mecánica, XIII Congreso Nacional de Ingeniería Mecánica de España. Año 12, Vol. 1, (1998). pp. 359 –367.
17. Maury H.: Aportaciones metodológicas al diseño conceptual: Aplicación a los sistemas continuos de manipulación y procesamiento primario de materiales a granel. Tesis Doctoral, Universidad Politécnica de Cataluña, (2000).
18. Chacrabarti, A. & Bligh, P. An Approach to Functional Synthesis Solutions in Mechanical Conceptual Design. Part II. Research in Engineeering Design, n8, (1996). pp. 52-62.
19. Chacrabarti, A. & Bligh, P.: An Approach to Functional Synthesis Solutions in Mechanical Conceptual Design. Part III. Research in Engineeering Design, n2, (1996). pp. 116-124.
20. Chacrabarti, A. & Bligh, P. An Approach to Functional Synthesis of Mechanical Design Concepts. Research in Engineering Design, n10, (1996). pp.313-331.
21. Riba, Carles.: Diseño Concurrente. Barcelona, Edicions UPC, (2002).
22. Kusiak & Szcerbicki.: A formal approach to specification in conceptual design. ASME, Vol. 114, (1992). pp. 659 – 665.
23. Kusiak, A. & Larson, N.: Decomposition and Representation Methods in Mechanical Design. ASME, Vol. 117, (1995). pp. 17 – 24.
24. Kusiak. A. & Wang J.: Decomposition of design process. ASME. Vol. 115, (1993). pp. 687 – 695.
25. Graedel T.E., Allenby B. R.: Design For Environment. N. J. Prentice Hall, Upper Saddle River, (1996).

Appendix

NO.1. LIST OF ABBREVIATIONS

R(TPP/L):	Receptor type pyramidal hopper for lateral / rear dump	M(B):	Ball Mill
R (TPV):	Receptor type pyramidal hopper for bottom dump	M(Tub):	Tubular mill
ED(P)	Plate Feeder	M(Co):	Combined mill
ED(VCB):	Vibratory feeder of low depth	M(Ce):	Ceramic carcass mill
ED(A):	Reciprocating feeder	M(A):	Autogenous mill
ED(C):	Belt feeder	M(V):	Vertical mill
ED(T):	Screw feeder	M(A):	Ring mill
ED(V):	Rotary feeder	R(EC+TC):	Belt conveyor – Reclaim
T(C):	Belt Conveyor	R(ECa+TCa):	Chain Conveyor – Reclaim
T(CF)	Flexowell conveyor	R(ET+TT):	Screw conveyor – Reclaim
T(L):	Plate Conveyor	R(EV+TT):	Rotary Valve – Screw conveyor Reclaim
T(B o E):	Bucket conveyor	R(EC+TC):	Rotary Valve – Belt conveyor Reclaim
T(CA):	Drag chain conveyor	R(EV+AD):	Rotary Valve – Air-slide Reclaim
T(T):	Screw conveyor	R(Rasc Lat+TC):	Scraper – Belt conveyor Reclaim
T(V):	Vibrating conveyor	CF(MBR):	Pivot mechanical ship-loader
T(A):	Reciprocating conveyor	CF(NBR):	Pivot pneumatic ship-loader
T(PN):	Pneumatic Conveyor	CF(MBRe):	Fixed pivot mechanical retractile ship-loader

T(AD)	Air-slide conveyor	CF(NBRe):	Fixed pivot pneumatic retractile ship - loader
A(SMNP):	Silos for non-sticky materials	CM(MBR)	Mobile pivot mechanical ship-loader
A(SMP)	Silos for sticky materials	CM(NBR)	Mobile pivot pneumatic ship - loader
A(SMF):	Silos for aerated materials	CM(MBRRe):	Mobile pivot mechanical retractile ship-loader
A(AC):	Cone stockpile	CM(NBRRe):	Mobile pivot pneumatic retractile ship-loader
A(AL):	Linear stockpile	DF(MBR):	Pivot mechanical ship-unloader
A(AR)	Radial stockpile	DF(NBR):	Pivot pneumatic ship-unloader
C(MMG):	Heavy screen	DF(MBRe):	Fixed pivot mechanical retractile ship-unloader
C(MF):	Medium screen	DF(NBRe):	Fixed pivot pneumatic retractile ship -unloader
C(F)	Light screen	DM(MBR):	Mobile pivot mechanical ship-unloader
TP(M)	Primary Jaw crusher	DM(NBR):	Mobile pivot pneumatic ship - unloader
TP(CG)	Primary gyratory cone crusher	DM(MBRRe):	Mobile pivot mechanical retractile ship-unloader
TP(R):	Primary roll crusher	DM(NBRRe):	Mobile pivot pneumatic retractile ship-unloader
TP(I):	Primary impact crusher	PC(V):	Vibratory pre-screener
TS(M):	Secondary Jaw crusher	PC(D):	Dish pre-screener
TS(CG)	Secondary gyratory cone crusher	PC(RE):	Elliptical roll pre-screener
TS(R):	Secondary roll crusher	D(TD):	Divergent tank
TS(I):	Secondary impact crusher	C(TC):	Convergent tank
M(B):	Rod Mill	C(RE):	Skirt board (feeding channel)
C(T):	Ending conduct (discharge chute)	C(INT):	Intermediate conduct (chute)
S(AC):	Air separator	S(ACVAC):	Controlled speed air separator
S(TS):	Turbo-separator	L(TG):	Twin screw washers
L(CE):	Spiral drum washer	L(N):	Wheel washer
L(TC):	Classification tank washer	CaET(DSE):	Silo truck-loader
CaET(RC):	Radial Conveyor Truck-loader	Mez(T):	Screw mixer
Mez(Ta):	Drum mixer	Mez(Pa):	Paddle mixer
Mez(SN)	pneumatic homogenizer and mixer silo		

Public Digital Collaboration in Planning

Stefanie Dühr, Nada Bates-Brkljac, and John Counsell

The University of the West of England, Bristol,Frenchay Campus,
Coldharbour Lane Bristol, BS16 1QY, UK
Stefanie.Duhr@uwe.ac.uk, Nada.Brkljac@uwe.ac.uk
John.Counsell@uwe.ac.uk

Abstract. It is becoming increasingly important to European democratic governments to increase public participation. One of the areas that gives rise to significant public dissatisfaction and feelings of dis-empowerment is that of environmental planning. Environmental planning is a diverse, complex and difficult process requiring consideration of wide ranging issues that include: urban function and form; cost and benefit; appearance and experience; and allocation and provision. The process is costly and so at any late stage difficult to deflect. This is not assisted by complex and even conflicting interpretations of the legislation. With broadening web access government bodies at various levels (national, regional, and city) are now turning their attention to web based approaches to easing access and empowering public involvement and debate on the issues. It is argued that replication on the web of existing plan and cartographic-based approaches to planning will tend to perpetuate the requirement for trained interpretation, thereby excluding the general public. More effective tools are required for collaborative citizen based examination and development of alternative proposals. This paper describes the basis for the coming three years of research in the Virtual Environmental Planning System (VEPS) project. VEPS proposes a new open standard web approach to enable citizens to view, analyse, interact with and respond to proposed changes, to collaborate and comment together and pose and test their own alternative solutions. This approach it is argued needs to be based on an interactive three-dimensional virtual reality visualisation that allows the viewer to experience the highly complex information without need for training, 'because they can see and experience what the visual impacts of the planned development will be and can see the environmental impact in the associated model'.

1 Introduction

Environmental planning is a diverse, complex and sensitive process that requires technical, economic, aesthetic and social considerations. These issues are in turn wide-ranging: 'technical' includes urban function and form; 'economic' addresses cost and benefit; 'aesthetic' relates to both appearance and experience; and 'social' involves allocation and provision. The planning process can be defined as: seeking advice; appraising the site; consulting the public and experts; briefing; developing technically acceptable proposals; negotiating; applying for and gaining legal consent and public acceptance. They all contribute to both "efficient process and an improved

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 186 – 193, 2005.

product"[1] A focus on planning procedures at different stages helps to recognise many of the difficulties inherent in planning; the problems of site and context, the need to compromise, the difficulties of communicating planning proposals and the effects of changes as well as the importance of people's involvement in the process. While the mechanisms differ from state to state in Europe, the problems as perceived by the public are often the same, including lack of meaningful consultation resulting in lack of influence over developments and changes, particularly those affecting the home and its immediate environment. These concerns have yet to be addressed by EC harmonisation.

2 VEPS

VEPS, the 'Virtual Environmental Planning System', is a newly commenced Interreg IIIB funded European project focused on the North West Europe region.[2] The project brings together partners from the UK, Germany and France, with skills in geographic visualisation and spatial planning. VEPS has received additional support from the UK government as a potential contribution to its e-Planning programme. VEPS aims to develop a common and transferable web based visualisation system that enables citizens to view, analyse, interact with and respond to proposed changes, to collaborate and comment together and pose and test their own alternative solutions. To increase participation it is considered essential that it be free at the point of use, so it will make use of the web and open source tools. It is intended to test this system in different demonstration case studies. The project is thus expected to improve future public participation procedures on planning strategies and development proposals by making planning information more accessible and easier to understand for the general public, while at the same time facilitating the decision-making structures of planning authorities. The proposed approach is to use an interactive three-dimensional (3D) virtual reality (VR) visualisation to enable the viewer to grasp highly complex information without the need for training, because users can see and experience both the visual impacts of the planned development and judge the environmental impact in associated (GIS based) simulations. A process of iterative testing by user groups and iterative refinement is being used to ensure that the final outcome meets user needs, based on the Experience and Application Research (EAR) Methodology advocated by ISTAG. [3]

2.1 Information Visualisation and the Planning Process

Communication and visualisation are already at the heart of the planning system. The rationale for visualisation in environmental planning and design, according to Langendorf, is based on three premises: To understand nearly any subject of consequence it is necessary to consider it from multiple viewpoints, using a variety of information; Understanding complex information about urban planning and design may be greatly extended if the information is visualised; Visualisation aids in communicating with others. [4] Traditionally, there is a strong relationship between plans and cartographic representations, and the planning discipline. Maps, plans, sketches, images or other cartographic representations are (besides language) communication media for planning. According to some researchers only these media are able to clearly demonstrate visually the complexity of different demands on space.[5] However, there has been little consideration given to the aspects of 'mapping' in the planning literature to date. Likewise, cartographic science has failed to connect theory and research to the real-world tasks of spatial planning, and the question of how communication through 'planning maps' could be improved has been given little attention.

Over the last thirty years, many perception and cognition studies have been undertaken in cartographical science, and the effects of visual variables have been investigated systematically.[6] [7] [8] However, there is growing recognition amongst cartographers worldwide that there will never be an all-embracing theory of map reading, or of cartographic communication in general. This is due to the fact that there are certain variables which are difficult to control or even to identify in the process, such as the map user's skills and capacities, and the purpose the map is used for in different circumstances. Rase for example states that 'we are not sure what really attracts the attention of the reader, how the essence of the map is extracted, how the content is stored in memory, or what makes a specific map type superior to another one under certain conditions'.[9]

As a result, empirical research and the theories proposed appear disjointed and unconnected. In the light of a missing comprehensive theoretical framework, many cartographers thus create maps and cartographic illustrations following their knowledge and experience rather than any 'universal' and comprehensible rules and guidelines. Cartography is thus often still applied as a craft discipline rather than a science, and knowledge of map acceptance and map perception is mostly based on intuition, assumptions and personal experience, and to a lesser extent on scientific evidence.

There is little previous work that explicitly addresses the relationship between planning and cartography, and the role and function of cartographic representations in the planning process.[10] The majority is published in the "home-country language" focused on regional, urban or local planning processes, and thus intended to inform 'domestic' policy-making, rather than provide a basis for comparative research. Neuman for Spain, Lussault for France, and Gabellini for Italy, for example, have investigated the communicative potential of visualisations in urban planning.[11] [12] [13] While Söderström's work has concentrated on understanding how visualisations structure the activity of planners in a Swiss town.[14] There is also some work on the communicative potential and significance of architectural drawings for architects themselves.[15] Hence it remains difficult to see how the localised interpretation of planning meaning using cartographic symbols can be made adequately transparent to the untrained user.

2.2 Computer Generated Visualisation Models

According to other research recent moves towards computer generated visualisation models reflect the acknowledgement that conventional techniques fail to communicate environmental information either effectively or clearly.[16] In regard to computer-generated visualisation, architecture has been greatly influenced by computer technologies such as computer-aided design (CAD) software packages. In parallel Geographic Information Systems (GIS) have been extensively developed with urban planning as a major area of application. The sheer diversity of computer technologies and the versatility of their application has encouraged an interest in their use for visualisation in planning. A UCLA (USA) group of researchers (Ligget, Jepson and Friedman) has pioneered exemplary information rich 3D modelling environments, while other researchers such as Day at Bath University have tested CAD generated computer visualisation of urban environments.[17] Findings by a research group at the Queensland University of Technology, Australia, led by Buccolo, tested computer visualisation on the design for the new town centre of Capalaba and the Brisbane Airport.[18] That group identified certain advantages in the use of computer visualisation when compared to the traditional static tool, but they suggested more systematic research be carried out to prove the reliability and validity of new technologies and tools. Similarly Mahmoud confirmed this conclusion and extended it by proposing more experimental research.[19] The importance of visualisation is being recognised as crucial for almost all environmental and planning professionals who need to represent, communicate and evaluate design ideas and planning proposals.[20] According to Sawczuk "the design and planning process revolves around client's needs and therefore the client should be part of the team…"[21] The findings revealed that while skilled participants appreciate traditional media, such as drawings, unskilled participants prefer photo realistic presentations. Similarly, it was reported that when lay-people were exposed to architects' drawings "plans had little meaning as the people could not understand what was represented".[22]

2.3 Collaborative Web Based 3D Visualisation and What If Exploration

The VEPS project is researching the extent to which users, who make decisions together about proposals for change of mutual concern in a digital environment, endorse those decisions when they are confronted by the final built outcome. The EAR methodology is being used to work with groups of users, in order to lead to user engagement in the research process and to user driven optimisation of the final system. Initial workshops held in the UK, France and Germany, focused on Planners, Architects, Citizens and Politicians, and invited open discussion about needs and issues. The workshops have identified among others the following requirements for collaborative use of such a system:

- Filter for non-public data needed to protect planners and other professionals;
- 'Citizens PC' in Council House, to ensure that access is available for all;
- Comments /objections visually represented;
- Search functions/queries;
- Email alerts sent to users when a new development is proposed in their area of interest;
- Users can set their own visualisation preferences;
- Enabling of accessibility features, such as force feedback on input devices;
- A guided tour around the features of the user interface, and featuring 'context-aware' help, such as speech bubbles popping up to help identify user interface features and similar;

- Public participation with complete argumentation chain or just results;
- Interaction with other users AND with Decision Makers;
- Leave comments; Rate / rank comments; Peer group moderation; Reply to comments;
- Run / choose and explore different options;
- While possibly facilitated by Decision Makers, must be seen as independent / unbiased.

Based on analysis of the outcomes of these workshops rich pictures and storyboards are being developed to illustrate a series of scenarios. These will be examined in a further series of workshops, but also presented to a wider user group using the Web with an associated questionnaire. Tools are being developed to enable an initial asynchronous collaborative discussion by that wider group, together with facilities to add comments and mark-up to images (storyboard and rich pictures). The rich pictures and storyboard illustration stage will be followed by digital mockups. Where possible these will be based upon existing digital city or countryside modelling in VRML or X3D, and tools are being developed to enable comments and markup to be appended as if overlaid on the viewport. (This is similar to the work undertaken by M.C. Stellingwerff in his research with Dutch Architects.[23]) HCI and interface issues, together with tools not easily enabled in VRML or X3D will be tested and user reaction and comments obtained from selected parts of current CAD and GIS systems deployed via thin client. These initial tools and scenarios will be used to guide the development of both the geo-portal and the eventual VEPS system. Other work is investigating how citizens can upload their own modelling into this context for exploration of 'what-if' alternatives without requiring CAD skills. A promising direction is being trialled by staff at Liverpool Hope University using a simple descriptive tool to assemble libraries of building elements (initially based on medieval church architecture), that is then compiled from XML into VRML for viewing and interaction. Initial tests with school children have found that they can produce quite sophisticated models within half an hour of introduction to the process.[24]

In tandem project partners are examining the modelling data and distribution standards for 3D City and environmental modelling. It is currently intended to build a geoportal using the Web 3D Service. The first instance of this is anticipated based on the Stuttgart City Model in September 2005. Two alternate emerging standards are being investigated. The first is a Geo extension to Industry Foundation Classes (IFC) called IFG, since it is acknowledged that some modelling will emanate from Architects or other professionals and this is the emerging interchange standard in Building Product Modelling (recently submitted to the ISO TC184/SC4 for accreditation). The second is an implementation schema within the Geographic Markup Language (GML3 based on ISO 19107) called CityGML that is the work of a German based working group, and that adds semantic classes such as Building or Wall to the geometric and topological classification in GML.[25] Within City GML a series of 5 levels of detail and accuracy of 'City Model' have been expressed. The simplest level 0 consists of a digital terrain model (DTM) classified by land use. Level 2 is a city model with buildings, roof shapes, and textures. Levels 3 and 4 are considered likely to have been obtained from CAD import and this is where a critical relationship needs to be established with IFCs.

The goal of the project is that groups of citizens can engage with each other, and possibly with 'experts', and discuss proposed changes in a VR context virtually on the web, sharing common interfaces and able to add comments and mark up. Beyond this project partners are developing tools to allow these groups to collaboratively develop and discuss the alternatives, and to upload their own models into the same context for exploration. While the full project results are not expected until 2008 initial results of the first iteration of tools and tests with collaborative discussion and comment on line by user groups are expected by the end of this year. These results will then be used to guide the refinement and further development of the system.

It is recognised that visualisation optimised for the web tends to offer lower levels of detail than CAD generated 3D models, however initial investigation questions the extent to which endorsable decisions are enhanced by photo-realism and to which ‘too much’ detail acts as a distraction to focused decision making at early proposal stages when it is more important to assess a range of potential interpretations. Little research has yet been found that addresses these issues or that establishes how citizens can make reliable informed decisions at an early stage in the planning process to guide rather than react to detailed proposals. It is suggested that neither photo-realism nor over-much detail are likely to assist in this process, so web based visualisation may not demand the bandwidth required by highly photo-realistic interactive large area models.

3 Conclusion

It is argued that replication on the web of existing plan and cartographic-based approaches to planning will tend to perpetuate the requirement for trained interpretation, thereby excluding the general public. Effective tools are held to be required for collaborative citizen based examination and development of alternative proposals. This paper has described the basis on which the recently commenced VEPS project will seek to establish a new open standard web approach to enable citizens to view, analyse, interact with and respond to proposed changes, to collaborate and comment together and pose and test their own alternative solutions. This approach it is argued needs to be based on an interactive three-dimensional virtual reality visualisation that allows the viewer to experience the highly complex information without need for training. The visualisation also needs to support easy to use upload of citizens own models, and to provide extensive support for shared comment and debate.

Acknowledgment. The project work presented in this paper is co-financed through the INTERREG IIIB North West Europe programme.

References

[1] Punter J., Carmona M. (1997), The Design Dimension of Planning, E& FN Spon, London

[2] Virtual Environmental Planning System funded Dec 2004 to 2007 by INTERREG IIIB NWE-ENO (a Community Initiative Programme funded by the European Regional Development Fund www.nweurope.org)

[3] ISTAG Report on Experience and Application Research "Involving Users in the Development of Ambient Intelligence" http://www.cordis.lu/ist/istag.htm

[4] Langendorf, R. (1992) The 1990's; information systems and computer visualisation for urban design, planning and management, Planning and Design (19) pp 723-738

[5] Lutterbach, D. (1998) Auswirkungen moderner Visualisierungstechniken auf die kartographische Kommunikation in Planungsverfahren. Kartographische Nachrichten, Vol. 48, No. 2, pp. 52-58

[6] Medyckyi-Scott, D., Board, C. (1991) Cognitive cartography: a new heart for a lost soul. Advances in cartography. Muller, J. C. (ed.), Essex: Elsevier Science Publishers Ltd, pp. 201-230

[7] Gill, G. A. (1988) Experiments in the ordered perception of coloured cartographic symbols. Cartographica, Vol. 25, No. 4, pp. 36-49

[8] Wenner, A. (1989) Welche Aussagen läßt die Blickregistrierung über die Gestaltung aktueller Planungskarten zu?. Wiener Schriften zur Geographie und Kartographie, No. 3, pp. 221-232

[9] Rase, W.D. (1991) Computer-assisted Map Design, in Muller, J.C. (ed) Advances in Cartography. Elsevier, Essex, pp 181-200.

[10] Dühr, S. (2003) Illustrating Spatial Policies in Europe. European Planning Studies, Vol. 11, No. 8, pp. 929-948.

[11] Neuman, M. (1996) Images as Institution Builders: Metropolitan Planning in Madrid. European Planning Studies, Vol. 4, No. 3, pp. 293-312

[12] Lussault, M. (1994) La Ville Clarifiée. Essai d'analyse de quelques usage carto et iconographiques en œuvre dans le projet urbain. EIDOS, Vol. 9-10, pp. 37-58

[13] Gabellini, P (1996) Il disegno urbanistico. Rome: La Nuova Italia Scientifica

[14] Söderström, O. (2000) Des images pour agir. Le visuel en urbanisme. Lausanne: Editions Payot.

[15] Robbins, E. (1997) Why Architects draw. Cambridge, MA; London: MIT Press

[16] Pietch S.M. (2000) Computer visualisation in the design control of urban environments: a literature review, Environment and Planning : Planning and Design, volume 21, pp 521-536

[17] Day, A. (1994) New Tools for Urban Design, Urban Design Quarterly: 51

[18] Buccolo et al. (2001),Client Expectation of Virtual Environments for Urban Design, Proceedings of IIV2001 Information Visualisation International Conference, IEEE 2001, California, pp690-694

[19] Mahmoud A.H. (2001) Can Virtual Reality Simulation Techniques Reshape the Future of Environmental Simulations? in Online Planning Journal

[20] McKechnie, G. E.(1997) Simulation techniques in environmental psychology in Stokols, D.(ed.) Psychological perspectives in environment and behaviour, Plenum Publishing Corp. New York,

[21] Sawczuk, B (1992) The management of the design process, in Nicholson, M. (ed) Architectural management, E&FN Spon, London, pp84-89

[22] Harrlichack M.A, (1993) The effect of rendering technologies on the evaluation of computer generated design simulation, Master thesis, College of Env. Science and Forestry, State University of New York, Syracuse, NY

[23] Stellingwerff M.C. (2005) Virtual Context - investigating the characteristics and opportunities of digital visualisation media for situated approaches to architectural design in an urban environment. Doctoral Thesis at Technische Universiteit Eindhoven on line at http://www.bk.tudelft.nl/users/stelling/internet/thesis/stellingwerff_thesis.pdf

[24] Farrimond B. Hetherington R. (2005) Compiling 3D Models of European Heritage from User Domain XML, in proceedings of IV05 9th Intl Conference on Information Visualisation London, edited by Banissi et al, published by the IEEE Computer Soc. Los Alamitos California. pp 163 - 174

[25] Ewald, K., Coors, V. (2005) Appraisal of Standards for 3D City Models, in proceedings of IV05 9th Intl Conference on Information Visualisation London, edited by Banissi et al, published by the IEEE Computer Soc. Los Alamitos California. pp 157- 162.

Designing Virtual Reality Reconstruction of the Koguryo Mural

Yongjoo Cho[1], Kyoung S. Park[2], Soyon Park[3], and Hyungtae Moon[4]

[1] Division of Media Technology, Sangmyung University, 7 Hongji-dong, Jongno-gu, Seoul 110-743, Korea
ycho@smu.ac.kr

[2] Digital Media Laboratory, Information and Communications University, 517-10 Dogok-dong, Gangnam-gu, Seoul 135-854, Korea
park@icu.ac.kr

[3] Department of Media Art, Jeonju University, 1200 Hyoja-dong, Wansan-gu, Jeonju, Korea
sopark@jj.ac.kr

[4] Department of Computer Science, Sangmyung University, 7 Hongji-dong, Jongno-gu, Seoul 110-743, S. Korea
moonsgoms@hotmail.com

Abstract. Digital Koguryo is a virtual reality reconstruction of the Koguryo mural tumulus called Anak No. 3. It was designed to help young children learn the cultural background and life style of the ancient Koguryo. This virtual environment brings the life aspects of the Koguryo culture to the users through a rich, interactive learning environment. In this paper, we present the collaborative design process among geographically distributed and multidisciplinary researchers, who worked to construct the Digital Koguryo project. We will also discuss design issues and lessons learned from this collaborative work.

1 Introduction

The Koguryo murals can be considered as the historical books or Koguryo-style televisions that vividly reflect the life and historical events of the Koguryo civilization and details of the custom special in Korea. Koguryo (37 BC – 668 AD) is one of the three kingdoms of ancient Korea. The cultural background of Koguryo is revealed to have been of considerable complexity with various layers of artistic, religious and cultural influences upon its neighbor countries including Paekje and Shilla and also upon the cultural development of Japan. Koguryo mural paintings are rich in color and tone, and they still retained their distinct colors for about fifteen hundred years. Women dancing, warriors practicing, birds singing in the sky, dragons, fish in rivers, beasts in forest, wind and clouds of the murals appear so real and fresh that they look like they would jump out of the wall at any time.

Digital Koguryo reconstructed a Koguryo's mural painting tomb, Anak No. 3, in virtual reality. The Anak No. 3 Tumulus is a large stone-built structure with multiple chambers and the mural paintings are drawn on the walls and ceilings. It is particularly famous for a 10-meter wide wall painting that depicts a long procession. Digital Koguryo propels visitors on a journey of discovery to the life of old country as they

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 194 – 201, 2005.

were in the age of Koguryo. The goal of the Digital Koguryo project was to create a learning environment that can enrich young children understanding of the past and the history, life and values of the Koguryo cultural heritage. Research in virtual reality and cultural heritage has shown considerable growth in recent years [2, 4]. Virtual heritage applications use the immersive and interactive virtual reality technology to give visitors access to computer reconstructions of historical sites that would normally be inaccessible due to location or fragile condition.

This paper presents a collaborative design effort among geographically distributed, multidisciplinary researchers, such as archeologists, artists, historians, and computer scientists as they worked on the construction of the Digital Koguryo project. The designers, while they were distributed in several research institutes, had worked closely together over six months from initial planning discussion to final deployment. During this collaborative design, we had encountered a number of design issues and often we had to compromise each other over the desire of building accurate digital reconstructions and creating real-time and interactive VR experiences. This paper first gives a brief overview of the Koguryo mural painting tombs with a special attention to the Anak No. 3 Tumulus. Then, it describes the collaborative design process of building a VR reconstruction of the tomb. It will also discuss some of the important design issues raised throughout the development process.

2 Digital Koguryo

There are thousands of Koguryo tombs spread in North Korea and China, and about a hundred so far discovered have wall paintings. The wall paintings found on the tombs offer unique insights into the Koguryo custom and life style. About 60 Koguryo mural tombs, located in the area of suburban Pyongyang and South Hwanghae province in North Korea, were listed as the UNESCO World Cultural Heritage in July 2004. Among the tumuli, the Anak No. 3 Tumulus is the oldest and the biggest. The tomb, dated 357 AD, is a tomb covered with an earthen mound and has four chambers and a corridor. The limited access is undertaken on this tomb for the conservation and protection.

In Digital Koguryo, the viewers can enter at the front entrance of the digital reconstruction of the Anak No. 3 tumulus. They can walk around the stone chambers and a corridor that have mural paintings on the walls and the ceilings. The mural paintings depict several historical items and events, such as the portraits of a king and a queen in the left side chamber, everyday life and culture of Koguryo people (e.g., a woman cooking in the kitchen, the three-legged crow on top of a roof, women pounding something in a tread mill, a women drawing water from a well, horses and carriages standing on a stable, etc) in the right side chamber, a musical performance in the inner middle chamber, and a long procession featuring as many as two hundreds warriors in the corridor. As the viewers approach to the paintings, the two dimensional painted figures become life-sized three dimensional characters and give narrations about historical items or ask the viewer questions to find information about the paints.

Fig. 1 shows the photographs of Digital Koguryo demonstrated at various exhibitions. A large number of people can participate in the VR experience at the same time by wearing special lightweight stereo glasses. A participant can walk through and

Fig. 1. Visitors interact with Digital Koguryo on a passive stereoscopic virtual reality system

interact in the space using a joystick. Audio is enabled through the use of loudspeakers to give narrations. Digital Koguryo is primarily designed to run on a passive stereoscopic virtual reality system, called GeoWall [3]. GeoWall provides passive polarized stereoscopic three-dimensional graphics using two low-cost projectors and a Linux or Windows PC. Circular polarizer is used to project both the left and right eye images simultaneously on the polarization-preserving screen. The observers wear low-cost polarizing three-dimensional movie glasses to see the immersive contents. If desired, an additional tracking system and pointing device can be incorporated to support intuitive three-dimensional interaction.

3 Design Process of Constructing the Digital Koguryo

Digital Koguryo was created in collaboration among geographically distributed, multidisciplinary researchers including an archeologist at Art and Technology Expression Center (ATEC) at Seoul Institute of Arts, media artist teams at Jeonju University, VR programmer and system engineers at Information and Communications University and Sangmyung University. Our primitive goal of this project was to create an educational content of the Koguryo cultural heritage. The Koguryo mural tombs are historical relics, which had large influence on the development of Eastern culture in the medieval ages for their outstanding architecture and astonishing painting techniques. In the absence of contemporary historical texts from the Koguryo kingdom, we can gain an insight into the social and cultural aspects of Koguryo through the mural paintings of the tumuli.

In this section, we will briefly discuss how the designers worked together to accomplish their tasks in the construction of Digital Koguryo. Fig. 2 illustrates the collaborative design process for developing this project. First, all designers were gathered in one place to discuss initiating this project. In this project-planning meeting, the archeologist explained the value of digital restoration of this tomb and also introduced the internal sketches of the tomb as well as several slides of the mural paintings. Then, the archeologist and the graphic designers worked together to collect detail information of the tomb, paintings, and historical documents. Next was to design the virtual environment in collaboration with the archeologist, graphic designers, and

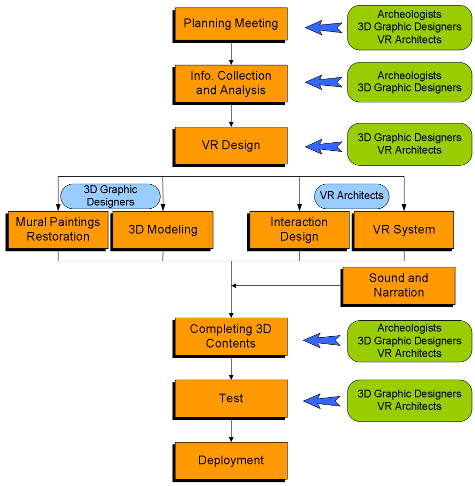


Fig. 2. Collaborative design process of developing the Digital Koguryo project

VR programmers, which included 2D paint restoration, 3D modeling, interaction and system design. In this stage, the designers have met several times via Microsoft's NetMeeting™ or on-site meeting to tune the contents of virtual environment, data formats, and other technical details. Throughout the meetings and testing, we found that 3D models had to be simplified so that they can be rendered in an interactive and real-time virtual environment. The digital restoration of the mural painting texture images was also reduced to smaller size to meet the real-time rendering needs. More design issues were brought up as programmers added user interactions in the virtual environment. Lastly, sounds and narrations created by the artists and the archeologist were added in the virtual environment.

3.1 Information Gathering and Archiving

While the Anak No. 3 tumulus exists in North Korea, little is known about the detail structure of the tomb and its mural paintings. Hence, it is a quite challenging task to reproduce the tomb in virtual reality. We use a variety of resources to find this information. First, we studied and collected the old documents, films, and slides. Fortunately, our archeology society had some slide cuts of the wall paintings and the tomb itself since the first excavation happened before South and North Korea got separated in 1953. We gathered some pictures and photographs of the wall paintings provided by the Koguryo expert historians, webs, about 50 papers, a few Koguryo books, and a TV documentary and a multimedia CD-ROM showing of the tomb. While we couldn't get the most up-to-date pictures or information about the tomb, we were still able to have enough information to build a virtual environment. We also selected some significant historical relics inside the mural paintings that might be of interest to the students in order to provide detail information about these items in the virtual environment. The archeologist visited the media artist teams at Jeonju University once to explain the tomb and its mural paintings and to discuss the restoration process. The collected information was stored in Webhard™ shared repository [7].

Fig. 3. The left image shows an original photograph taken from the mural painting of a queen in Anak No. 3 tomb; the middle is the illustration; the right shows after the restoration.

3.2 2D Image Restoration

Next was the two dimensional digital restoration of mural painting figures and backgrounds through color correction and filling. As shown in Fig 3, the original mural paintings appear to be in fairly good condition considering the fact that they have been exposed in a damp environment for about 1500 years. However, the existing surfaces of the paintings are damaged and corroded in various places, and hence, do not provide enough details. In particular, the mural painting of musical performance in the inner middle room has only 30 percent of the original shape left. In this case, the rest 70 percent obscure figures and erased parts were reconstructed by analogy with Koguryo related documents and historic relics.

The digital restoration for two-dimensional images of the mural paintings was done through the following process: scanning photographs taken from the murals and books, followed by creating illustrations for each wall painting to make the base image, and then correcting color, hue, and intensity of the digital images and filling the abrasion parts. In this stage, we had difficulty in inferring the color tone of mural paintings and reproducing the cuts and abrasions on mural paintings. For the color tone problem, we received help from the best person in the area of Koguryo mural painting research. The restored colors were inferred from analyzing the polygenetic natural dyes of the Koguryo period and how the colors would be changed with moisture. For the cuts and abrasions problem, we had to analogize the entire painting out of the remains by interring them from the texture of the painted walls. In this stage, the archeologist and media artists worked collaboratively using instant messaging systems, a shared repository, and phone conferencing for acquiring documents and images.

3.3 3D Modeling

We tried to digitally restore the Anak No. 3 tomb as if it exists today, to enrich learners to understand historical and artistic value of the wall paintings. The three-dimensional modeling of the tomb structure and cultural relics were completed based on historical documents that we collected from historians and archeologists. The 3D modeling for reconstructing the existing cultural heritage required tight collaboration among artists, archeologists and computer scientists due to many challenging design issues. First, we had to depend on the old documents and photographs because we could not visit to North Korea to measure the actual size of the tomb and artifacts. We also needed to compromise between the accurate restorations to the original form and the real-time rendering issues for virtual reality.

The artists created the 3D models of the tomb and relic objects with the inspection from archeologists. The first step was to measure the size of the tomb from a map on a scale of one to twenty to make it as a real tomb. Then, precisely measured walls, pillars, and other historical objects were placed in the virtual tomb. In the three-dimensional modeling, we had to consider making object models lighter for real-time rendering, i.e., the models should have as fewer polygons as possible to make the graphics rendering faster in a virtual environment. It was challenging to produce the virtual objects look similar to the real objects while reducing the number of polygons.

The modelers worked hard to manually optimize the number of polygons of objects because the automatic optimization tool offered by the 3D modeling tool often changed the shape of the objects while reduction. In the modeling of a pillar, for example, the number of polygons was originally 6302. The automatic optimization shrunk the number of polygons down to 2313, but it made the object looked completely different from the original shape. Thus, the modelers had to create the object manually from the scratch with the consideration of lighter but not sacrificing the original shape, and by doing so the number of polygons shrunk down to 1928.

The texture maps were used for mapping the restored mural painting images on the virtual tomb walls. The artists and archeologists, at first, attempted to create the texture map images of the wall paintings as closely as possible to the original. For an example, the artists created the texture map of the 'long procession' mural painting that had about 3000 by 3000 pixels size. They wanted to present the images as clear as possible to the viewers in a virtual environment because this painting was the highlighting among the other paintings. While high quality of the texture map images enhanced the VR experience, this is a source of degrading the real-time rendering performance, especially the PC-based VR system. Hence, after testing and long discussion amongst the artists and VR software engineers over instant messaging systems, the size of the texture map images got shrunk down to 512 by 512 pixels or less to make the textures more manageable and light.

We also added about ten items of three-dimensional models for historical relics in the virtual tomb to make the virtual environment look more interesting and encourage viewers to come to know these objects. The items included a king's crown called 'Baeklakwan', a queen clothing, a cooking pot, a carriage, a horn, a Geomoongo (Korean lute), and so on. In the creation of 3D models for ten items, there was conciliation among the designers, which also affected the quality of digital restoration. Digital Koguryo asks students to find out the virtual objects matched to the items shown on the wall paintings. Fig. 4 shows the wall paintings and the virtual objects placed in front of the wall, which could be interacted by the user. The archeologists wanted the artists to make the objects exactly the same as the excavated objects shown in the right image of Fig. 4. However, these models had to be simplified for real-time rendering and user interactions in virtual reality.

Fig. 4. The left image shows the three-dimensional models of cooking pots reconstructed from the mural painting as experienced in Digital Koguryo. The right shows the pot excavated from the tomb.

3.4 VR Interaction Programming

Digital Koguryo has been written using a high-level VR toolkit called Ygdrasil (YG) [5]. Ygdrasil is a set of C++ classes built around SGI's Performer graphics library [6], CAVElib VR library [1], and the CAVERNsoft networking toolkit [5]. Ygdrasil is designed to simplify the construction of behaviors for virtual objects using re-usable components; and on sharing the state of an environment through a distributed scene graph mechanism. Ygdrasil focuses on constructing dynamic, interactive, collaborative virtual worlds, so in addition to the basic graphical data as used in SGI Performer, its scene graph nodes can have behaviors attached to them. Individual nodes are compiled into dynamically shared objects (DSOs), so that they can be rapidly added to a world. The system includes a number of pre-made classes that implement common capabilities in VR, such as audio, avatars, navigations, and triggers that detect when a user enters an area.

The modelers and VR programmers worked closely to shape out the virtual environ ment and adding some user interactions. We wanted to use the models constructed from the modeling tools to create a virtual environment, but the reality was not so sim pleThe models had to be adjusted in the format to match with the unit defined in the virtual environment. Hence, the modelers and the VR programmers had to meet sever al times off-line and online to test the models for creating the VR world, such as the si ze of the tomb that allowed users to interact easily and the model format appropriate for YG scripting. After several changes in configurations, we finally decided the size, placement, and interaction of the virtual tomb and objects.

4 Conclusion

The Digital Koguryo project was a collaborative effort among archeologists, historians, media artists, computer scientists located in different research institutes. Digital Koguryo is a virtual reality reconstruction of the mural tomb, Anak No. 3. The aim of this project is to help students and scholars learn the spirits and life aspects of the Koguryo culture through a rich, interactive, and narrative virtual environment. This paper described our experiences in collaboratively designing and developing Digital Koguryo. The digital restoration and reconstruction of cultural heritage in immersive virtual reality is difficult due to technical and performance restrictions for real-time rendering. Hence, close collaboration among designers was required to reach the aforementioned architectural details and relic objects that create believable and interactive experience. This paper illustrated our experience of designing the digital restoration of cultural heritage in concert with other collaborators at several production phases using various collaboration tools.

References

1. CAVElib. http://www.vrco.com/
2. Gaitatzes, A., Christopoulos, D., Voulgari, A. & Roussou, M. Hellenic Cultural Heritage through Immersive Virtual Archaeology. In Proc. 6th International Conference on Virtual Systems and Multimedia, Ogaki, Japan, October 3-6, 2000, pp.57-64
3. Geowall Consortium. http://www.geowall.org/
4. Pape, D., Anstey, J., Carter, B., Leigh, J., Roussou, M., Portlock, T., Virtual Heritage at iGrid 2000, In the proceedings of INET 2001, Stockholm, Sweden, 5-8 June 2001.
5. Park, K., Cho, Y., Krishnaparasad, N., Scharver, C., Lewis, M., Leigh, J., Johnson, A., CAVERNsoft G2: a toolkit for high performance tele-immersive collaboration, In the proceedings of the ACM Symposium on Virtual Reality Software and Technology 2000, Oct22-25, 2000, Seoul, Korea, pp. 8-15.
6. SGI Performer. http://www.sgi.com/products/software/performer/
7. Webhard http://www.webhard.co.kr/

A Cooperative System Environment for Design, Construction and Maintenance of Bridges

Nobuyoshi Yabuki[1], Tomoaki Shitani[2], and Hiroki Machinaka[3]

[1] Associate Professor, Department of Civil Engineering and Architecture, Muroran Institute of Technology, Muroran, Hokkaido, Japan
yabuki@news3.ce.muroran-it.ac.jp
http://yb3.ce.muroran-it.ac.jp
[2] Doctor Course Student, Division of Civil and Environmental Engineering, Muroran Institute of Technology, Muroran, Hokkaido, Japan
shitani@news3.ce.muroran-it.ac.jp
[3] Master Course Student, Division of Civil Engineering and Architecture, Muroran Institute of Technology, Muroran, Hokkaido, Japan

Abstract. This paper proposes a new cooperative system environment for design, construction and maintenance of civil structures, especially bridges. We identified seven tools required to improve the productivity of design, construction and maintenance of bridges. They are product models, visualization system, communication tool, task systems with agents, database system, radio frequency identification (RFID) tags, and data reuse facility. These tools are to be integrated and used by owners, design consultants, contractors, manufacturers, etc., who can exchange and share structural, construction and inspection data, information and knowledge with this system.

1 Introduction

In the process of design, construction and maintenance of civil structures, many engineers in heterogeneous disciplines from owners, design consultants, contractors, etc. perform various tasks in a distributed manner in terms of time and space. Since each task is done based on the contract, it is not easy to exchange broad range of information among clients, consultants and contractors. Furthermore, as the data interoperability among application systems is quite low, data sharing and reuse are hard to achieve so far. Thus, the productivity of construction industry is usually lower than other manufacturing industries. In order to solve this problem, extensive basic research has been done concerning product models, cooperative and concurrent design and engineering. However, not so much has been done to integrate not only design but also construction and maintenance of civil structures. In this paper, we discuss the required elements for developing a cooperative and integrated design, construction and maintenance system for bridges and describe the system development.

2 The System Architecture

In order to improve the productivity by using the cooperative system environment in the long-term lifecycle of bridges, we believe that the following seven tools are

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 202 – 209, 2005.

necessary. They are 1) product models, 2) visualization system, 3) networked communication tool in a distributed system, 4) task systems and agents, 5) data collection and database, 6) linkage between the ‘virtual’ world (data system) and the ‘actual’ world (bridges), 7) data reuse. The conceptual layout and relationship of the seven tools are shown in Fig. 1.

The product model is the key element in the total system environment. The product model is a generalized data model for representing geometric information of products and structures as well as various property information of each member, element and equipment based on the object-oriented paradigm. Most data concerning the structures and facilities are generated by application systems locally. Then, the local data is transformed to the interoperable data based on the standardized product model specification by the corresponding converting program so that data can be used by other application systems by using converting programs for them.

The visualization system displays the bridge and its related facilities two, three or four dimensionally to the user. Although the 3D CAD system is more suitable to represent complicated structures such as bridges than 2D drawings, 2D drawings are commonly used in business and 3D CAD is used only for special purposes at the moment. However, in the near future, we believe that virtual reality system that the user can feel immersed in the virtual world will provide much better user interface and will be used for design and visualizing the bridge structure and construction process.

As the owner, design consultants, contractors, manufacturers, etc. are usually geographically distributed, a networked communication tool is necessary to share and exchange data safely and without stress. This tool should facilitate the communication among the stakeholders so that in case of design change or various troubles in design, construction or maintenance all the necessary information should be sent to all the related recipients. And further, all the related stakeholders should be able to have a virtual meeting using this networked communication tool at each member’s place without visiting an actual meeting place.

Owners, design consultants, contractors, manufacturers, etc. are responsible for their own tasks such as total management, design, analysis, estimation, schedule control, documentation, etc. They use various application systems for these tasks. Each application system usually has its own function completely and contributes to the efficiency improvement significantly. However, due to the poor interoperability of data, cooperation among these application systems is not good and the total efficiency has not been improved yet. Thus, data is converted and shared by conversion programs. Further, to link various systems and convert data, agents play an important role. They move as an agent of the user around the network and systems autonomously, collect data, respond to the change of the system environment, and execute appropriate tasks corresponding the change or the user’s operation.

The data about the project and structure are stored in the database in the form satisfying the requirements of the specification of the product model. The data collection system should support the user for ingesting the generated data and information of the project at various stages of the project efficiently and stress-freely. The database should provide data retrieval function that the user can query the database easily.

The cooperative system treats data and information concerning civil structures such as bridges. Apparently, the data and information are abstract and virtual but not ‘real,’ while the actual bridge is a real thing. Since the real thing has not been constructed yet in the design process, the virtual data will do. However, in construction and maintenance, the actual objects are made and the linkage between the actual object and data about the object in the virtual world is weak. This tends to generate the estrangement between the virtual and actual worlds and hinders various minute engineering tasks. We believe that the radio frequency identification (RFID) technology will link these two different worlds more smoothly than now. Sensor networks will also provide data regarding the actual structures and help the user link the two worlds.

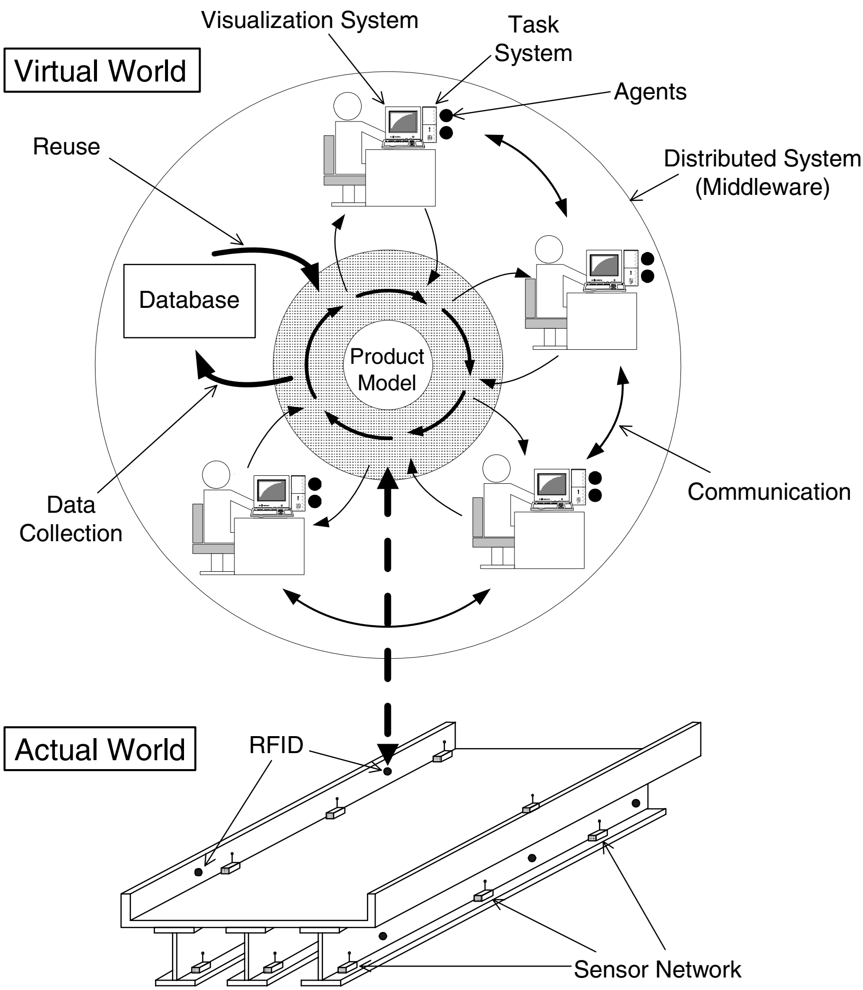


Fig. 1. Conceptual scheme of the cooperative system environment for bridges

A large amount of data is generated during the lifecycle of the project and stored in the database. The data is shared and exchanged among the stakeholders, which will improve the efficiency and productivity. Furthermore, if the data is reused in the future projects, this will generate much more effect. In order to reuse data effectively, a system with intelligence such as case-based reasoning (CBR) should be developed. In CBR, the system retrieves a similar case from the database and modifies it if necessary and produces a new case and stores it in the database for future use. And for retrieving or extracting new meaningful knowledge from the large-scale database, data mining would be effective.

In this section, the seven tools and their functions were briefly described. We are currently developing these tools and four of them, i.e., 1) product models for bridges, 2) a visualization system, 3) task systems with agents, and 4) an RFID linkage system, have mostly been developed. In the following section, these fours tools will be described.

3 The System Tools

3.1 Product Models for Bridges

We have been developing product models for bridges for several years based on Industry Foundation Classes (IFC) of International Alliance for Interoperability (IAI) [1]. IFC is an object-oriented data model for representing buildings and other construction related information. IFC contains not only physical properties of structures such as its parts and members, their dimensions, materials, shapes, locations, etc., but also spatial concepts such as floors, rooms, and abstract concepts including projects, organizations, etc. IFC has been developed to enable the interoperability among CAD and non-CAD application systems by IAI.

Our first version product model for prestressed concrete (PC) bridges was published in 2002 [2]. This product model was based on IFC2x (Version 2x) and we added new classes for concrete slabs, reinforcing bars, prestressing strands, anchoring device, sheaths, and voids for PC bridges to IFC2x. The detail of this model can be found in papers [3] [4]. We then developed a product model for steel girder bridges based on IFC2x and then modified it based on IFC2x2 (Version 2x Edition 2). We added new classes for representing plate girders, steel built-up members, steel plate elements, steel shape elements, etc. The detail of this model can be found in papers [5]. To implement the product model schema and instances, ifcXML (Extensible Markup Language) was used.

IAI French Speaking Chapter developed the product model called IFC-BRIDGE [6] at around the same time we developed our PC bridge product model. We discussed with them and agreed to merge our product model into IFC-BRIDGE to enhance it by international collaboration. To step forward the collaboration project, we organized Civil Engineering Group in IAI Japan Chapter in 2004.

3.2 Visualization System

The visualization system displays the data in a virtual world so that the user can imagine how the structure and facilities look like when they are built. The virtual reality technique can support the user to view and understand the structure more in detail and with more reality than 2D drawings or 3D CAD images on a flat display monitor. We developed a virtual reality CAD system, using OpenGL, GL4Java, liquid crystal shutter glasses (CrystalEYES3 of StereoGraphics), and a sensor system (FASTRAK of Polhemus) before [5]. We have been applying various virtual reality technologies to the bridge system. We have tried a head mounted display and a liquid crystal display monitor that the user can feel immersed in the virtual world without wearing special glasses as well.

The strength checking agent checks whether the section designed by the user is appropriate or not for the given loads by checking the conformance with the design code. As a design code, Japanese Highway Bridge Specification was used.

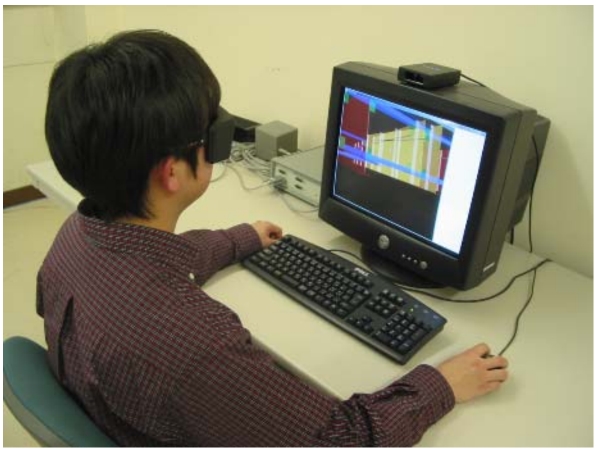


Fig. 2. Virtual reality CAD system with liquid crystal shutter glasses and infrared rays emitter

The design situation checking agent checks whether the design satisfies situational constraints such as: whether the selected section of the girder has the same height as the adjacent girder and whether the designed section is economical or not. This agent retrieves the location and section data of the related given by the product model data and checks the situation based on the knowledge given by design engineers.

Those agents work in the background and check the design autonomously when the user determines the section dimensions or modifies them. Multi-agents may make conflicts at some occasions, where some mediator agents may be necessary. However, the user plays as a mediator in this system at the moment. This may increase the opportunity for the user to consult with other designers and engineers about conflicts. We intend to develop agents for mediating such conflicts in the future.

3.3 Task Systems and Multi-agents

Task systems include 2D/3D CAD system, automated design system, design conformance checking system, structural analysis system, cost estimation system, construction schedule management system, contract documentation system, etc. To automatically share and exchange the data generated by these systems, conversion programs must be developed in accordance with the specification of the product model. As at the moment no standardized product model for bridges exist, we developed conversion programs by ourselves. However, commercial software vendors should make and attach conversion programs when the product model is to be standardized.

One of the conversion programs we developed so far is a two-way CAD conversion program called CAD2PM and PM2CAD. CAD2PM can generate product model data of PC bridge objects as an ifcXML instance file, while creating 3D shapes of the objects with AutoCAD 2002. PM2CAD can retrieve data from an instance file of the product model by using an XML parser and render the 3D model in AutoCAD automatically.

Various agent programs have been developed in our research. For PC bridges, interference checking agent, rebar cover checking agent and rebar space checking agent have been developed. The interference checking agent can find interferences of members, such as reinforcing bars, prestressing strands, sheaths, which are buried in concrete (Fig.3). If interference is found, the agent informs the user and shows the interfered parts on the CAD display monitor. The rebar cover checking agent can detect reinforcing bars that lack required cover autonomously and informs the user. The rebar space checking agent can check the spaces of reinforcing bars in conformance with design codes such as Japan Highway Bridge Specification.

For steel girder bridges, strength checking agent and design situation judging agent were developed [5]. The strength checking agent can check if the section designed by the user is appropriate or not for the given load in terms of structural mechanics. The design situation checking agent checks whether the design satisfies situational constraints such as consistency of member height and clearance.

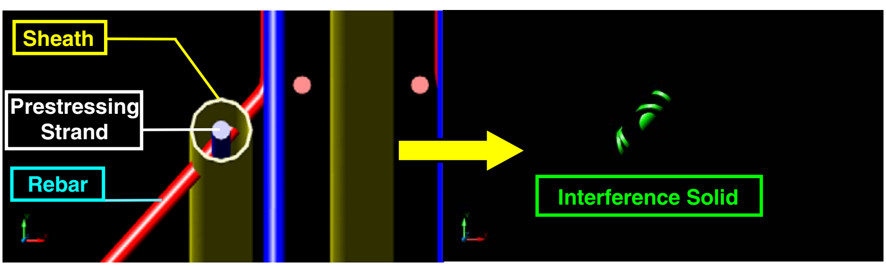


Fig. 3. Interference among sheath, prestressing strand and reinforcing bar

3.4 Linkage Between Data and Structures with RFID

There are two types of RFID. One is an active RFID, which has a battery, and the other is a passive RFID, which does not have a battery. In our research, to link the object data in the database and the actual object itself, we use passive RFID. For network sensors, we use active RFID. A passive RFID consists of an IC chip containing a memory and an antenna. RFID is usually resin molded for protection. By using a reader/writer, which consists of an antenna and a controller and which is attached to a computer or a personal digital assistant (PDA) the user can input data such as ID number and text data into the RFID tag and can read information contained in the tag (Fig. 4).

Fig. 4. RFID and reader/writer (left) and an inspector retrieving data from RFID by PDA (right)

We put 68 RFID tags to various members, equipments and measuring devices at Haneji Dam in Okinawa, Japan for supporting inspectors [7]. Our experience showed that inspection support system developed by us with RFID, PDA and database improved the efficiency of the inspection task, decreased the number of errors and enhanced the technical transfer by promoting and encouraging the users to output, exchange, and share various information and knowledge with each other. We are going to put a large number of RFID tags to a bridge for the inspection purposes this year.

4 Conclusion

In this paper, we proposed a cooperative system environment for design, construction and maintenance of bridges. The system is an integration of seven tools: product models, visualization system, networked communication too in a distributed system, task systems and agents, data collection and database, linkage between data system and bridges, and data reuse. In this paper, we described the four tools that have mostly been developed, which are product models for bridges, the visualization system, task systems and agents, and linkage between data and structures with RFID. As this is an ongoing research, other tools are now under development. We intend to integrate these tools and apply the system to a real bridge for evaluation and verification.

References

1. International Alliance for Interoperability: http://iai-international.org.
2. Yabuki, N. and Shitani, T.: Development of 3D Product Model for PC Hollow Slab Bridges Based on IFC (in Japanese), Journal of Civil Engineering Information Processing System, Vol.11, pp.35-44, 2002.

3. Yabuki, N. and Shitani, T.: An IFC-based Product Model for RC or PC Slab Bridges, Proc. of CIB w78, International Council for Research and Innovation in Building and Construction, Waiheke Island, New Zealand, pp.463-470, 2003.
4. Yabuki, N. and Shitani, T.: Development and Application of an IFC-Based Bridge Product Model, Proc. of the Ninth International Conference on Civil and Structural Engineering Computing, Egmond aan Zee, the Netherlands, B.H.V.Topping (Ed), Civil-Comp Press, Paper 93, pp.1-14, CD-ROM, 2003.
5. Yabuki, N., Kotani, J. and Shitani, T.: A Cooperative Design Environment Using Multi-Agents and Virtual Reality, Proc. of First International Conference, CDVE 2004, Palma de Mallorca, Spain, Y. Luo (Ed.), pp.96-103, 2004.
6. IFC-BRIDGE: http://www.iai-france.org/bridge/
7. Yabuki, N., Yamashita, T., Shimada, Y. and Sakata, T.: Application of RFID Tags to an On-site Inspection Support System for Large Dams, Proc. of the third Civil Engineering Conference in the Asian Region, Seoul, Korea, pp.397-400, 2004.

A Concurrent Approach to Design of Reconfigurable Machine Tools to Process Bamboo

R. Carles Riba[1], R. Roberto Pérez[2], A. Jorge L. Sánchez[2], María D. Domínguez[2], S. Joaquín Aca[3], and G. Arturo Molina[3]

[1] Industrial Equipments Design Centre. Mechanical Engineering Department. Polytechnical University of Catalonia. Diagonal 647, 08028, Barcelona, Spain
carles.riba@upc.edu
[2] University of Holguín, Mechanical Engineering Department, Cuba
{roberto.perez, jorge.sanchez, mdominguez}@facing.uho.edu.cu
[3] CSIM – ITESM. Campus Monterrey, México
{aca, armolina}@itesm.mx

Abstract. Today, manufacturing systems must compete in increasingly diverse and demanding markets. Customers demand quick response to their needs, and products have shorter life cycles than ever before. In response to this context, the next generation of machine tools should be reconfigurable and intelligent. Reconfigurability allows for the reduction of machine design lead time, machine set-up and ramp-up time. This paper shows the development of an approach to concurrent design of reconfigurable machine tools to process bamboo, based on a concurrent design reference model to Reconfigurable Machine Tool development. In response, a methodology for the development of this type of machine tools is outlined.

Keywords: Reconfigurability, Concurrent Design, Machine Tool, Bamboo.

1 Introduction

Nowadays, the increasing in the speed of changes which take places in the different places of manufacturing in the global economy and the constant aggressiveness of the competition, lead to the creation of mechanical systems which are highly adjustable to the requirements of the environment. These conditions demand the development and application of approaching which allows to convert the production to new models capable of integrating new technologies and of increasing the variety of products.

Over the last few years, considerable efforts have been made to develop and implement Reconfigurable Manufacturing Systems (RMS) that provide a cost effective answer to such market demands [1]. In principle, RMS are designed to adapt rapidly to changes in product mix and volume. The key is to rapidly update the structural design of the equipment; both in terms of hardware and software. The art and science of RMT (Reconfigurable Machine Tools) development are still at their developmental stage.

A goal of the work presented in this paper is to provide a methodology for the design of RMT to obtain pressed wood out of bamboo, departing from a reference model. In particular, this paper proposes a design process for RMT to obtain press

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 210 – 217, 2005.

wood out of bamboo, and outlines a methodology for the designs of modules reconfigurables for RMT. On section 2 a literature review of research in Reconfigurable Machine Tools (RMT) is presented. A Reference Model used to configure a first approach of the methodology to design RMT is described in Section 3. Section 4 presents experiences obtained from implementation of this methodology, and describes the bottlenecks that are particular to RMT design. Finally, Section 5 deal with the methodology developed for the reconfigurable design of bamboo processing machines that need to be integrated in the library of the Reference Model to make it applicable to RMT design.

2 Related Previous Studies

The characteristics that determine the ease of reconfigurability are the following: modularity, integrability, customization, convertibility and diagnosability [2]. Table 1 summarizes the major studies about the reconfigurability.

Table 1. Summary of studies about the reconfigurability

Fields of research	Authors
– Modularity of machine from the functional and constructive point of view	Stone, Robert B. [3] Fixon, S.K. [4]
– Modular design of a machine tool related with the environment and control modules	Garro and Martin [6]
– Machine tool with reconfigurable spindle	Katz and Chung [7]
– Machine tool where the type, position and location of spindle can be adjusted depending on the requirements	Koren and Kota [2]
– Production requirements impact in the design of reconfigurable machines	Lander [8]
– Develop of systematic design tools and analysis of dynamic stiffness of a machine tool using pre–calculated component information	Zatarain *et al.* [9]
– Methodology to evaluate the dynamic characteristics of RMT	Yigit and Ulsoy [10]
– Kinematic synthesis of machine tools starting from a mathematical description of the machining tasks	Moon and Kota [11]

Modularity allows for a fast and reliable integration during the change of structure of the RMT from given production capacity to a different production capacity for a desired amount of products from a given family. The studies of the modularity in the RMT are sparse, and for the most part dedicated to the modular selection [5].

All of these studies address specific topics related to RMT design and construction. However, they do not address the design process as a whole.

3 Reference Model

To develop a methodology for RMT design, this research project proposes the use of a Reference Model that allows the companies to create a Particular Model to set-up successful Integrated Product, Process and Facility Development Processes (IPPFD). The model is independent of the industrial sector of a company, but focuses on specific issues of the company like market opportunities, technological constraints and declared goals. This model has been tested for different product development scenarios [12], [13].

The Reference Model developed is defined as the complete representation of activities to complete the Product Life Cycle. The proposed Reference Model is described through three axes: Processes, Stages and Activities (see Figure 1).

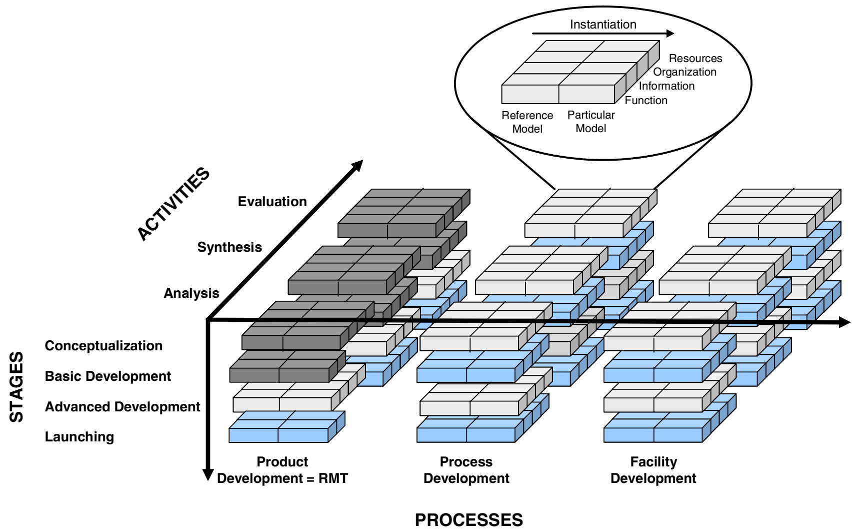


Fig. 1. Reference Model for IPPFD that was applied to RMT design to obtain a formal methodology

— Processes to be developed during engineering projects: Product, Process and Facility Development. Facility Development involves three sub-processes: (a) Product Transfer; (b) Technology Transfer; and (c) Machine Design: if the component to be manufactured is not standard and the technology to manufacture this product is not available, then it is necessary to design a new machine or product to get the component.

— Stages are indicators of evolution level for processes: Conceptualization, Basic Development, Advanced Development and Launching.

— Activities are specific tasks that must be executed in order to complete a stage. These are classified in three types: (a) Analyses Activities are oriented to diagnose, define and prepare information; (b) Syntheses Activities are oriented to laying together elements to produce new effects and to demonstrate that these effects create an overall order; and (c) Evaluation Activities are oriented to test those solutions against the goals and requirements.

The properties of this Reference Model are:

— Reusability / Configurability: ability to be configured in a Particular Model to get a specific goal in the Product Life Cycle focusing on specific issues of the company like: market, knowledge and technology.

— Robustness: based on a proven library of methods and tools to ensure information flow among product development stages and avoid the lack of collaboration between design engineers and manufacturing engineers.

— Integral: due of its structure, the model is able to adopt new methods and tools from different disciplines (e.g. mechanics, electronics) and integrate them allowing for the development of particular models to different industries.

Based on this Reference Model a Particular Model for RMT development was prepared. The details of the methodology for the Conceptualization and Basic Development stages to develop a RMT (product development) for bamboo processing and the experiences obtained from this process are presented in the following sections.

4 Bamboo Processing and Reconfigurable Machine Development

The methodology for the development of RMT was prepared and tested around a specific design case. The Cuban eastern region stated the need for a machine that could be adapted to process different types of local bamboos. A team of two master ship students, working under the supervision of a Professor was responsible for applying the reference model and deploy the design methodology using this case. The scope of the process was limited to the stages of Conceptualization, Basic Development and Advanced Development. Some details of this process are presented in the following paragraphs. In Figure 2 the conventional process to obtain pressed wood out of bamboo and its related machines are shown.

Conceptualization: during this stage of the methodology the objective is to define the project to be developed and delegate responsibilities for project execution. For this case study the customer request a reconfigurable machine able to process different types of bamboo from the eastern region of Cuba, following the path of the conventional process show in Figure 5. Two master ship students of the Faculty of Engineering at the University of Holguín were responsible to follow the methodology to develop the machine in a period of three months.

Basic Development: during this stage the objective is to capture customer requirements, establish target specifications and generate and select a feasible concept to be developed in next stage. To capture customer requirements and translate them to technical specifications, Competitive Benchmarking and Patent Analysis techniques were used. The requirements and specifications for this type of machine were: high speed (40 m/s); medium precision (+/- 1 mm); quick change and set up time (0.5 hr maximum); work envelope of 0.8 m x 1 m; maximum wood thickness of 150 mm; and maximum wood length of 3 m.

After the specifications were defined, a functional structure was obtained for each operation mode of each one of the machines which make up the process of wood processing. In Figure 2 the obtained functional structure for the case of the planning machine in its main operation mode: planning is show.

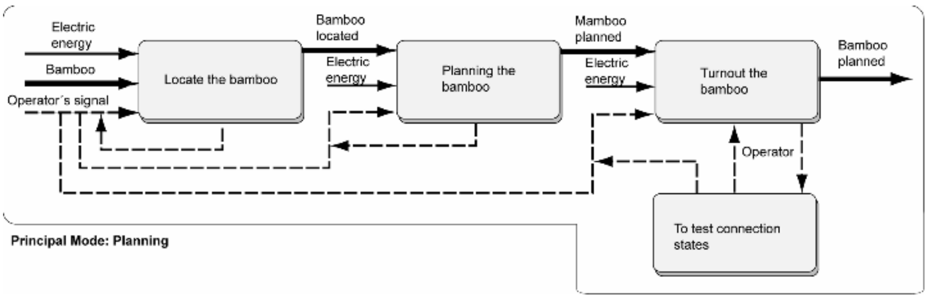


Fig. 2. Functional decomposition to the principal machine mode of operation

Advanced Development: Mathematic models were used for the selection of components such actuators, drivers and machine elements. Selected components were modelled using a commercial CAD system (Pro/Engineer).

Figure 3 shows the structure of the RMT machine that was designed to polishing and planning bamboo, which was design and built in the Machine Developing Centre (CEDEMA) (Holguín, Cuba). The important features in this design are the reconfigurable modular construction of the tool holder. Both of them can be rapidly changed when changing from one process to the other (polishing or planning).

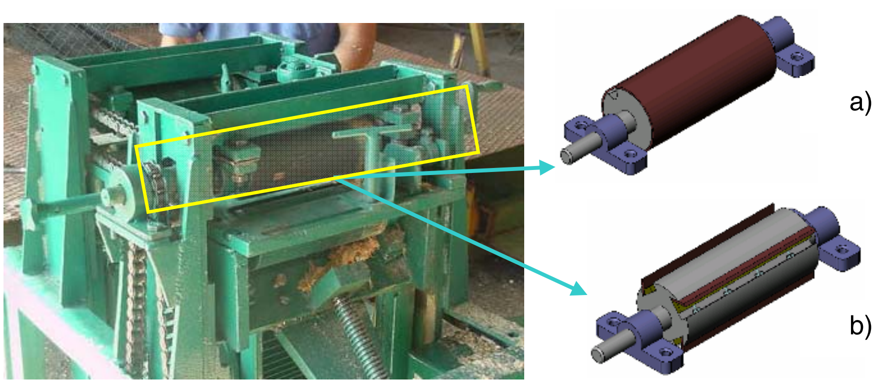


Fig. 3. Reconfigurable machine prototype to polishing (a) and planning (b) bamboo

The methodology was analyzed to identify the strengths as well as the bottlenecks that were exposed during the solution of the test case. This methodology allowed departing from the different functional structures in each machines and for each operation mode, the identification of the possible reconfigurable modules of the machines which provided the design of a reconfigurable machine for polishing and planning wood made of bamboo.

The biggest difficult found in developing and applying this methodology was to identify which where the most recommendable reconfigurable modules to be used in the group of machines which make up the process. The test case exposed the need for a tool to bridge this gap, that is, between the functional structure to be performed and reconfigurable modular components to facilitate the concurrent design of a RMT.

5 Reconfigurable Design of Bamboo Processing Machines

The general characteristics of a methodology based on reconfigurability are proposed now to fill the gaps that were found in the test case. In general, modular design methodologies are based on two perspectives: functional oriented design and constructive oriented design. In the functional approach, modules are designed in terms of the functions that each module is to perform, and satisfies the customer's objectives by the simple addition or subtraction of modules. On the other hand, the constructive approach produces a modular decomposition that facilitates fabrication, assembly and transportation during the product life cycle.

The fundamental characteristic of RMT's is their modular construction, which is neither completely functional, nor completely constructive, but rather possesses reconfigurability oriented modularity. *Reconfigurable modularity* refers to a set of principles and rules that define a family of machines or products, to the manner and shape in which modules are standardized and their respective interfaces. These modules allow modifications to the structural and functional configuration of the machine, in such a way that significant changes to product form and demand can be accommodated.

The proposed methodology for the modular development of RMT's to process bamboo establishes four domains (see Figure 4) which go from the requirements of the machine tool builder to the definition of the reconfigurable modularity. Each layer established by the methodology constitutes a formalization of the knowledge that is obtained in any given step.

The first domain allows the identification of the machine tool builder's requirements that affect not only the development of the machine but also its ability to be reconfigured. This domain is developed in the general framework. Each functional requirement contains a set of variables that facilitate their analysis in later stages of the model. This step helps to establish the requirements layer, which is simply a representation of the intentions of the machine tools builder about the reconfigurability, as well as basic information of the machine's builders, useful in the second step.

The methodologies developed about the functional structures present two basic problems [3], [4]. First, there is no clear methodological procedure for the definition of the ideal sub-functions for a given set of requirements. Second, there is no method that guarantees a coherent decomposition of the functions into sub-functions. The step of Functional Structure's Domain, is a way to solve the previous problem, using a combination of a hierarchical functional decomposition, heuristic rules and knowledge about the machine (see Figure 4).

Once all functions have been decomposed to a given level, the last layer will represent the sub-functions that are more related to the final structure of the machine.

This process facilities the decision making process about the reconfigurability of the machine. Once this step is concluded, the layer of the modular structure is obtained, which is only a transformation of the information that is contained in the layer of functional requirements into more concrete information with the modularity of the product.

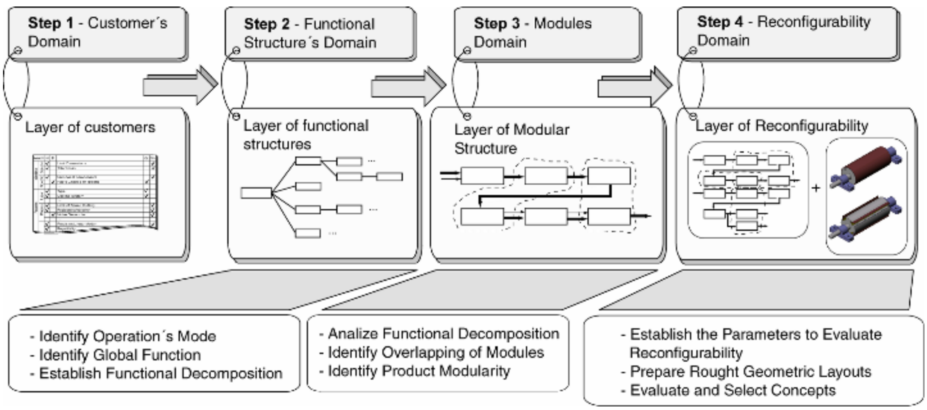


Fig. 4. Overview of the developed design methodology

Four activities are recommended within this step (see Figure 4). The first one consists of identifying which are the parameters that characterize the reconfigurability of the machine, or family of machines. The second activity consists of generating preliminary geometric schematics that allow one to determine configurations for the machine from the modular structure layer. To determine possible variants for the concepts, heuristics criteria are used.

In Figure 5, the integration of the planning and polishing function in only one function is shown. This methodology is applied as a way of integrating the reconfigurability to machine concurrent design.

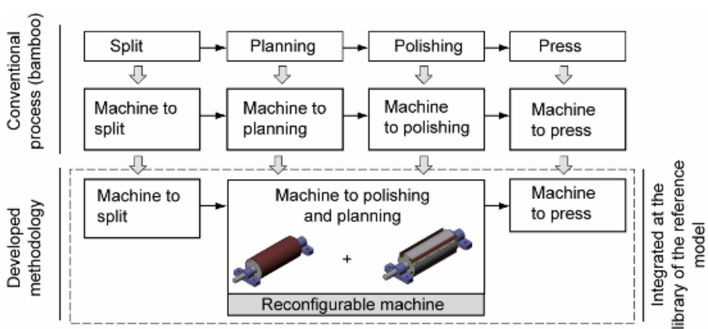


Fig. 5. Functional integration departing from developed methodology

The work presented can facilitate the integration of the system to the model of reference and can facilitate the concurrent design of machine tools to process bamboo.

6 Conclusions

This article has presented a framework for Concurrent Design of Reconfigurable Machine Tools to process bamboo. The developed methodology provides an aid in the identification of the reconfigurable modules starting from the knowledge of the requirements of the machine tool builder. This methodology was integrated to the library of techniques and tools of the references model.

Acknowledgement

This research is part of the Research Chair in Concurrent Engineering at Mechanical Engineering Department at the University of Holguín (Cuba), the Technical Department of KARIME workshop (Cuba) and the Industrial Equipments Design Centre at Polytechnical University of Catalonia (Spain). The authors want to acknowledge the support of this grant in the preparation of the manuscript.

References

1. Koren, Y., et al., Reconfigurable Manufacturing Systems, Annals of the CIRP, Vol. 48/2 (1999) 527–540
2. Koren, Y. and Kota, S., Reconfigurable Machine Tools, U.S. Patent 5,943,750 (1999)
3. Stone, Robert B., Towards a Theory of Modular Design, Doctoral Thesis, University of Texas at Austin, United States of America (1997)
4. Fixon, S.K., Three Perspectives on Modularity – A Literature Review of a Product Concept for Assembled Hardware Products, Massachusetts Institute of Technology, ESD-WP-2001-06 (2001)
5. Moon, Y-M., Kota, S., Design of Reconfigurable Machine Tools, Journal of Manufacturing Science and Engineering, Vol. 124, May, (2002) 480-483
6. Garro, O. and Martin, P., Towards New Architectures of Machine Tools, International Journal of Production Research, Vol. 31, No. 10 (1993) 2403–2414
7. Katz, R. and Chung, H., Design of an Experimental Reconfigurable Machine Tool, Japan–USA Symposium on Flexible Automation, Ann Arbor, Michigan, (2000)
8. Landers, R.G., A New Paradigm in Machine Tools: Reconfigurable Machine Tools, Japan–USA Symposium on Flexible Automation, Ann Arbor, Michigan, (2000)
9. Zatarain, M., Lejardi, E., and Egana, F., Modular Synthesis of Machine Tools, Annals of the CIRP, Vol. 47/1 (1998) 333–336
10. Yigit, A.S. and Ulsoy, A.G., Application of Nonlinear Receptance coupling to Dynamic Stiffness Evaluation for Reconfigurable Machine Tools, Proceeding IME, Part B: Journal of Engineering Manufacture, Vol. 216, No. B1 (2002) 87-101
11. Moon, Y-M., Kota, S., Generalized Kinematic Modeling of Reconfigurable Machine Tools, Journal of Mechanical Design, Vol. 124, March, (2002) 47-51
12. Aca, J., Reference Model and Methodology to Configure/Reconfigure Integrated Product, Process and Facility Development Processes, Master Thesis in Manufacturing Systems, ITESM, Campus Monterrey, (2004)
13. Pérez R., R., *et al.*, A Modularity Framework for Concurrent Design of Reconfigurable Machine Tools, Lecture Notes in Computer Science, Volume 3190, (2004) 87-95

Model for an Integrated Analysis of a Building's Life Cycle

Edmundas Zavadskas, Arturas Kaklauskas, Algirdas Andruskevicius, Povilas Vainiunas, and Nerija Banaitiene

Vilnius Gediminas Technical University
Sauletekio al. 11, LT-10223 Vilnius, Lithuania
property@st.vtu.lt

Abstract. In order to design and realize a high-quality project, it is necessary to take care of its efficiency from the brief stage to the end of its service life. The entire process should be planned and executed with consideration of the goals aspired to by the participating interested parties and the micro and macro environmental levels. In order to realize the above purposes an original Model of an Integrated Analysis of a Building's Life Cycle was developed by the authors to enable users and the parties involved in the project to analyze a building's life cycle, as well as to be able to see its micro and macro environments as one integrated entity. To achieve the above-mentioned aims new multiple criteria analysis methods and a Multiple Criteria Decision Support System for Building's Life Cycle were developed.

1 Introduction

A thorough building's life cycle analysis is quite difficult to undertake, because a building and its environment are a complex system (technical, economical, social, ecological, etc.), where all sub-systems influence the total efficiency performance and where the interdependence between sub-systems play a significant role.

Over the last decade, there have been extensive research and development activities in many countries regarding service life prediction (SLP) methods for building materials, products and components. Internationally, the outcome of many of these efforts have been reported and applied within two bodies: the working commission CIB W80/RILEM 175-SLM "Service life methodologies" and the standardization technical committee ISO/TC 59/SC 14 "Design life". Since 1999, the commission CIB W80/RILEM 175-SLM has been working on describing three levels of SLP methods (probabilistic based methods, engineering methods, and the simple factor method) [7].

Lacasse and Sjöström [15] outlined advances in three areas of estimating the service life of building products, components or systems. These include a factorial method, an engineering design approach and reliability-based methods. Performance based methods for service life prediction focuses on two approaches for estimating the service life: the factor method, and the engineering method.

It can be noticed that researchers from various countries engaged in the analysis of a building's life cycle but its stages did not consider the research's object as was analyzed by the authors of the present investigation. A life cycle of a building may be de-

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 218 – 226, 2005.

scribed as follows: the parties involved in its design and its realization as well as the micro and macro environments, having a particular impact on it and making an integral whole. A complex analysis of the research's object that was formulated and made with the help of methods multiple criteria project analysis specially developed by authors for this purpose and is presented in this paper.

The paper is structured as follows. Following this introduction, Section 2 explains a model for an integrated analysis of a building's life cycle. In Section 3 we have provided a practical realization of this model. Section 4 describes testing the developed systems and model. Finally, some concluding remarks are provided in Section 5.

2 A Model for an Integrated Analysis of a Building's Life Cycle

In order to design and realize a high-quality project, it is necessary to take care of its efficiency from the brief stage to the end of its life's service. The entire process should be planned and executed with a consideration of the goals that are aspired to by the participating and interested parties and the micro and macro environment levels. In order to realize the above purposes, an original model of a complex analysis of a building's life cycle was developed by the authors, enabling one to analyze a building's life cycle, the parties involved in the project and its micro and macro environment as one integrated entity. A Model for an Integrated Analysis of a building's life cycle has been developed in the following stages:

- Comprehensive quantitative and conceptual description of a research object;
- Multi-variant design of the life cycle of a building;
- Multiple criteria analysis of the life cycle of a building;
- Selection of the most rational version of the life cycle of a building;
- Development of rational micro and macro levels of the environment.

3 Practical Realization of a Model for an Integrated Analysis of a Building's Life Cycle

The investigation carried out by the authors of this paper under Framework 5 and 6 (BRITA in PuBs [1], LIFETIME [4], CONSTRINNONET [5], PeBBu [6]), Phare-ACE programmes [2, 3], and other projects [9, 11, 21] helped to identify and describe major trends of the life cycle of a building's development as well as the development and practical realization of a Model for an Integrated Analysis of a Building's Life Cycle. A practical realization of a Model for an Integrated Analysis of a Building's Life Cycle was developed in the following five stages (Fig. 1):

- *Stage 1.* A comprehensive quantitative and conceptual description of the life cycle of a building, its stages, interested parties and the environment.
- *Stage 2.* Development of a complex database, based on quantitative and conceptual descriptions of the research's object.
- *Stage 3.* Development of new methods of multiple criteria analysis to carry out multi-variant designs of a building's life cycle, to determine the utility degree of the obtained alternative versions and set priorities.

- *Stage 4.* Creation of a multiple criteria decision support system to be used in computer-aided multi-variant design of a building's life cycle, and so determining the utility degree of the obtained alternative versions and setting priorities;
- *Stage 5.* Analysis of micro and macro level environment factors that influence a building's life cycle and the possibilities to alter them in a desired direction.

The stages mentioned above will be now described in more detail.

The authors performed a comprehensive quantitative and conceptual description of the life cycle of a building, its stages, interested parties and the environment during the first stage in the following steps:

- A system of criteria characterizing the efficiency of the life cycle of a building was determined by means of using relevant literature and expert's methods.
- Based on a system of criteria, a description of the present state of the life cycle of a building is given in conceptual (textual, graphical, numerical, etc.) and quantitative forms.

A quantitative and conceptual description of the research's object provides information about various aspects of a building's life cycle, i.e. economical, technical, technological, infrastructure, qualitative, legislative, and social, etc. Quantitative information is based on criteria systems and subsystems, units of measure, values and initial weight as well as data on the alternative project's development. A conceptual description of a building's life cycle presents textual, graphical, visual (videotapes) information about the projects and the criteria used for their definition, as well as presenting reason for the choice of a particular system of criteria, their values and weight. This part also includes information about the possible ways of completing a multi-variant design.

In order to perform a complete study of the research's object, an integrated evaluation of its economic, technical, qualitative, social, legislative, infrastructure and other aspects is needed. The diversity of aspects being assessed should follow the diversity of ways of presenting data that is needed for decision-making. Therefore, the necessary data may be presented in numerical, textual, graphical, formula, videotape and other forms. Quantitative and conceptual descriptions of a building's life cycle and its stages are used as a basis for developing a complex database containing the overall information about it and allowing one to carry out its multi-variant design and a multiple criteria analysis. Since the efficiency of any constituent part of the project depends on a particular party in its execution, only the complex design of a building's life cycle that involves close co-operation of all the interested parties can yield good results.

Alternative building's life cycle versions include different aspects such as the cost of a plot and a building, maintenance costs as well as various architectural, aesthetic, space planning, comfort characteristics, infrastructure and environmental pollution. Particularly interested parties often have their own preferential rating of these criteria and this result, in different values of the qualitative characteristics. Besides, the designing of a building's life cycle allows for the development of many alternative versions at particular stages. This causes many problems in determining the most efficient project.

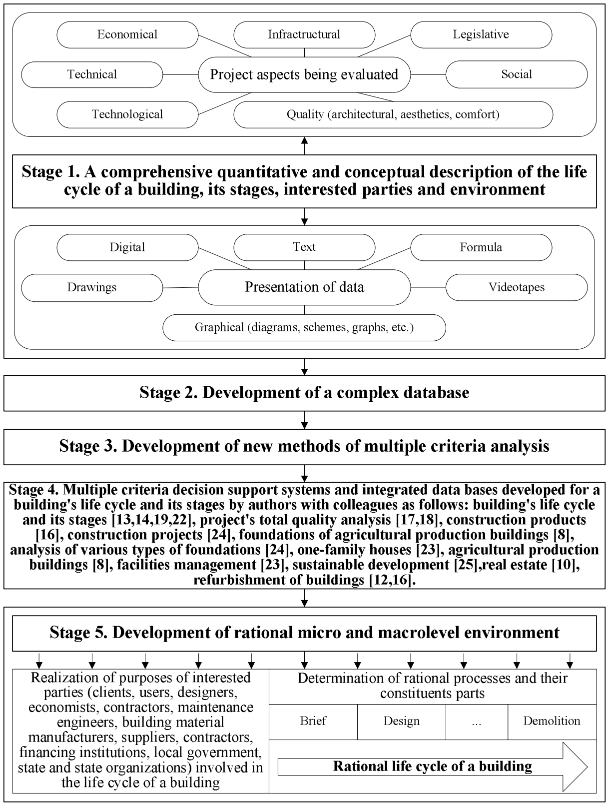


Fig. 1. Practical realization of a model for a complex analysis of a building life cycle

During the second stage the integrated databases, based on quantitative and conceptual descriptions of the research's object were developed. Integrated databases consist of the following parts:

- Initial database that contains the initial data provided by various interested parties, allowing one to carry out a complex design of the whole project or its parts.
- Evaluation database, containing comprehensive quantitative and conceptual information provided by interested parties and allowing one to receive a full description of the alternative variants. Based on the evaluation database, a multiple criteria analysis of a building's life cycle and its stages is then carried out.
- Multi-variant design database consisting of comprehensive quantitative and conceptual information about the possible combinations of the available alternative variants.

These integrated databases can contain data on theoretical and practical experiences of the interested parties, some additional facts, as well as recommendations as to how to avoid previous mistakes. For example, a Multiple Criteria Decision Support System for Building's Life Cycle (DSS-BLC) as developed by the authors [10, 20] and by using these databases, can help one to compare the project being designed or executed, with an alternative or already realized project, in order to find its disadvantages and provide recommendations so as to increase its efficiency. In this way, the use of integrated databases enables the user to take into account an expert's knowledge (including building owners and users, financing organizations, architects, engineers, contractors, etc.). Expert advice in various fields plus previous experiences gained in developing similar projects can then be applied to the currently developed project. For more efficient projects, this information should also be used at an early stage such as when first meeting with a client because it could save the repetition of prior mistakes and lead to a more advanced and efficient project. In making a complex building's life cycle architects, designers, utility engineers, economists, contractors, suppliers and users can solve common problems more efficiently. Better problem solving results in lower project costs and building time, as well as increasing its quality.

During the third stage the authors developed the following methods of multiple criteria analysis [19, 20]:

1. A new method of complex determination of the significance of the criteria taking into account their quantitative and qualitative characteristics was developed.
2. A new method of multiple criteria complex proportional evaluation of the projects enabling the user to obtain a reduced criterion determining complex (overall) efficiency of the project was suggested. This generalized criterion is directly proportional to the relative effect of the values and significances of the considered criteria on the efficiency of the project.
3. In order to find what price will make an object being valuated competitive on the market a method for determining the utility degree and market value of objects based on the complex analysis of all their benefits and drawbacks was suggested. According to this method the object's utility degree and the market value of an object being estimated are directly proportional to the system of criteria that adequately describe them and the values and significances of these criteria.
4. A new method of multiple criteria multivariant design of a building's life cycle enabling the user to make computer-aided design of up to 100,000 alternative project versions was developed.

To develop and analyze thousands of alternative variants based on dozens of criteria, each having specific values and weight would hardly be possible without the use of intelligent systems. Only the development of intelligent systems can help to solve this complicated problem. To achieve the above aims, during the fourth stage, a DSS-BLC and other systems (see Fig. 1) consisting of integrated databases, a database management system, a model base, a model base management system and the user's interface was created to be used for a building's life cycle and the design of its constituent parts and a multiple criteria analysis.

By interacting with the database a user can also receive more detailed or integral information on the considered object. Given this opportunity and by using data from integrated databases as well as being provided with DSS-BLC [10, 20], the user can then find an effective project variant in a comparatively short time. In this way, a project best satisfying the needs of the client can be found and also save the client and designer's time. Two versions of DSS-BLC have been developed. The first version was created in 2000 and described in different authors publications [20]. Now a second Web-based DSS-BLC version is under development (see http://dss.vtu.lt/lifecycle/).

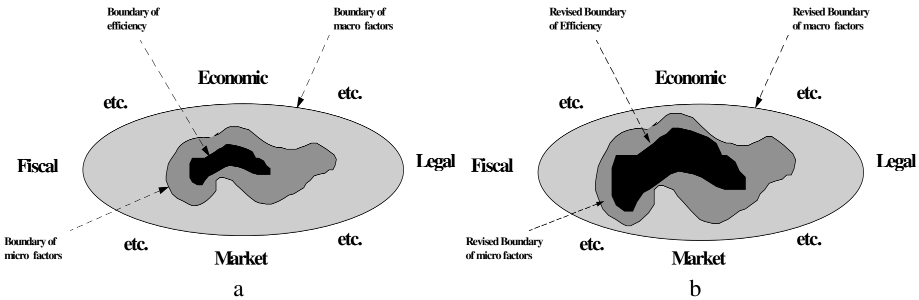


Fig. 2. Macro and micro factors influencing the efficiency of a building's life cycle

The authors performed an analysis of micro and macro level environment factors that influence a building's life cycle and the possibilities to alter them, in a desired direction during the fifth stage in the following six steps:

- Identifying the global development trends (general regularities) of the micro and macro level environment factors that influence a building's life cycle.
- Identifying a building's life cycle differences between developed countries and Lithuania.
- Determining pluses and minuses of these differences for Lithuania.
- Determining the best practice for Lithuania as based on actual conditions.
- Development of some of the general recommendations as to how to improve the efficiency levels for micro and macro levels.
- Submission of particular recommendations.

In order to assure the efficiency of a project's life cycle, it should be implemented within certain bounds, which are determined by the micro and macro environments. The fact is that these factors are different in every country, so are the possibilities for an efficient realization of projects (see Fig. 2a) and will vary.

Figure 2a indicates diagrammatically factors at macro and micro levels that can impinge on the efficiency of the building's life cycle. This means that to be efficient the building's life cycle must operate within certain boundaries as imposed by the macro and micro factors. Recognizing that in each country the factors will be different, this diagram will also vary accordingly. It is necessary to utilize knowledge and experience about the macro and micro level factors, so as to increase the efficiency level of the environment in each country that is being considered. This was performed by analyzing the experience and knowledge of advanced industrial economies and by applying this to Lithuania.

Using taxation as an example, it can be appreciated that if the level of real estate taxation is high, the volume of construction output decreases. Similarly, if the real estate tax level is lowered, volume of construction production increases.

Such changes in taxation will alter the boundary of efficiency of the building's life cycle. Similar micro and macro environmental changes can shift this boundary, i.e. the area within the boundary of efficiency expresses the total satisfaction level of needs of all interested parties. For example, the Lithuanian government (in order to solve the most important problems within Lithuania's society) can abolish VAT on new residential buildings in order to promote investment in housing. Thus, the boundary of efficiency is extended to include new development from the former situation. After the development of the Lithuanian financial sector, the boundary will alter again (Figure 2b illustrates a revised level of efficiency as an example of how to take alterations into account).

4 Testing the Developed Systems and Model

In order to test the usefulness of the multiple criteria decision support systems and integrated data bases developed for a building life cycle and its stages by authors with colleagues, final semester master degree students from the Construction Economics at Vilnius Gediminas Technical University solved more than 50 practical examples during the final thesis. These students work in various construction and facilities management sector companies in Vilnius. In order to check the correctness of the developed systems, the whole of its solution process has been more than once gone through manually. The results of manual and computer calculations matched. Besides, all separate working stages of the systems as well as all complex calculations have been coordinated with experts in this field – i.e. the essence of the calculations has been found to be in conformity with their logical reasoning. Owing to suggestions of these experts, some useful changes have been introduced in the developed systems. The stages of a Model for an Integrated Analysis of a Building's Life Cycle were tested in different EU and Lithuanian programmes and projects [1–6, 11, 21].

5 Conclusions

As developed by the authors, the Model for a Building's Life Cycle Complex Analysis, methods of multiple criteria analysis and Multiple Criteria Decision Support System for Building's Life Cycle can define an efficient building's life cycle when many various parties are involved. The alternative project versions add up to hundreds of thousand and the efficiency changes with alterations to the environment's conditions and the constituent parts of the process in question are then also possible.

References

1. Bringing Retrofit Innovation to Application in Public Buildings (BRITA in PuBs). Project Reference: EcoBd/503135/2003 (2003)
2. EC Phare-ACE Programme. Contract Number: P96-6708-F. Total Life Analysis, Modelling and Forecasting of Construction in Lithuania (1996)
3. EC Phare-ACE Programme. Contract No ACE-92-0250-R. Achieving Quality Management in Construction (1992)

4. Framework 5. Lifetime Engineering of Buildings and Civil Infrastructures (LIFETIME). Contract GTC1-2001-43046 (2001)
5. Framework 5. Promoting Innovation in Construction Industry SMEs (CONSTRINNONET). Contract IPS 2000-00002 (2000)
6. Framework 5. Performance Based Building (PeBBu). Contract GTC3-2001-63013 (2001)
7. Hovde, P. J.: The Factor Method – a Simple Tool to Service Life Estimation. In: 10DBMC International Conference on Durability of Building Materials and Components. Lyon, France (2005) CD
8. Kaklauskas, A.: Multiple Criteria Evaluation of Technological Decisions of Agricultural Buildings. PhD Dissertation, Dnipropetrovs'k Civil Engineering Institute, Dnipropetrovs'k, Russia (1990) (in Russian)
9. Kaklauskas, A.: Total Life Analysis, Modeling and Forecasting of Construction in Lithuania. Research output, EC Phare-ACE Programme, Vilnius. Project No.: P96-6708-F (1998)
10. Kaklauskas, A., Zavadskas, E.K.: Web-based Decision Support. Technika, Vilnius (2002) (in Lithuanian)
11. Kaklauskas, A., Zavadskas, E., Banaitis, A.: Systematization of Housing Construction and Renovation Goals. Order of the Ministry of Urbanistics and Construction. Vilnius (1997) (in Lithuanian)
12. Kaklauskas, A., Zavadskas, E. K., Raslanas, S.: Multivariant Design and Multiple Criteria Analysis of Building Refurbishments. Energy and Buildings 37(4) (2005) 361–372
13. Kvederyte, N., Zavadskas, E.K., Kaklauskas, A.: Multiple Criteria Alternative Design of Building Life Cycle. Statyba (Journal of Civil Engineering and Management) 6(2) (2000) 128–142 (in Lithuanian)
14. Kvederyte, N., Zavadskas, E.K., Kaklauskas, A.: Multiple Criteria Analysis of Semidetached House Life Cycle. Statyba (Journal of Civil Engineering and Management) 6(3) (2000) 179–192 (in Lithuanian)
15. Lacasse, M.A., Sjöström, C.: Advances in Methods for Service Life Prediction of Building Materials and Components – Final Report – Activities of the CIB W802005. In: 10DBMC International Conference on Durability of Building Materials and Components. Lyon, France (2005) CD
16. Trinkunas, V., Kaklauskas, A., Zavadskas, E.K.: Selection of Rational Construction Products Regarding Building Refurbishments. Property Management 6(2) (2002) 74–82
17. Zavadskas, E., Kaklauskas, A.: Project Quality Increase by Application of Quality Expert System. Proceedings of 4th International Conference "Modern Building Materials, Structures and Techniques". Technika, Vilnius (1995) (in Lithuanian)
18. Zavadskas, E., Kaklauskas, A.: Total Quality Project Assessment System. Deutsh-Litauisch-Polnische kolloquium zum Baubetriebswesen. Hochshule für Technik, Wirtschaft und Kultur in Leipzig. 3 Jahrgang, Sonderheft1 (1996) 36–41
19. Zavadskas, E., Kaklauskas, A.: Multiple Criteria Analysis of Buildings. Technika, Vilnius (1996) (in Lithuanian)
20. Zavadskas, E.K., Kaklauskas, A., Banaitiene, N.: Multiple Criteria Analysis of a Building Life Cycle. Technika, Vilnius (2001) (in Lithuanian)
21. Zavadskas, E., Kaklauskas, A., Banaitis, A.: Perspective and Economical Validity of Refurbishment of Large-Panel Housing (1959–1970). Order of the Ministry of Urbanistics and Construction. Vilnius (1997) (in Lithuanian)
22. Zavadskas, E.K., Kaklauskas, A., Banaitis, A., Kvederyte, N.: Housing Credit Access Model: the Case for Lithuania. European Journal of Operation Research 155(2) (2004) 335–352

23. Zavadskas, E.K., Kaklauskas, A., Lepkova, N.: Multiple Criteria Analysis of Facilities Management Alternatives. Property Management 6(1) (2002) 31–39
24. Zavadskas, E., Peldschus, F., Kaklauskas, A.: Multiple Criteria Evaluation of Projects in Construction. Technika, Vilnius (1994)
25. Zavadskas, E.K., Rutkauskas, A.V., Kaklauskas, A.: Development of Rational Model for Lithuanian SUD. Statyba (Journal of Civil Engineering and Management) 5(2) (1999) 123–134 (in Lithuanian)

Cooperative Shared Learning Objects in an Intelligent Web-Based Tutoring System Environment

Sylvia Encheva[1] and Sharil Tumin[2]

[1] Stord/Haugesund University College, Bjørnsonsg. 45, 5528 Haugesund, Norway
sbe@hsh.no
[2] University of Bergen, IT-Dept., P. O. Box 7800, 5020 Bergen, Norway
edpst@it.uib.no

Abstract. This paper focuses on an intelligent Web-based tutorial system composed of reusable learning objects (LOs). Reusable LOs lay down the foundation for a large scale collaboration among educational organizations. LOs developed at different organizations can be used by other organizations thus providing high quality courses to all students. Expert instructors at each organization maintain a repository of LOs. For organizations owning LOs an audit trail for each LO can be easily gathered for billing purposes.

1 Introduction

An intelligent tutoring system (ITS) is an educational software system containing machine intelligence components. An ITS adapts instruction to a student's individual learning needs by collecting information on that particular student's performance. Such a system makes inferences about strengths and weaknesses, and suggests additional work. One of the most common solutions for student diagnosis is testing [6].

A critical aspect in the development of an ITS is how related knowledge is represented and how reasoning for problem solving is accomplished [7]. Some single knowledge representation schemes are used in [5] and [16], and hybrid representations in [8] and [13]. Learners tend to learn better and more deeply if they are motivated by an internal interest and desire to master the material, as opposed to extrinsic rewards and punishments such as grades [9].

In this paper we discuss an intelligent Web-based tutorial system (IWBTS) composed of reusable LOs. Reusable LOs lay down the foundation for a large scale collaboration among educational organizations. LOs developed at different organizations can be used by other organizations, thus providing high-quality courses to every student. Expert instructors at each organization maintain a repository of LOs. Such a flexible framework is not trivial. It is in fact difficult to design and implement. However, by carefully defining LOs and relations among them, together with careful analysis of integration among supporting subsystems, the framework can be realized.

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 227–234, 2005.

IWBTS provides a blueprint for solving each particular problem while introducing a new term or concept. The support for subsequent problems is limited by giving Socratic hints. The number of hints required for the student to solve each problem serves as a measure of learning efficiency. IWBTS also provides different levels of presentation with respect to problem solving. Each level has tutorial material generated for it, since it is important to target tutorial tasks at the student's ability. The tutorial model in IWBTS involves Socratic dialogs that foster critical thinking on course content through a series of questions and answers that enable students to explore and develop understanding. Corrective and elaborative aspects are used for tutoring in IWBTS.

The assessment part of IWBTS contains a polytomous Item Response Theory (IRT) based model [6] for handling data from several response options following each question in a test. IRT models for polytomous items not restricted to two response alternatives are applied by IWBTS to both multiple choice items that possess several response alternatives and to free-response items.

2 Related Work

A survey of current approaches in the area of technologies for electronic documents that are used for finding, reusing and adapting documents for teaching or learning purposes is presented in [17].

LOs are defined as any digital resource that can be reused to support learning [18] and any digital, reproducible, and addressable resource used to perform learning activities or learning support activities, made available for others to use [10]. These definitions do not discuss relationships among LOs in a reuse environment. Our model is based on independent LOs, where a LO is an atomic entity, i.e. one that cannot be decomposed to lower LOs.

A personalized intelligent computer-assisted training system is presented in [14]. A system mapping classroom lectures into Web-based educational lessons is described in [4]. A Category-based Self-improving Planning Module for a tutor agent that utilizes the knowledge learned from automatically-derived student categories to support efficient on-line self-improvement is presented in [11].

ITSs use a tutoring model to guide students' learning with Socratic hints. A Socratic hint is a cognitive and motivational strategy aimed at evoking the student's curiosity and requiring further thought from him [12]. The corrective behavior [15] of the software rejects a wrong value and conveys that it was incorrect. If there is a second consecutive mistake, then the elaborative behavior [15] comes into action, suggesting an appropriate relationship by which the value can be derived. The suggestion depends on what the student has done so far.

IRT, also called latent trait theory, is the study of test and item scores based on assumptions concerning the mathematical relationship between abilities (or other hypothesized traits) and item responses [2]. Polytomous IRT based models are capable of handling data from several response options while dichotomous IRT based models consider only two possible scored responses such as true/false, correct/incorrect, endorsed/not endorsed, etc. [1].

A framework for sharing protected Web resources among independent organizations is presented in [3]. Very important components of the model are a collaborative, distributive management of user membership in a group in one organization, resource management by the service provider, and security enhancement.

3 System Framework

IWBTS is based on a flexible framework that provides the instructors with possibilities in designing courses using shared, reusable LOs and provides the students with an ITS environment.

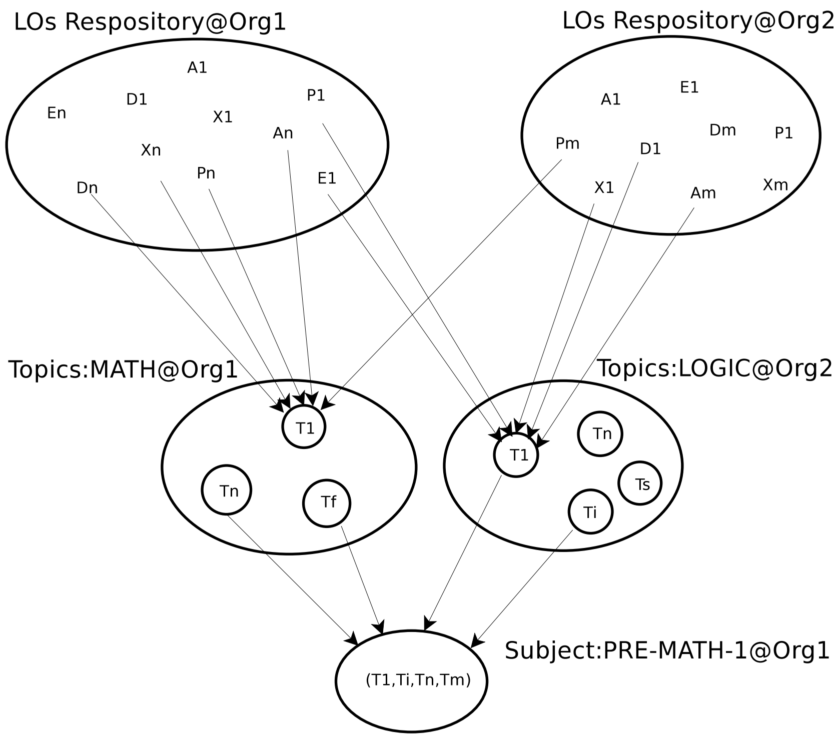


Fig. 1. Learning objects

Course design costs a significant part of a WEB-based learning project budget. A big cost reduction is possible if as many LOs as possible are reusable and are shared among collaborative organizations. Instructors can just define which LOs should be included in their courses. Using drills and students responses to questionnaires in a WEB-based course can provide students with an ITS environment. This will make the course building more flexible and reduce costs for both students and organizations.

A subject $\mathcal{S}_i = (\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_n)$ is defined as a strictly sequential list of topics $\mathcal{T}_j$ by a course builder (Fig. 1). Each $\mathcal{S}_i$ definition is saved in the database in XML. The subject's definition contains all the metadata needed to uniquely describe the subject in question. These metadata include server, domain, title, topics list, and other operational parameters.

A topic $\mathcal{T}_j = \{\mathcal{P}_\alpha, \mathcal{E}_\beta, \mathcal{X}_\gamma, \mathcal{D}_\delta, \mathcal{A}_\epsilon\}$ is composed of theoretical parts $\mathcal{P}_\alpha$, exercises $\mathcal{E}_\beta$, examples $\mathcal{X}_\gamma$, drills $\mathcal{D}_\delta$, and assessments $\mathcal{A}_\epsilon$, where some of them can be empty. Any $\mathcal{P}_\alpha, \mathcal{E}_\beta, \mathcal{X}_\gamma, \mathcal{D}_\delta, \mathcal{A}_\epsilon$ is a LO designed by expert instructors. These LOs are WEB documents (HTML, PDF, Flash, Java-applet, WEB-form, etc.).

Each $\mathcal{T}_j$ is defined by the course builder in XML and saved in the database. The topic's definition contains all the metadata needed to describe the topic content, drill, and assessment policies. These documents can be used to provide learners with dynamic HTML pages for each invocation of a topic depending on its LO's set.

Topics are defined dynamically by diagnostic components of the system. The drills' design contains inference rules analyzing students' answers to carefully prepared tests. This provides students with an environment supporting an ITS. Students' test scores and wrong responses are used as input parameters to a diagnostic component in the $\mathcal{D}_\delta$, which provides recommendations to alternative personal topic trails containing relevant LOs in order to solve the current topic's drill.

Assessments $\mathcal{A}_\epsilon$ are used for both formative and summative evaluations. Exercises $\mathcal{E}_\beta$ give a list of unsolved problems for students to practice on, while examples $\mathcal{X}_\gamma$ can be a list of solved problems, Flash, or Java-applets for students to work with. Since each LO is self-contained, they can be developed independently at different organizations. Expert instructors at each organization maintain a repository of LOs containing $\mathcal{P}_\alpha, \mathcal{E}_\beta, \mathcal{X}_\gamma, \mathcal{D}_\delta$, and $\mathcal{A}_\epsilon$. For students at different collaborative organizations, they will have the advantage that all LOs are accessible to them. For organizations owning the LOs, an audit trail for each LO can easily be gathered for billing purposes.

3.1 System Model

RPC based agents provide system integration and specialized services (Fig. 2). They communicate with each other by a request-respond mechanism in which remote procedure calls are done among different agents providing users with a dynamic and personalized learning environment.

XML subject and topic definition documents retrieved from the database are converted into an internal datastructure by an agent before the final HTML documents are presented to students. By using the user's response data and the user's current status, the system invokes appropriate agents to provide each learner with a corresponding HTML page.

Agents providing the system with specialized functions are developed and maintained separately. Different communication protocols such as SOAP and XML-RPC can be used. Agents can be implemented and can run on a multi-host and multi-OS depending on implementation architecture.

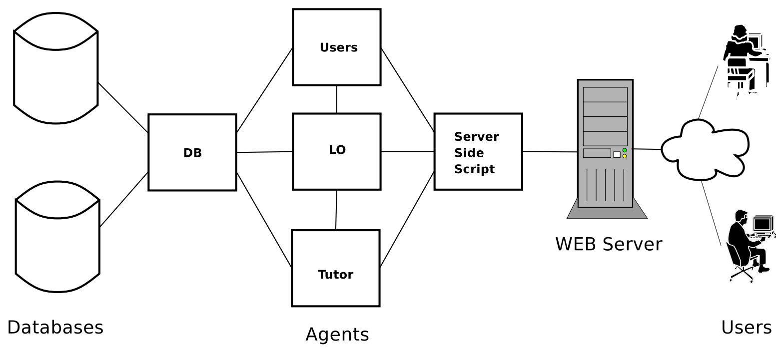


Fig. 2. Agents

3.2 System Architecture

The system uses an Apache WEB server and a PostgreSQL relational DBMS that provide the back-end database. The Python script programming language is used for all the agents. These three system software components are high quality, well proven, reliable, and free open source software.

XML-RPC implemented in Python provides the necessary tools for implementing multi-platform agents. Python encourages modular software engineering methodology. Many core modules are freely available and continually maintained by Python's Web communities.

Functions provided by Python's xmlrpclib are used to export agents' data structures to XML and to import XML back to agent's data structures. These agents' data states are then saved in the database in XML and are used as extended memory for running agents in the system for a particular user at a specific time. Such saved data states are subject and topic definitions, user's profile and status, user's test scores, and topic trails.

Server-side scripts (SS) and these agents are core components in IWBTS providing the following functionalities: create dynamic HTML pages using template files, thus producing a consistent page style and a structured navigational menu for the users; process users' GET/POST variables from Web forms and manage Web cookies; call appropriate agents in responce to users' interaction with the system.

A User agent contains user and instructor modules. A user module contains, among other things, user registration, user administration, user authentication, and authorization. To use the system a user needs to register. A user needs to fulfill certain criteria to be able to subscribe to a subject. Users are authenticated by consulting a users' domain Directory server via Lightweight Directory Access Protocol (LDAP) on logon. Users are authorized to access a subject if subscribed. Instructor modules provide an interface for the instructors for defining topics and subjects, students' status reports, and messaging.

An LO agent is responsible for dispatching LOs, creating dynamic menu structure, and providing a user's current profile and status. It is a core system component for providing a personalized learning environment for the users. It keeps track of users' navigational activities, guiding users through the course in responce to their previous actions, responses, and drill scores. SS calls the LO agent, and it in turn calls a DB agent. User profile and/or status data provided by the DB agent gives enough information for the LO agent to render dynamic HTML for the user via SS. This agent also collects users' audit trails.

A DB agent provides database connectivity for the other agents. This agent receives Structured Query Language (SQL) statements from other agents as call parameters. If the SQL statements are 'select' statements then the DB agent will return the calls with data arrays read from the database. This agent provides database service abstraction. Other agents need not consider the lower database client functions to do SQL queries. The DB agent acts as a proxy database client on behalf of others agents.

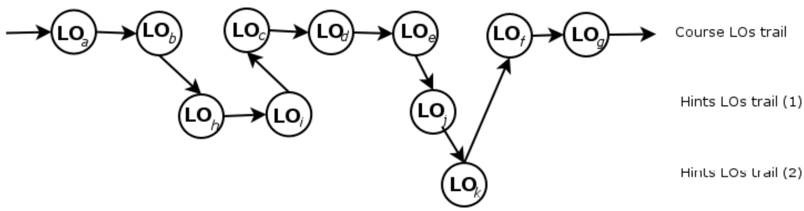


Fig. 3. Trails

A tutor agent responds to users' answers from tests and drills. A pedagogically crafted scheme with a set of questions and answers provides the agent with a programmed intelligence, in which wrong answers to a drill lead learners to appropriate hints and advice. The students can then subscribe to those LOs suggested by the hints or try the drill again and continue with the current course. Students can jump back to the current course trail at any time while following trails suggested by the tutor agent. The agent calculates scores, shows result status, and keeps track of assessments taken by each student. After each assessment the agent gives recommendations on how to proceed with the course. The tutor agent's instructor modules provide the instructors with an interface for students' progress reports and diagnostics.

4 Conclusion

A Web-based ITS should be able to respond to students' needs as they progress through the system. IWBTS provides a structured learning system environment, suitable for producing automatic links through simple logic-based reasoning mechanisms and guiding students in their navigation. Such an approach helps to overcome a major concern about a cognitive overload in Web enhanced courses. A mutually respectful environment and probing questions are essential elements of effective tutorial dialogs provided by IWBTS.

An instructor is presented with a complete report about every student's activities by IWBTS. The multi-agent architecture tracks student's responses, performs a qualitative reasoning on them, and delivers immediate diagnostics and advice. A qualitative reasoning engine in IWBTS evaluates each student separately, compares students' results, and then creates their learning profiles.

IWBTS is composed of reusable LOs. A large scale collaboration among educational organizations can provide high quality courses to their students by reusing LOs. For organizations owning LOs, an audit trail for each LO can easily be easily for billing purposes.

References

1. Dodd, B.G., DeAyala, R.J., Koch, W.R.: Computerized adaptive testing with polytomous items. Applied Psychological Measurement **19**(1) (1995) 5–22
2. Embretson, S., Reise, S., Reise, S.P.: Item Response Theory for Psychologists. Lawrence Earlboum associates, Inc., NJ (2000)
3. Encheva, S., Tumin, S.: Collaborative Role Management for Sharing Protected Web Resources. Lecture Notes in Computer Science, Vol. 3190. Springer-Verlag, Berlin Heidelberg New Jork (2004) 221–229
4. Friedland G., Knipping L., Rojas R. and Tapia E.: Web Based Education as a Result of AI Supported Classroom Teaching. Lecture Notes in Artificial Intelligence, Vol. 3190. Springer-Verlag, Berlin Heidelberg New Jork (2003) 290–297
5. Guin-Duclosson, N., Jean-Danbias, S., Norgy, S.: The AMBRE ILR: How to use Case-Based Reasoning to teach methods. In Cerri, S.A., Gouarderes, G., Paraguacu, F. (eds.). Lecture Notes in Computer Science, Vol. 2363. Springer-Verlag, Berlin Heidelberg New Jork (2002) 782-791
6. Guzmàn, E., Conejo, R.: A model for student knowledge diagnosis through adaptive testing. Lecture Notes in Computer Science, Vol. 3220. Springer-Verlag, Berlin Heidelberg New Jork (2004) 12–21
7. Hatzilygeroudis, I., Prentzas, J.: Knowledge representation requirements for intelligent tutoring system. Lecture Notes in Computer Science, Vol. 3220. Springer-Verlag, Berlin Heidelberg New Jork (2004) 87–97
8. Hatzilygeroudis, I., Prentzas, J.: Using a hybrid rule-based approach in developing an intelligent tutoring system with knowledge acquisition and update capabilities. Journal of Expert Systems with Applications **26** (2004) 477–492
9. Johnson, L.W., Rizzo, P.; Politeness in tutoring dialogs: "Run the factory, that's what I do". Lecture Notes in Computer Science, Vol. 3220. Springer-Verlag, Berlin Heidelberg New Jork (2004) 67–76
10. Koper, E.J.R.: Combining re-usable learning resources and services to pedagogical purposeful units of learning. In A. Littlejohn (Ed.), Reusing Online Resources: A Sustainable Approach to eLearning London Kogan Page (2003) 46–59
11. Legaspi, R., Sison, R. and Numao, M.: A Category-Based Self-Improving Module. Lecture Notes in Computer Science, Vol. 3220. Springer-Verlag, Berlin Heidelberg New Jork (2004) 554–563
12. Lepper, M.R., Woolverton, M., Mumme, D., Gurther, G.: Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S.P. Lajoie, S.J. Derry (Eds.): Computers as cognitive tools. LEA, Hillsdale, NJ (1993) 75–105

13. Magoulas, G.D., Papanikolaou, K.A., Grigoriadou, M.: Neuro-fuzzy synergism for planning the content in a Web-based course. Informatica **25** (2001) 39–48
14. Pecheanu E., Segal C. and Stefanescu D.: Content modeling in Intelligent Instructional Environment. Lecture Notes in Artificial Intelligence, Vol. 2774. Springer-Verlag, Berlin Heidelberg New Jork (2003) 1229–1234
15. Self, J.: Student models: What use are they?. In P. Ercoli., and R. Lewis, Artificial intelligence tools in education. Amsterdam: North Holland. (1988) 73–86
16. Vanlehn, K., Zhendong, N.: Bayesian student modeling, user interfaces and feedback: a sensitivity analysis. International journal of AI in Education **12** (2001) 155–184
17. Vercoustre, A. and McLean, A.: Reusing educational material for teaching and learning: Current approaches and directions. International Journal on E-learning, a special issue on Technologies for Electronic Documents. **4**(1) (2005) 57–68
18. Wiley, D. (Ed.): The Instructional Use of Learning Objects. Agency for Instructional Technology and the Association for Educational Communications and Technology (2002)

Cooperative Integrated Web-Based Negotiation and Decision Support System for Real Estate

Arturas Kaklauskas, Edmundas Zavadskas, and Algirdas Andruskevicius

Vilnius Gediminas Technical University
Sauletekio Ave. 11, LT-10223 Vilnius, Lithuania
property@st.vtu.lt

Abstract. Construction and the real estate sectors have low efficiency and high fragmentation compared to other branches of the building industry. Scientists, practitioners and politicians from various countries input much attention, effort and time to eliminate drawbacks such as efficiency and fragmentation. Currently construction and real estate are characterized by the intensive creation and use of information, knowledge and automation (software, knowledge and decision support systems, neural networks, etc.) applications. It is commonly agreed that a better integration of information, knowledge and automation applications would significantly speed up construction and real estate sector processes, improve the value of decisions made and decrease the overall cost of a building's life cycle. The authors' objective in integrating decision support with knowledge systems is to improve the quality and efficiency of decision support systems. In order to demonstrate the integration decision support and knowledge systems a Cooperative Integrated Web-based Negotiation Decision Support System for Real Estate will be considered as an example.

1 Introduction

Construction and the real estate sectors were always one of the most important branches in Europe. However, Information Technologies (IT) is less used in construction and real estate when compared to other industries, and IT is rarely integrated. A number of EU countries consider IT applications in construction and the real estate sector as one of the paramount tasks to be completed for successful results. For example, programs Framework 5 and 6 prefer to fund projects related to IT application in construction and real estate rather than to traditional construction and real estate projects.

The authors of this paper participated in the project Framework 6 "Intelligent Cities" (INTELCITIES). The main objective of INTELCITIES is to create a new and innovative set of interoperable e-government services to meet the needs of both citizens and businesses. One of the INTELCITIES's goals (on the Lithuanian side) was to develop a Cooperative Integrated Web-based Negotiation Decision Support System for Real Estate (NDSS-RE) that can use best practice, explicit and tacit knowledge. Real estate brokers from Vilnius city tested the system and used it to for different broker activities.

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 235 – 242, 2005.

Under the traditional system, the real estate agent offers a package of services: showing homes, advising sellers on how to make the house more marketable, assessing current market conditions, providing information about home values and neighborhoods, matching buyers and sellers, negotiating the sale price, signing contracts, arranging for inspections, and assisting with closings, and so on.

The Internet and intelligent technologies can disaggregate the above services: the Internet searches for real estate, finds alternatives and prepares comparative tables, databases that provide information about real estate, their values and neighborhoods, match buyers and sellers, negotiate the sale price, assist with real estate selection, and lender selection, provides smart software for boilerplate contract's language, and personalized websites that manage complicated transactions.

In a rapidly changing real estate market conditions and the large demand and supply for real estate, it is difficult to adequately orient all the above in the existing situation and to make rational decisions without the help of intelligent systems, knowledge and data bases.

This paper is structured as follows. Following this introduction, Sect. 2 describes the possibilities of the integration decision support and knowledge systems. In Sect. 3 we have provided a comparative description of Cooperative Integrated Web-based Negotiation Decision Support System for Real Estate. Testing the developed System is presented in Sect. 4. Finally, some concluding remarks are provided in Sect. 5.

2 Integrating Decision Support and Knowledge Systems

Integration of neural networks, multimedia, knowledge-based, decision support and other systems in construction and the real estate sectors has a very promising future in scientific research. Recently, much effort has been made in order to apply the best elements of multimedia, neural networks, and knowledge-based and other systems to decision support systems. For example, the use of artificial intelligence in DSS systems does not influence the persons who make decisions; however, it allows optimization of a DSS's possibilities.

Scientists and practicing users call the integrated knowledge-based and decision support systems - intelligent DSS, knowledge-based management support systems, expert DSS, expert support systems and knowledge-based DSS. Various forms of the integration of these systems were investigated and several architectures of systems were offered. Some authors in their work have suggested the idea to integrate knowledge-based and decision support systems. All these systems have a typical and common feature, i.e. in integrated systems the knowledge-based part performs auxiliary and an advisory role and chooses decision alternatives, data and their sources, problem solution tools, methods and models, and organizes flexible interaction between the user and the system. Knowledge-based and decision support systems are related, but they treat decisions differently. For example, knowledge systems are based on previously obtained knowledge and rules of problem solving, and a decision support system leaves quite a lot of space for a user's intuition, experience, and outlook. Knowledge systems form a decision trajectory themselves, while decision support systems perform a passive auxiliary role, though a situation might occur when decision support systems suggest further actions to the decision maker.

Calculation and analytical DSS models can be applied to process the information and knowledge that is stored in the knowledge base. For example, some DSS models can be applied to prepare recommendations by referring to the knowledge in the knowledge base. Decision support systems can also facilitate the search, and an analysis and distribution of the explicit knowledge.

In order to demonstrate the integration decision support and knowledge systems in the construction and real estate sector a Cooperative Integrated Web-based Negotiation Decision Support System for Real Estate will be considered as an example.

3 Cooperative Integrated Web-Based Negotiation Decision Support System for Real Estate

NDSS-RE is a Cooperative Integrated Web-based Negotiation Decision Support System for Real Estate and can be found at the following web address: http://dss.vtu.lt/realestate/. The NDSS-RE consists of a Decision Support Subsystem (DSS-RE) and Expert Subsystem (ES-RE). DSS-RE consists of a database, a database management system, model-base, a model-base management system and a user interface (see Figure 1).

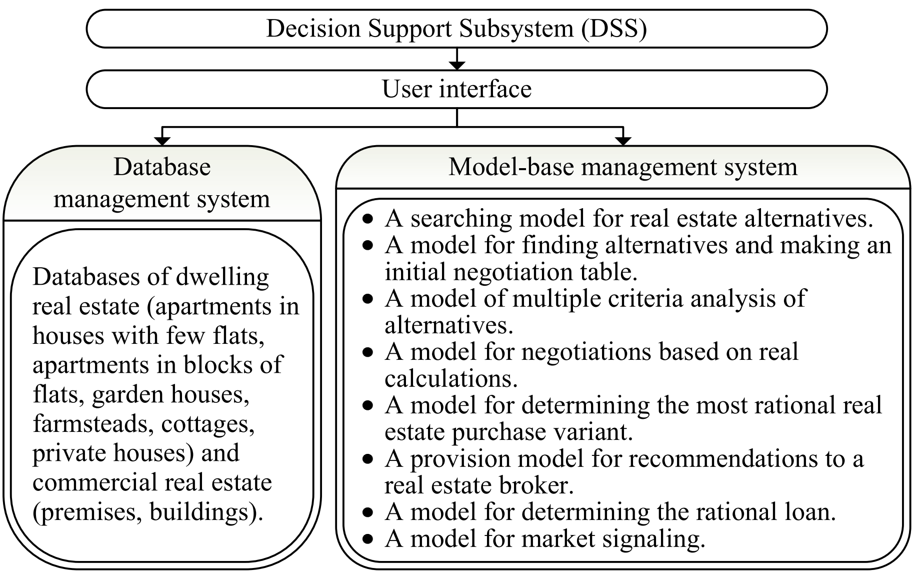


Fig. 1. The components of Decision Support Subsystem

Real estate listings are an interface for a seller to post listings. The system provides forms for sellers or real estate brokers to fill in information about their real estate. Real estate brokers wishing to present information on their objects must receive permission from DSS-RE administrator. Having this permission the broker inserts all the necessary information about real estate objects under sale in the DSS-RE databases according to the system's requirements (i.e. system of criteria, values and weights of criteria). Access to the databases developed personally by brokers is provided only to the broker and to the DSS-RE administrator. At present the developed DSS-RE allows for the performance of the following five main functions: search for real estate alternatives; finding out alternatives and making an initial negotiation table; analysis of alternatives; negotiations; determination of the most rational real estate purchase variant. In order to throw more light on the DSS-RE, a more detailed description of some of the above-mentioned Subsystem functions follows.

A consumer may perform a search for real estate alternatives from databases from different brokers. This is possible because the forms of data submissions are standardized at a specific level. Such standardization creates conditions that can be applied to special intelligent agents that are performing a search for the required real estate in various databases, and the gathering information/knowledge.

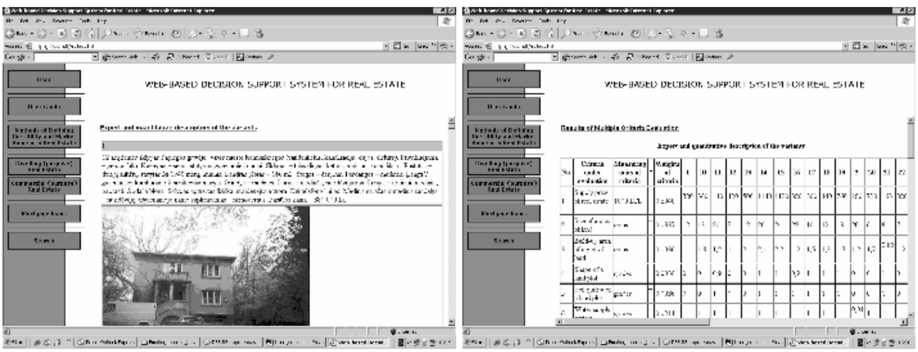


Fig. 2. Search results for a specific real estate are submitted in: textual, photo/video and graphical information on the real estate (*left*); expert and quantitative description of the real estate's alternatives (initial negotiation table) (*right*)

Consumers specify requirements and constraints and the system queries the information of a specific real estate from a number of online brokers. The system performs the tedious, time-consuming, and repetitive tasks of searching databases, retrieving and filtering information/knowledge, and delivering it back to the user. Search results for a specific real estate are submitted in a textual, photo/video and graphical information on the real estate's alternatives and the initial negotiation table (see Fig. 2), which may include direct links to a Web page of brokers. When submitting such a display, the multiple criteria comparisons can become more effectively supported. By clicking the link "Expert and quantitative description of variants" (see Fig. 2 (*left*)), the expert and quantitative description of private houses' alternatives is presented (see Fig. 2 (*right*)). Each alternative described by the quantitative information (system of criteria, weights of criteria and values) has a number (see Fig. 2) that coincides with the verbal and photographic information describing the mentioned alternative (see Fig. 2).

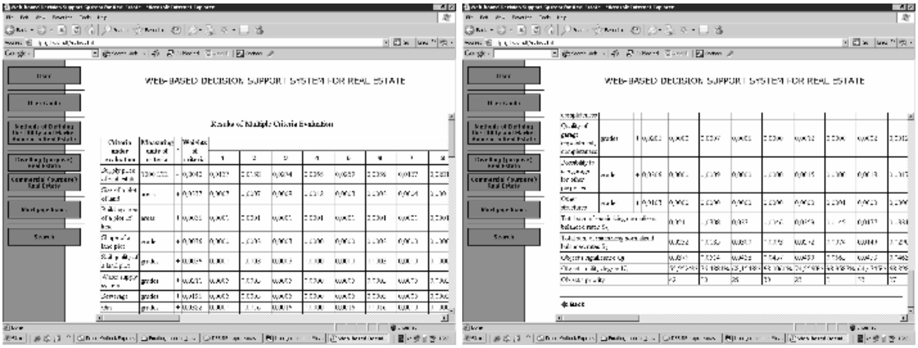


Fig. 3. Results of multiple criteria evaluation of the private house's alternatives: upper part of the matrix for obtained results (left); lower part of the matrix for obtained results (*right*)

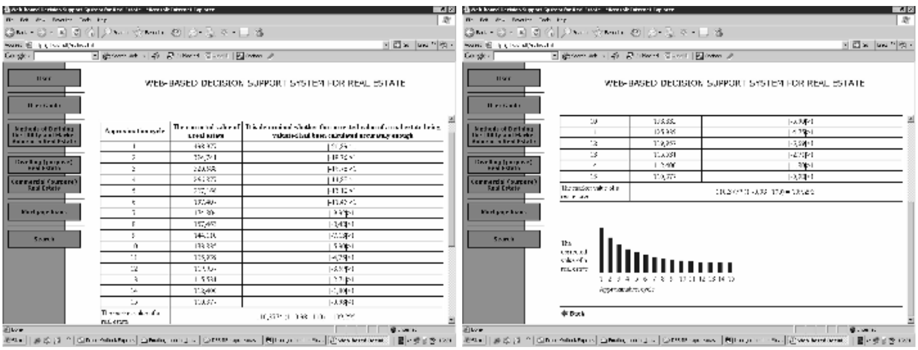


Fig. 4. Calculation of the market value: presentation of the market value's calculations in numerical form (*left*); presentation of the market value's calculations in graphic form (*right*)

While going through the purchasing decision process a customer should examine a large number of alternatives, each of which is surrounded by a considerable amount of information/knowledge (economic, quality-architectural, aesthetic, comfort, infrastructure, technical, legal, technological), and other factors. Following on from the gathered information and knowledge, the multiple criteria analysis is then carried out. By using multiple criteria methods [2, 8] as was developed by the authors, the buyer (broker) determines the initial priority, utility degree and market value of the analyzed real estate's alternatives During this analysis. Clicking the link "Results of Multiple Criteria Evaluation" (see Fig. 2 (*left*)), the results of the multiple criteria evaluation of the private house's alternatives are thus demonstrated (see Fig. 3). In the lower part of the obtained result's matrix the calculated significance of the real estate's alternatives, their priority and utility degree are presented (see Fig. 3). The upper part of the obtained result's matrix shows the numbering of the real estate's alternatives (see Fig. 3 (*left*)). By clicking these blue underlined numbers it is possible to calculate the market value of a certain alternative (see Fig. 4). The table presented in Fig. 4 (*left*) shows the iterations made during the calculation of the real estate's market value. The same information, only in graphical form is presented in Fig. 4 (*right*). And by moving a mouse above any column of the graphical part, the numerical value of the column can be seen. For example, the market value of the eighth alternative was calculated by making 15 iterations (see Fig. 4 (*left*)).

A buyer performing a multi-criteria analysis of all real estate alternatives selects the objects for starting the negotiations. For that purpose he/she marks (ticks a box with a mouse) the desirable negotiation objects (see Fig. 5 (*left*)). A negotiations e-mail are created by the Letter Writing Subsystem and sent to all real estate sellers after the selection of the desired objects is made and then *Send* is clicked.

During negotiations the buyer and the seller with the help of DSS-RE may perform real calculations (the utility degree, market value and purchase priorities) of the real estate. These calculations are performed on the basis of characteristics describing the real estate's alternatives obtained during negotiations (explicit and tacit criteria system, criteria values and weights). According to the results received, the final comparative table is then developed. Following on from the developed final comparative table the multiple criteria analysis and selection of the best real estate buying version is carried out by using DSS-RE.

After a variant of the real estate is selected, most often a purchaser has to borrow part of money from a bank. The Loan Analysis Subsystem is created for this purpose (see Fig. 5 (*right*)).

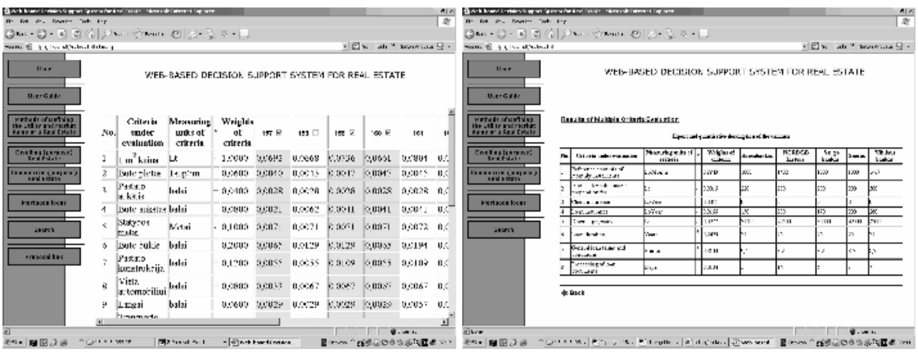


Fig. 5. Selection of real estate objects for automated negotiations (*left*) and analysis of the loan alternatives offered by certain banks (*right*)

There are two main categories of rules and procedures in the Expert Subsystem:

- Development of suggestions as to what brokers to use and for what reasons further negotiation should be carried out. With the help of the DSS-RE having determined the sequence of priority, the degree of utility and the market value of the real estate, the rules of the expert's subsystem suggests what brokers to use and for what reasons further negotiation should be carried out.
- Composition of comprehensively reasoned negotiation e-mail for each of the selected brokers. By using information inherited from the previous DSS-RE calcula-

tions and predefined rules and procedures, the expert's subsystem composes of negotiation e-mail for each of the selected brokers, where it reasonably suggests that the price of the real estate should be decreased. The e-mail includes references to the calculations performed by DSS-RE.

4 Testing the System

In order to test the usefulness of the system, final semester master degree students from the Real Estate Management and Evaluation program at VGTU collected more than 250 listings. These students work as brokers in various real estate companies in Vilnius. They have placed information about real estate objects that they were selling at the time into the database. This system has been tested by eighteen students for areas that could be improved, e.g. process, interface, navigation, search for alternatives from different brokers' databases, multiple criteria evaluations, calculation of market value and negotiation. A testing of NDSS-RE was also performed by a designed questionnaire that included four organizations from real estate brokers in Vilnius. The letter was attached to the questionnaire was as follows "We would like you to draw on your experience and expertise to help us to test whether the NDSS-RE can also meet *your* needs as a user. Please read through the following questions circling your response". A more complete study is underway to study the satisfaction of users and the current real estate agents do in order to survive. If not, then what are the issues that prevent people from using such approach.

A first functional NDSS-RE prototype was developed in 2003, using existing information in order to start testing the main characteristics of the System together with prospective user groups in order to determine. Real estate brokers performed tests to ensure the desired results are achieved. When additional or improved modules were ready, they were included/exchanged in the NDSS-RE.

We have some lessons to learn about the social impacts of implementing such kinds of e-broker systems. Several experts from Vilnius broker companies estimated that about forty percent of employees in real estate broker firms may lose their jobs.

The most important issues are as follows, and we will further explore and collect the needed knowledge to develop the following NDSS-RE subsystems:

- Provision of recommendations to a real estate broker. A model of consumer needs analysis is still under creation for NDSS-RE. This sub-system will accumulate information about the popularity of real estate alternatives that are placed into the database. The popularity is determined on the basis of the number of consumers analyzing a certain object and on the basis of the time-spent watching. A seller will offer to reduce or increase the value of the real estate being sold on the basis of such information; other pieces of advice will also be provided.
- Market Signaling Subsystem. Links to sites with exhaustive information and analytic reviews about the situation in the Lithuanian real estate market and about its development perspectives will be provided. Also, in future this Subsystem can send selected messages with required information to subscribers.

5 Conclusions

Integration of knowledge-based and decision support systems have a very promising future in scientific research. The authors of this paper participated in the project Framework 6 “Intelligent Cities“ (INTELCITIES). One of INTELCITIES’s goals (on the Lithuanian side) was to develop a Cooperative Integrated Web-based Negotiation Decision Support System for Real Estate (NDSS-RE) which consists of a Decision Support Subsystem and an Expert’s Subsystem and can create value in the following important ways: search for real estate alternatives, find out alternatives and make an initial negotiation table, complete a multiple criteria analysis of alternatives, make negotiations based on real calculations, determine the most rational real estate purchase variant, and complete an analysis of the loan alternatives offered by certain banks.

References

1. Emerging Construction Technologies. Hybrid Computerized Decision Support System for Infrastructure Assessment. URL: www.new-technologies.org/ECT/Other/imageprocess.htm (current as of March 7, 2005)
2. Kaklauskas, A., Zavadskas, E. K., Raslanas, S.: Multivariant Design and Multiple Criteria Analysis of Building Refurbishments. Energy and Buildings, Vol. 37. No 4 (2005) 361–372
3. Kvederyte, N., Zavadskas, E. K., Kaklauskas, A.: Multiple Criteria Alternative Design of Building Life Cycle. Statyba (Journal of Civil Engineering and Management), Vol. 6, No 2 (2000) 128-142
4. Kvederyte, N., Zavadskas, E.K., Kaklauskas, A.: Multiple Criteria Analysis of Semidetached House Life Cycle. Statyba (Journal of Civil Engineering and Management), Vol. 6, No 3 (2000) 179-192
5. Mobile Integrated Communication in Construction (MICC). URL: ica.cordis.lu/search/index.cfm?fuseaction=proj.simpledocument&PJ_RCN=1428168&CFID=1724563&CFTOKEN=94987971 (current as of March 7, 2005)
6. Support Centres Network for Information Technology in Construction (SCENIC). URL: ica.cordis.lu/search/index.cfm?fuseaction=proj.simpledocument&PJ_RCN=1510268&CFID=1724563&CFTOKEN=94987971 (current as of March 7, 2005)
7. Technology and Processes for Integrated Construction Project Documentation Production and Management (CONDOR). URL: ica.cordis.lu/search/index.cfm?fuseaction=proj.simpledocument&PJ_RCN=2172179&CFID=1724563&CFTOKEN=94987971 (current as of March 7, 2005)
8. Zavadskas, E.K., Kaklauskas, A., Banaitis, A., Kvederytė N.: Housing Credit Access Model: the Case for Lithuania. European Journal of Operation Research, Vol. 155, No 2 (2004) 335-352

A Directed Evolution Modularity Framework for Design of Reconfigurable Machine Tools

Horacio Ahuett, Joaquin Aca, and Arturo Molina

Instituto Tecnológico y de Estudios Superiores de Monterrey,
Center for Innovation in Design and Technology ,
Av. Eugenio Garza Sada 2501 Sur, 64849, Monterrey, N.L. Mexico
{horacio.ahuett, aca, armolina}@itesm.mx
http://cidyt.mty.itesm.mx/web/

Abstract. Reconfigurable machine tools (RMT) have emerged as a potential solution to meet the demand for rapid adaptation in the next generation manufacturing systems. While a significant amount of research in RMT design is available, formal design methodologies are still under development. In previous work, a modularity framework technique for the development of RMT modules was outlined. The technique focused on module functionality and constructability. While in general these characteristics are critical for the survival of any product, specific characteristics of RMT were not easily addressed by the proposed methodology. In particular, convertibility, in the form of upgradeability and adaptability, was not addressed. This article presents an enhanced modularity framework that takes into consideration the intended evolution lines of the reconfigurable machine tool. The basic principles are outlined and applied to the design of a machine tool for metal working. Recommendations for future enhancements to the framework are made, in particular, the possibility of developing a convertibility index to facilitate evaluation of candidate designs is discussed.

1 Introduction

Reconfigurability is a key feature in the design and construction of machine tools of the next generation manufacturing systems. While the concept of reconfigurable machine tools has been researched for about a decade, integrated design methodologies for reconfigurable machine tools (RMT) are still being developed.

Experience shows that modularity is a key ingredient of RMT. Modular design presents some advantages that are well suited to machine tool development, for example: economies of scale, increased feasibility of product/components change, increased product variety, reduced lead time, decoupling of tasks, and ease of product upgrade, maintenance, repair and disposal [1].

A significant amount of work has been done for the development of methodologies for module design, particularly for consumer products. Some of these techniques are currently being applied to the development of RMT. As experience from test cases becomes available, a modularity framework for RMT design and construction has begun to emerge [2].

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 243 – 252, 2005.

In this paper, *upgradeability* and *adaptability* are identified as key drivers behind the modular construction of an RMT. An argument is made as to the importance of these parameters, in particular within the context of current trends in RMT development, such as the evolution of Reconfigurable Manufacturing Systems (RMS). A case is presented in which upgradeability and adaptability concepts are applied to support decisions in the design of a machine tool. The experience from this case shows that these parameters not only drive the design of the modules, but also the evolution of the machine tool. The need to develop an index that can be used to compare alternate designs is exposed. In addition, a PLM environment was established as means to share information among team members (designers, customers, suppliers and constructors), evaluate technical feasibility of the design and coordinate efforts of team members working in different modules of the machine.

This article is organized as follows: in Section 2, a brief literature review is presented. Research issues in modular design applied to RMT are made explicit from this review. The concepts of upgradeability and adaptability of RMT are introduced in Section 3. An industrial case that illustrates the use of these concepts to drive the design of a machine tool is then presented and the experiences are analyzed in Sections 4 and 5. The final section presents some conclusions and proposed future work.

2 Literature Review

Timely, affordable, one-of-a-kind products will require rapid product and manufacturing process designs. These demands have created a need for reconfigurable manufacturing systems [3]. While a significant amount of research has been conducted in this field, RMT design methodologies are still being developed [2]

As stated before, modularity facilitates RMT design and construction. Bi [1] presents a taxonomy for modularity that considers manufacturability. He points out that cellular manufacturing is based on flexible equipment that performs specific functions and establishes that optimal modular system configuration may need to be modeled as a multiple optimization problem. Moon [4] proposes a design methodology that integrates modules and Huan [5] presents a design methodology for teams that are dispersed throughout the world, using the World Wide Web as communication tool. In all of these cases, a methodology for the definition of the modules is not presented, that is, these studies assume that the modules are already defined.

Newcomb [6] proposes a methodology for establishing modules in terms of reciclability of materials and introduces two parameters with which modularity can be evaluated for any given design. They claim that changes proposed by their methodology match those changes that would have been proposed by intuition and experience. However, it is not clear how well the algorithm might work when more than one dimension is analyzed, for example, ease of service and reciclability.

Rogers [7] propose the use of standardized modular machines to develop reconfigurable, modular production systems. The concept of reconfigurable modular systems in which machine tools represent the modules within the system, is introduced in this work. It is interesting to note that, with the falling prices of machine tools, such systems are beginning to spread within important sectors, such as the automotive industry.

Marshall [8] presents the impact that modularity has on the characteristics of the industry that adopts it, and explains the need to address modularity at three levels: systems engineering, methodologies for product modularization, and the process of modular product development. Little details are given about any specific level, although they point out the close relationship that must exist between functions and modules.

Zatarain et. al [9] points out a tendency towards modularization of machine tool design and construction. In their work, they propose a methodology for modeling the dynamic response of a machine tool of modular construction. To reduce computation time and effort, pre-calculated stiffness and mass machine tool modules are assembled for finite element analysis. Their work introduces a formulation for modeling the interfaces between the structural modules of the machine tool. The design of the actual modules is not discussed in this paper.

As the review illustrates, most studies start from the fact that modules have already been designed. In very few instances are the algorithms for designing modules actually discussed. The authors' previous work [2] focused on the deployment of a modularity framework for the design of RMT. The previous study introduced a general design methodology that was applied to a specific RMT design case. Bottlenecks were identified in this process and as a result the authors proposed a modularity framework for RMT design. Four domains were established. Details were given about three of them but little was proposed about tasks at the front end, which constitutes the customer's (i.e. the machine tool builder) domain. This issue is now addressed.

3 Upgradeability and Adaptability: Definition and Relevance

Reconfigurable machine tools evolve in well known directions. Since its conception as a technology, reconfigurability has been characterized in terms of the modularity, integrability, customization, convertibility and diagnostibility embedded in the design of a particular machine tool [10]. From the perspective of the machine tool builder, an important issue is how these characteristics are accounted for in the conception and evolution of the RMT design.

In previous work [2], the authors proposed an approach to define modules for RMT design and construction. The framework, which has been represented schematically in Figure 1, provides a roadmap for the development of modular reconfigurable machines. Four domains are established leading from the requirements of the machine tool builder, through the definition of functions and finally to the definition of the reconfigurable modularity. In principle, each layer formalizes the knowledge developed at any given stage. As shown in the figure, steps two through four rely on well known techniques such as function analysis, plus a reconfigurability concept matrix, a tool introduced in the original work by Perez [2].

In Step one of the methodology, the intentions of the customer, i.e. the machine tool builder, are captured. In the original work, little information was known as to the nature of these intentions. It is now argued that upgradeability and adaptability are of utmost importance for a reconfigurable machine, and should be accounted for throughout the design process.

From the perspective of the RMT manufacturer, the design of the RMT must allow for the adaptation of new technology as its customer's demands evolve. Convertibility of the RMT becomes an issue of *upgradeability and adaptability*. These characteristics reflect the capacity of the machine tool to adapt in response to changes induced from a variety of conditions.

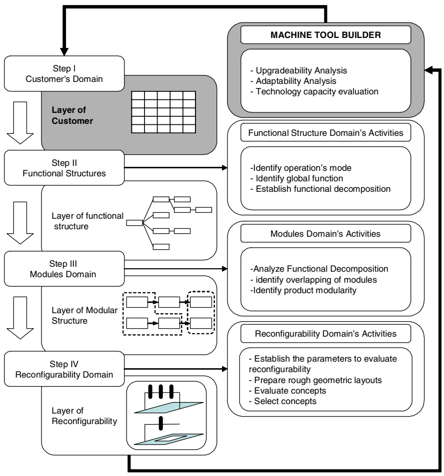


Fig. 1. Overview of the modular reconfigurabilty design methodology under the perspective of the machine tool builder, expanded from [2]. The shaded area represents the front end concerns that are intended to drive the evolution of the machine tool.

Upgradeability and adaptability are important drivers for the design of an RMT. The model proposed by Perez et al, can be enhanced as illustrated in Figure 1. In principle, upgradeability refers to the ability of the manufacturer to introduce improved technology into the machine. RMT should be designed with upgradeabilty in mind; the design of the RMT should be adjusted to allow for predefined evolutionary steps. Adaptability refers to the capacity of the RMT to suit the customer's desires. The customer's demands will likely shift in response to changes in the business environment: financial conditions, competitiveness issues, labor concerns, etc. In fact, customers select a particular level of technology based on all of these parameters. At the RMS level, modularity, customization and integrability are addressed by the use of flexible equipment from the same vendor. Convertibility and adaptability becomes a matter of adding or removing machines.

There are other reasons that justify the need to build upgradeability and adaptability into a machine design. For example, during machine tool development, an important decision involves the amount of automation that can be built into the machine in such a way that it can fulfill its mission without incurring in prohibitive costs. There is a limit as to how much the technology can increase performance. If the budget or the design mission does not allow for a current version of a particular machine to be automated to gain full advantage of current technology, modules should be designed in such a way that they can be replaced with improved technology. In the next section, a test case that illustrates the relevance of these ideas is presented.

4 Case Study

The ideas expressed in the previous section were put to test in an industrial case. Because of confidentiality issues, the description that follows focuses on experiences associated with the application of the methodology, rather than on the details of the equipment that was developed.

A company that manufactures high performance components for the power generation industry required a metal cutting machine to replace a manual process that caused excessive operator fatigue. The goal was therefore to design a machine that reduced operator fatigue without sacrificing productivity. A first generation machine was designed and built. The design was not modular, and could be updated only with significant modifications to the machine's structure. While this machine reduced operator fatigue, it had very limited capacity: product changeover required significant effort and the machine could accommodate only a limited variety of products. A second generation machine that could offer more flexibility was needed.

A PLM environment was established to share and manage information among team members. PLM is a technology stack that brings together all product-related data into a single place for use across a vastly larger community than the engineering and design labs. By using PLM, product stakeholders — both internal and external to the organization — can share, comment on, modify, and track a product throughout its lifecycle. PLM solutions enable manufacturers to explore a variety of business strategies — including virtual manufacturing and engineer-to-order — to establish efficiencies and maintain competitive product advantage [11]. A simple engineering environment was implemented based on functional tools and PDM tools. CATIA and DELMIA 3D were used for the design, simulation and integration of different modules of the machine. A Concurrent Version System (CVS) was also utilized to share information among engineers, suppliers and constructors. CATIA and DELMIA were used for kinematics evaluation of the machine and its individual modules, for interference and assembly analysis and to analyze the ergonomic process of the operator interaction with the machine. CVS was selected because it required only qualified supervision during installation. Because of the efficiency and autonomy of the software, it practically does not require maintenance.

At the beginning of the project, a decision was made to apply reconfigurabilty concepts for the design of the second generation machine. The framework / methodology was used to design the modules. From the start of the design process, the need to establish the technological limits of the machine was evident. A minimum limit was established to meet the technical goals of the project, within the allocated budget and delivery time. In addition, the modules were designed in such a way as to allow for technological enhancements to be easily adopted. The direction of the evolution of the machine was therefore set from the beginning of the design process (with a time horizon of a couple of years).

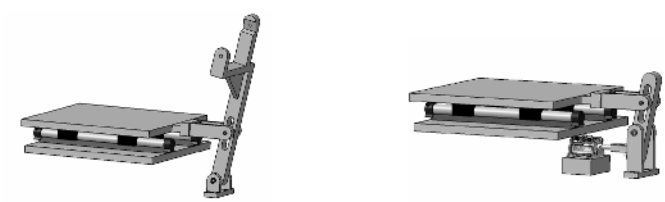
While the original modularity framework allowed for the definition of the modules, a set of principles had to be adopted to support decisions about the technology built into each module and to trace the evolution of the machine. This type of information was not accounted for in the original framework.

The design of the control module illustrates typical decisions that had to be made. In general, two alternatives are available for machine control: PLC or computer control. A PLC offers a relatively simple but rather limited solution. PLC's are very effective when all process parameters and control functions are well established. However, changes in process conditions are not very easily adopted. Set-up time is slow and capacity is limited. As a consequence, upgradeability and adaptability are hindered. This option was used in the first generation machine.

Computer control is more difficult and expensive to implement: algorithms to control process and machine variables, as well as interfaces to communicate the computer with the user and the electric system of the machine tool need to be developed. On the other hand, computer control allows for a greater number of products and variables to be processed. Upgradeability and adaptability of the machine tool to facilitate new technologies and adapt to changes in process conditions is much easier with a PC than with a PLC.

Similar issues had to be addressed from the perspective of the mechanical design of the machine. For example, the conceptual design of the machine required the deployment of at least four axes to perform its functions. The manual process was analyzed to determine which movements needed to be automated in such a way that operator fatigue could be reduced without the need for complex control equipment. The decision was to automate two axes, those that produced most of the stress on the operator. These axes were designed to be computer controlled in position and speed. The remaining degrees of freedom were controlled manually by the operator.

As stated before, the evolution of the machine was defined from the start of the project. Provisions were made to allow the manual control mechanisms to be replaced with minimal changes to the rest of the machine. While the actuation system was not completely predetermined before the prototype was designed, the manual actuation mechanism was built in a package that could be easily removed. A mechanical interface was designed in such a way that a replacement of the manual control mechanism with a commercial hydraulic or electric actuator could be designed and implemented. This concept is illustrated in Figure 2.



(a) (b)

Fig. 2. (a) Hand operated control lever; and (b) Automatic control lever. The modular design of the machine allows for easy replacement of the manual assembly.

5 Analysis and Discussion

The selected CVS fulfilled requirements during initial stages of the project (mechanical design). However, as more information from other disciplines (electric, electronic and control) were integrated to the project the capacity of the CVS was exceeded and the integration of the information was based on knowledge and skills of engineers more than on capacity of the software. For this reason it was necessary to migrate to a more powerful tool, in this case SMARTEAM. This option requires an IT manager to set-up and maintains the environment. One person was required to transfer the information from CVS to SMARTEAM. At the moment this paper was written all the information had been transferred to SMARTEAM and engineers has been trained in the use of this tool.

Table 1 compares different versions of the second generation machine as planned during its conception. Four different possible versions were analyzed. The table presents the modules as well as the elements that constitute each module. All versions have the same number of modules. The degree of technology in each module constitutes the difference between each version. To save space, columns were added to the table for each module that was part of the machine. For example, version 1 has one automatic and one manual axial linear motion system, while version 2 of the machine has 2 automatic linear axes

Table 1. Comparison of different versions of the second generation machine (M = manual)

MODULES	TYPICAL ELEMENTS	2nd GENERATION MACHINE VERSION							
		1		2		3		4	
Machine Frame	Structural Elements								
Linear Motion	Structure								
	Guides								
	Actuator		M						
	Ball screw								
	Driver								
	Encoder								
Rotation	Structure								
	Transmission								
	Actuator		M						
	Driver								
	Encoder								
Work piece Holding	Conventional								
	Quick Change								
	Tailstock								
Control	PLC								
	PC								
	Interface								
Monitoring	Temperature								
	Pressure								
TOTAL COST (USD)		$ 18,500.00		$ 23,800.00		$ 25,300.00		$ 29,100.00	
FATIGUE REDUCTION		70%		80%		85%		90%	
PRODUCTIVITY INCREASE		12%		22%		27%		32%	
VALUE ADDED		51%		61%		66%		71%	

The *Total* Cost reported at the bottom of the table is the actual cost of each version of the machine and takes into account the cost of parts and labor associated with the fabrication of the machine. *Value added* by the technology is calculated based on the two possible effects sought by the machine: fatigue reduction and productivity increase. These effects account for 67% and 33% of the parameter *value added*, respectively. The target of 100% *value added* means that cycle times are reduced by 50% and idle time due to rest periods are reduced to zero. Fatigue reduction is measured based on the actual hand operated process and the cycle time. As the number of manual operations that cause fatigue is reduced by the introduction of technology, the fatigue reduction factor is increased proportionally. Increase in productivity results from the reduction in the idle time. In manual operations the operator needs to rest periodically. As more automation is available, periods of rest are reduced. Automation can also reduce set-up time and cutting time.

Figure 3 summarizes the cost and the value added by each version of the machine. Major strides are made with the initial design of the machine. As more technology and cost is added, improvements become marginal because benefits obtained from the implementation of the technology are not proportional to the investment. An ideal design window can be established for the evolution of the machine. This information can be used to establish the first version of the machine, the nature of the most advanced version, and the stages that the machine design can go through as it evolves from one to the other. As a reference, the figure shows where the first generation machine would lie with respect to the RMT-second generation machine. It is represented as a point because this version cannot be upgraded or adapted.

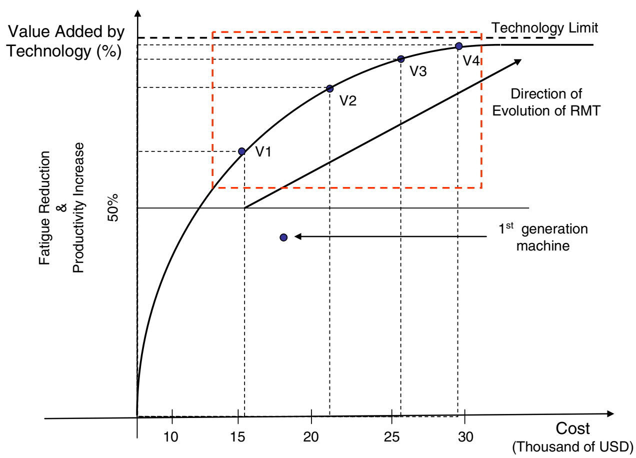


Fig. 3. Value added vs cost of the technology associated with the RMT. A design window that traces the evolution of the machine can be established from this chart.

It is important to note that although a formal methodology was not utilized to facilitate upgradeability or adaptability, the direction of the evolution of the machine was predetermined from the moment the design process began. Once the design of the machine had reached the stage in which modules were to be specified, the morphology of these modules was determined with upgradeability and adaptability in mind.

The case showed that the adoption of these criteria from the beginning established the direction in which the machine would evolve. However, the evaluation of prospective module designs was done based on the experience of the design team. The development of and upgradeability – adaptability index would be highly desirable to capture and facilitate the design process. This index could be used to compare and evaluate prospective module designs in terms of the ease with which the machine can be upgraded.

6 Conclusions

This work discussed the concept of directed evolution associated with the design of reconfigurable machine tools. It was stated that upgradeability and adaptability are dimensions that must be addressed at the top of the modularity framework that had been proposed in a previous study. A test case that used these concepts was presented. The case illustrated how the technological evolution of a machine tool can be traced as the user's demand change. Finally, the case study showed that there is a need to develop an upgradeability / adaptability index to allow machine tool builders to evaluate and make decisions about prospective module designs.

Acknowledgements

This research is part of the Research Chair in Mechatronics at ITESM Campus Monterrey. The authors wish to acknowledge the support of this grant in the preparation of the manuscript.

References

1. Bi, Z. M. and Zhang, W.J.: Modularity Technology in Manufacturing: Taxonomy and Issues. Int. J. Adv. Manuf. Technol. (2001) 18:381-390
2. Perez, R., Aca, J. Valverde, A., Ahuett, H., Molina, A., Riba, C.: A modularity framework for Concurrent Design of Reconfigurable Machine Tools, In: Y. Luo (eds.): Cooperative Design, Visualization and Engineering. Lecture Notes in Computer Science, Vol. 3190. Springer-Verlag, Berlin Heidelberg (2004) 87-95
3. Engineering Research Center for Reconfigurable Manufacturing Systems, National Science Foundation (ERCRMS), http:// ercrms.engin.unmich.edu/Research/ Research.html, 2002
4. Moon, Y.M. and Kota, S.: Design of Recofigurable Machine Tools, Transactions of the ASME: Journal of Manufacturing Science and Engineering, (2002), 124:480-483
5. Huan, C.C. and Liang, W.Y.: A formalism for Designing with Modules, Journal of the Chinese Institute of Industrial Engineers, Vol. 18, No. 3, (2001), 13-90
6. Newcomb, P.J., Brass, B., Rosen, D.W.: Implications of Modularity on Product Design for the Life Cycle, Transactions of the ASME: Journal of Mechanical Design, (1998), 120: 483-490.
7. Rogers, G.G. and Bottaci L.: Modular Production Systems: a new manufacturing paradigm, Intl. J. of Intelligent Manufacturing, Vol. 8, No. 2, pp. 147-156, Apr. 1997.

8. Marshall, R. and Leaney, P. G.: A system engineering approach to product modularity, Proceedings of the Institution of Mechanical Engineers, 213, Part B, 847-851.
9. Zatarain, M., Lejardi, F., Egana, F.: Modular Synthesis of Machine Tools. Annals of the CIRP. Vol. 47, 1, 1998.
10. Koren, Y. and Kota, S., Reconfigurable Machine Tools, U.S. Patent 5,943,750 (1999)
11. Aberdeen Group, PLM — It's for Any Manufacturer Striving for Product Excellence, Executive White Paper, July 2002, Boston, Massachusetts

Study of the Intelligent Turning Machining Method Based on Geometrical Feature Recognition

Guofeng Qin[1,2] and Qiyan Li[1,2]

[1] The CAD Research Center, Tongji University, Shanghai 20092, P.R. China
gfqing@yahoo.com.cn
qylcad@sina.com
[2] The Chinese National CAD Application Technology,
Training Center For Engineering Design
Tongji University, ShangHai 20092, P.R. China

Abstract. Feature recognition is a key technology of the automatic intelligent turning machining method. This paper provides an intelligent turning machining method based on geometrical feature recognition, including putting up viewpoint, direction and cylindrical co-ordinate. For real-loops and unreal-loops in projection, an algorithm is taken forward to recognize the straight faces, turning circle (cone, frustum and cylinder) and slot, and creates a relationship functions among all the loops, determine the machinable turning cutting regions respectively according to the relationship functions. The study resolves the machinable problem in design stage, a method is supported to cooperative design and decreases the cost and save time. a hierarchy architecture of the center controller is constructed including integration process planning.

1 Introduction

This paper covers an intelligent manufacture method based on geometrical feature recognition, which will run in the EdgeCAM automatic manufacture system.

Feature-based CAD/CAM can runs process planning automatically. It reduces the interactions between designers and fabricators, and decreases the complexity of operation decomposition and process planning time, and minimizes the number of design/fabricator interactions. Recently, there are much study concerning algorithms for defining optimal tool paths for various tools, part geometries, and machines. In CAD/CAM systems, major challenges still are process planning and operation decomposition.

Three kinds of geometric models, namely, wire-frame models, surface models, and solid models support most CAD/CAM system. There are three kinds of express methods for 3-D geometries, namely the BRep method, the CSG method, the compound model of both.

In this paper, based on EdgeCAM, strategies and algorithms are taken forward according to the features of a geometric model. Turning features, accessible from a predefined direction, are gained by manual or automatic feature extraction.

Y. Luo (Ed.): CDVE 2005, LNCS 3675, pp. 253 – 262, 2005.

2 The Center Control Hierarchy Architecture

In EdgeCAM system, The adaptive intelligent process control strategy and algorithms will be used. The center control is made up of the graphics generator, the center control cell, and the NC code controller[6-8]. The graphics generator is used for creating and displaying the graphics, which can import the graphic file by import interface in Fig.1, of course the operator can design geometric graphics within the EdgeCAM system by himself. The center control cell deals with the 3-D Geometric model for feature recognition and process planning decomposition. The NC code controller copes with the NC code generation and file management; NC Wizard is a post dispose generator of files, The end of NC Wizard generates a post dispose model file, then a NC code file is produced by the NC code constructor.

3 Feature Recognition

In the intelligent automatic machining process, feature recognition is very important for process planning and operation decomposition. Optimal strategy and algorithm can decrease the process route of machining, reduce the cost, and shorten the time. In short, feature recognition is used to research the areas of machining, and to avoid collision and interference.

Visibility is very important in feature recognition [9,10]: the method of visibility will be used to determine from a viewpoint in the visible space which is accessible to an observer located outside the convex envelope of the object. The set of visible directions for a point is referred to as the visible area. We compute the visible area for a representative set of sample points distributed uniformly in the entire visibility space. Visibility gives us a "first perception" about how to orient the cutting tool during cutting. Of course, visibility cannot account for the diameter of the tool, the tool-holder or the spindle, this job will be done by the machining NC system. In the following, there will describe turning and milling feature recognition, respectively, and provide corresponding algorithms and cases.

Turning feature recognition will be used to search the area of turning operation to avoid collision and interference for the turning tool. Collision avoidance can be viewed alternatively as access assurance. An interference free tool can be assured by an access direction[11,12].

3.1 Creating Visibility Space for Turning Feature Recognition

The access direction may be not unique, but the method is unique. For turning feature recognition, firstly, we should create a viewpoint for visibility space, as shown in Fig.1. The viewpoint is aligned with the center line outside the geometry model.

After the viewpoint created, a direction will be given for view in visibility space, and the direction shall be parallel with the center line. The direction is shown in Fig.1.

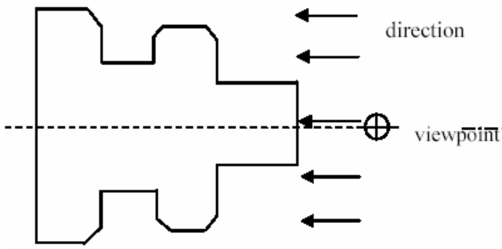


Fig. 1. Viewpoint for visibility space

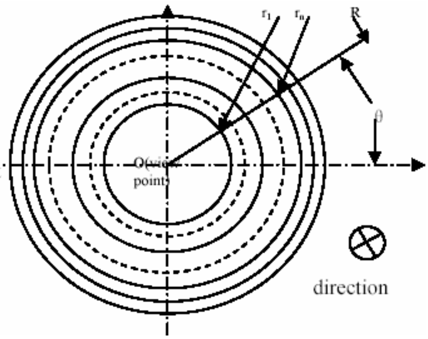


Fig. 2. Cylindrical co-ordinate and projection

3.2 Background and Projection for View Direction in Visibility Space

The background and projection for view direction is used to find the machinable area and size. The turning cutting visibility space should be continuous. If the unit vectors that represent directions in R3 are viewed as position vectors, the surface of a cylinder is defined C2. Clusters of visibility or accessibility directions become continuous patches on the surface of the cylinder and referred to as visibility areas and accessibility areas. Access can also be generalized for regions. A access area for a region R can be defined formally in terms of point-access areas as follows:

$$A(R) = \bigcap_{p \in R} A(p) \tag{1}$$

Where: A is access area;
P is point.

After defined them, create a cylindrical co-ordinate shown in Fig.2. The vector of a point P is positioned by Z Axes, radius ‘r’, and angle ‘θ’, namely, (r, θ, Z). The original point is positioned at the viewpoint. Along the direction from the view point, the background and projection shown in fig.2. can be attained .

$$P=\Phi(r, \theta, Z) \tag{2}$$

3.3 Algorithm of Turning Feature Recognition

The visibility space, including the viewpoint, the view direction, and the cylindrical co-ordinate have been defined, then the areas of the machinable turning will be found. At first, obtain the largest outline of every real-loop li, the radius of the outline ri, and then the length (ΔZ) or the relationship function Φ(r,Z)=f(r,Z) between r and Z of every real-loop. If Δr is zero, then the extend of the real-loop is a cylindrical solid; if ΔZ is zero and Δr is not zero, then the patch is a face between two real-loops; otherwise, it is a turn surface, for example, a cone or a frustum of a cone. The regions between the outside the cylindrical solid, cone or frustum and inside the value of the radius R of visibility space should be cut, which is shown in fig.3.

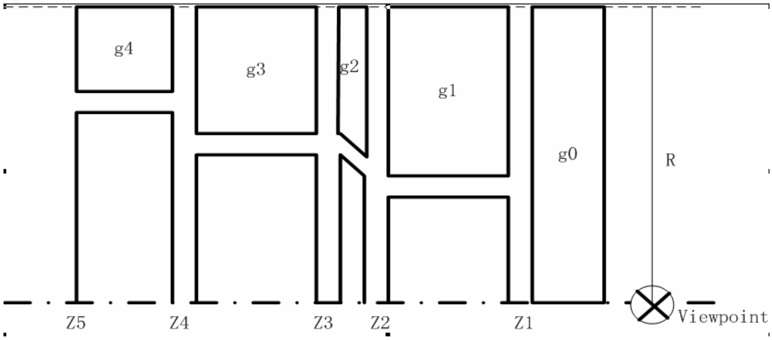


Fig. 3. Regions of turning cutting

The overall area outside profile of real loops (T) in Fig.4 is clustered by the every machinable turning region gi, as follows:

$$T=\sum_{i=0}^{K} g_i \tag{3}$$

Where: T is the overall area outside profile of real loops.

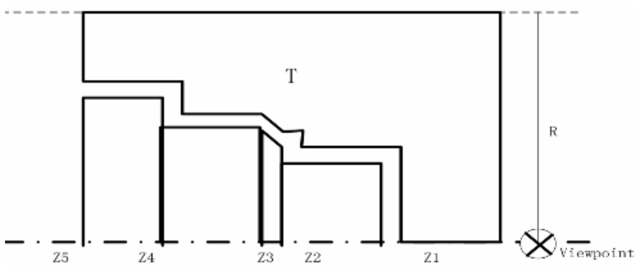


Fig. 4. Machinable turning cutting area outside profile for real loops

After the T finished, the profile of the related unreal loops is a slot, the Jj and the cluster area Fh can be gained.

The procedure is as follows:

(1) Initial the parameter, n is the number of real loops, k is the number of unreal loops;

(2) for n in 1 to N, for m in 1 to M;

(3) The nth real loop is hid, if the mth unreal loop changes into real loop, then find out the g_i as the method for real loop, then $J_j \leftarrow g_i$; otherwise n=n+1, go to (2);

(4) Output J_j;

(5) If J_j is continuous, then cluster it to F_h,

$$F_h = \sum_{j=1}^{E} J_j \tag{4}$$

Where: E is the number of continuous regions.

J_j is the regions outside profile of unreal loops

If J_j is not continuous, then $F_h \leftarrow J_i$;

$$F_h = Ji \tag{5}$$

(6) End the procedure.

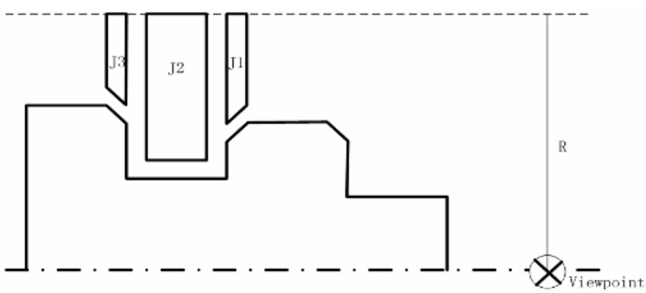


Fig. 5. Overall machinable Turning cutting area outside profile unreal loops

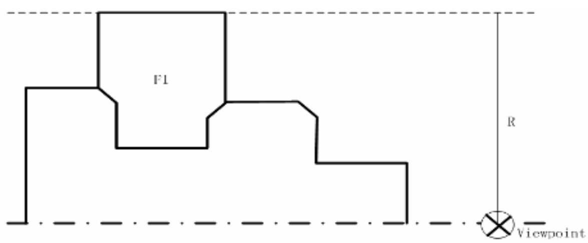


Fig. 6. Overall machinable Turning cutting area profile of unreal loops

The whole area Q of machinable turning cutting area outside profile of real loops and unreal loops is expressed as fellows:

$$Q=T+\sum_{h=1}^{H} F_h \tag{6}$$

Where H is the number of machinable areas of the unreal loops.

The relationship function between two adjoined loops is constructed as fellows:

$$\Phi(r,Z) = \begin{cases} 0 & no \quad relationship \\ 1 & \quad relationship \end{cases} \tag{7}$$

The algorithm of Q as fellows:

A machinable space search function is defined in visibility space.

$$f=\begin{cases} f(\alpha, r_i, r_j \mid \alpha \in (r, R), r = r_i) & \Phi_{ij}(r,Z)=0 \\ q(\alpha, r_i, r_j \mid \alpha \in (\Phi(r,Z), R), r \in (r_i, r_j)) & \Phi ij(r,Z) \neq 0 \end{cases} \quad (8)$$

Where: α is a machinable access radio.

If $\Phi_{ij}(r,Z)=0$, straight feed in cutting region g_i from the former real loop to the latter real loop on Z Axes; if $\Phi_{ij}(r,Z) \neq 0$, straight feed in cutting region g_i from the former real loop to the latter real loop on the turning surface defined by the function $\Phi_i(r,Z)$ between outside the turning surface and inside the R value of the cylindrical co-ordinate.

Input: initial

Specify the viewpoint, direction, and R value of Cylindrical CO-ordinate;

Import 3-D geometry model for other CAD system, or one designed in the EdgeCAM.

Output:

T----the overall area outside profile of real loops, including g_i ;

F_h---the area outside profile of unreal loops, including J_i;

Q----the whole area of real loops and unreal loops;

Algorithm:

Procedure 1: For real loops

Compute the number for all loops, including real and unreal loops according their Z axes value of them, then code them.

$l_i \leftarrow$ {real loops, unreal loops};

M←the number of real loops;

N←the number of unreal loops;

$m_i \leftarrow$ {real loops};

$n_i \leftarrow$ {unreal loops};

Extract the real loops in the projection view, and the relationship functions Φi(r,Z) between every two adjoined real loops.]

For $m_i \leftarrow 1$ to M do {

if(Φ_{ij}(r,Z)=0),then $f(\alpha, r_i, r_j \mid \alpha \in (r, R), r = r_i)$;

Else if(Φ_{ij}(r,Z) ≠0),then $q(\alpha, r_i, r_j \mid \alpha \in (\Phi(r,Z), R), r \in (r_i, r_j))$;

}

For j←1 to k do {

T←g_i +T

}

End the procedure.

Procedure 2: For unreal loop

Compute the number for all loops, including real and unreal loops according their Z axes value of them, then code them.

l_i←{real loops, unreal loops};

M←the number of real loops;

N←the number of unreal loops;

m_i←{real loops};

n_i←{unreal loops};

For n_i←1 to M do {

Extract the ni-th unreal loop, the former loop and the latter loop using the projection view, and the relationship functions Φi(r,Z) among the three loops.

If the former loop is a real loop, then obtain the relationship function Φi(r,Z) between it and the former loop, do

{

if(Φ_{ij}(r,Z)=0),then $f(\alpha, r_i, r_j \mid \alpha \in (r, R), r = r_i)$;

Else if(Φ_{ij}(r,Z) ≠0),then $q(\alpha, r_i, r_j \mid \alpha \in (\Phi(r, Z), R), r \in (r_i, r_j))$;

}

Else if the latter loop is a real loop, then obtain relationship function Φi(r,Z) between it and the latter loop, do

{

if(Φ_{ij}(r,Z)=0),then $f(\alpha, r_i, r_j \mid \alpha \in (r, R), r = r_i)$;

Else if(Φ_{ij}(r,Z) ≠0),then $q(\alpha, r_i, r_j \mid \alpha \in (\Phi(r, Z), R), r \in (r_i, r_j))$;

}

Else the three loops are unreal loops, reserve the ni-th unreal loop and the latter unreal loop, then obtain relationship function Φi(r,Z) between it and the latter unreal loop, do

{

if(Φ_{ij}(r,Z)=0),then $f(\alpha, r_i, r_j \mid \alpha \in (r, R), r = r_i)$;

Else if(Φ_{ij}(r,Z) ≠0),then $q(\alpha, r_i, r_j \mid \alpha \in (\Phi(r, Z), R), r \in (r_i, r_j))$;

}

For j←1 to l do {

If the region J_i is continuous, then do

F_h ←F_h + J_i

Else F_h ←J_i

End the procedure.

Q←T+ F_h

End the algorithm.

The whole machinable turning cutting area H is shown in Fig.7.

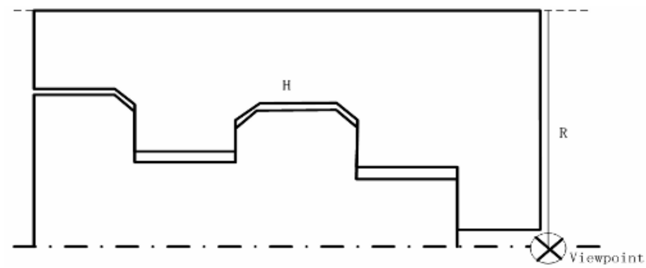


Fig. 7. Whole machinable turning cutting area in visibility space

4 Turning Process Plan and Strategies

4.1 Turning Process Planning and Operation Decomposition

After completing the turning feature recognition and find the overall machinable turning cutting area, start the turning process planning and operation decomposition.

For turning cutting, the turning process planning is divided to rough, half-finish and finish operations. In every operation, the users should choose the turning tools for every machinable turning cutting area from the tool store. The system will generate the machining paths, list of process and list of material. A flexible process plan is variables, alternatives, and possible omissions. Omissions include omitting tool-changing and coolant commands. EdgeCAM uses the method of man-machine interaction to avoid these omissions. Using flexible plans solves the common problem in the discrete parts and other domains where the formulation of plans be reusable. A flexible plan is expected to be usable for an extended time under a variety of conditions whose variability is embodied in the variables and omissions of the plan. In addition to providing for changeable conditions, the users can change the turning toll and machining pattern for any machinable turning cutting area.

Three-stage planning will be used during turning process planning and operation decomposition. The use of two-stage planning is driven by the desirability of cutting flexible plans, combined with the requirement of doing hierarchical planning.

A stage-one plan is in charge of choosing the machining area, turning tool from the tool store, and deciding the method of machining. The stage-one plan will be executed not only automatically, but also by man-computer interaction. The stage-two searches the machinable regions or areas. The tool paths according to the parameters of the turning tool is computed, and the process of turning machining is stimulated in stage-three. When the plan is executed, it is traversed one step in every time. For each step of the plan, one or more executable operations are executed by physical action or by sending commands to a subordinate. By waiting for each command to be executed before selecting the next step to run, the center control cell allows for using up-to-down values of plan variables. In the center control cell, each executes the plan to carry out the operation and then tells the subordinate to execute the plan it just made. If either the subordinate's planning or its execution is unsuccessful at any point, execution of the superior's plan fails and it is stopped for the measure of security.

The disadvantages of executing a flexible plan are as following:

For each command sent to a subordinate, it may be necessary for the subordinate to make a plan from scratch for carrying out a command received from its superior. if the subordinate planner cannot make such a plan, execution of the superior plan fails.

If it is executed more than once under the same conditions, after the first execution, the subordinate will be needlessly making the same plans over again, entailing needless cost, and wasting time.

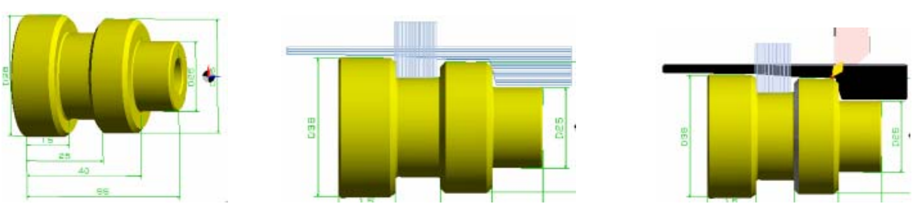
In order to avoid these disadvantages, the superior's stage-one plan is traversed, producing a set of tasks from among alternatives in the stage-one plan, the subordinate makes a plan for carrying out each of the superior's executable operations. The superior makes an ordered list of pairs of the executable operation, each pair consist of the subordinate plan name and the name of the plans of the executable operation, including list of process, tools and materials. In order to support optimizing plans across a hierarchy including the strategy and method of turning machining, the stage two plan is used

4.2 Strategies of Turning Machining

For roughing, half-finishing and finishing, they are adaptive to five similar strategies of turning machining. The five strategies are Turning, Straight Turn/Face, Threading, Grooving and Parting off.

5 Cases of Paths of Tools and Simulation

In general, Turning machining and turning tool from the tool store will be used for the shaft parts. For the different machinable turning cutting areas or regions, different tools shall be selected as per user's actual requirement, the system will generate paths of turning tool, and stimulate the process of machining. Fig.8(a)is a 3-d model of the shaft part,(b) is a tool paths of its turning machining, (c) is simulation of process of its turning machining.



（a）3-d model （b）tool paths of its turning machining （c）simulation of process

Fig. 8. Sample turning for 3-d model

6 Conclusion

Feature recognition is a key technology of the automatic intelligent turning machining method. This paper provide an intelligent turning machining method based on geometrical feature recognition, including putting up viewpoint, direction and cylindrical co-ordinate. For real-loops and unreal-loops in projection, an algorithm is taken forward to recognize the straight faces, turning circle (cone, frustum and cylinder) and slot, and to create the relationship functions among all the loops, and to determine the machinable turning cutting regions respectively according to the relationship functions. For the real-loops, if the regions are continuous then clustered; it is similar for the unreal-loops. The aim is to decrease the cost and save time. For roughing, half-finishing and finishing of different regions or areas, Five different turning machining strategies and methods can be used. They are: Turning, Straight Turn/Face, Threading, Grooving and Parting off. Procedures and parameters of every strategy and method are taken forward. In addition, a integrated CAM system of many CAD system is given, and a hierarchy architecture of the center controller is constructed. The architecture consists of the Graphics Generator, the Center Control Cell and the NC Code Controller. At last, the system will generate a NC program automatically.

References

[1] T.R. Kamer, H. Huang, E. Messina, F.M. Proctor, H. scott. A feature-based inspection and machining system. Computer-Aided Design 33 (2001): 653-669.

[2] M. Balasubramaniam, P. Laxmiprasad, S. Sarma, Z. Shaikh. Generating 5-axis NC roughing paths directly from a tessellated representation. Computer-Aided Design 32 (2000): 261-277.

[3] Chih-Ching Lo. Two-stage cutter-path scheduling for ball-end milling of concave and wall-bounded surfaces. Computer-Aided Design 32 (2000): 597-603.

[4] X. Yan, K.Yamazaki, J. Liu. Recognition of machining features and feature topologies from NC programs. Computer-Aided Design 32 (2000): 605-616.

[5] B.H. Kim, B.K. Choi. Guide surface based tool path generation in 3-axis milling: an extension of the guide plan method. Computer-Aided Design 32 (2000): 191-199.

[6] S.C. Park, B.K. choi. Uncut free pocketing tool-paths generation using pair-wise offset algorithm. Computer-Aided Design 33 (2001): 739-746.

[7] A. Flutter, J. Todd. A machining strategy for toolmaking. Computer-Aided Design 33 (2001): 1009-1022.

[8] S.C. Park, B.k. Choi. Tool path planning for direction-parallel area milling. Computer-Aided Design 32 (2000): 17-23.

[9] Ke-Zhang Chen, Xin-An Feng, Quan-sheng Lu. Intelligent dimensioning for mechanical parts based on feature extraction. Computer-Aided Design 33 (2001): 949-965.

[10] Ganping Sun, Carlo H. Sequin, Paul K. Wright. Operation decomposition for freeform surfaces in process planning. Computer-Aided Design 33 (2001): 621-636.

Author Index