JYVÄSKYLÄ STUDIES IN COMPUTING 50

Jouni Huotari

Integrating Graphical Information System Models with Visualization Techniques

Esitetään Jyväskylän yliopiston informaatioteknologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Agora-rakennuksessa (Ag Aud. 3)
helmikuun 25. päivänä 2005 kello 12.

Academic dissertation to be publicly discussed, by permission of
the the Faculty of Information Technology of the University of Jyväskylä,
in the Building Agora, Ag Aud. 3, on February 25, 2005 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2005

Integrating Graphical Information System Models with Visualization Techniques

JYVÄSKYLÄ STUDIES IN COMPUTING 50

Jouni Huotari

Integrating Graphical Information System Models with Visualization Techniques

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2005

Editors
Seppo Puuronen
Department of Computer Science and Information Systems, University of Jyväskylä
Pekka Olsbo, Marja-Leena Tynkkynen
Publishing Unit, University Library of Jyväskylä

URN:ISBN 9513920488
ISBN 951-39-2048-8 (PDF)

ISBN 951-39-2077-1 (nid.)
ISSN 1456-5390

Jyväskylä University Printing House, Jyväskylä 2005

ABSTRACT

Huotari, Jouni
Integrating Graphical Information System Models with Visualization Techniques
Jyväskylä: University of Jyväskylä, 2005, 56 p. (+ included articles)
(Jyväskylä Studies in Computing
ISSN 1456-5390; 50)
ISBN 951-39-2048-8
Finnish summary
Diss.

Graphical information system (IS) models comprise of several diagram types representing both static structure and dynamic behavior at many levels of detail. Critical design information is distributed across a collection of diagrams, which share some common elements such as objects and their properties. The examination of multiple diagrams slows search performance and could result in reading errors that later cause omissions and inconsistencies. We need better methods and tools, which help understanding the "whole" model and how different parts relate to each other. Especially, during the reviewing process, reviewers need representations that help exploring the model and enable seeing inter-relationships between diagrams. In this thesis, we examine visualization and cognitive aspects of representing design information. We start this by synthesizing the main contributions and research methods in the HCI field. Then we propose a framework for evaluating visualization tools. As a third step, we introduce initial findings and suggestions for using advanced visualization techniques in computer-aided software engineering (CASE) environments. Our first research prototype was used in a laboratory experiment, where we examined the impact of using a large screen, an elision technique, connecting lines, and three-dimensional visualization on readability and recall of a set of graphical IS models. Our second prototype is able to produce 3D UML diagrams, to examine the model, and to show how diagrams are related to each other. We found out that visual integration techniques decrease the designers' cognitive efforts to read and integrate diagrams and significantly reduce errors in both search and recall tasks, especially with respect to individuals with low spatial visualization ability. The research results improve our understanding of how design information is represented. It helps understanding relationships between different types of diagrams and between different levels of detail. This conclusion is applicable to other domains, where information is scattered in various visual forms and levels of detail.

Keywords: information visualization, CASE environments, diagrammatic representation, visual search

ACM Computing Review Categories

D.2.2	Software Engineering: Design Tools and Techniques
	Computer-aided software engineering (CASE)
E.2	Data Storage Representations
	Linked representations
H.1.2	Models and Principles: User/Machine Systems
	Human information processing
	Human factors
H.5.2	Information Interfaces and Presentation: User Interfaces
	Screen design
	Graphical user Interfaces
	Theory and methods
I.3.6	Computer Graphics: Methodology and Techniques
	Graphics data structures and data types

Author's address Jouni Huotari
Jyväskylä Polytechnic
School of Information Technology
Piippukatu 3, 40100 Jyväskylä, Finland
E-mail: jouni.huotari@jypoly.fi

Supervisor Professor Kalle Lyytinen
Department of Information Systems
Case Western Reserve University, USA

Reviewers Professor Harri Oinas-Kukkonen
Department of Information Processing Science
University of Oulu, Finland

Dr. Jonathan C. Roberts
Computing Laboratory
University of Kent, UK

Opponent Professor Keng L. Siau
Department of Management
University of Nebraska-Lincoln, USA

ACKNOWLEDGEMENTS

This research took place at the Department of Computer Science and Information Systems, at University of Jyväskylä, and in the School of Information Technology, at Jyväskylä Polytechnic, and I am grateful to the staff for their support. I am also grateful to COMAS Graduate School and the INFWEST.IT program, which funded a great part of my studies.

I would like to express my gratitude to my supervisor, Professor Kalle Lyytinen. His insightful comments directed the research as well as inspired me to think how visualization aspects are considered within the information system area. I learnt a lot in our discussions, especially during the writing process of our coauthored paper.

Professor Harri Oinas-Kukkonen from University of Oulu and Dr. Jonathan C. Roberts from University of Kent have acted as external reviewers (examiners) of the dissertation. Their constructive comments and suggestions helped me to improve the work in many ways.

Professors Juhani Iivari and Seppo Puuronen encouraged me to continue my studies when I needed some reinforcement. Prof. Puuronen did very careful work in finding many mistakes before the manuscript was printed. I also want to thank Professors Pertti Järvinen and Antero Malin for their comments on my research plan.

I appreciate the ways all the members of the MetaPHOR research group supported each other e.g. by commenting the papers and discussing other research issues. Personally, I found this support invaluable, especially in the beginning of my studies. My fellow researchers during my stay in the MetaPHOR group were Steven Kelly, Matti Rossi, Juha-Pekka Tolvanen, Pentti Marttiin, Minna Koskinen, Janne Kaipala, Janne Luoma, Risto Pohjonen, Zheying Zhang, Eleni Berki, and Kalle Korhonen, whom all I wish to thank here for their contribution. Especially Janne Kaipala, with whom I have written two joint papers, deserves my gratitude. Later, I was fortunate to meet a fellow researcher Marketta Niemelä from the field of cognitive psychology. I particularly admire her critical thinking and statistical skills, which were certainly needed in our joint articles. I also appreciate her comments concerning the different versions of this thesis and the research setting with the second prototype.

In addition, I am indebted to Lauri Koutaniemi, Joonas Hemmilä, and Ilkka Maasola for implementing the two research prototypes. Zoltan Farkas wrote a Bachelor's thesis about the evaluation of the second research prototype. Juho Ristola helped me to improve the figures in the last two papers. I would also like to thank my sister Merja Huotari, my colleague Marjatta Puolakka, Steve Legrand, and various other friends for their support in checking the language. From the INFWEST.IT program, Juha Knuuttila, Sanna Hirvola, and Sanna Paara helped with the practical arrangements.

I have received a lot of support from my superiors and colleagues. Especially Heikki Saastamoinen, Petri Maaranen, Tapani Äijänen, Kari Niemi, Maritta Pirhonen, Esa Salmikangas, and Sampo Kotikoski have helped me with their encouraging comments about my research work.

Finally, I wish to express my special thanks to my family: Leena, Lyydia, and Amanda. Without their support this thesis would never have been completed.

Jyväskylä
January 2005

Jouni Huotari

LIST OF FIGURES

FIGURE 1 Information system development (Tolvanen 1998, 34, based on Lyytinen 1987) ... 17
FIGURE 2 Deficiencies might exist between a) external representations, b) external and internal representations, c) views (internal representations), d) views and real world, and e) external representations and real world ... 18
FIGURE 3 Classification of visualization ... 28
FIGURE 4 A multi-methodological approach to IS research (Nunamaker et al. 1991, 94) ... 34
FIGURE 5 The research method, data gathering techniques, and the timeline of the research process ... 36

LIST OF TABLES

TABLE 1 The concepts of data, information, and knowledge on different layers ... 24
TABLE 2 The main features of a comprehensive information visualization tool (extended from Shneiderman (1998)) ... 29
TABLE 3 The relation between research questions and papers ... 35
TABLE 4 The relation between papers and used research approach ... 36

CONTENTS

ABSTRACT
ACKNOWLEDGEMENTS
LIST OF FIGURES
LIST OF TABLES
CONTENTS
LIST OF INCLUDED ARTICLES
LIST OF ACRONYMS

1 INTRODUCTION 13

2 INFORMATION SYSTEMS, MODELS, AND TOOLS 16
2.1 Information Systems Development 16
2.2 Modeling 17
2.2.1 Information modeling and potential deficiencies 17
2.2.2 Quality factors 19
2.3 CASE tools 20
2.4 Summary 21

3 INFORMATION VISUALIZATION 23
3.1 Data, Information, and Knowledge 23
3.2 Visualization, Perception, and Representation 25
3.3 Information Visualization 27
3.3.1 Data visualization, scientific visualization, and information visualization 27
3.3.2 Classification of information visualization 28
3.4 Visualization Techniques and Applications 30
3.5 Summary 31

4 RESEARCH PROBLEM AND METHODOLOGY 32
4.1 Research Problem 32
4.2 Research Methodology 33
4.3 Application of the Methodology 35

5 SUMMARY OF THE ARTICLES 37
5.1 Review of HCI Research Focus on Cognitive Aspects and Used Research Methods 37
5.1.1 Research problems and methodology 37
5.1.2 Research results 37
5.2 Supporting User's Understanding of Complex Information Spaces by Advanced Visualization Techniques 38
5.2.1 Research problems and methodology 38

5.2.2 Research results 39
5.3 Towards Advanced Visualization Techniques in CASE: Initial Findings and Suggestions 39
5.3.1 Research problems and methodology 39
5.3.2 Research results 39
5.4 Enhancing Graphical Information System Models with VRML 40
5.4.1 Research problems and methodology 40
5.4.2 Research results 40
5.5 Improving Graphical Information System Model Use with Elision and Connecting Lines 41
5.5.1 Research problems and methodology 41
5.5.2 Research results 42
5.6 Integrating UML Views with Visual Cues 43
5.6.1 Research problems and methodology 43
5.6.2 Research results 43

6 CONCLUSION 45
6.1 Contribution of the Thesis 45
6.2 Limitations of this Study 47
6.3 Directions for Further Research 48

REFERENCES
FINNISH SUMMARY
ORIGINAL ARTICLES

LIST OF INCLUDED ARTICLES

1. Huotari, J. & Kaipala, J. 1999. Review of HCI Research – Focus on Cognitive Aspects and Used Methods. In T.K. Käkölä (Ed.) Proceedings of IRIS 22 "Enterprise Architectures for Virtual Organisations", Vol. 2. Jyväskylä: University Printing House, 435–445.

2. Huotari, J. 1998. Supporting user's understanding of complex information spaces by advanced visualisation techniques. In P. Koikkalainen and S. Puuronen (Eds.) Proceedings of STep'98 - Human and Artificial Information Processing. Helsinki: Finnish Artificial Intelligence Society, 41–50.

3. Kaipala, J. & Huotari, J. 1998. Towards Advanced Visualisation Techniques in CASE: Initial Findings and Suggestions. In J. Zupancic, W. Wojtkowski, W.G. Wojtkowski and S. Wrycza (Eds.) Proceedings of Information Systems Development - ISD'98. New York: Kluwer Academic / Plenum Publishers, 742–752.

4. Huotari, J. & Niemelä, M. 2002. Enhancing Graphical Information System Models with VRML. In A.D. Williams (Ed.) Proceedings of the Sixth International Conference on Information Visualisation (IV'02) London: IEEE Computer Society, 670–676.

5. Huotari, J., Lyytinen, K. & Niemelä, M. 2004. Improving Graphical Information System Model Use with Elision and Connecting Lines. ACM Transactions on Computer Human Interface (TOCHI) 11(1), 26–58.

6. Huotari, J. 2004. Integrating UML Views with Visual Cues. In J. Roberts (Ed.) Proceedings of the Second International Conference on Coordinated and Multiple Views in Exploratory Visualization - CMV 2004. London: IEEE Computer Society, 84–92.

LIST OF ACRONYMS

AI	artificial intelligence
CASE	computer-aided systems/software engineering
CSCW	computer supported cooperative work
DFD	data-flow diagram
ERD	entity-relationship diagram
FRISCO	framework of information system concepts
HCI	human-computer interaction
ICT	information and communication technology
IR	information retrieval
IS	information system
ISD	information system development
IV	information visualization
LOD	level-of-detail
RW	real world
SVA	spatial visualization ability
UML	unified modeling language
VRML	virtual reality modeling language

1 INTRODUCTION

During the last decades we have seen a rapid growth in the use and importance of information and communication technology (ICT). ICT has many implications in the society. One implication is the increasing amount of data in an electronic form. The rapid growth of electronic data has lead to an increasing difficulty to understand what data actually means and how it relates to other data.

These complex relationships between design data exist also in information system development (ISD) and computer-aided systems/software engineering (CASE) environments, which consists of many interrelated diagrams and documents ranging from requirement specifications to documentations of code. We create information system specifications for the same reason as architects create blueprints for houses: they help communication before, during, and after the actual "thing" (house or information system) is built. As with houses, there are many types of information systems, e.g. manufacturing information systems, management information systems, transaction processing systems, decision support systems, experts systems, and human resources information systems (Gupta 2000). A common issue with these systems is that they all deal with data, which is usually stored in one or more databases. This thesis focuses mainly on design data, which exist in design databases called repositories.

Although the aim of using ICT is to help people to be more productive in their work, this does not always happen, and there are many examples of failures. The risks and problems experienced with these systems originate from several factors, such as the large size of the project, complexity of the problem domain, project members being unfamiliar with new technology, unstable information requirements, and difficulties in integrating different component systems into a composite system (Ewusi-Mensah 1997). Consequently, when people develop information systems for other people to use, there is always a possibility of an error, omission, or inconsistency. The increasing demand for high-quality systems that should be delivered to market at a fast rate often results in over-the-budget, poorly documented systems that are difficult to maintain. The importance of correct, complete, and consistent specifications is increasing. Therefore, we need tools that help reviewing information system models and finding deficiencies as early as possible.

I propose several solutions to address the problems mentioned above. To begin with, we need guidelines and standards, which tell what information to include, where to store that information, how it is done, how it is achievable, and who does what and when. People need to be motivated to follow those guidelines. We should therefore explain why this is important and teach how to do it. In addition, we need also technical solutions, such as an efficient repository, which integrates parts of the design documents to one manageable entity. Lastly, we need a tool, which supports finding, integrating, and visualising information. In this thesis, we[1] are concentrating on the last issue, especially on its cognitive and visualization aspects.

The visualization and conceptualization of complex information spaces has been considered important in several studies (Monarchi and Puhr 1992; Spence 1993; Noik 1994; Andrews 1995; Card and Mackinlay 1997; Kim et al. 2000; Roberts 2001; Shneiderman 2002). The growing number of books (Tufte 1983; Tufte 1990; Tufte 1997; Blasius and Greenacre 1998; Shneiderman 1998; Wildbur and Burke 1998; Card et al. 1999; Chen 2000; Ware 2000; Fayyad et al. 2001; Spence 2001; Bederson and Shneiderman 2003; Geriomenko and Chen 2003; Chen 2004; Hansen 2004), journals (e.g. Information Visualization since 2002) and conferences related to data and information visualization indicate increasing interest towards the subject. However, not many studies deal with the combination of information system area and visualization area.

Information visualization shows great promise for communicating information properties. For instance, through visualization the context and the structure of information space can be explicitly represented, which essentially supports different ways of using the information, such as searching, browsing, interacting with, and thus understanding its complexity. Moreover, it can reduce cognitive overload by filtering unnecessary information, as well as enhance users' perception and exploration of the content and structure of the information space. Thus, relevant information and overall structure can be shown in a meaningful way. Since advanced visualization techniques and virtual environments are emerging at a rapid rate, cognitive aspects must be considered before inappropriate designs and practices become common.

Even though the processing power and storage capacity have increased by orders of magnitude, the screen size has remained quite small. There is simply not enough space on a typical computer screen to visualize the rich information content characteristics of today’s computers. Big screens and virtual environments could help seeing, exploring, and understanding information.

A large portion of the brain is devoted to visual processing. People have well-developed languages for visual communication and a capability to see things in two (2D) and three dimensions (3D). Vision is therefore argued to be the most important of human senses (Bear et al. 2001). Now, when computers are capable of providing advanced visualization techniques fluently, including

---

1 I am using “we” because most of the papers are a joint effort.

3D and virtual environments (Ware and Franck 1996), their use should be carefully studied in order to utilize their full potential. Simultaneously, the characteristics of the internal representations (mental models) must be considered in order to make consistent and rapidly comprehensible depictions.

This thesis identifies representational and, in dealing with the design information, integration problems that information system designers and reviewers confront. We seek an answer to the question how to improve seeing inter-connections in graphical information system models. We introduce some suggestions on how to solve those problems with a special interest placed on utilizing visual cues, advanced visualization techniques, and 3D. We examine how visualization techniques could enable users to extract important elements from large amounts of abstract data more easily, effectively, and/or rapidly.

Our aim is to find ways and means that would support IS designers, reviewers, and other stakeholders in grasping better the idea behind a complex set of specifications. It is one of the steps towards CASE tools that would better address the needs of designers, particularly by improving the design information visualizations during an ISD process. The findings do not merely help CASE tool builders to improve the existing CASE tools, but all information seekers can benefit from them. For example, existing information in the Web, in databases, or in multimedia applications can be visualized by applying suggestions and techniques explored in this thesis.

The thesis focuses on information, visualization, cognition, and design work. These aspects are studied in many research fields, such as information systems, human-computer interaction (HCI), artificial intelligence (AI), databases, computer graphics, ergonomics, arts, cognitive science, neurology, cognitive psychology, and computer supported cooperative work (CSCW). These areas emphasize different aspects of visualization: databases include storing and retrieval of data, computer graphics visualizing data on the screen, arts proposes aesthetic values, cognitive psychology and neurology dive inside the human brain, and CSCW puts the emphasis on collaboration aspects. We have examined all these areas and strive for a multi-faceted, integrating, and interdisciplinary research.

In the next two sections we will look at these themes, define the basic terminology concerning information system models and information visualization, and review related research. Then we will formulate the research problem and applied research methodology. A short summary of each paper in the thesis is presented in Section 5, followed by a brief overall conclusion, discussion about limitations, and suggestions for future research.

2 INFORMATION SYSTEMS, MODELS, AND TOOLS

In this Section, we present the basic terminology concerning information systems development, modeling, quality factors, and CASE tools. We are concentrating on these issues, because they all relate to the main focus of the thesis: improving integration of graphical information system models. The improvement of IS models and tools is a part of improving the quality of IS development, during which modeling takes place.

2.1 Information Systems Development

Developing information systems is a complex process. It involves communication between different stakeholders, such as developers, users, managers, and administrators. ISD is often aiming at a moving target, because requirements usually change during the development process. Information systems development is a change process taken with respect to an object system in an environment by a development group using tools and an organized collection of methods to produce a target system (Welke 1981; Lyytinen 1987, 6, see Figure 1). This definition takes into consideration technical, conceptual, and organizational aspects. Consequently, the concept of a system is wide, including code, a database, or a method. Even an organization itself can be seen as a system.

An ISD method defines the way of carrying out the change process. There is no unified definition about what constitutes a method. In this thesis, I propose a quite broad definition: a method is an organized collection of concepts, techniques, beliefs, values, and normative principles (Hirschheim et al. 1995). Information modeling method can be defined as an approach to perform modeling, based on a specific way of thinking, consisting of directions and rules, and structured in a systematic way (Brinkkemper 1996). A central part of a method is the technique which explains the procedure of how to perform a task to accomplish a desired state (Welke 1983).

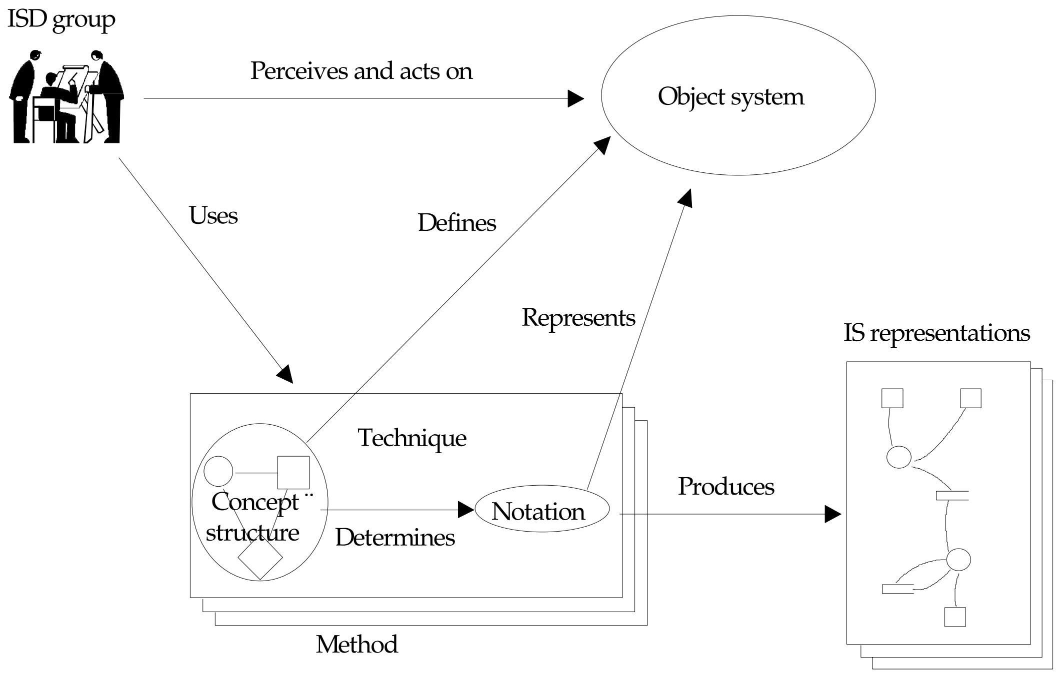

FIGURE 1 Information system development (Tolvanen 1998, 34, based on Lyytinen 1987)

The improvement of the ISD process is one of the critical development areas in software development companies. Standardizing the ISD process may solve many of the design time problems, but it does not, however, help people to understand those descriptions (specifications, design decisions etc.) that emerge during the ISD process (Shu 1988). One of the key issues in this thesis is to address the problem of how to integrate and visualize effectively those IS representations that form the output from different modeling techniques.

2.2 Modeling

Modeling is a central part of the ISD. In this Section, I describe the reasons why modelling is important. I also introduce some quality factors, which can be used as evaluation criteria for information system models.

2.2.1 Information modeling and potential deficiencies

Siau (1999, 44) defines information modeling as "the process of formally documenting the problem domain for the purpose of understanding and communication among the stakeholders". The outcome of the modeling process is one or more information system models, which are constructed by using an information modeling method such as Modern Structured Analysis (Yourdon 1989) or a modeling language such as Unified Modeling Language (UML) (Booch et al. 1999) . These methods and languages usually contain one or more diagram types (e.g. data flow diagrams, entity-relationship diagrams, use cases, sequence diagrams, class diagrams), which depict the system from one or more perspectives. In this thesis, we are focusing on the use and integration of different types of diagrams and thus we are dealing with graphical information system models.

Selonen et al. (2001) studied the relationships and transformation operations of different UML diagram types. They show that it is possible to utilize certain rules in order to transform UML diagrams from one type to another. Their findings help to understand how different UML diagram types relate to each other.

Designers build models for several reasons: to communicate the desired structure and behavior of an information system, to visualize and control its architecture, and to better understand the system (Booch et al. 1998). Early detection of faults through reviews, inspections, prototyping, or simulation will improve reliability. In addition, computer supported modeling often helps simplifying, reuse, and managing risks. It is also important to check that the model is consistent and meets the requirements. Even though the requirements' traceability is difficult to implement, it is important to document to which requirements a single diagram is related. We could, for example (see paper 3, Figure 6), visualize the links from requirements to diagrams and even further to code.

We can identify at least five types of deficiencies (errors, omissions, or inconsistencies). These deficiencies exist in and between three "places": the real world, human minds (internal representations or views), and external representations (see Figure 2).

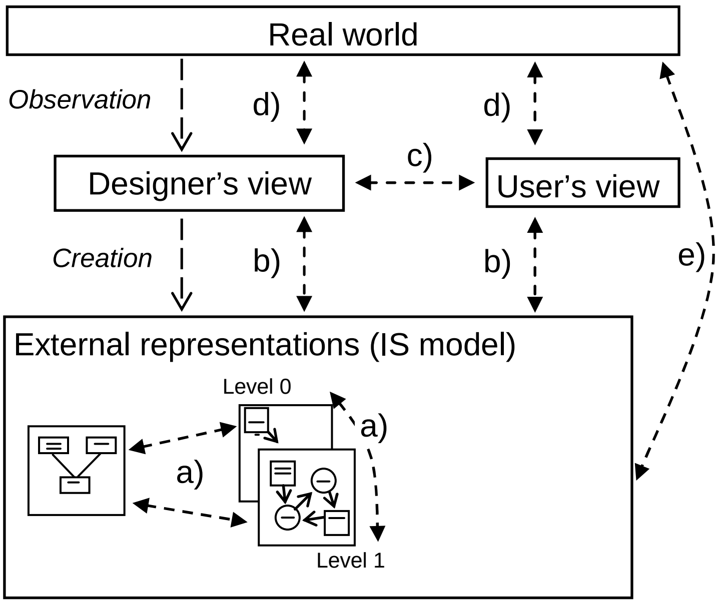

FIGURE 2 Deficiencies might exist between a) external representations, b) external and internal representations, c) views (internal representations), d) views and real world, and e) external representations and real world

Deficiencies of the first type exist between external representations, which might contain deficiencies both between different views or types of diagrams and between levels of detail (Figure 2, a). A model is internally consistent if there are no contradictions between the elements in the different diagrams that comprise the model (McGregor 1998). There can also be contradictions between the elements in a diagram, and this is likewise regarded as belonging to the first type of deficiencies.

The second area of concern is between external and internal representations (Figure 2, b). Internal representation denotes designers' and users' view or mental model about how things work, what things mean, how they relate to each other etc. Scaife and Rogers (1996) outlined some of the central properties of the relationship between external and internal representations. They identified five key conceptual design issues: explicitness and visibility, cognitive tracing and interactivity, ease of production, combining external representations, and distributed graphical representations.

Third, there might be contradictions between users' and designers' views (Figure 2, c). Fourth, deficiencies can exist between real world and people's views (Figure 2, d). The fifth area concerns deficiencies between real world and external representations (Figure 2, e). This area is close to representation deficiencies, which are defined in terms of the difference between the view of the real-world system as inferred from the information system and the view that is obtained by directly observing the real-world system (Wand and Wang 1996). In fact, deficiencies between the real world and external representations (Figure 2, e) are often due to lack of communication between stakeholders (Figure 2, c) or incomplete understanding about the real world (Figure 2 d). In this thesis, I concentrate mainly on the first issue, i.e. helping to find deficiencies from graphical information system models.

As the complexity of systems increases, so does the importance of good modeling and visualization techniques. Grasping the idea behind an existing IS from its documentation is often time-consuming, partly because of the complexity involved in interpreting how different parts of the graphical information system model relate to each other. Therefore, more efficient ways to visualize the design information are needed.

2.2.2 Quality factors

Wand and Wang (1996) summarize the most often cited data quality dimensions based on a comprehensive literature review (Wang et al. 1995). The seven most often mentioned quality attributes were accuracy, reliability, timeliness, relevance, completeness, currency, and consistency. Although there is no general agreement on the criteria for evaluating a model, three common quality factors can be found: correctness (accuracy), completeness, and consistency (IEEE 1989; Wand and Wang 1996; McGregor 1998).

McGregor (1998, 21) defines *correctness* as follows: "a model is correct if it is judged to be equivalent to some reference standard that is assumed to be an infallible source of truth". For example, the standard can be a domain expert who evaluates the model. According to McGregor (1998, 21), "a model is *complete* if no required elements are missing". This is iteratively evaluated by determining if the model sufficiently matches the goals for the current increment and all necessary values are included. A model is *consistent* "if there are no contradictions among the elements within the model" (McGregor 1998, 22). These contradictions may be due to, for example, multiple representation of the same concept or differing cardinalities.

Wang (1998) continued the classification of quality factors by introducing information quality (IQ). He categorizes accuracy as an intrinsic IQ, which states that information has quality in its own right. Relevancy, timeliness, and completeness belong to contextual IQ, which emphasizes that information quality must be considered within the context of the task at hand. Consistency along with interpretability, ease of understanding, and concision belong to representational IQ, which highlights the importance of information systems. Wang treats information as a product that moves through an information manufacturing system (Wang 1998). We can apply his ideas also in CASE environments, where the design data is produced, analyzed, and maintained. In fact, the quality factors mentioned above form the basis of our information search tasks given to the subjects in our laboratory experiment and pilot study.

2.3 CASE tools

Even though the problems in ISD are often non-technical – the commitment of the stakeholders as an example (Iivari 1994) – CASE tools, CASE shells (Bubenko 1988), or metaCASE tools (Alderson 1991) can increase productivity, improve system's quality, shorten the development lifecycle, and result in reusable methods, method components, and code. CASE, in the large, includes any computer support from designing specifications to automatic generation of database and code. A CASE tool typically supports only one part of the development process, whereas a CASE environment[2] is aimed to support a large part of the ISD process.

The research in metaCASE environments has natural connections to CASE environments. The MetaPHOR group, in which I also participated, developed a metaCASE tool called MetaEdit+. The participants of the MetaPHOR group have published over 100 papers in various conferences and journals, e.g. (Lyytinen et al. 1994; Kelly and Rossi 1997; Tolvanen 1998). MetaEdit+ can be used also as a CASE tool. Problems and solutions found in metaCASE environments apply often in CASE environments, and vice versa. Therefore, I have reviewed also metaCASE literature extensively.

We can identify four problem areas where research and tools should be extended: the representational, conceptual, methodological, and implementation (Kelly and Smolander 1996) area. Current commercial CASE (and metaCASE) tools and environments fall short in many respects (Kelly 1997):

- Lack of different representational paradigms (diagrams, matrices, tables, etc.)

[2] CASE environments have many different names, such as Integrated Project Support Environment (IPSE) and Software Engineering Environment (SEE).

- Only simplistic checking and visualization of the contents of the repository
- Only partial graphical support for the (meta)modeling process
- Difficulty in seeing how the modifications of the metamodel influence the corresponding model
- Lack of support for finding and seeing reused and reusable components
- Poor horizontal and vertical method integration
- Several methods can not be used in an integrated manner
- Insufficient integration of different tools, such as project management tools, from different levels and phases of ISD.

From this list, we address most of the issues, especially integration and visualization of the contents of the repository. Even though we are discussing model visualization, our findings are applicable also for the metamodels.

Henderson and Cooprider (1994) created the functional CASE technology model (FCTM), which consists of 98 functions. These functions are assigned to three categories: production, coordination, and organizational technology. Production technology is further divided to representation, analysis, and transformation. Coordination technology consists of control and cooperative functionality. Organizational technology contains support and infrastructure. From these three dimensions of IS planning and design technology, we are especially interested in production technology.

The 98 functions can be used as the evaluation criteria for CASE tools. Henderson and Cooprider (1994) asked expert users of a CASE tool to evaluate its ease of use with respect to the functions. According to the detailed listing of CASE tool functionality, all the eight CASE tools that were evaluated fell short in many respects. For example, under the functionalities of analysis, there are functions “Identify the design impact of proposed changes in a design” and “Search the design for similar objects". The support for these functions were absent in all eight CASE tools (Henderson and Cooprider 1994).

2.4 Summary

Successful development of IS requires effective communication between all stakeholders. To improve communication and to achieve a more comprehensive understanding of the target, designers should produce meaningful and understandable representations (IS specifications). According to our literature review and examination of current CASE tools, there is not enough support for designers and reviewers to enable them to see how different parts of information system models relate to each other. This support is important to spot deficiencies (errors, omissions, and inconsistencies) as early as possible during information system development. Support is also needed to understand whether a part of an information system model can be reused and how a change (e.g. renaming an object) affects the other parts of the model. In the next Section, I discuss the way that information visualization could help seeing those interconnections and thus support designers and reviewers (as discussed further in papers 2–6).

3 INFORMATION VISUALIZATION

Here I try to answer the question: what is information visualization? Choosing an appropriate definition for the terms is difficult, because different research fields or interest groups differ in their points of view. For example, the term "visualization" can be seen as a process or a product, emphasizing either human (mental) or technical aspects. Here I aim at a broad, holistic view of defining basic terminology.

I introduce here also classification frameworks for information visualization. This gives us a checklist for building a visualization tool, which supports designers or reviewers in finding deficiencies from a graphical information system model.

3.1 Data, Information, and Knowledge

What is data, what is information, and what is knowledge? These questions have long exercised the minds of philosophers. The FRISCO[3] report (Falkenberg et al. 1998), defines the term *data* as any set of representations of knowledge (i.e. meaningful symbolic constructs, such as numbers, characters, or images), expressed in a language. The report treats *information* as the knowledge increment, brought about by a receiving action in a message transfer, i.e. it is the difference between the conceptions interpreted from a received message and the knowledge before the receiving action. Here the word "message" may have several meanings, depending on the context. We can think of the physical appearance of a message, its syntax and semantics, and its pragmatics in a social context. These semiotic layers are helpful also when thinking about terms such as visualization so that we can talk about data visualization, information visualization, and knowledge visualization. Because the meanings of data and information are often mixed, both terms are used and the context reveals the meaning of the term. In Table 1 the terms data, information, and knowledge, are

---

3 FRISCO is an acronym for FRamework of Information System COncepts

positioned on these semiotic layers. We can also identify a counterpart for them at different abstraction levels (Iivari 1989).

TABLE 1 The concepts of data, information, and knowledge on different layers

Main concept	Semiotic layer	Explanation	Abstraction level
Data	Physical	The physical appearance, the media	Technical
	Syntactical	The language, the structure, and the logic used	
Information	Semantical	The meaning and validity of what is expressed	Conceptual
Knowledge	Pragmatic	The intentions, responsibilities, and consequences behind the expressed statements	Organizational

Thinking with these layers or levels, we have been seeing a shift of research effort from technical and conceptual to wider, social context direction (Kuutti and Bannon 1993). In this thesis, we are focusing mostly on the middle level, while considering also the other levels.

CASE and visualization tool developers should decide what is the most appropriate visualization technique for different levels of abstraction. The challenge is to visualize the design data so that different people understand the meaning of a particular design data. As an example from the conceptual level, understanding the meaning of the concept of a 'Customer' requires often more than one view (e.g. a use case diagram, a class diagram, and a sequence diagram in UML). Without efficient visualization techniques, which show how the concept is used in different views (diagrams), it might take long to understand the relationship with other concepts (structure and behavior) or it might lead even to misunderstanding.

As in the case of abstraction levels, Juhani Iivari (1989) identifies three domains: Organizational, Universe of Discourse (UoD), and Technical. For each domain, three types of abstractions can be distinguished: structure, function, and behavior (Iivari 1989). Another effort to organize design information is introduced by John A. Zachman (1987). Zachman's Framework consists of six layers or levels of abstraction from business environment and enterprise model through system and technology model to detailed representations and functioning enterprise. In addition, Zachman's Framework provides answers to six types of questions (and corresponding artifact): what (data), how (function), where (place), who (people), when (time), and why (motivation).

The apportionments mentioned above can be used also as selection criteria for choosing a particular perspective through which the information system is viewed. For example, one might want to see what functional abstractions exist between the organizational and conceptual level. A CASE or visualization tool should manage these views so that a person reviewing an information system model could understand the relationship between different levels of detail.

One might also want to change the representational paradigm, e.g. from a table to a diagram, on one abstraction level. In theory, we could even choose the method according to which the representation changes, e.g. the conceptual schema can be shown as an entity-relationship diagram (ERD) or as a class diagram in UML. In practice, mappings for such method conversions are difficult to implement. In all, it is difficult to map the real world (RW) system to the information system and to compare representations of the RW system and IS (Wand and Wang 1996).

One of the most fundamental reasons for the poor visualization support is the lacking of features that enable the integration and further showing of concrete links between different objects, e.g. structural and behavioral diagrams. To enable this integration, (meta) data model must be made more powerful to enable the linking and representation of all the kinds of information necessary (Kelly 1997). In addition to showing associations to other objects, the associated properties of the design information, such as creator of an object, creation date, or motivation for creation/modification could be of value when deciding further actions (reusing, deletion, renaming) for an object. This metadata could be used also as categorizing and further visualizing design information.

3.2 Visualization, Perception, and Representation

The four concepts – perception, conception, representation, and visualization – are difficult to define separately without referring from one to another. They all have many definitions, depending on the research field. The researchers from the computer science field emphasize *visualization* as a product or a technique. As an example of product-oriented definition, visualization may be seen as

> the visual representation of a domain space using graphics, images, animated sequences, and sound augmentation to present the data, structure, and dynamic behavior of large, complex data sets that represent systems, events, processes, objects, and concepts (Williams et al. 1995).

Researchers from many other fields view it more as a cognitive process, for example visualization can be seen as mechanisms by which humans perceive, interpret, use, and communicate visual information (McCormick et al. 1987). Card et al. (1999, 6) synthesize these “hard” and “soft” views by regarding visualization as “the use of computer-supported, interactive, visual representations of data to amplify cognition”. In addition, they characterize visualizations as adjustable mappings from data to visual form to the human perceiver. All of these definitions refer to the term *visual*, which is commonly related to seeing, but we can use all other senses as well, such as touching and hearing. Visualization is thus a multi-faceted “product” or process, and the meaning of the term is often revealed from the context.

Ware (2000), McCormick et al. (1987), and Tufte (1983) list several advantages of visualization:

- Visualization enables the viewing, comparing, and comprehending of huge amounts of data
- Visualization allows the perception of emergent properties and thus fosters profound and unexpected insights
- Visualization reveals hidden problems, deficiencies, or errors
- Visualization eases hypothesis formulation and enriches the process of scientific discovery
- Visualization facilitates understanding of both large-scale and small-scale features of the data.

Visualization can thus be invaluable in quality control and perception of patterns linking local features (Ware 2000). Friedhoff and Kiley (1990) argue that graphically rendered information is assimilated at a much faster rate. Robertson (1991) emphasizes the selection of an appropriate representation, because it affects the observer's mental model and further subsequent analysis, processing, or decision-making.

As noted before, people have remarkable perceptual abilities for visual information. *Perception* covers all human senses, but vision is the most important one for this thesis, and forms the focus of this study. Human perception and the interpretation of received information is a complex process. It involves several levels of processing, ranging from low-level sensory mechanisms to higher-level cognitive mechanisms. Perceptual process can be seen as akin to scientific process (Bruner 1957). According to Bruner, the perceptual process consists of finding clues, formulating a hypothesis about what the clue is, verifying it, and then either accepting the hypothesis or reformulating it and trying again. Bruner also thinks that perceptual experience is the end product of a categorization process. Moreover, Bruner states that concepts or categories have to be defined before any recognition is possible. In addition, subjects' knowledge and expectations influence how one interprets the perceived stimulus. Here I adopt this constructivist view, although I agree with Gibson (1979) that “direct” pickup of relevant information is also possible.

Representation is one of the central concepts related to human cognition. Marr (1982, 20) defines a representation as “a formal system for making explicit certain entities or types of information, together with a specification of how the system does this”. Here the system refers to human brain, which processes the received information. Representation thus share many aspects with visualization and it can also be tackled internally or externally. Internal representations or models are often referred as a conceptual or mental map. These maps can contain deficiencies in the same manner as external representations, as noted in the Section 2.2.1.

In information system models, an external representation is typically a labeled graphical symbol showing the values of properties of a given conceptual model component (Kelly 1997). Representation can contain exact positioning information (coordinates) of a particular object in a diagram. Other examples of the properties of a representation are size, color, and font (Spence 2001). These representational elements constitute a so-called notation. A representation can contain also some "intelligence", e.g. when, where, and how the property values are displayed.

Distinction between conceptual and representational aspects of information is important. This can be seen as a dimension in CASE (Smolander et al. 1991), another dimension being type-instance. Kelly (1997) gives a good example of maintaining the same conceptual information and displaying it in several different representations, e.g. as a diagram (possible with different layouts), matrix, or a table. According to this example (p. 24),

> a Class Diagram graph could have two different diagram representations, one stressing the inheritance hierarchy and the other the aggregation hierarchy: the underlying conceptual graph would be the same for both. Most of the conceptual objects would then have representations in both diagrams, whereas the conceptual relationships would mostly have a representation in only one diagram or the other.

Tweedie (1997) introduces three aspects to representation: data "behind" the representation (meta data/raw data); forms of interactivity (direct/indirect); and input and output information (is it represented and in what direction, e.g. Input → Output). From these points of view we can infer that a representation contains more or less hidden information.

3.3 Information Visualization

3.3.1 Data visualization, scientific visualization, and information visualization

Defining information visualization is difficult because of the various, somewhat contradictory or confusing uses of the term. One of the reasons is the mixed use of the words data and information, as noted before. As can be seen from Figure 3, *Data visualization* or data graphics consists of scientific visualization and information visualization. *Scientific visualization* is mainly concerned with phenomena that are based on the physical world. The data for it is collected about the earth, buildings, molecules, or other topics. *Information visualization* (IV), in contrast to scientific visualization, is concerned with visualization of large volumes of abstract, non-physical data (Card et al. 1999). It is noteworthy that the distinction of the terms data and information here is different from previously expressed definitions.

Kamada and Kawai (1991, 36) present one of the early definitions of information visualization, which is "translation from textual or internal representations into pictorial representations". Furthermore, they regard the visualization process as translation from textual languages into two- or three-dimensional visual languages and call this process *translation into pictures*. More recently, Card et al. (1999, 7) define information visualization as "the use of computer-supported, interactive, visual representations of *abstract* data to amplify cognition". This broader view is more applicable for our purposes, even though non-visual presentation of information also can facilitate understanding.

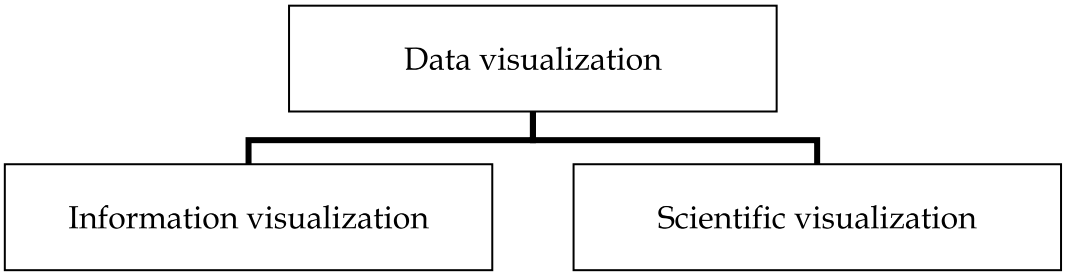

FIGURE 3 Classification of visualization

In the field of information visualization, a diverse range of dynamic and interactive visual environments have been developed. These exploratory tools enable a user to investigate and search the information space to develop a better understanding of the underlying information. Such investigative environments often utilize many different views of the same data, so the user understands the information from different perspectives; the views are also tightly coupled together to allow rapid coordinated investigation and exploration (Roberts 2004).

3.3.2 Classification of information visualization

There are several classifications and frameworks for visualization and visual representations. These classification frameworks are needed, when we are building a visualization tool which shows design data effectively. According to Lohse et al. (1994), functional classifications focus on the intended use and purpose of the graphics, whereas structural categories focus on the form of the image. The latter is closer to the previously presented definition of a representation, because it is intended to correspond to an underlying representation in memory.

Shneiderman (1998) and Card et al. (1999) introduce data type and use of space (1D, 2D, 3D, temporal, multi-D, hierarchy, and network) as the basis for classification. The categorization is sometimes difficult, because the visualization might contain both hierarchical and network structure and the display might contain many types of use of space. Card et al. (1999) mention the document lens as an example, which involves 1D data (the sequence of document pages), a 2D visual structure (the surface), and a 3D spatial substrate (for the distortion). They classify the document lens as a 2D visualization that uses a 3D distortion to increase the amount of information. In many cases the actual visualization is a combination of different uses of space. One possible simplification to categorize visualizations is to think of just two issues: dimension (1D, 2D, 3D, multi-D, and time) and structure (hierarchy and/or network). With this categorization, we could talk about 2D hierarchies, 3D networks etc.

Ben Shneiderman (1998) identifies seven tasks that can be executed with the visualization. Those tasks are overview, zoom, filter, details-on-demand, relate, history, and extract (Table 2). Basing on the observations, he proposes an information-seeking mantra, which is "overview first, zoom and filter, then details-on-demand". Our review in paper 2 supports Shneiderman's ideas, but we have modified and extended the list by combining zoom and filter as one of the features and by adding two other features: search strategy and representation. This list contains some of the issues mentioned in the Nielsen's usability heuristic list (1993), but it lacks some of the basic user interface design issues such as providing help and providing user control and freedom. In summary, both the extended Shneiderman's list and the Nielsen's list should be used for building and evaluating visualization tools.

TABLE 2 The main features of a comprehensive information visualization tool (extended from Shneiderman (1998))

Feature	Explanation
Overview	A clearly organized overview map is provided; if 3D is used, it should deal effectively with occlusion
Search strategy	Searching is possible by using keyword(s), index, browsing, and/or agent (both group of objects and single nodes can be browsed)
Zoom and filter	Zoom in on items of interest and filter out uninteresting items (option: global context retained)
Details-on-demand	The actual information source can be achieved; logic behind the representation is shown or explained
Relate (focus + context)	The relationships among items and the place of selected information are shown (in the overview)
Representation (layout and structure)	Simple and clear representation, which is appropriate considering the task in hand; viewer's attention is not drawn from the essential information; the Gestalt laws are considered
Navigation aids	Landmarks, history, and backtracking facilities are provided; traversal path can be seen if desired
Extract, customizability	Current state of the representation and user options can be saved; dynamic links can be created (hypertext functionality)

Shneiderman's list has been expanded later. For example Card et al. (1999) list more than ten other tasks (read fact/comparison/pattern, manipulate, create, delete, reorder, cluster, class, promote, average, abstract, instantiate, compose, and organize) that a user may want to do with the visualization. Recently, Itoh et al. (2004) suggest a list of features that a visualization technique should include. The features are: efficient use of display spaces, no overlaps between nodes, an aspect ratio of subspaces, a flexible placement of arbitrarily shaped nodes, similarity (one of the Gestalt laws), and semantics of placement. In our list (Table 2), we emphasize these representational issues, but we have a bit different emphasis: visual layout and structure should be simple and clear; the viewer's attention should not be drawn away from the essential information; the Gestalt laws (simplicity; familiarity; similarity; good continuation; proximity; common fate; and connectedness) are considered; and crossing lines in a diagram and between diagrams should be managed effectively. All of these features have been considered in our research prototypes (papers 4, 5, and 6), even though not all of them have been implemented yet.

3.4 Visualization Techniques and Applications

Although there are a number of research papers and projects which offer promising solutions, these solutions are rarely commercialized. This implies the difficulty of implementing a comprehensive support for information visualization. Because the research on visualization techniques and applications is currently advancing rapidly, we do not introduce current solutions here in more detail, we just refer to them in appropriate situations. One of our research aims is to select the appropriate solutions and extend their use in CASE.

For visualizing large information structures, especially hierarchies (a tree structure), researchers have proposed a number of visualization techniques: fisheye views (Furnas 1986; Sarkar and Brown 1994; Turetken et al. 2004), interactive graphical documents (Feiner 1988), the cone tree (Robertson et al. 1991), the space-filling tree-map (Johnson and Shneiderman 1991), elision (Parker et al. 1998), and the hyperbolic browser (Lamping and Rao 1996), to name just a few. Elision technique, where detailed parts of the model are collapsed into small icons, reduces visual clutter and enables better focus over designs and a more efficient management of the representation space. When a collapsed node is opened, it expands, or explodes, and shows its more detailed content. Selective aggregation shares characteristics with elision. Both are focus+context techniques, as is the original idea of fisheye views.

For visualizing networks, alternatives vary from straightforward connecting lines to brushing (Becker and Cleveland 1987). A network of connecting lines (arcs) and semantically related elements (nodes) form technically a graph. Research on diagrams or graphs in general has a long history. One of the most acknowledged researchers is Jacques Bertin, who has done extensive work on graphical semiotics and graphic information processing (Bertin 1983). Mackinley (1986) formalizes Bertin's ideas and proposes an automatic design of graphical presentations of information.

Kaipala (1997) and Oinas-Kukkonen (1997a) have studied how to connect diagrams located in separate windows by using a feature called linking-ability (LA). It enables creating hyperlinks between diagrams and shows explicitly an object containing links. It also shows the actual link as a line. Hypermedia applications (Garzotto et al. 1996; Oinas-Kukkonen 1997b) offer solutions for linking parts of (design) documents together. HyperTree (Salampasis et al. 1997) and MICROCOSM (Fountain et al. 1990) are two typical examples. The Brain (see Figure 3 in paper 2) is an example of a patented network visualization application, which effectively displays connections between "thoughts" and related documents (Harlan 2000).

Roberts has studied multiple-view and multiform visualization in several papers (Roberts 1998; Roberts 2000; Roberts 2001). One of the challenges is to coordinate multiple views so that the comparison of data sets is possible. Usually there is a window for every view, which shows data from different perspectives. It is also possible to integrate multiple views in one "world", as we show in paper 6. Egyed presents the view integration framework (Egyed 1999) and UML/Analyzer tool (Egyed 2002), which automatically integrates architectural views of UML. The tool supports model transformation and consistency checking. Although his findings are important, the automatic integration is not always certain.

3.5 Summary

There are several categorizations of information visualization, which can be used when thinking of appropriate visualization tools. A CASE or visualization tool should manage different views so that a person reviewing an information system model would understand the relationship between different levels of detail. Several visualization techniques have been proposed during the last two decades. Surprisingly, these have not been implemented in CASE tools, yet. This finding has motivated us to examine the possibilities of building support for reviewers of an IS model.

4 RESEARCH PROBLEM AND METHODOLOGY

Having introduced the background and terminology for the research, we now propose our research problems, questions, and a research methodology that directs the way we address these problems and questions. We also describe how we apply the research method and data gathering techniques to different phases of the research process.

4.1 Research Problem

Research in this thesis aims towards better support for designers and reviewers (people checking an IS model) to see and understand how different diagrams and objects within them relate to each other. We assume that this will help finding deficiencies from graphical information system models and thus will improve the quality of those models. From this, the general research problem of this study is:

> *How can we improve seeing inter-connections in graphical information system models?*

Designers use CASE tools for building graphical information system models. A conventional CASE tool shows diagrams in separate windows. In fact, it is very difficult to show connections from one window to other windows which share similar design elements. Because current CASE tools lack effective support for integrating and visualizing diagrams and objects within them, we are interested in how to improve CASE tools by utilizing information visualization techniques. This means that both the hierarchical and the network structure in and between IS diagrams and the objects' use in different diagrams should be visible.

Card et al. (1999, 23) mention three characteristics for effective mapping from data to visual form: faster to interpret, conveys more distinctions, or leads to fewer errors than other mapping. There are also three basic criteria for evaluating a model: correctness, completeness, and consistency (see Section

2.2.2). By using these characteristics, the previously stated research problem can be reformulated as:

> *How to utilize visualization techniques so that graphical information system models are faster to interpret, convey more distinctions, or lead to fewer errors, omissions, and inconsistencies than in the conventional CASE tools?*

To answer that question, we need to consider cognitive aspects as well as technical aspects for information visualization. Cognitive aspects include acquiring, storing and using knowledge, i.e. attention, perception, learning, memory, and problem solving (Wærn 1989). Technical aspects concern mainly choosing appropriate visualization techniques for representing relationships between design data. Therefore, the following two research questions are asked:

> *RQ1) what cognitive and visualization aspects should be considered when building support for IS designers and reviewers?*

> *RQ2) how can visualization techniques support the user's understanding of graphical information system models? Especially, how can we improve existing CASE tools to facilitate seeing interrelationships between diagrams and design elements within them?*

These research questions have been addressed in many studies (e.g. Wærn 1989, Shneiderman 1998, and Card et al. 1999), but mainly from a general perspective and not specifically in the IS research field, as I have pointed out in the Sections 2 and 3. Here we are applying the ideas from other research fields to the IS research and we are trying to focus on an individual designer or reviewer. Although we are concentrating on an individual user, the results should be applicable to a group of designers or reviewers.

4.2 Research Methodology

This research learns and extends ideas from many disciplines and research areas. Visualization, representation, perceptualisation, navigation, and other relevant issues are studied in many fields of science, which all have their own applicable research methods.

The objective of research in information systems is "to acquire knowledge and understanding that enable the development and implementation of technology-based solutions to heretofore unsolved and important business problems" (Hevner et al. 2004, 84). Behavioral science and design science characterize much of the research in the IS discipline, which studies organizations, people, and technology (March and Smith 1995). The behavioral science paradigm has its origins in natural science research methods, while the design science has its roots in engineering and the sciences of the artificial (Simon 1969; Hevner et al. 2004). An IT artifact, e.g. an IS model or an instantiation (an implemented or a prototype system), is often the object of study in both sciences (Hevner et al. 2004). IS behavioral science research seek to develop and justify theories explaining or predicting phenomena that occur "with respect to the artifact's use (intention to use), perceived usefulness, and impact on individuals and organizations (net benefits) depending on system, service, and information quality" (DeLone and McLean 1992, 2003; Seddon 1997; Hevner et al. 2004, 77). Design science approaches the goal of research in IS "through the construction of innovative artifacts aimed at changing the phenomena that occur" (Hevner et al. 2004, 84).

Our research approach is based on the design science paradigm. We utilize the multimethodological research framework (Nunamaker et al. 1991) that has been successfully used in many previous studies, e.g. (Kelly 1997; Oinas-Kukkonen 1997a; Marttiin 1998; Rossi 1998). The main reason for choosing this framework as a research methodology is motivated by the fact that the research in information system area is still relatively young and thus constructive approach is needed. The framework consists of four strategies: observation, theory building, systems development, and experimentation. As can be seen from Figure 4, these strategies may be used in any preferred order.

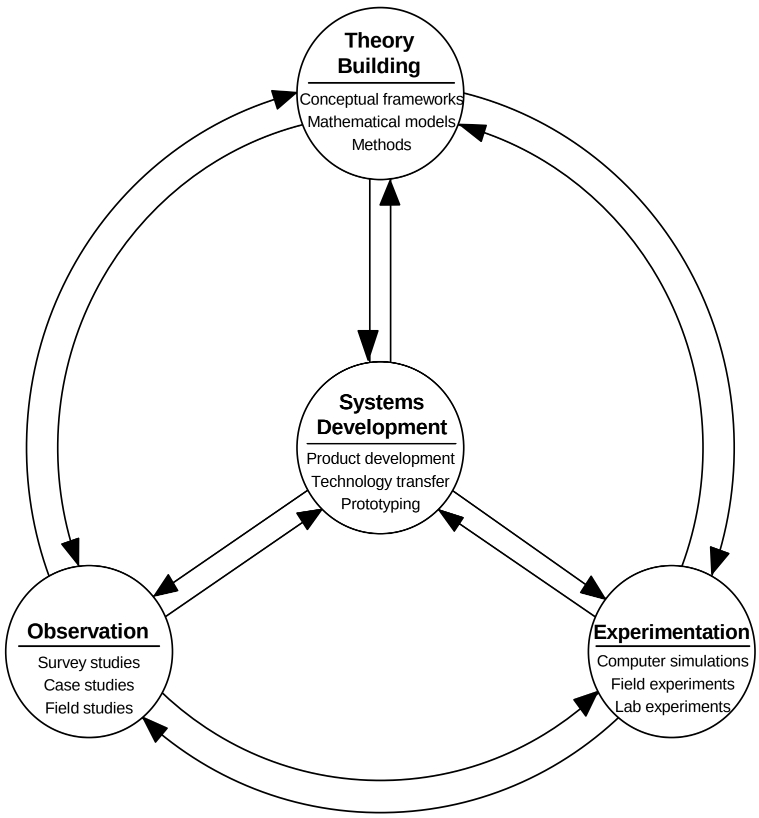

FIGURE 4 A multi-methodological approach to IS research (Nunamaker et al. 1991, 94)

Observation includes research methods such as case studies, field studies, and surveys. It is often used when relatively little is known in a research area, to help researches to formulate specific hypotheses to be tested through experimentation, or to arrive generalizations that help focus later research.

Theory building includes the development of new ideas and concepts, and the construction of conceptual frameworks, new methods, or models (e.g. mathematical models, simulation models, and data models). Theories may be used to suggest research hypotheses, guide the design of experiments, and conduct systematic observations.

Systems development consists typically of five stages: concept design, architecture construction, prototyping, product development, and technology transfer. Systems development is, fundamentally, a problem solving activity (Vessey and Glass 1998; Hevner et al. 2004).

Experimentation includes research methods such as laboratory and field experiments as well as computer simulations. Results from experimentation may be used to refine theories and improve systems.

4.3 Application of the Methodology

The research was performed as follows. We started from *observation*, where we looked at the state of the practical and research field of visualization and CASE (paper 1). Parallel to that, we observed user actions and collected user opinions about the usability of the implemented systems and potential problems faced by an information systems designer. The results from these observations and questionnaire forms are not reported fully in this thesis, but they motivated us to continue on this topic. As a part of *theory building*, we also built a framework, which was applied when visualizations were evaluated (paper 2). Based on our observations, we developed new ideas and suggested how CASE tools could benefit from advanced visualization techniques (paper 3).

As the result of the *system development*, we constructed two research prototypes (papers 4 and 6). They include some of the suggested visualization techniques. Finally, we conducted a laboratory *experiment* (paper 5) and a pilot study (paper 6). There, implementation was examined by giving specific tasks and recording the number of errors (papers 5 and 6). In addition, we collected user opinions, expressed our conclusions, and gave suggestions for the future (all papers). The relation between research questions, papers, and prior research approach is summarized in the Tables 3 and 4. Figure 5 illustrates the research process of this thesis with a timeline.

TABLE 3 The relation between research questions and papers

Question	Paper
1) What cognitive and visualization aspects should be considered when building support for IS designers and reviewers?	1, 2, and 3
2) How can visualization techniques support the user's understanding of graphical information system models? Especially, how can we improve existing CASE tools to facilitate seeing interrelationships between diagrams and design elements within them?	3, 4, 5, and 6

TABLE 4 The relation between papers and used research approach

Paper	Detailed explanation of the research approach	Aim
1	Observation (literature review: surveys, case studies, field studies, and other research methods)	Looking at the state of the HCI research
2	Theory building (evaluation framework) + observation (comparison of visualization tools)	Developing criteria for visualization tools
3	Observation (literature review); theory building	Developing new ideas for integration support
4	Systems development (first research prototype) and laboratory experiment	Developing prototype to test our ideas
5	Theory building and a laboratory experiment (with the first research prototype),	Validation of hypotheses (experiment with novices)
6	Systems development (second research prototype); observation (six interviews and video recording during a pilot study)	More insight of the usefulness of integrating diagrams (pilot study with experts)

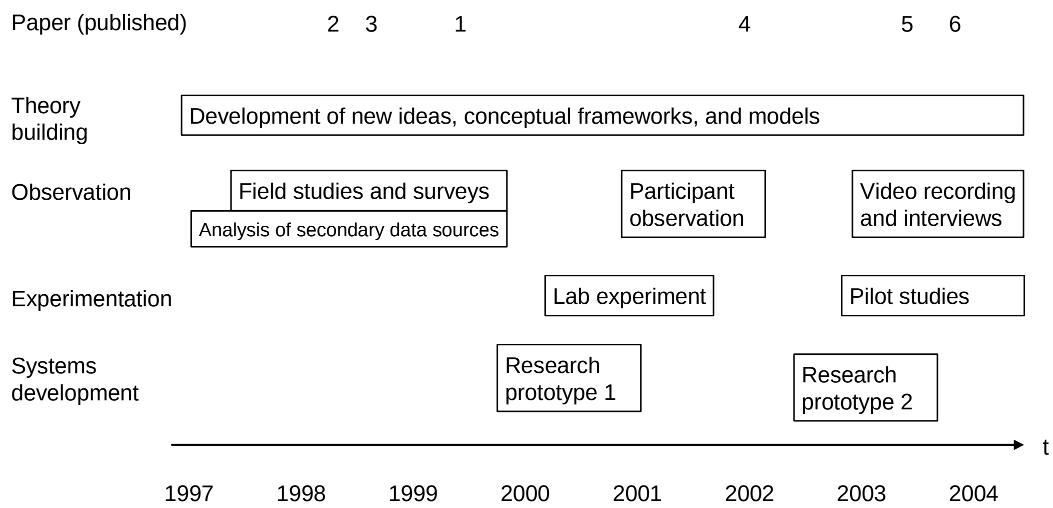

FIGURE 5 The research method, data gathering techniques, and the timeline of the research process

5 SUMMARY OF THE ARTICLES

In this Section, I shortly describe the six papers, the problems addressed, the research methodology used, and the results of each. The publication details of the papers and authors are listed for each paper, and my personal contribution of the joint-articles is reported at the end of each coauthored paper. Note that the order of the papers is logical rather than chronological.

5.1 Review of HCI Research Focus on Cognitive Aspects and Used Research Methods

Proceedings, IRIS'22 Conference, 1999.
Jouni Huotari and Janne Kaipala

5.1.1 Research problems and methodology

This paper investigates the scientific work within the field of Human-Computer Interaction (HCI) with focus on cognitive aspects. We list some of the most influential concepts, theories, methods, and techniques used in HCI research. We apply a three-level research framework proposed by Kuutti and Bannon (1993) in order to categorize and clarify the important underlying arguments in HCI. Our review analyses and synthesizes the main contributions of the field and takes a critical view for some of the influencing methods and theories related to cognition. We also summarize some theories used in reference disciplines and their implication for HCI. Moreover, we survey the research methods used in HCI literature dealing with cognition related concepts.

5.1.2 Research results

Based on the literature review, HCI is the most distinctive field in IS research that deals with cognitive aspects. Most theories used in reference disciplines such as cognitive psychology belong to the conceptual interaction level, while the emphasis is shifting more to the social / contextual interaction level.

According to the examination of the structure of two major IS literature classification systems (ACM and MISQ), the common categories related to HCI are human factors, cognitive style, human information processing, ergonomics, interface design, human/computer interaction, and graphical user interface. The results from examining the research methods used indicate that empirical work and non-empirical work are almost evenly emphasized, and that most of the empirical research has been experimental. Despite a trend of applying cognitive task analysis and other user-centered system design methods, issues of human cognition and human information processing still need more attention in the IS research.

The division of work was equal in the write-up of the paper. I was responsible for Sections 1 (Introduction) and 4 (Review of the selected literature). In addition, I collected the classification data and created an Excel macro program, which added the missing level codes and combined the classification systems. I also summarized the HCI related keywords in the Appendix.

The idea of writing this paper and paper 3 (Section 5.3) came from the long-term discussions with Janne Kaipala, with whom I worked very closely for three years. Janne is interested in the same issues as I am, i.e. HCI, cognitive aspects, and visualization. It is very difficult to point out exactly, which of the ideas is mainly conceived by me or Janne or both.

5.2 Supporting User's Understanding of Complex Information Spaces by Advanced Visualization Techniques

Proceedings, eighth Biannual Conference on Artificial Intelligence, Human and Artificial Information Processing, 1998.
Jouni Huotari

5.2.1 Research problems and methodology

In the beginning of my Ph.D. studies, I had several ideas for my research. One of those ideas was improving existing information system development tools to visualise graphical information system models so that it would be easier to understand the "big picture" and also to see how a small part relates to other parts. In order to evaluate existing tools I needed a framework, which could be used as the evaluation criteria.

This paper examines how users' cognitive capabilities and usability issues should be considered in visual representations. The focus is on supporting users' understanding of complex information spaces by advanced visualization techniques. The main data gathering technique is literature review. As a result a framework is created for evaluation of information visualization tools.

5.2.2 Research results

The first of the findings was that in recent years the computing power has increased so much that applications using advanced visualization techniques can be implemented. They support users' understanding by providing graphical overviews, interaction, and efficient search facilities.

The main contribution is the proposal of an evaluation framework which can be used as a checklist to develop a visualization tool showing relationships between abstract data effectively. Interestingly, my original list of features is very similar to Shneiderman's (1998) list of tasks that need to be supported. These lists are combined in Table 2 (Section 2.3). We used this list when defining the requirements for the research prototypes.

We used our evaluation framework to test both the framework and check the completeness of the available information visualization tools. I found that although there are several novel information visualization applications and research prototypes available, none of them meets all the suggested evaluation criteria.

5.3 Towards Advanced Visualization Techniques in CASE: Initial Findings and Suggestions

Proceedings, Seventh International Conference of Information Systems Development - ISD'98.
Janne Kaipala and Jouni Huotari

5.3.1 Research problems and methodology

We discuss how representations in CASE can be improved using advanced visualization techniques. While CASE tools allow creating different representations supporting perceptual cues, they have largely ignored the fact that a problem in a design situation can be the amount of irrelevant information. CASE tools also lack overview representations that provide a holistic or focused view on the whole design repository or part of it. Moreover, decomposition structure (hierarchy) is often difficult to see. In addition, CASE tools seldom provide traceability, and if they do, traces are not clearly visible.

We started our study by reviewing CASE and information visualization literature. In addition, we examined some CASE tools and identified problems that relate seeing relationships between hierarchical levels and different representations (diagrams, matrices, and lists).

5.3.2 Research results

We show that advances in research of visualization can be utilised in solving the above problems. We give some suggestions on how to improve existing CASE tools to facilitate seeing interrelationships between design elements in a more comprehensible manner. One of the suggestions is to link corresponding objects from different types of design objects with connecting lines. Another suggestion is to utilize distortion (fish-eye view) in order to show context while maintaining focus on a selected item. Third suggestion is to utilize graphics for showing the decomposition structure. These suggestions were considered when we designed the two research prototypes.

My contribution to the paper includes the whole of Sections 2 (Visual representations in design work) and 3 (Overview on visualization techniques). In addition, I commented the text in Section 4, and wrote the Introduction and Conclusions together with Janne Kaipala. Both of us have quite an extensive understanding of a (meta)CASE tool called MetaEdit+ and we have also examined other CASE tools. I suggested distortion and cone tree for visualising design data, and Janne created the Figure 6, which illustrates one possible solution to show a trace of a requirement.

5.4 Enhancing Graphical Information System Models with VRML

Proceedings, Sixth International Conference on Information Visualization - IV'02.
Jouni Huotari and Marketta Niemelä

5.4.1 Research problems and methodology

This paper introduces a research prototype, which was developed for a laboratory experiment (presented in more detail in paper 5). Our VRML implementation integrates different types of diagrams in one whole. Three aspects are especially emphasised. Firstly, our solution preserves structure, which can have semantic value. Secondly, it provides both focus and context in order to make it easier understand how detailed information relates to other design elements. Thirdly, it enables tracing between diagrams and other design documents. We applied elision technique for decomposition of data flow diagrams (DFD) and added visible lines to link parts of DFD to entity-relation diagrams (ERD). In our laboratory experiment, we collected users' subjective opinions and evaluated their performance in information search tasks.

5.4.2 Research results

The preliminary results from the laboratory experiment indicated differences between the conditions. The error rate of the no-context conditions was higher than of the context conditions. This suggests that context-providing methods, elision and explicit lines, helped the users to be more accurate in searching information in an IS model visualization. The detailed results are presented in the fifth paper. It is noteworthy that in the actual experiment we also had fifth condition: 3D with stereo effect (3Ds). Although we received very encouraging comments from the people that participated in our study using stereo glasses, we could not take the results from the experiment into the consideration (in paper 5) because of the technical difficulties. One interesting result concerns the number of perfect answers. An answer was defined as perfect if the subject did not make any mistakes when answering to the question. The best condition in this respect was 3Ds (2.1 perfect answers on average; the maximum was 10). The control condition, diagrams printed on white paper with black ink and attached to the paperboard, was evaluated significantly lower than the other conditions.

We also noted that it is possible to create a VRML representation from IS models without any manual work. By utilizing VRML browsers and adjusting settings for the level-of-detail (LOD), a person can dive into the model and see how details are smoothly revealed while preserving the context. In addition, we managed to keep text readable during the rotation of the model by using VRML's Billboard. We also implemented information hiding and hypertext functionality, even though these were not utilized in the laboratory experiment. This research prototype was the basis for the second research prototype where we used UML diagrams instead of ER and data-flow diagrams.

For this paper, I wrote the Sections 1, 2, and 3. Sections 4, 5, and 6 were a joint effort. I have cooperated with Marketta Niemelä for five years now. We have had many discussions and managed to find a common interest from examining the use of two and three dimensions, large screen, and visualisation techniques when representing design data to the user. I estimate that 90 % of the writing up the paper was my work. Marketta's main contribution was performing the statistical analysis. The idea of using CAVE environment and VRML was mine, but the actual VRML implementation was created together with a skilful research assistant, Lauri Koutaniemi.

5.5 Improving Graphical Information System Model Use with Elision and Connecting Lines

Journal, ACM Transactions on Computer Human Interface (TOCHI), Vol. 11 (1), 2004.
Jouni Huotari, Kalle Lyytinen, and Marketta Niemelä

5.5.1 Research problems and methodology

We conducted a literature review in order to find out critical issues in improving information integration in graphical IS models. The first issue is to decrease the cognitive "distance" caused by separate displays or windows. This can be addressed by showing diagrams side by side on a large screen. Two other issues are to integrate separate diagrams and hierarchically organized diagrams. We applied two visualization techniques, elision and connecting lines, to resolve the problem of integrating and interpreting design information from two diagram types, entity-relationship diagram (ERD) and data-flow diagram (DFD). To reduce visual clutter and improve symbol appearance, we introduced 3D to map additional semantic information into spatial configuration of ERD and DFD. In addition, the impact of visuo-spatial memory and spatial visualization ability (SVA) on search performance had to be considered. We conducted a laboratory experiment where we studied the impact of the visualization techniques used and the use of a large screen on search performance (diagram legibility) in designs. Subjects' spatial visualization ability was measured as a possible covariant.

The research setting was aimed to resemble a normal reviewing situation. We gave simple search tasks and more complex tasks that involved integrating two types of diagrams. A control group examined paper-based diagrams while the rest of the groups examined a large screen.

5.5.2 Research results

The error rate in the "Paper" condition was significantly higher than in the three large-screen conditions. Thus, hypothesis "Displaying diagrams on a large screen with colors improves search accuracy when compared to traditional paper diagrams" is supported.

The error rates were significantly smaller in the conditions with visual integration techniques (4.8% in "2D visual integration" and 6.4 % "3D visual integration") than in the "No integration" condition (9.6 %). We take this as support for hypothesis "Connecting lines and the elision with DFD and ERD improve search accuracy with large screens when compared to multiple separate diagrams on a large screen".

"2D visual integration" and "3D visual integration" conditions did not differ significantly in error rates. We did not succeed in implementing three-dimensional objects and layout on a large screen so that they would have improved the search accuracy of DF and ER diagrams when compared to two-dimensional implementation on the same medium.

Our results supported the hypothesis "The visual integration techniques improve speed and accuracy of recall of the object layout in diagrams when compared to traditional multiple separate diagrams on paper or with large screen. We also found support to the fifth hypothesis "The visual integration techniques improve the search accuracy of individuals with low SVA more than the performance of individuals with high SVA".

Concerning the examination of printed diagrams in the controlled laboratory experiment, one could argue that it is easier to read diagrams from a printed paper than from the large screen partly because a human can use natural zooming, i.e., leaning towards those objects that are of interest. Our results show that this speeds the process of finding deficiencies, but that the tradeoff is an increased error rate. In practice, people often review information system models together and therefore our research setting with a large screen is more relevant in the real world.

In summary, our results show that visualizing relationships reduces errors in search tasks. In addition, memory of location was improved when semantic organization was implemented with the elision technique. This technique effectively manages the amount of information on the screen and thus helps location learning. We also found that search accuracy correlated positively with memory performance.

The paper was of an equal contribution. In this paper, I wrote and commented parts of all sections, especially introduction and discussion. In addition, I was responsible on putting the paper together as a whole.

We worked over four years from the research design to the publication of this paper. My estimation is that I have put nearly the same number of working hours on this single paper as on the other five papers, partly due to the fact that the laboratory experiment was so laborious. Note that this paper relates to the paper 4, which is a kind of a technical description about the implementation.

5.6 Integrating UML Views with Visual Cues

Proceedings, Second International Conference on Coordinated and Multiple Views in Exploratory Visualization - CMV 2004.
Jouni Huotari

5.6.1 Research problems and methodology

I continue on introducing problem areas and potential solutions on how to visualize interrelationships between graphical IS models. In this paper, I concentrate on UML diagrams, especially on integrating use case, sequence, and class diagrams. I examine the possibilities to help a reviewer to understand how the use case and logical views (static structure and dynamic behavior) could be integrated and visually presented to the reviewer.

In order to test our ideas in practice, we developed a research prototype. It enables exploration within the model, utilizes intelligent zooming (hides and shows details automatically), and shows how diagrams and objects within them could be related to each other by using visual cues. I introduced this prototype to people from three software companies and Jyväskylä Polytechnic. I conducted a pilot study, where two subjects from two software companies and one subject from business field of study participated. In this study, I recorded the actions of the subjects to a videotape and a log file. In addition, the answering time and the results of twelve tasks were saved.

5.6.2 Research results

The preliminary results suggest that our research prototype helps understanding relationships between different types of UML diagrams. Especially the comments of the subjects and other people that were interviewed before, during, and after using and showing the research prototype were encouraging. Two teachers noted that the prototype could be used as a learning tool: a student can learn how different types of UML diagrams relate to each other. The experts from three software companies noted the benefits of seeing and reviewing diagrams in one application and work space. The point that deserves attention here is that no separate application (a CASE tool) is needed and there is no need to export the diagrams to a separate document.

Our examination of the videotape and the log files revealed that the usability of the prototype should be improved. This will be done according to the list introduced at the end of the paper. We will also revise the reviewing tasks and conduct a controlled laboratory experiment in a near future.

The idea of integrating UML views with visual cues came partly from the external reviewer of my licentiate thesis. I had the same idea myself, because UML is now the de-facto standard. My Ph.D. supervisor suggested me to concentrate more on the external validity (in the previous experiment the subjects were students, see Section 5.5.). Due to this, I asked subjects from software companies to participate in the research design. During the implementation, I was the project leader in the so-called ISVIS project (see http://sinuhe.jypoly.fi/~huojo/projects/ISVIS/ for details). The project group consisted of three persons, of whom two persons (Joonas Hemmilä and Ilkka Maasola) implemented the research prototype. Even though most of the fundamental ideas in the prototype are mine, all the persons in the project group contributed towards the actual implementation. Some of the ideas listed in the end of paper 6 were given by the representatives of the software companies.

6 CONCLUSION

Here I summarize the contributions of the thesis, discuss the limitations of the study, and introduce some directions for further research.

6.1 Contribution of the Thesis

The main contribution of this thesis is to introduce some new aspects for visualizing and integrating graphical IS models. Our general interest was directed towards CASE tools and environments, which help designers and reviewers to understand interrelationships between diagrams and thus decrease the number of errors, omissions, and inconsistencies in graphical information system models.

We started our studies by examining literature and current applications, focusing on cognitive and information visualization aspects. We found that issues of human cognition and human information processing still need more attention in the IS research (RQ1; paper 1). We also introduce evaluation criteria for applications that visualize information (RQ1; paper 2). This criterion includes features such as overview map, support for different search strategies, interaction, focus+context, visual layout and structure, natural metaphor and navigation aids, and customizability. According to our evaluation, current applications do not utilize possible visualization solutions comprehensively. We also found that current CASE tools and environments still fall short in representing design information, especially showing interrelationships between design elements (RQ2; papers 3, 4, 5, and 6). We give suggestions on how to improve CASE tools, for example, by utilizing visualization techniques to show the decomposition structure or adding traceability links from requirements to IS diagrams and even to code (RQ2; paper 3). These suggestions should help CASE tool designers and vendors to improve the tools, especially those used in the reviewing process.

Based on these observations and the evaluation framework, we started building two research prototypes as a possible solution for representing multifaceted design information (papers 4 and 6). We used these prototypes in a laboratory experiment and a pilot study and found that they really help people to see connections between different diagrams and between different levels of detail within one diagram type (RQ2, papers 4, 5, and 6).

Our finding adds to the body of knowledge also in respect of visual search. This is well documented in the dissertation of Niemelä (2003). Historically, the elision method has been found to influence task completion time (Schaffer et al. 1996). We found that visual integration techniques had a positive impact also on the search accuracy – especially among individuals with low spatial visualization ability (SVA). Our results show that separate representations can have complex relationships that should be made more explicit in order to support persons with a low SVA. This is in line with, and extends findings from, previous studies of search in hierarchical structures.

Our findings correspond with Hahn and Kim (1999). They studied the integration of information across multiple diagrams by using visual cues and contextual information with system diagrams during problem solving. Even though we used a different integration method from that of Hahn and Kim for representing the corresponding objects, we came to a similar conclusion: diagram presentation significantly influences information integration and appropriate visual cues facilitate information extraction and integration across diagrams.

We agree with the contingency theoretical view (Katz 1984), which emphasizes selecting the most appropriate actions for specific situations. Applying this view, there is no one best representation type or visualization tool for all situations, and all representation types or visualization tools are not equally good in all situations. Optimal representation is produced quickly and cost-efficiently. The external representations must fit / match with the designers' internal representation, and should result in shared understanding, thus facilitating communication between all stakeholders. Our tool helps this communication by visualizing connections between different parts of the IS model.

Our research has also some practical contributions. Based on the statistical analysis and comments of the subjects in our laboratory experiment and pilot study, our implementations help users to see interrelationships between different diagrams and between different levels of detail within one diagram type. Many novice designers (students) mentioned that they did not understand the meaning of consistency and actual idea of data flow and ER diagrams until they saw those diagrams side by side and the corresponding objects were visualized explicitly. In the pilot study with the second research prototype (UML diagrams visualized with the ISVIS tool), experienced designers evaluated the ISVIS tool to be better than a commercial CASE tool both in the usefulness and usability for reviewing graphical IS models. We thus argue that the two research prototypes help the users to better understand graphical information system models. This is at least partly due to the use of advanced visualization techniques that help finding errors, omissions, and inconsistencies. The designers in software companies and students should benefit from using the ISVIS tool for checking the models before submitting them to reviewers or teachers. Furthermore, reviewers or teachers do not necessarily need a separate commercial tool for the examination of the model delivered by the designer or student. Both tools are applicable for teaching how different diagram types are related to each other.

In general, our findings are important, because we need efficient ways to present complex interrelationships between increasing amounts of digital data. Our ideas are applicable, for example, for representing relationships between web pages, text documents, presentation slides, or organizational hierarchies and networks. We also expect the CASE tool vendors to implement some of the ideas presented in this thesis.

6.2 Limitations of this Study

We recognize many limitations in this study, including methodological and technical ones. The chosen research method did not explicitly mention qualitative research methods such as action research or interviews. Nevertheless, we interviewed representatives of two CASE tool vendors, three software companies, and two colleagues (paper 6). In addition, although thorough understanding of information visualization requires the user to perceive and create imaginary models, it was beyond our knowledge and resources to measure brain activities, eye movements, or other physiological phenomena. Therefore, some insightful knowledge that could have been gained was not obtained.

We can list many shortcomings in our research setting with the first research prototype (papers 4 and 5). First, we did not control the effect of using colors and the large screen separately. The first condition was with black and white diagrams attached to the paperboard and this was compared to colored diagrams on a large screen. Second, we did not control separately the conditions of using connecting lines and elision. Therefore, we cannot be sure which of the methods was more helpful when spotting corresponding objects from other diagrams. Based on the discussions with the subjects after the experiment, we can argue that both methods help integrating diagrams: connecting lines horizontally between different types of diagrams and elision between the levels of detail. The examination of the error rates suggests that especially elision decreased the amount of errors.

In our laboratory experiment, we were only examining novice designers. In the pilot study, the subjects were expert designers, but due to the small number of subjects no statistical conclusions can be drawn. The fact that I was present during the subjective evaluation might have affected the evaluation.

Because of the controlled laboratory experiment and the research setting in general, we do not know for certain how practical our suggestions are. In addition, our technical solution was a research prototype and as such, we could not include all the possible improvements in one application. We need to develop the prototype into a product in order to test its applicability in a real-world setting.

6.3 Directions for Further Research

In this study, we have only scratched the surface of information visualization in CASE environments. There are still many elements to be discovered in relation to human vision and how to transform abstract data into visual-spatial forms (Chen 2002). For example, some questions related to the properties and the structure of data need to be answered, among them:

- What visual characteristics (e.g. color, animation) to attach to the representation?
- What combination of visualization techniques is the most appropriate in different situations?
- How can sound, touch, or even smell be used to depict information properties and structure?
- How to automate the clustering of data, e.g. generating automatic views to information space?
- How to mix data visualization with information visualization?

Considering the research framework presented in paper 1, we have focused mainly on the conceptual interaction level. In the future, also the physical/technical interaction level, i.e. sensory, perceptional, and (eye) movement coordination studies should be addressed as well as the organizational level, where social and contextual interaction takes place. Concerning the technical support for visualization, we could utilize augmented reality more, e.g. embedding virtual information in the physical world by using see-through displays, real 3D (stereo-optic 3D, where no special glasses would be needed), or by using wearable or mobile user interfaces that would make information available anytime and anywhere, also with small screens.

We have many ideas about how to continue our research. Because we still need more understanding on how to support reviewers during IS model exploration, we will conduct a laboratory experiment also with the second research prototype. We are addressing the question: can visual integration methods improve finding related and missing objects from a graphical IS model and help reviewing the model? We are investigating to what extent 2D diagrams in 3D environment with visual integration methods improve search performance and reviewing of an IS model when compared with a traditional solution where diagrams span multiple pages. We do not want to compare the superiority of any specific visualization technique as they meet separate needs in improving the integration of UML diagrams. While hierarchical relationships (vertical dimension) could be displayed by techniques developed for representing hierarchical structure such as elision or tree view, horizontally related diagrams demand other visualization techniques such as connecting lines or brushing-and-linking.

We are also seeking software companies or research groups which are willing to participate in case studies with real design data. The ISVIS prototype is available for research purposes or it can be licensed to real reviews. I hope that my efforts seeking cooperative partners will yield a more generalized solution as the end result.

I see a future, where all information system documents and components, including all project documents and design decisions during the design process, are linked together and easily reachable. This information space is navigable, clustered (grouped) and visualized in a meaningful way, and enables seeing the development from the idea to the product. This might be called as visualization-in-the-large.

In order to automatically link documents, we need applications, which utilize both data and metadata. This metadata (e.g. id or name of the creator) is partly created automatically during the creation or modification of a document or an object in a document. The Semantic web, which is based on metadata and where links can be visualized, is one potential application area. Perhaps a better file system is also needed: adding files to folders is not the best of the solutions, especially if the document belongs to many folders or categories. ZigZag (Nelson 2004) with the ability to visualize source data and its links to documents might be a potential solution to this problem. In all, we are just taking our first steps towards comprehensive information visualization within CASE environments.

REFERENCES

Alderson, A. 1991. Meta-CASE Technology. In A. Endres and H. Weber (Eds.) Software Development Environments and CASE Technology, Proceedings of European Symposium, Königswinter, June 17-19. Berlin: Springer-Verlag, 81-91.

Andrews, K. 1995. Visualizing Cyberspace: Information Visualization in the Harmony Internet Browser. In N. Gershon and S. G. Eick (Eds.) IEEE Symposium in Information Visualization (InfoVis'95), Atlanta, Georgia. IEEE Computer Society, 97-104.

Bear, M. F., Connors, B. W. and Paradiso, M. A. 2001. Neuroscience: Exploring the Brain. Lippincott: Williams & Wilkins.

Becker, R. A. and Cleveland, W. S. 1987. Brushing Scatterplots. Technometrics 29(2), 127-142.

Bederson, B. B. and Shneiderman, B. (Eds.) 2003. The Craft of Information Visualization - Readings and Reflections. The Morgan Kaufmann Series in Interactive Technologies: Morgan Kaufmann, 432.

Bertin, J. 1983. Semiology of Graphics. Milwaukee, Wis.: University of Wisconsin Press.

Blasius, J. and Greenacre, M. 1998. Visualization of Categorical Data. London: Academic Press.

Booch, G., Jacobson, I. and Rumbaugh, J. 1998. The Unified Modeling Language User Guide. Reading Mass.: Addison-Wesley.

Booch, G., Rumbaugh, J. and Jacobson, I. 1999. The Unified Modeling Language Reference Manual. Reading Mass.: Addison-Wesley.

Brinkkemper, S. 1996. Method engineering: engineering of information systems development methods and tools. Information & Software Technology 38(6), 275-280.

Bruner, J. S. 1957. On perceptual readiness. Psychological Review 64(2), 123-152.

Bubenko, J. A. 1988. A Method Engineering Approach to Information Systems Development. In N. Prakash, C. Rolland and B. Pernici (Eds.) IFIP WG8.1 Working Conference on Information Systems Development Process, Amsterdam. Elsevier Science Publishers, 167-186.

Card, S. K. and Mackinlay, J. D. 1997. The structure of the information visualization design space. In N. Gershon (Ed.) IEEE Symposium on Information Visualization 1997, Phoenix, Arizona, USA. Computer Society Press, 92-99.

Card, S. K., Mackinlay, J. D. and Shneiderman, B. (Eds.) 1999. Readings in information visualization: using vision to think. San Francisco: Morgan Kaufmann Publishers, Inc.

Chen, C. 2000. Individual Differences in a Spatial-Semantic Virtual Environment. Journal of the American Society for Information Science 51(6), 529-542.

Chen, C. 2002. Information Visualization. Information Visualization 1(1), 1-4.

Chen, C. 2004. Information Visualization: Beyond the Horizon. New York: Springer-Verlag.

DeLone, W. H. and McLean, E. R. 1992. Information Systems Success: The Quest for the Dependent Variable. Information Systems Research 3(1), 60-95.

DeLone, W. H. and McLean, E. R. 2003. Model of Information Systems Success: A Ten-Year Update. Journal of Management Information Systems 19(4), 9-30.

Egyed, A. 1999. Integrating Architectural Views in UML. Los Angeles, Center for Software Engineering, University of Southern California. USC/CSE-99-TR-514.

Egyed, A. 2002. Automated Abstraction of Class Diagrams. ACM Transactions on Software Engineering and Methodology (TOSEM) 11(4), 449-491.

Ewusi-Mensah, K. 1997. Critical Issues in ABANDONED Information Systems Development Projects. Communications of the ACM 40(9), 74-80.

Falkenberg, E. D., Hesse, W., Lindgreen, P., Nilsson, B. E., Oei, J. L. H., Rolland, C., Stamper, R. K., Assche, F. J. M. V., Verrijn-Stuart, A. A. and Voss, K. 1998. A Framework of Information System Concepts: The FRISCO Report (Web edition). Leiden, The IFIP WG 8.1 Task Group

Fayyad, U., Grinstein, G. and Wierse, A. 2001. Information Visualization in Data Mining and Knowledge Discovery. San Francisco, CA: Morgan Kaufmann.

Feiner, S. 1988. Seeing the Forest for the Trees: Hierarchical Display of Hypertext Structure. In R. Allen (Ed.) COIS88 - Conf. on Office Information Systems, Palo Alto, California. New York, NY, USA: ACM Press, 205-212.

Fountain, A. M., Hall, W., Heath, I. and Davis, H. C. 1990. MICROCOSM: An Open Model for Hypermedia With Dynamic Linking. In A. Rizk, N. Streitz and J. Andre (Eds.) Hypertext: Concepts, Systems and Applications, INRIA, France. Cambridge University Press.

Friedhoff, R. and Kiely, T. 1990. The eye of the beholder. Computer Graphics World 13(8), 46-59.

Furnas, G. W. 1986. Generalized Fisheye Views. In M. Mantei and P. Orbeton (Eds.) CHI'86 Conference on Human Factors in Computing Systems. New York, NY, USA: ACM Press, 16-23.

Garzotto, F., Mainetti, L. and Paolini, P. 1996. Navigation in Hypermedia Applications: Modelling and Semantics. Journal of Organizational Computing and Electronic Commerce 6(3), 211-237.

Geriomenko, V. and Chen, C. (Eds.) 2003. Visualizing the Semantic Web: Xml-Based Internet & Information Visualization. New York: Springer-Verlag, 224.

Gibson, J. J. 1979. An Ecological Approach to Visual Perception. Boston: Houghton Mifflin.

Gupta, U. 2000. Information Systems: Success in the 21st Century. New Jersey: Prentice Hall.

Hahn, J. and Kim, J. 1999. Why are some diagrams easier to work with? Effects of diagrammatic representation on the cognitive integration process of system analysis and design. ACM Transactions on Computer-Human Interaction 6(3), 181-213.

Hansen, C. 2004. Visualization Handbook. Orlando, FL, USA: Academic Press.

Harlan, H. M. 2000. Method and apparatus for displaying a thought network from a thought's perspective. US, Natrificial. 6031537; 6037944.

Henderson, J. and Cooprider, J. 1994. Dimensions of IS Planning and Model Aids: A Functional Model of CASE Technology. In T. Allen and M. Scott-Morton (Eds.) IT and the Corporation of the 1990's: research Studies. New York: Oxford University Studies Press, 221-248.

Hevner, A., March, S., Park, J. and Ram, S. 2004. Design Science in Information Systems Research. MIS Quarterly 28(1), 75-105.

Hirschheim, R., Klein, H. and Lyytinen, K. 1995. Information Systems Development and Data Modeling, Conceptual and Philosophical Foundations. New York, NY, USA: Cambridge University Press.

IEEE. 1989. IEEE Guide for the Use of IEEE Standard Dictionary of Measures to Produce Reliable Software. New York, USA, The Institute of Electrical and Electronics Engineers, Inc. Std 982.2-1988.

Iivari, J. 1989. Levels of abstraction as a conceptual framework for an information system. In E. D. Falkenberg and P. Lindgreen (Eds.) Information System Concepts: An In-depth Analysis. Amsterdam: North-Holland, 323-352.

Iivari, J. 1994. Object-oriented information systems analysis: A comparison of six object-oriented analysis methods. In A. A. Verrijn-Stuart and T. W. Olle (Eds.) Methods and Associated Tools for the Information Systems Life Cycle (A-55). New York, NY, USA: Elsevier Science, 85-110.

Itoh, T., Yamaguchi, Y., Ikehata, Y. and Kajinaga, Y. 2004. Hierarchical Data Visualization Using a Fast Rectangle-Packing Algorithm. IEEE Transactions on Visualization and Computer Graphics 10(3), 302-313.

Johnson, B. and Shneiderman, B. 1991. Tree-Maps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures. In G. M. Nielson and L. Rosenblum (Eds.) IEEE Visualization '91, San Diego, California. Los Alamitos, CA, USA: IEEE Computer Society, 284-291.

Kaipala, J. 1997. Augmenting CASE Tools with Hypertext: Desired Functionality and Implementation Issues. In A. Olivé and J. A. Pastor (Eds.) Proceedings of CAiSE '97, Barcelona, Catalonia, Spain, June 16-20. Berlin: Springer-Verlag. 1250, 217-230.

Kamada, T. and Kawai, S. 1991. A General Framework for Visualizing Abstract Objects and Relations. ACM Transactions on Graphics 10(1), 1-39.

Katz, R. H. 1984. Transaction Management in the Design Environment. In G. Garderin and E. Gelenbe (Eds.) New Applications of Databases, Cambridge, UK. San Diego, CA, USA: Academic Press Professional, 259-273.

Kelly, S. 1997. Towards a Comprehensive MetaCASE and CAME Environment: Conceptual, Architectural, Functional and Usability Advances in MetaEdit+. Jyväskylä Studies in Computer Science, Economics and Statistics. Jyväskylä: University of Jyväskylä.

Kelly, S. and Rossi, M. 1997. Differences in Method Engineering Performance with Graphical and Matrix Tools: A Preliminary Empirical Study. In K. Siau, Y. Wand and J. Parsons (Eds.) Proceedings of 2nd CAiSE/IFIP8.1 International Workshop on Evaluation of Modeling Methods in Systems Analysis and Design, EMMSAD'97, June 16-17, Barcelona, Spain.

Kelly, S. and Smolander, K. 1996. Evolution and Issues in MetaCASE. Information and Software Technology 38(4), 261-266.

Kim, J., Hahn, J. and Hahn, H. 2000. How do we understand a system with (so) many diagrams? Cognitive integration processes in diagrammatic reasoning. Information Systems Research 11(3), 284-303.

Kuutti, K. and Bannon, L. J. 1993. Searching for unity among diversity: exploring the "interface" concept. Human factors in computing systems, Amsterdam, The Netherlands. New York, NY, USA: ACM Press, 263-268.

Lamping, J. and Rao, R. 1996. Visualizing Large Trees Using the Hyperbolic Browser. In M. J. Tauber (Ed.) ACM CHI '96, Vancouver, Canada. New York, NY, USA: ACM Press, 388-389.

Lohse, G. A., Biolsi, K., Walker, N. and Rueter, H. H. 1994. A Classification of Visual Representations. Communications of the ACM 37(12), 36-49.

Lyytinen, K. 1987. A Taxonomic Perspective of Information Systems Development: Theoretical Constructs and Recommendations. In R. J. Boland and R. A. Hirschheim (Eds.) Critical Issues in Information Systems Research. Chichester: John Wiley & Sons Ltd., 3-41.

Lyytinen, K., Kerola, P., Kaipala, J., Kelly, S., Lehto, J., Liu, H., Marttiin, P., Oinas-Kukkonen, H., Pirhonen, J., Rossi, M., Smolander, K., Tahvanainen, V.-P. and Tolvanen, J.-P. 1994. MetaPHOR: Metamodelling, Principles, Hypertext, Objects and Repositories. Finland, Department of Computer Science and Information Systems, University of Jyväskylä. TR-7.

Lyytinen, K., Smolander, K. and Tahvanainen, V.-P. 1989. Modelling CASE Environments in Systems Development. Proceedings of the first Nordic Conference on Advanced Systems. Stockholm: SISU.

Mackinley, J. D. 1986. Automating the Design of Graphical Presentations of Relational Information. ACM Transactions of Graphics 5(2), 110-141.

March, S. and Smith, G. 1995. Design and natural science research on information technology. Decision Support Systems 15, 251-266.

Marr, D. 1982. Vision: A computational investigation into the human representation and processing of visual information. San Francisco: W.H. Freeman.

Marttiin, P. 1998. Customisable Process Modelling Support and Tools for Design Environment. Jyväskylä Studies in Computer Science, Economics and Statistics. Jyväskylä: University of Jyväskylä.

McCormick, B. H., Defanti, T. A. and Brown, M. D. 1987. Visualization in Scientific Computing. IEEE Computer Graphics & Applications 21(6), 61-70.

McGregor, J. D. 1998. Testing Models: The Requirements Model. Journal of Object-Oriented Programming 11(3), 20-23,31.

Monarchi, D. E. and Puhr, G. I. 1992. A Research Typology for Object-Oriented Analysis and Design. Communications of the ACM 35(9), 35-47.

Nelson, T. H. 2004. A Cosmology for a Different Computer Universe: Data Model, Mechanisms, Virtual Machine and Visualization Infrastructure. The Oxford Internet Institute, Project Xanadu and Operation ZigZag.

Nielsen, J. 1993. Usability Engineering. Boston: Academic Press Professional.

Niemelä, M. 2003. Visual Search in Graphical Interfaces. Department of Information Systems. Jyväskylä: The University of Jyväskylä.

Noik, E. G. 1994. A space of presentation emphasis techniques for visualizing graphs. In W. A. Davis and B. Joe (Eds.) Graphics Interface'94, Banff, Alberta, Canada. Summerland, BC: New Horizon Printing, 225-233.

Nunamaker, J. F., Chen, M. and Purdin, T. D. M. 1991. Systems Development in Information Systems Research. Management Information Systems 7(3), 89-106.

Oinas-Kukkonen, H. 1997a. Improving the Functionality of Software Design Environments by Using Hypertext. Department of Information Processing Science. Oulu: University of Oulu.

Oinas-Kukkonen, H. 1997b. Towards Greater Flexibility in Software Design Systems through Hypermedia Functionality. Information & Software Technology 39(6), 391-397.

Parker, G., Franck, G. and Ware, C. 1998. Visualization of Large Nested Graphs in 3D: Navigation and Interaction. Journal of Visual Languages and Computing 9(9), 299-317.

Roberts, J. C. 1998. On Encouraging Multiple Views for Visualization. In E. Banissi, F. Khosrowshahi and M. Sarfraz (Eds.) IEEE International Conference on Information Visualization '98, London, UK. IEEE Computer Society Press, 8-14.

Roberts, J. C. 2000. Multiple-View and Multiform Visualization. In R. Erbacher, A. Pang, C. Wittenbrink and J. Roberts (Eds.) Visual Data Exploration and Analysis VII, San Jose, California, USA. SPIE, 176-185.

Roberts, J. C. 2001. Issues of Dataflow and View Presentation in Multiple View Visualization. CISST Annual Conference. CSREA Press, 177-183.

Roberts, J. C. 2004. CMV 2004: Call for Papers. International Conference on Coordinated & Multiple Views in Exploratory Visualization, London UK. IEEE Computer Society.

Robertson, G. G., Mackinlay, J. D. and Card, S. K. 1991. Cone trees: Animated 3D visualizations of hierarchical information. SIGCHI Conference on Human Factors in Computing Systems. ACM, 189-194.

Robertson, P. K. 1991. A Methodology for Choosing Data Representations. IEEE Computer Graphics and Applications 11(3), 56-68.

Rossi, M. 1998. Advanced Computer Support for Method Engineering: Implementation of CAME Environment in MetaEdit+. Jyväskylä Studies in Computer Science, Economics and Statistics. Jyväskylä: University of Jyväskylä.

Salampasis, M., Tait, J. and Hardy, C. 1997. HyperTree: A Structural Approach to Web Authoring. Software - Practice and Experience 27(12), 1411-1426.

Sarkar, M. and Brown, M. H. 1994. Graphical Fisheye Views. Communications of the ACM 37(12), 73-84.

Scaife, M. and Rogers, Y. 1996. External cognition: how do graphical representations work? International Journal of Human-Computer Studies 45, 185-213.

Schaffer, D., Zao, Z., Greenberg, S., Bartram, L., Dill, J., Dubs, S. and Roseman, M. 1996. Navigating hierarchically clustered networks through fisheye and full-zoom methods. ACM Transactions on Human-Computer Interaction 3(2), 162-188.

Seddon, P. B. 1997. A Respecification and Extension of the DeLone and McLean Model of IS Success. Information Systems Research 8(3), 240-253.

Selonen, P., Koskimies, K. and Sakkinen, M. 2001. How to Make Apples from Oranges in UML. HICCS 34, Hawaii. IEEE Computer Society.

Shneiderman, B. 1998. Designing The User Interface: Strategies for Effective Human-Computer Interaction. Reading, Massachusetts: Addison Wesley Longman.

Shneiderman, B. 2002. Inventing Discovery Tools: Combining Information Visualization with Data Mining. Information Visualization 1(1), 5-12.

Shu, N. C. 1988. Visual programming. New York: Van Nostrand Reinhold Company Inc.

Siau, K. 1999. Information Modeling and Method Engineering: A Psychological Perspective. Journal of Database Management 10(4), 44-50.

Simon, H. A. 1969. The Sciences of the Artificial: The M.I.T. Press.

Smolander, K., Lyytinen, K., Tahvanainen, V.-P. and Marttiin, P. 1991. MetaEdit - A Flexible Graphical Environment for Methodology Modelling. In R. Andersen, J. A. B. jr. and A. Solvberg (Eds.) Advanced Information Systems Engineering, Proceedings of the Third International Conference CAiSE'91, Trondheim, Norway, May 1991. Berlin: Springer-Verlag, 168-193.

Spence, R. 1993. A taxonomy of graphical presentation, Imperial College of Science. Information Engineering Section report 93/3.

Spence, R. 2001. Information Visualization. Harlow, UK: Pearson Education Limited.

Tolvanen, J.-P. 1998. Incremental Method Engineering with Modeling Tools: Theoretical Principles and Empirical Evidence. Jyväskylä: University of Jyväskylä.

Tufte, E. R. 1983. The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press.

Tufte, E. R. 1990. Envisioning Information. Cheshire, CT: Graphics Press.

Tufte, E. R. 1997. Visual Explanations. Cheshire, Connecticut: Graphics Press.

Turetken, O., Schuff, D., Sharda, R. and Ow, T. T. 2004. Supporting systems analysis and design through fisheye views. Communications of the ACM 47(9), 72-77.

Tweedie, L. 1997. Characterizing Interactive Externalizations. CHI'97. ACM, 375-383.

Wærn, Y. 1989. Cognitive aspects of computer supported tasks. Oxford, England: John Wiley & Sons.

Wand, Y. and Wang, R. Y. 1996. Anchoring Data Quality Dimensions in Ontological Foundations. Communications of the ACM 39(11), 86-95.

Wang, R. Y. 1998. A Product Perspective on Total Data Quality Management. Communications of the ACM 41(2), 58-65.

Wang, R. Y., Storey, V. C. and Firth, C. P. 1995. A framework for analysis of data quality research. IEEE Transactions on Knowledge Data Engineering 7(4), 623–640.

Ware, C. 2000. Information Visualization: Perception for Design. San Francisco: Morgan Kaufmann Publishers.

Ware, C. and Franck, G. 1996. Evaluating stereo and motion cues for visualizing information nets in three dimensions. ACM Transactions on Graphics 15, 121-140.

Welke, R. J. 1981. IS/DSS DBMS Support For Information Systems Development. Hamilton, Faculty of Business, McMaster Univ., Ontario, Canada. ISRAM WP-8105-1.0.

Welke, R. J. 1983. IS/DSS DBMS Support For Information Systems Development. In A. B. W. C. W. Holsapple, Reidel Publishing Company (Ed.) Database Management, 195-250.

Vessey, I. and Glass, R. 1998. STRONG Vs. Weak Approaches to Systems Development. Communications of the ACM 41(4), 99-102.

Wildbur, P. and Burke, M. 1998. Information Graphics: Innovative Solutions in Contemporary Design. London: Thames & Hudson.

Williams, J. G., Sochats, K. M. and Morse, E. 1995. Visualization. In Annual Review of Information Science and Technology (ARIST) 30, 161-207.

Yourdon, E. 1989. Modern Structured Analysis. Englewood Cliffs, NJ, USA: Prentice-Hall.

Zachman, J. A. 1987. A framework for information systems architecture. IBM System Journal 26(3), 276-292.

FINNISH SUMMARY

Elektronisessa muodossa olevan tiedon määrä kasvaa jatkuvasti. On yhä vaikeampaa ymmärtää yksittäisen tiedon suhdetta kokonaisuuteen ja muihin tietoihin. Sama ongelma esiintyy myös tietojärjestelmäkuvauksissa, joita tarvitaan lähinnä kommunikoinnin helpottamiseksi ja monimutkaisuuden vähentämiseksi. Kuvaukset ilmentävät tulevan tai nykyjärjestelmän rakennetta ja käyttäytymistä eri tasoilla. Tasoja voidaan tarkastella muun muassa yleisestä yksityiskohtiin tai vaihejaon mukaan analyysistä suunnitteluun ja toteutukseen.

Suunnittelijalle tai arvioijalle voi olla ongelmallista ymmärtää, kuinka kuvaukset liittyvät toisiinsa. Tavoitteena on ymmärtämisen lisäksi löytää mahdollisimman varhain virheet tai puutteet, joita kuvauksiin on saattanut jäädä, koska virheiden korjaaminen toteutetusta järjestelmästä on kallista. Vaikka tietokoneavusteisessa systeemityössä käytetyllä CASE-välineellä voidaan kuvauksista löytää puutteita ja virheitä, eivät nykyiset välineet tue tarpeeksi mm. semanttisten virheiden löytämistä. On esimerkiksi mahdollista, että suunnittelija on käyttänyt väärää tietoa (esimerkiksi luokkaa) tai määrittänyt jonkin yhteyden väärin. Sen vuoksi tarvitaan työkalu, joka helpottaa kuvausten välisten yhteyksien ymmärtämistä ja auttaa löytämään virheitä, puutteita ja ristiriitaisuuksia.

Tutkimuksen tavoitteena oli selvittää, voidaanko visualisointitekniikoilla auttaa graafisia tietojärjestelmäkuvauksia (erilaisia kaavioita) katselmoivia henkilöitä löytämään mahdollisia virheitä, puutteita tai ristiriitaisuuksia. Tutkimustyön aikana perehdyttiin mm. tietojärjestelmien kehittämisongelmiin, informaation visualisointiin ja kognitiivisiin tekijöihin. Työn aikana laadittua visualisointityökalujen arviointikehikkoa hyväksikäyttäen rakennettiin kaksi tutkimusprototyyppiä. Laboratorio-olosuhteissa suoritetussa kokeessa saatiin selville, että kuvausten välisten yhteyksien visualisointi auttaa virheiden löytämisessä ja kuvausten muistamisessa. Toinen tutkimusprototyyppi lukee XMI (XML Metadata Interchange) -muotoisessa kuvauksessa olevat UML (Unified Modeling Language) -peruskaaviot ja visualisoi kaavioiden ja niiden sisältämien objektien väliset yhteydet. Käyttäjä näkee esimerkiksi sen, missä kaavioissa jotakin yksittäistä luokkaa on käytetty tai onko yksittäistä käyttötapausta tarkennettu toisilla kaavioilla.

Saatuja tuloksia voidaan hyödyntää CASE-välineiden kehittämisen lisäksi opetuksessa: ohjelmiston käyttäjä oppii ymmärtämään, kuinka erityyppiset tietojärjestelmäkuvaukset liittyvät toisiinsa. Koska kuvauksia voidaan luoda muistakin kuin tietojärjestelmistä, kehitettyä tutkimusprototyyppiä ja tutkimustuloksia voidaan hyödyntää yleisesti havainnollistettaessa staattisen rakenteen ja dynaamisen käyttäytymisen välistä suhdetta: kuinka yksittäinen objekti esiintyy ja käyttäytyy eri kuvauksissa. Työssä esitetyt ratkaisut ovat eräs vaihtoehto parempien tietojärjestelmien kehittämiseksi.