









## ABSTRACT

Honkaranta, Anne

From Genres to Content Analysis. Experiences from Four Case Organizations

Jyväskylä: University of Jyväskylä, 2003, 90 p. (+ included articles)

(Jyväskylä Studies in Computing

ISSN 1456-5390; 31)

ISBN 951-39-1649-9

Finnish summary

Diss.

People in contemporary organizations are dealing with multiple information systems storing information in varying forms and units, including word processing documents, web pages, video, graphical files, email messages, and data elements in databases. The term content refers to the information in different forms enacted upon by various systems. Content analysis is needed for planning new solutions for information management in organizations. The environments may be complex, and new solutions should fit with the existing ones. The content may be scattered into multiple systems, and enacted by different means with varying technologies and media. It may even be difficult to identify and name the units of content to be managed. This dissertation aims to explore how the theory of *genres* - typified communicative actions used in recurrent situations – could be applied for content analysis. Findings of a literature review and four action research studies in four organizations are reported. In these organizations, content analysis was needed to elicit requirements for a new content management system, to specify assembly of new content from units of old content, and to migrate a part of the content to the Web. Units of content smaller than, for example, that of a document were analyzed. A mediator between concepts and content units of the user world and those of the technologies, media and systems was needed, and the theory of genres was utilized as such. The genre-based approach in content analysis supported the collaborative design of content, and encouraged people to define the contemporary and future content with their own terminology. Use of domain-oriented concepts allowed identification and analysis of the content that was vague by its features, regardless of users' limited knowledge of technologies or media. The two genre-based methods developed adopted consensus-creating workshops, wall-diagram technique, mind maps, and jointly filled questionnaire forms as collaborative techniques for content analysis. Use of these techniques was observed to help in avoiding communication breakdowns, and support users on making their requirements explicit.

Keywords: content analysis, genre theory, organizations, documents

## ACM Computing Review Categories

- H.1.0. Information Systems: Models and Principles: General
- I.7.1 Computing Methodologies: Document and Text Processing, Document and Text Editing  
*Content analysis*  
*Requirements analysis*
- K.6.1 Management of Computing and Information Systems: Project and People Management  
*Systems analysis*

**Author's address** Anne Honkaranta  
University of Jyväskylä  
Dept. of Computer Science and Information Systems  
P.O. BOX 35 (Ag C 514.2), 40014 Jyväskylän yliopisto  
Finland. E-mail: anne.honkaranta@cc.jyu.fi

**Supervisors** Professor, Ph.D., Airi Salminen  
Dept. of Computer Science and Information Systems  
University of Jyväskylä  
E-mail: airi@cs.jyu.fi

Professor, Dr., Pasi Tyrväinen  
Dept. of Computer Science and Information Systems  
University of Jyväskylä  
E-mail: pasi.tyrvainen@cc.jyu.fi

**Reviewers** Associate Professor, Dr., Kevin Crowston  
School of Information Studies  
Centre for Science and Technology  
Syracuse University, NY, USA

Assistant Professor, Ph.D., Lisa Murphy  
Information, Statistics, & Management Science  
Commerce & Business Administration  
The University of Alabama  
Tuscaloosa, AL, USA

**Opponent** Assistant Professor, Ph.D., Ulrike Schultze  
Dept. of Information systems & Operations Management  
Cox School of Business  
Southern Methodist University  
Dallas, TX, USA

## ACKNOWLEDGEMENTS

During the years this research was carried out, I was lucky to be supported by so many people. My husband Jorma has been my mentor, and a dear friend. I've been guided by two "hairy psychologists", Donna and Elvis. They have taught me the most important things in life: to appreciate the scent of flowers and to enjoy the sun. My parents, Maija and Kaarlo, provided support in many different ways.

My supervisors, Professor Airi Salminen and Professor Pasi Tyrväinen have provided invaluable support, feedback, and ideas. They patiently put up with my ambiguous and fuzzy texts, and helped me forward step by step. The reviewers of the thesis, Ph.D. Lisa Murphy (University of Alabama) and Dr. Kevin Crowston (Syracuse University) have provided insightful comments and suggestions.

My colleagues in the Information Technology Research Institute, and in the METODI project, especially Tero Päivärinta and Tuomo Peltola, introduced me to the theory of genres. Our enthusiasm grew as we got together and took our first steps in genre-based research. Ever since, these colleagues of mine have provided guidance, support, and insights, as well as collaborated with me in later studies. My other colleagues in the Department of Computer Science and Information Systems, and in the inSGML project, especially Virpi Lytikäinen and Pasi Tiitinen have helped me with my writing, and been fantastic companions in the project work. In addition, Riitta Kuisma, Marketta Niemelä, Hannakaisa Isomäki, Anja Mursu, Nazmun Nahar, and Kikka Kärkkäinen have provided support during the research and created enjoyable, relaxing moments outside the office.

I am in great debt to the numerous people in collaborative organizations that have made this research possible. I have been able to learn from them, and with them. I am especially grateful to Olavi Isola, Seppo Pakkanen, Anne Leinonen, Hannu Similä, Reijo Heinämäki, Ari Ekblom, Kari Kotirinta, Riikka Helin, Riitta Ahovaara and Vesa Pylvänäinen for their enjoyable collaboration.

Finnish Technology Development Centre has funded a large part of this research, together with the participant companies. Jyväskylä Graduate School in Computing and Mathematical Sciences has provided funding which enabled me to report this study. INFWEST.IT programme, as well as the Faculty of Computer Science and Information Systems supported the research by providing financial support for attending conferences, obtaining literature sources, and for language checking. Steve Legrand has checked the language of most of the articles that are included in this dissertation. Remaining errors and clumsy sentences are all due to my personal limitations.

*Anne Honkaranta*  
Jyväskylä 24.11.2003

## CONTENTS OF THE INTRODUCTION

1	DOCUMENT AND CONTENT MANAGEMENT IN ORGANIZATIONS	13
1.1	Document Management Research	14
1.2	From Documents to Content Management	17
2	THEORY OF GENRES AND RELATED RESEARCH	20
2.1	Genres of Organizational Communication	20
2.2	Genre-based Research	23
2.3	What is the "Genre Lens"?	27
3	RESEARCH OBJECTIVES AND METHODS	30
3.1	Research Objectives	32
3.2	Research Approach and the Methods	33
3.3	Research Process and Relationship to the Included Articles	35
4	OVERVIEW OF THE INCLUDED ARTICLES	39
4.1	Article 1: "Genre-Based Metadata For Enterprise Document Management"	39
4.2	Article 2: "Possibilities and Constraints for Managing and Reusing Information Content of Structured Documents: The Case of Operation and Maintenance Manuals"	41
4.3	Article 3: "Bridging The Gap Between Hard And Soft Information Genres"	42
4.4	Article 4: "Developing Document And Content Management In Enterprises Using A "Genre Lens"	44
4.5	Article 5: "Evaluating A 'Genre Lens' For Analysing Requirements For Content Assembly"	45
4.6	Article 6: "Designing Training In Manufacturing Organizations Using The Genre-Based Method"	47
4.7	Article 7: "Operationalizing A Genre-Based Method For Content Analysis: A Case Of A Church"	48
4.8	About the Joint Articles	49
5	THE CASE OF FCP	51
5.1	The Finnish Centre for Pensions	51
5.2	The Content Management Development Initiative	52
5.3	The Content Analysis Process	54
5.4	Evaluation	58
5.5	Preliminary Findings and Implications	60



6	RESULTS, IMPLICATIONS AND SHORTCOMINGS .....	63
6.1	Research Themes Revisited .....	63
6.2	The Genre-based Methods .....	69
6.3	Implications.....	71
6.4	Shortcomings of the Study.....	73
7	CONCLUSIONS AND AVENUES FOR FURTHER RESEARCH.....	76
7.1	Concluding Summary .....	76
7.2	Avenues for Further Research.....	78
	REFERENCES .....	80
	FINNISH SUMMARY .....	90

## LIST OF INCLUDED ARTICLES

- 1 Karjalainen, A., Päivärinta, T., Tyrväinen, P. & Rajala, J. 2000. Genre-Based Metadata for Enterprise Document Management. In R. H. Sprague (Ed.) Proceedings of the 33rd Annual Hawaii International Conference on System Sciences (HICSS), 4-7 January, Maui, Hawaii, U.S.A. Los Alamitos CA: IEEE Computer Society. CD-ROM (\*)
- 2 Honkaranta, A. & Tyrväinen, P. 2001. Possibilities and Constraints for Managing and Reusing Information Content of Structured Documents: The Case of Operation and Maintenance Manuals. In A. Mørch, A. L. Opdahl, B. Solveig & R. Moe, E. (Eds.). Proceedings of Information Systems Research Seminar in Scandinavia (IRIS) 24 Conference, 11-14 Aug., Ulvik in Hardanger, Norway. Bergen: University of Bergen, Department of Information Science, 381-396.
- 3 Karjalainen, A. & Salminen, A. 2000. Bridging the Gap between Hard and Soft Information Genres. In M. Khosrowpour (Ed.) Challenges of Information Technology Management in the 21st Century. Proceedings of 2000 Information Resources Management Association International Conference, 21-24 May, Anchorage, Alaska, U.S.A. Hershey: Idea Group Publishing, 92-95. (\*)
- 4 Honkaranta, A. 2003. Developing Document and Content Management in Enterprises Using a "Genre Lens". In O. Camp, J. Filipe, S. Hammoudi & M. Piattini (Eds.). Proceedings of the 5th International Conference on Enterprise Information Systems (ICEIS), 23-26 April, Angers, France. Setúbal, Portugal: Escola Superior de Tecnologia do Instituto Politécnico de Setúbal, 334-340.
- 5 Honkaranta, A. 2003. Evaluating a "Genre Lens" for Analysing Requirements for Content Assembly. In K. Siau, J. Krogstie & T. Halpin (Eds.). The 8th Caise/IFIP 8.1 International Workshop on Evaluation of Modeling Methods in Systems Analysis and Design (Emmsad03), 16-17 June, Velden/Klagenfurt, Austria. Velden/Klagenfurt, Austria: IFIP 8.1 WG, 11p.
- 6 Honkaranta, A. & Tyrväinen, P. 2003. Designing Training in Manufacturing Organizations Using the Genre-Based Method. Journal of Educational Technology and Society (JETS). Special Issue on Digital Contents for Educations, edited by Aedo, I. and Landoni, M. 6 (4), 73-85.

List of included articles continues on the following page

List of included articles, continued from the previous page

- 7 Honkaranta, A. & Lyytikäinen, V. 2003. Operationalizing a Genre-Based Method for Content Analysis: A Case of a Church. In W. Abramowicz & G. Klein (Eds.). *The Proceedings of the 6th International Conference on Business Information Systems*, 4-6 June, Colorado Springs, U.S.A. Poznan, Poland: Department of Management Information Systems at Poznan University of Economics, 108-116.

(\*) Articles 1 and 3 were written under the author's maiden name, Karjalainen.

# INTRODUCTION

# 1 DOCUMENT AND CONTENT MANAGEMENT IN ORGANIZATIONS

*“Please, tell me what, by and large, is a document to be managed?” (a manager of a large, interorganizational electronic document management project; Tyrväinen & Päivärinta, 1999, p. 4.)*

Content management is focused on managing a variety of content, such as documents or Web sites. This dissertation is focused on content management in organizations, and especially on the analysis phase of content management development. A great deal of the content to be managed is comprehended and enacted as documents. Hence document management covers a wide spectrum of the content management field. In this section, document management in organizations and related research is introduced, and reasons for favoring the term content management over the term document management are presented.

Document management has come under increasing attention in the field of information systems during the past decades. A *document* is a unit of content, which can be identified and handled as a unit, and understood as information pertaining to a topic (Salminen, 2000; Meier & Sprague, 1996). Documents are flexibly structured for human consumption (Murphy, 2001) and can be considered as units of communication meant and organized primarily for humans. It has been estimated that in a typical organization up to 80-90 % of recorded information that is not designed to be managed in databases, can be enacted as documents (Sprague, 1995), and that information workers spend a quarter of their work time in distributing, filing, and retrieving documents (Gordon, 1997). Data that can be managed by database management systems occupies a much smaller part of organizations' needs than does textual content (*ibid.*). According to Gartner Group as much as 90 % of a corporation's information is contained in its documents (Blair, 2002a).

Documents play a remarkably important role in organizations' business activities. Business contracts can act as legal evidence. Validity, consistency and long-term access are particularly important requirements for business docu-

ments. For example, technical documents contain information that is essential for using and maintaining machines that can be very large and complex, and used for several decades. The failure to locate a relevant document in time, or an error in its content can lead to dramatic effects in an organization. Blair (2002) has provided four examples of such occasions: 1) A major utility company that had lost repair instructions was required to shut down four nuclear reactors, at a loss of \$ 2m per day, 2) The US Department of Defense has estimated that missing or inaccurate technical information has caused about half of all military accidents, 3) A major airline was fined \$ 10k per take-off because of out-of-date maintenance information, and 4) the entire R&D investment of a drug company was lost because it failed to provide timely documentation.

Documents play sometimes inherent but important roles in organizational work settings. According to Brown and Duguid (1996) documents do not only convey but actually create meaning for communities using them. They can be used for patrolling and controlling the borderline of a community. Documents can act as a glue, which ties the members of the community together, strengthening the feeling of “belonging together”, yet create boundaries, which prevent entrance for outsiders (Brown & Duguid, 1996). Documents can also take the role of a mediator, convey and negotiate meanings and transfer knowledge between different communities (Wenger, 1998; Star & Greismer, 1989). Roles of documents in communities of practice can be studied in three dimensions (Murphy, 2001). Firstly, documents act as objects of practice, bringing in the information needed and structuring the work as well as information. Secondly, they act as reifications of practice by making the practical work visible and clarifying the boundaries between the outside world and that of the members. Thirdly, they work as boundary objects, i.e., as inanimate brokers between the community and the other communities in the surrounding world. By using these three dimensions of documents, the work and work practices within communities of practice can be scrutinized. The importance of documents for an organization can hardly be overemphasized.

## 1.1 Document Management Research

Documents are involved in numerous information systems research and application areas. For example, office automation, computer-supported collaborative work, executive information systems, enterprise resource planning, electronic publishing and e-Business involve processing of digital documents. Library and Information Science (e.g., Blair, 2002b; Hertzum & Pejtersen, 2000; Vakkari, 1999; Savolainen, 1999; 1993; Dervin, 1992; Ingwersen, 1992) as a well-established discipline provides findings for document management research. History and background for document management research has been introduced in Päivärinta (2001).

According to Sprague (1995) *document management* involves the creation, organization, use, manipulation, and disposition of documents for organizational

purposes. In *electronic document management* these activities are performed by utilizing information technologies and software applications. Research on document management in organizations can have various types of focuses, such as standardization (Salminen, 2003a; 2000), metadata (Murphy, 1998), requirements elicitation (Lyytikäinen, 2003) or information retrieval (Blair, 2002; Salminen, 2000; Gordon, 1997), to mention a few. Research on *structured documents*, such as documents in SGML (Standard Generalized Markup Language; Goldfarb, 1990) and XML (Extensible Markup Language; Bray, Paoli, Sperberg-McQueen et al., 2000) formats focus on the exploitation and use of the logical units of content within documents. In SGML and XML, a logical unit of content is indicated in document instances by markup. The structure of units of content is described as a hierarchy of named elements. Further information can be attached to elements by attributes. A *document type definition* (DTD) defines the markup vocabulary and the structure for a class of SGML or XML documents. A class of XML documents can also be defined by an XML schema (Fallside, 2001), even though XML does not require a DTD or XML schema to be used. In contrast to SGML and XML, the HyperText Markup Language HTML (Raggett, Le Hors & Jacobs, 1999), widely used for Web publishing, uses a pre-defined vocabulary, prescribed by a SGML DTD. The explicit structure and the definition capabilities as characteristic features of structured documents offer special areas for research, for example, document transformations (Lindén, 1997), document assembly (Heikkinen, 2000), document design (Fallside, 2001; Maler & ElAndaloussi, 1996) and SGML/XML standardization (Salminen, 2003a; 2000).

The selection of document management systems available is wide. For example, Doculabs (1998) compared the features of 16 document management systems available, and Ovum (Staunton, 1998) listed 26 document management systems providers. Document management in a system can be supported by a wide range of technologies and offer a variety of functions, for example, optical scanning, character recognition and forms processing, folder-file hierarchies and views, COLD (computer-output-to-laser-disk) technologies, document profiling, versioning, workflow support, user group views, compound documents, Web/Intranet publishing, and various search and retrieval capabilities. Most of document management systems are purchased by departments of corporations (Staunton, Tougard, Richardson et al., 1998). These may have limited expertise and resources for system design and implementation. Many organizations may choose to implement a document management system off-the-shelf, or to combine document management system components available instead of designing and implementing one. Comparing systems can be complicated and can require a thorough examination of technologies and functionalities provided by individual systems (Karjalainen, 1999b). Requirements elicitation may also be intricate. The users' assumptions about what is required may be based on limited knowledge about systems and technologies. As a consequence new innovations, such as models and ways of work, might not come to light (Koulopoulos & Frappaolo, 1995).

In organizations document management is not an issue of technologies, but rather a tool for supporting business processes and tasks where humans act

as stakeholders in different roles and groups (e.g., Murphy, 2001; Salminen, 2000; Päivärinta, 1999; Gordon, 1997; Sprague, 1995). Although a great deal of the research literature about document management concerns tools and technologies rather than documents in their organizational context (Murphy, 2001), a multitude of frameworks, approaches and methods for document management in organizations has been introduced. Examples of these are the method of Sutton (1996), the Stair Step method (Koulopoulos & Frappaolo, 1995), the RASKE methodology (see, e.g., Salminen, 2003a; Salminen, 2000; Salminen, Lyytikäinen & Tiitinen, 2000; Tiitinen, Lyytikäinen, Päivärinta et al., 2000; Salminen, Kauppinen & Lehtovaara, 1997), the genre-based approach and method (Päivärinta, 2001) and the method of Barry (1993) for requirements definition. Päivärinta (2001) has compared the methods for document management and considers the RASKE methodology as the most comprehensive representative of the methods, with elements necessary for organizational document management as well. Conventional information systems design methods, such as Yourdon's Modern Structured Analysis can be adapted for document management development as well (Päivärinta, 2001).

*Document analysis* may be considered as an activity in which the documents in the target area are identified and scrutinized by their properties. Yet the range of analysis can be very different. The RASKE methodology (Salminen, 2003a; 2000) and the method of Maler and ElAndaloussi (1996) are both intended for document analysis, however, the RASKE methodology considers not only the document structures, but also the organizational context of documents. In the methodology, interviews of domain experts and users, meetings, literal sources, and example documents are used as means for data gathering. For analysing as well as for designing the content and logical units within a class of similar documents, i.e., for a document type, the methodology utilizes the Maler and ElAndaloussi (1996) method. In this method, tables and graphical Elm models are used for visualizing schemas.

*Metadata* can be characterized as "information about information" (Saarela, 1999, p.14), "information associated with a piece of data" (Curtis, Foster & Stentiford, 1999, p. 2), or as description of information objects, which is targeted for locating, evaluating, and accessing appropriate sets of objects (Murphy, 1998). For document management in organizations, the role of metadata is essential (Murphy, 1998). Metadata associated with a document in an organization provide information, for example, about the content of the document and about the context where the document has been created, about the format of the document, or about the uses of the document. There are both particular metadata schemas, such as Dublin Core (1999), and more general models and schema definition languages such as RDF and RDF schema (Lassila & Swick, 1999; Saarela, 1999), that may be adopted or adapted in an organization.

Term metadata may also be used to refer to data that is concerned with the behaviour of data needed in organizations (Tozer, 1999). This dissertation is concerned mainly with two kinds of metadata: metadata about objects – such as documents - and metadata about organizational context. Metadata about organizational context may describe, for example, people and their roles in an organi-



zation and processes and activities related to documents in an organization. The metadata about organizational context may – but does not have to - be used for defining workflows, access rights, or user groups for a document management system implementation. It may also be used for understanding organizational aspects of document use, albeit not stored as such in a system. The metadata about the documents is focused on aiding people to search for, categorize, and provide information about a document instance or a set of document instances in a system.

## 1.2 From Documents to Content Management

The units of content meant and structured primarily for humans have so far been enacted and recognized mainly as documents. A concept of “a document” has thus provided a profound conceptual base for recognizing, naming, and analyzing these units of content in organizations. Changes in media and technologies used for enacting the content have blurred this concept. As early as 1996, Schamber (1996) pointed out that the archetypal concept of a “document” as a “book” or as a “paper print” is insufficient to deal with a multitude of media and diversity of electronic formats such as journals and web content. The traditional concept of a document ignores many characteristics of electronic formats and how people use them. Furthermore, the characteristics of documents have been bound to technologies: perceptions of traditional documents have been shaped by print technologies, and perceptions of electronic documents are being shaped by electronic technologies (ibid.).

A study revealed that in real-life workplace settings it has become increasingly difficult to decide whether some content unit is a document or should be considered as such (Tyrväinen & Päivärinta, 1999). A possible consequence is that content not considered or recognized as a document might not be identified nor included in the analysis and development of content management. According to Boiko (2002), the obstacles caused by the blurred nature of the concept of a document can be avoided by a shift of emphasis in content analysis from documents to units such as a block or a component (Boiko, 2002). However, these concepts may seem equally vague to a community of users.

*Content management* has been characterized as a variety of tools and methods used for collecting, processing and delivering content of diverse types (McIntosh, 2000). Content management development efforts may be focused on Web site management (Boiko, 2002), document management, or integration of diverging content sources (Phillips & Greenfield, 2002). Albeit that content management can consider use of databases (ibid.), the main focus still resides on the content primarily meant for human consumption, as it does in document management. One can say that document management is content management, while content management also concerns the content that might not be considered, recognized or manipulated as documents due to the use of variety of technologies and media. The role of metadata relating to a unit of content is

equally important for content management as well as for document management in an organization. The kinds of metadata about a document may be applied to content units, too.

In this dissertation, *content analysis* is regarded as an activity within content management in organizations. In the content management development activities, analysis precedes design, implementation and maintenance of content management solutions. Content analysis may, for example, involve identification of the content units, analysis of the contemporary and future manipulation of the units, and elicitation of requirements for content management in the target domain. Detailed and technical design of future target systems takes place in a design phase following the content analysis. In this dissertation the term content analysis does not refer to methods for analyzing research data, such as the content analysis method of Krippendorff (1980) or the MCD method of Silverman (1993).

Boiko (2002) and McIntosh (2000) have introduced methods for content management. According to Boiko, content management is intended for managing Web publications, and for being able to reuse the content of Web pages across them. Boiko's viewpoint is that document management is focused on managing files, whereas content management is focused on managing *components*, i.e., units of content. Components and related metadata allow workflow management and creation of *aggregations*, which are characterized as collections of existing content, such as documents or parts of them. A *template* defines the components, scripts, and static units of content (such as company logos or navigation toolbars) that are combined when Web content is created. The content management method by McIntosh (2000) focuses on managing a mixture of documents and content stored in databases, as an assembly of existing content. His method is based on adopting the Rational Unified Process® for content management. The concepts of aggregation and assembly are akin to each other. The concept of aggregation seems to be favored on discussions about the collections of content on the Web (Boiko, 2002). The concept of *content assembly* (e.g., Heikkinen, 2000; McIntosh, 2000) may be preferred about a process, in which the content results from a collection of various types of content from various sources. The assembly process may be more complex than that of an aggregation.

It seems that document management systems are on their transition toward content management systems (Medina, Meyers, Bragg et al., 2002; West, Brook & Klima, 2002). It might be possible to employ and enhance the methods developed for document management for content management, too. On the other hand, at least some of the document and content management methods utilize conventional information systems analysis and design methods (e.g., the RASKE methodology, the method of McIntosh). Hence, conventional information systems development methods may be adoptable for content management, too.

Content developers may choose between the method of Boiko (2002) or McIntosh (2000) Rational Unified Process®, or enhance document management or conventional information systems analysis methods for content management. These may be operationalized for a certain kind of content analysis, but

may not cover a variety of content analysis needs. For example, the method of Boiko is only concerned with Web publishing. Neither the method of Boiko nor the method of McIntosh covers requirements analysis thoroughly. Conventional information systems analysis methods may prove inappropriate for the various types of content to be analyzed. For example, Overmyer (2000) has pointed out that requirements engineering for Web sites is different from that of conventional interactive applications. In the same way the requirements for content analysis may differ from the requirements for information systems analysis. The objectives for the requirements and content analysis can also depart from the requirements for information systems analysis. In content analysis, the objective may be on eliciting requirements and analyzing the units of content for a system to be implemented off-the-shelf rather than designing and implementing one from the start. The methods developed for document management may not be suitable for environments where the content is more vague. For example, it may be questioned whether the methods developed for document analysis may be utilized in a kind of an organization, in which Tyrväinen and Päivärinta (1999) encountered difficulties while trying to define what are the documents to be managed.

Possible vagueness of the content is not the sole factor that may hinder the analysis of the content. Content analysis researchers have reported that content analysis and requirements elicitation may also be obfuscated by communicational elements. For example, Byrd, Cossick and Zmud (1992) found that due to the lack of common background and language between end users and systems analysts, communication breakdowns can occur. Users may have difficulties in relating themselves to their work on an abstract and conceptual level; thus highly personal requirements are expressed, with ambiguous terms (Mumford, 2000). Common information system analysis methods, as well as those for document analysis, utilize individual user interviews as the prominent mode of eliciting information from the users. Enhancements or new techniques may be needed for overcoming the multiple communicational limitations on requirements elicitation.

New tools for content analysis and management have appeared. After Yates and Orlikowski (1994; Orlikowski & Yates, 1992) suggested using the theory of genres as an analytical tool or as a "lens" for studying organizational communication, it has been successfully applied for many purposes, including document and content management (Honkaranta, 2003). This dissertation adopts the direction provided by the theory of genres. The following section introduces the theory of genres and related research.

## 2 THEORY OF GENRES AND RELATED RESEARCH

*“The genres through which information is shaped and shared for particular purposes (reports, spreadsheets, meetings, or teleconferences) are no longer merely an aspect of organizational work; rather, they are the organizational work.” (Orlikowski & Yates, 1994, p. 572).*

Genres can be thought of as prototypical models for communication (Swales, 1999). Genres are not bounded by technologies or media used. The concept of a genre operates on a different level from that of a document or a unit of content. It provides a human-crafted schema that can be instantiated as a document or as any other type of content unit. Hence the theory of genres can be applied for studying organizational communication regardless of technologies and media used (cf. Yates & Orlikowski, 1992; Article 4). Due to its emphasis on organizational communication the theory of genres may have potential for preventing language and communication breakdowns. This Section introduces the theory of genres from the viewpoint of genres of organizational communication. An account of previous research that provided implications and drove the interpretations about applying the theory of genres in this study is presented. Both of the former were combined as the “Genre Lens” that gradually surfaced along the advancement of this study.

### 2.1 Genres of Organizational Communication

Even if the theory of genres were unfamiliar to many, the genres are used and recognized in everyday life. The difference between an academic paper and an operation manual is evident for most of the people. From the perspective of the genre theory, academics writing academic papers and technicians using manuals form separate communities of discourse. A *community of discourse* (Swales, 1999) can be juxtaposed with a more common term “specific interest group” (ibid.). Genres are the assets used by the communities of discourse for charac-

terizing and classifying their prototypical forms of communication. A community of discourse can develop and use a language of its own. Use of acronyms, tone of speech, and adaptation of community-specific terminology can tie the members together, enforcing the communicative power possessed by them as a community (Swales, 1999). On the other hand, the specialized language and other carefully crafted features of community's genres can provide a barrier against outsiders (*ibid.*). For example, a genre of an academic paper may use acronyms and jargon, which may only be fully understood by the researchers in that particular field of research.

In an organization, there may be one or several communities of discourse, which enact genres of organizational communication - such as memos, budgets, or manuals. Yates and Orlikowski (1992) have characterized a *genre* of organizational communication as a typified communicative action that is acted as a response to a recurrent communicative situation. Figure 1 visualizes the key concepts related to the genres of organizational communication.

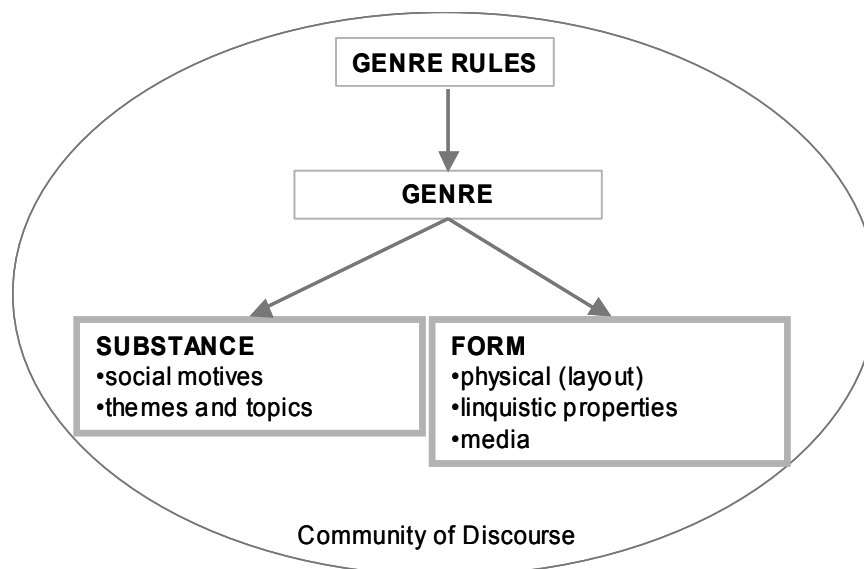


FIGURE 1 Key Concepts Related to Genres of Organizational Communication

The key concepts of genres of organizational communication are illustrated with rectangular shapes, whereas the arrows depict a chain of influence between the concepts in the figure. A recurrent communicative situation in an organization may be akin to a work task or work-related activity, such as a meeting. For example, a secretary takes part in the meeting, and identifies the need to produce a memo of it. The identification of the typified communicative situation, and production of a memo genre is governed by *genre rules* (Miller, 1994). The genre rules enforce the adoption of an appropriate genre and apply appropriate features of substance and form with regard to the communicative situation. In the last phase of the communicative action, the secretary may produce a memo document, print it, and distribute copies of it to the attendees of the meeting.

According to Yates and Orlikowski (1992), a genre has similar features of substance and form. *Substance* of a genre refers to social motives, such as the purpose of the communication. For example, the purpose of a memo may be to record the decisions made. However, the communicative purpose of a genre may not always be easily recognizable, and may become visible only when the producers and users of genres are studied (Askehave & Swales, 2001; Askehave, 1999). Substance of a genre is organized according to some kind of internal schema. It may consist of themes and topics (e.g., Swales, 1999; Berkenkotter & Huckin, 1995; Bazerman, 1994) or of a set of rhetorical moves (Bhatia, 1993). In this dissertation, *topic* is characterized as a name of a unit of substantial content within a genre. *Form* of a genre refers to the physical and linguistic features that can be observed. Layout features, language (e.g., formal way of speaking, technical or legal jargon), and media used can constitute a form for a genre (Yates & Orlikowski, 1992).

People tend to rely on their existing knowledge on genres. Thus when a community of discourse is born, members bring their genre knowledge with them and start using genres familiar to them (Swales, 1999; Orlikowski & Yates, 1994; Miller, 1994). With time, the *repertoire* of genres that are habitually used becomes established for a community (Swales, 1999; Orlikowski & Yates, 1994; Miller, 1994). Hierarchical and other relationships between genres can be observed, and a genre as a concept can be extended to multiple directions for studying the *hierarchy* of communicational content constituted (Crowston & Williams, 1997). Figure 2 visualizes genre-based concepts with regard to a hierarchy of communicational content in a community of discourse.

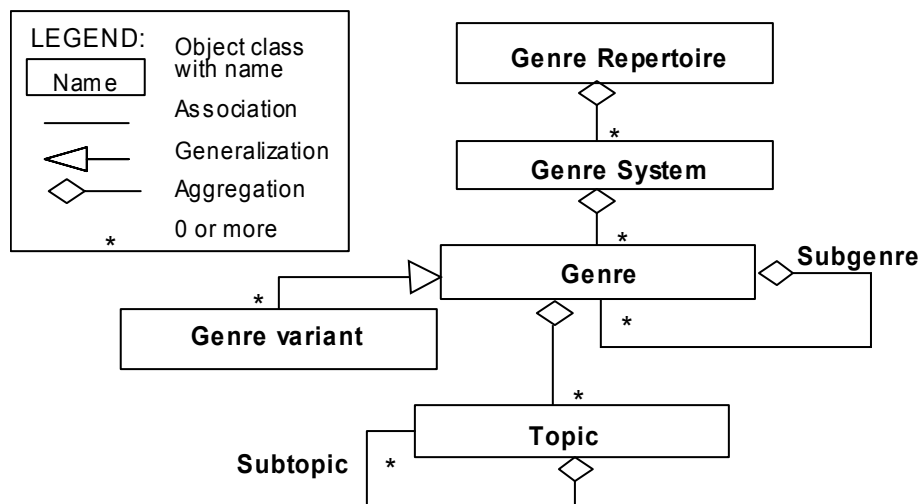


FIGURE 2 The Hierarchy of Genre Concepts

Figure 2 uses Unified Modelling Language class diagram (UML-1.1, 1997) notation. In the figure, the concepts of a genre hierarchy are visualized as rectangles, and reified as object classes. The figure considers only one community of dis-

course; hence there is only one repertoire of genres visualized. Genres that are used in tandem, such as an agenda and a memo, constitute a *genre system* (Bazerman, 1994). As illustrated in the figure, the genre repertoire may consist of multiple genre systems, or of none at all. A genre may be a part of a genre repertoire either directly or indirectly; as a member of a genre system. Genres that have both similar and different features with respect to each other are considered as *genre variants* (Crowston & Williams, 1997). In this dissertation, a genre the substance of which is similar to a topical content of another is considered as a *subgenre* of the broader one. As illustrated in the figure, a genre may consist of a number of topics. A topic may belong to several genres.

As an example of genre variants and topics, consider a research and development group, which uses two kinds of agendas and memos. Agendas and memos used within the group may contain informal language and acronyms related to products developed, and their layout may be delimited by the fact that genre instances are delivered via e-mail as a text in the e-mail body section. An agenda may consist topics such as “date”, “time”, “place” and “list of issues”. A memo may cover the same topics, and “decisions” and “announcements” in addition. The group has to inform the product manager about their development each month. Hence once a month they use variants of their agenda and memo genres. The language used is more formal. These agendas and memos are produced using a word-processing program, and a company’s document template. There is an additional topic “review of product development” in an agenda and in a memo. Copies of more formal agendas and memos are delivered to the product manager for informing her/him about their development.

The theory of genres provides the means for scrutinizing genres as prototypical models of communicational units of content. The use of the hierarchy of genre concepts provides means for expanding the inspection to the whole repertoire of content in an organization, or a relevant part of it. The features of genres, i.e., the substance and form, as well as genre rules do not only interact through a genre; they also provide information about organizational context of a genre or a genre system. The substance and form can be considered as names for metadata categories. The metadata about genres of an organization may be used for: 1) analyzing the relationships and differences of individual genres or genre systems with respect to each other, 2) considering if a genre acts as a variant or a subgenre for another, and 3) for producing groupings of genres with regard to metadata values. The metadata categories may be adjusted and expanded according to the focus of the study, and metadata about organizational context as well as of individual genres may then justifiably be gathered for content analysis and even for a content management system.

## 2.2 Genre-based Research

Genres have been studied in multiple research disciplines and areas, such as applied linguistics (e.g., Maingueneau, 2002; Askehave & Swales, 2001; Swales,

1999; Bhatia, 1993) and rhetoric (Miller, 1994), from a perspective of organizational communication (Yates & Orlikowski, 1992), and with regard to their use on systemic management strategy (Yates, 1989), to mention a few. There are two philosophical schools in which genres are studied (Freedman & Medway, 1994). The North-American school of genres is oriented towards change with the attitude that the reality is constantly produced and reproduced by humans (ibid.). The viewpoint adopted by this school is that genres of a discourse community are in a flux: some genres are born, some evaluated, and some left dormant or even disposed. Yet the genre repertoire of a community of discourse remains quite stable (Orlikowski & Yates, 1994). The development of the theory of organizational genres has been influenced by the structuration theory developed by Giddens (1984). According to Yates and Orlikowski (1992), and Berkenkotter and Huckin (1995) the structuration theory is focused on constant production, reproduction and transformation of social institutions. The process is enacted through individuals' use of social rules, which shape the action taken by the individuals in organizations. Consequently, the enactment of rules by strengthening or evaluating them leads the social institutions to an ongoing, recursive interaction. As proposed by Miller (1994), production and reproduction of genres are communicative acts. The relationship between the community of discourse and its genres and genre rules can be seen in a similar fashion. A community of discourse produces and reproduces its genres and rules for enacting them. The enactment and evaluation of genres and genre rules may, in turn, affect the community using them in a reciprocal manner. As a consequence of the adaptation of the structuration theory, at least two kinds orientations with respect to the genre theory may be identified: 1) emphasis on the interrelated, close relationship between the community of discourse and its genres and genre rules, 2) emphasis on the genres being in a state of flux; as stabilized-for now. The theory of genres of organizational communication, as presented by Yates and Orlikowski (1992) is rooted to the North-American school of genres. On the other hand, the Sydney school (Freedman & Medway, 1994) is focused on educational applications of functional linguistics. They have adopted a static view of a genre, implying that genres *exist*, regardless of the community using them. Consequently, genre-based research carried out has used a diversity of research approaches, which affect the selected methods, techniques and tools used in genre-base inquiries.

The North-American school's influence is evident in research findings, which emphasize the intertwined nature of genres and their user communities. With regard to the extent of the community of discourse, i.e. the normative scope taken, genres can be considered as *common* or *local*. For example, genres used on the Web may be considered as common, while genres used by a research and development group of an organization may be considered as local (Yates & Orlikowski, 1992). As pointed out by Yates and Orlikowski (1992) genres should be studied in the context of their use. If common, abstract genres are used as a base for analyzing content in the organization, there is a risk of misjudgment. The genre used in an enterprise can be strikingly different from the common one even if they both are called by the same name (Berkenkotter &



Huckin, 1995). For example, a “business letter” genre is commonly known despite its vague substance and form. By contrast, a localized genre, “letter from a Utah bank promoting a new savings program”, can be more fully defined, including the topics and subtopics such as interest rates, tax benefits, etc. of its content (ibid.). Swales (1999) has anticipated that users also wish to name and categorize the genres they use, if allowed. These findings imply that when a community of discourse for study is being selected, efforts should be focused on that particular community and genres used in it. For example, if a research and development group of an organization is being studied even outsiders of the organization may belong to their community of discourse. Another implication is that one should scrutinize the genres used by this particular community. For example, if one needs to define the topical content of a “letter from a Utah bank promoting a new savings program” one should study the instances of the genre in the Utah bank, instead of focusing on a common type of “business letter”, or to instances of business letters in a Sampo bank in Jyväskylä.

One should also be aware that exemplars of genres vary in their prototypicality (Swales, 1999). In other words, instances of genres do not all follow the genre features and rules in an orthodox way. Maingueneau (2002) has proposed that there are five levels of diversity of genres with respect to their accordance to a genre model; from instituted genres that allow very little variation on a genre instance with respect to its pre-defined model, to genre instances that may even be difficult to relate to a genre. The level of detail in which genres and genre rules are defined may be varying, too. Terms hard and soft genre refer to the extent and level of detail of the genre features and rules in the user community (Schulze & Boland, 1997). If a genre is widely recognized in the community using it, if its features are defined in detail, and the rules for enacting it are clear and widely known, it can be said that the genre is *hard*. If a genre is poorly recognized, if its features are abstract or only recognized by a few, and the rules for enacting it are fuzzy, the genre may be referred to as *soft*. Hardness or softness of a genre cannot be measured by metrical instruments; these qualities can be observed and reasoned about by a community using the genre. In their study Schulze and Boland (1997) noticed that for successful communication (assisted by a software) genres and their rules need to be “hard enough”, and that people do not necessarily adopt the genres provided by the system designer (Schulze & Boland, 1997). These findings imply that the genres intended to be enacted using computer systems need to be hard enough by their features and rules that apply to them. As hardness is a feature of user recognition, users should participate in defining or designing the genres they use.

Genre researchers have exhorted the researchers and practitioners to consider the equilibrium between the user community, their genre repertoire and the media or software used (see, e.g., Crowston & Williams, 1997; Schulze & Boland, 1997; Yates & Sumner, 1997). The equilibrium implies that a change in one leads to an excessive reliance on the others. If the technology or system changes, the users may expect to see genres that are familiar to them. For example, although the electronic newspaper gradually evolved away from the newspaper-in-print (Eriksen & Ihlström, 2000), the readers of pioneering elec-

tronic newspapers insisted on having replicates of newspaper-in-print in an electronic form (McAdams, 1995). The users of a new collaborative system did not adopt knowledge-sharing capabilities provided, since the genres and rules for enacting them, provided by the system designer, were unfamiliar to the users (Schulze & Boland, 1997). Albeit users can adapt a new technology, such as e-mail, and evaluate the genre features and rules for a better fit, they seem to like to steer the process themselves (e.g., Schulze & Boland, 1997; Yates & Orlikowski, 1992). The finding may be due to the fact that users prefer being in charge instead of being lead in a deliberate way, or users might not be able to adapt since the genres and rules for enacting them are not comprehensive enough for their comprehension. Implications for researchers and practitioners are, that when new media or technologies are being developed, the users quite likely find it easier to adopt genres familiar to them (e.g., Crowston & Williams, 1997; Yates & Sumner, 1997; McAdams, 1995). On the other hand, if a use of new genres needs to be forced, the users should be carefully trained for the use of them, and the rationale behind the change should be negotiated and explained with the users.

Means for scrutinizing the genres in an organization are needed. According to Swales (1999), the active members possess the greatest knowledge over genres; they usually dominate the process of nomenclature and categorization over the repertoire of genres, too. Expert members of a discourse community can thus be a valuable source for an inquiry. However, Tyrväinen and Päivärinta (1999) found that no single expert in an organization was capable of defining the whole repertoire of genres used. As genres are “by-products of a history of negotiation among social actors” (Yates & Orlikowski, 1992, p. 305) some sort of group activity on defining the genres may be needed.

The features of genres and genre rules can provide information of an organizational context for scrutinizing the genres. The *5W1H* (Yoshioka, Yates & Orlikowski, 2001; Orlikowski & Yates, 1998) aspects of genres extend the aspects of genre rules, substance and form, and can be thought of as a genre-based metadata framework, which can be adopted for studying genres of an organization. The *5W1H* was originally intended for studying the similarities and differences of genre systems with regard to each other. The *5W1H* aspects may also be applied as names of metadata categories for defining genre systems or individual genres of interest. The *5W1H* considers the *why* (socially recognized purpose of communication), *what* (expected content), *how* (media, type of language), *who(m)* (who communicates, to whom), *when* (e.g., time schedules or deadlines, duration) and *where* (physical or virtual places; such as company building or URL) aspects of genres (Yoshioka, Yates & Orlikowski, 2001; Orlikowski & Yates, 1998). For example, the “who” can be used as a name for a metadata category. As the values pertaining genres of interest are gathered, the genres may be inspected and categorized with regard to their producers. The *5W1H* framework can be enhanced by defining more detailed or adjusted metadata categories. For example, the “how” can be extended to multiple metadata categories, such as categories named as “technology(-ies) used”, and “software used”. The “when” category may be expanded to cover “frequency”

for measuring the frequency of production and use of a genre with respect to others in an organization (e.g., Tyrväinen & Päivärinta, 2003). Table 1 provides an example of the use of the 5W1H framework for metadata about two genres called “unofficial memo” and “official memo”, used on the research and development group of an organization.

TABLE 1 Example of Metadata Using the 5W1H Framework

<b>Community of discourse: R &amp; D group</b>		
	<b>Unofficial memo</b>	<b>Official memo</b>
<b>Why</b>	Record /remind about task allocation	Record task allocation, past & future development with regard to a plan
<b>Who(m)</b>	From-to: Members of R & D group	From: Members of R & D group To: members of R & D group, product manager
<b>What</b>	Style: informal language, lots of acronyms Topics: Time Date List of items Announcements	Style: more formal language, company document template. Topics: Time Date List of items Announcements Review of product development
<b>Where</b>	R & D department	R & D department+ product department
<b>When</b>	Weekly; after meeting	Once a month, after meeting
<b>How</b>	Text processing in e-mail program, via e-mail, in the e-mail body	Word text processing program, via e-mail; document attached to an e-mail message

In the table, the values of the six metadata categories are listed under the names of genres. For example, the metadata category “where” has the value “R & D department” for the genre “unofficial memo”.

Despite the numerous findings with regard to genres and their use in organizational settings, there seems to be a lack of certain type of genre studies. Berkenkotter and Huckin (1995) noticed that genre researchers tended to use reifications of common genres in their studies. They pointed out that studies in which the users are allowed to define their genres themselves are needed. There seems to be a dearth of studies considering the local genres, and studies in which the topical content within genres is scrutinized. For example, Swales (1999) has defined the topics of academic paper genre, but these kinds of studies are still rare. This type of research might provide findings for the research area of content analysis, too.

### 2.3 What is the “Genre Lens”?

A theory can be instantiated in multiple ways. A researcher may be more or less knowledgeable about the theory, and the ways for applying an analytical lens provided by a theory can be varying. Researchers may be familiar with a larger or a smaller portion of previous research, and acknowledge the findings and

implications to a large or small extent in their research settings. While an analytical tool is being applied there is room for interpretations and reasoning. This subsection describes the “genre lens” used in this dissertation research. It consists of three components: 1) the selected concepts of the genre theory and relationships between them, 2) our selection of, and understanding about the implications from previous research for content analysis, and 3) the philosophical assumptions which affect the choices for techniques and types of inquiry used.

The “genre lens” surfaced gradually as the researcher gained more insights about the theory of genres itself, on the implications provided by previous research, and with respect to the content analysis needs identified on the course of the study.

Concepts of the theory of genres, and the relationships between them were discussed in the subsection 2.1. The concepts that provided a conceptual base for operationalizing the “genre lens” in this dissertation are: genre hierarchy, genre repertoire, genre system, genre, genre rules, genre variant, subgenre, and topic (cf. Swales, 1999; Crowston & Williams, 1997; Miller, 1994; Bazerman, 1994; Yates & Orlikowski, 1992). We have used a slightly different definition of subgenre from that provided by Yates and Orlikowski (1992), and added a concept of topic, based on many researchers’ (e.g., Swales, 1999; Bazerman, 1994) ideas about “topical content” of a genre. The genre features, such as substance and form or the aspects of the 5W1H framework (Yoshioka, Yates & Orlikowski, 2001; Orlikowski & Yates, 1998) can be considered as having potential for defining genre-based metadata categories. They can be enhanced to fit to the needs of research or practical objectives. For example, the 5W1H aspect of “who” can be enhanced to include groups of users, such as in user roles. We consider that the genre-based metadata values provide an organizational context for a genre. By scrutinizing genre-based metadata values, we can also consider the differences and similarities of genres and genre systems. For example, we can study the genres produced by a certain group of users, or genres produced in a recurrent communicative situation.

Findings from the previous research were multiple. Genres should be studied in organizational contexts, and the users should be allowed to name and categorize (Swales, 1999), as well as to define the genres they use themselves (Berkenkotter & Huckin, 1995). It is important to set the normative scope of the study right, i.e. to define the community of discourse of our interest. The findings related to hard and soft genres (cf. Schulze & Boland, 1997) seem to imply that in order to implement a system, the genres to be enacted in it should be sufficiently hard. As the softness or hardness is a measurement based on the understanding of the people using the genres, means for acknowledging and expanding their genre knowledge should be sought.

Many of the findings imply that techniques that support collaborative group activities for defining the genres by their users are needed in organizations: genres are produced, used, and negotiated by *a group of users* (Yates & Orlikowski, 1992), the hardness and softness is a measure that can be applied by *a community of users* (Schulze & Boland, 1997), a single expert user of genres may not be capable for defining the whole repertoire her- or himself (Tyrväinen &

Päivärinta, 1999), and that the genres should be named and categorized (Swales, 1999), and defined by *the users*, i.e., the *community* of users, themselves (Berkenkotter & Huckin, 1995). Use of collaborative group techniques may ensure that a consensus and knowledge about the genres being defined is negotiated and shared among the users.

Use of inquiry techniques that support group work could also help to overcome the caveat of communication breakdowns (Byrd, Cossick & Zmud, 1992) identified on the area of content analysis. The use of collaborative group technique could also provide a setting in which people of the community may more easily orient themselves for defining their needs and tasks in a more abstract way, since people can reflect and negotiate about their needs together (cf. Mumford, 2000). On the other hand, the content analyst or developer should participate in the group activities, or observe the target organization closely enough so that possible communication breakdowns between the analyst and community of users could be avoided, or minimized. A researcher who acts as participant or close observer may be able to adopt the specific language and jargon used by the community (cf. Swales, 1999).

It seems that research on local genres, and topical content of them is quite rare. For example, Swales (1999) studied the topical content of academic articles, and Crowston and Williams (1997) observed that a genre might be embedded to other, thus acting as a subgenre for it. Rehm (2002) envisioned that content schemas for genres could be designed as a combination of genre modules. A genre could be considered as a collection of genre modules. Each module may cover one or several topics of substance (ibid.).

It may have become evident that this research adopts the underlying philosophy of the North-American school of genres (Freedman & Medway, 1994). The next section moves on from the related research and research area of the dissertation, and discusses the objectives of this dissertation.

### 3 RESEARCH OBJECTIVES AND METHODS

*“The real purpose of scientific method is to make sure Nature hasn't misled you into thinking you know something you don't actually know”. (Pirsig, 1974, p. 94)*

In this section, the background for the dissertation study and case organizations involved are introduced first. Then research objectives, approaches and methods are discussed. The last part of the section is an overview of the research process with respect to included articles.

The research included in the dissertation was carried out in two document and content management development projects, METODI and inSGML, (Salminen, 2003b), and in two action research studies succeeding them. The research was carried out in the University of Jyväskylä during 1998-2003. METODI and inSGML projects were funded by the Finnish Technology Development Centre and the collaborating companies. A possibility of applying genre theory to content management arose during the METODI project, even though the concept of content management was not yet commonly established. The inSGML project was focused on technical documentation and XML (Bray, Paoli, Sperberg-McQueen et al., 2000) language. At first it seemed that research on the genre theory would not fit the objectives of the inSGML project. It was discovered that content analysis methods developed for structured documents, i.e., the RASKE (Salminen, 2003a; 2000) methodology and the method of Maler and ElAndaloussi (1996) were not applicable for analyzing the training content of a collaborative organization as such. The content was found to be too vague. The possibility of operationalizing the theory of genres and findings from genre-based research were reconsidered. Hence, the research process was guided by the researcher's interest on the theory of genres, and also by the practical problems posed by the collaborating companies.

A characteristic feature of the research is its empirical nature. The results have been derived from data collected through active participation in development projects in four different kinds of organizations: a non-profit national sci-

entific organization, an international industrial company, a church, and a public administration organization acting as the central body for private sector pension institutions. In the following, the collaborating organizations and their reasons for collaboration are described.

The Center for Scientific Computing (CSC) is a non-profit organization owned by the Ministry of Education in Finland. It provides services in several areas of expertise related to scientific computing and communications (CSC, 2003a). For instance, CSC provides the services of FUNET, the national university and research network in Finland. CSC is organized into several expertise groups. In 2002 there were approximately 120 employees in CSC, and its annual revenue was approximately 13 million euros (CSC, 2003b). The objective for the study at CSC was to define the contemporary document management in CSC, and to elicit requirements for a future document management system.

A manufacturing company dubbed BIRD is an international producer of production lines for process industry. BIRD's customers are typically factories. Each production line consists of several machines and components tailored according to the needs of individual customers. BIRD has tens of manufacturing units and tens of other types of units worldwide. The number of employees of the whole corporation, of which BIRD forms a part is nearly 30 000. The study was carried out with a unit of BIRD located in Finland. The unit consists of multiple departments in addition to a factory, in which many of the machines of the production line are prepared. There are a few thousand employees in this unit. The objective of BIRD was to study reuse of content from its Operation and Maintenance (O&M) manuals as a source for preparation of training materials. Each production line is a one-of-a-kind product; hence the O&M manuals and the training materials are prepared individually for each customer according to the specific features of a production line. The size of O&M manuals can exceed over 50 folders of paper when printed out. For the purpose of the research, both O&M manuals and training materials needed to be analyzed.

The collaborative church is a medium-sized church located in Finland. It has 9 employees and 1,700 members, most of whom work as volunteers in different positions. Many of their activities are organized by groups of volunteers. The church had established a project for improving the production and communication of information about its services. As a part of the project, a requirements analysis was carried out (Lyytikäinen, 2003). It was observed that a set of documents and other information needed to be defined in greater detail.

The Finnish Centre for Pensions (FCP) works under the supervision of the Ministry of Social Affairs and Health and the Insurance Supervision Authority. It acts as the central body for the multiple private-sector pension institutions. These take care of pension provisioning for employees of specific branches of work (ETK, 2003; FCP, 2003). In 2002, there were nearly 400 employees in FCP. The objective for FCP was to develop a Web service for their directives and circulars, improve the quality of their directives' and circulars' content, and find enhancements for their production and manipulation processes. For the purpose, a large development initiative was founded. As a part of this work, a consultant explored how to instantiate a genre-based method developed by the author.

During the study, two genre-based methods were developed. The first one (Method 1, Article 4) was developed in the METODI project, in the study carried out at CSC, by a group of researchers including the author. Another member of that group, Tero Päivärinta, elaborated this method further as a genre-based method for information systems planning (Päivärinta, Halttunen & Tyrväinen, 2001), and explicated a genre-based approach to developing electronic document management in organizations (Päivärinta, 2001). The second genre-based method was targeted for content analysis (Method 2, Article 5). The author developed it during the inSGML project, in the study at BIRD. The method 2 was operationalized and evaluated in two succeeding studies at Church and at FCP.

### 3.1 Research Objectives

Genres have been extensively investigated within various disciplines. Yates & Orlikowski (1992) have pointed out that genres should be studied in the context of their use, i.e., in real-life organizations. Berkenkotter and Huckin (1995) have noticed that studies in which the users are allowed to define the genres they use are rare, but needed (*ibid.*). Hence techniques supporting group-based activities for identifying and defining the genres used by a community are needed (cf. Tyrväinen & Päivärinta, 1999; Berkenkotter & Huckin, 1995). Only a few studies were found in which the topical content of a genre (e.g., Rehm, 2002; Swales, 1999) was studied. There seems to be a lack of genre studies that focus on local genres, or on the hierarchical relationships (cf. Crowston & Williams, 1997) between them.

The area of content analysis being new, there is a need for empirical, explorative research shedding light on content analysis in organizations. Methods and techniques for preventing communication breakdowns between users and analysts (Byrd, Cossick & Zmud, 1992), and helping the users to express their requirements despite their limited knowledge about the technological potential have been addressed (Mumford, 2000; Koulopoulos & Frappaolo, 1995). In this dissertation, requirements analysis and content assembly are considered as specific types of content analysis tasks.

This research is focused on exploring the potential uses of genre theory for content analysis, and on finding examples of the types of content analysis needed in organizations. The focus of this study is characterized by the following questions:

- What kinds of content analysis objectives there are in organizational development initiatives?
- What means could be used for hindering the communication breakdowns, and helping the users to express their requirements in a more coherent and explicit way in content analysis?
- Which aspects of the genres, and in what detail, should be scrutinized in content and requirements analysis?
- What are the findings, if any, for the use of structured documents?
- How the genre theory and related research could support content analysis?



By exploring the use of the genre theory and the field of content analysis together, this study aims to report findings that are relevant for the researchers and for the practitioners as well.

### 3.2 Research Approach and the Methods

Research can be characterized and categorized from multiple perspectives, for example, with respect to the research approach used, or the methods instantiated. According to Galliers (1991) a *research approach* is a way of going about one's research. An approach may embody a particular style and may adopt different methods or techniques. A *method* consists of directions and rules of action according to some systematic ordering (Hirschheim, Klein & Lyytinen, 1996). It might be considered as a procedure or mode of inquiry, which consists of intertwined or integrated *techniques* as procedures of actions (Blokdiik & Blokdiik, 1987, p. 5).

Burrell and Morgan (1979) have identified two main categories for research approaches: the subjectivist approach, and the objectivist approach. The classification is based on differences in ontology, epistemology, human nature, and the research methodology or method used (Burrell & Morgan, 1979). The subjectivist approach considers reality to be constructed through human actors and maintains that the social world can only be understood via humans involved on the focus area of the study. The objectivist approach considers the social world as the object independent of humans, and stresses systematic research techniques for explaining and discovering regularities, patterns of actions, and causal relationships (Järvinen, 2001). Orlikowski and Baroudi (1991) have proposed a categorization into positivist, interpretive, and critical research approach. According to Järvinen (2001), the positivist approach category of Orlikowski and Baroudi (1991) is akin to the objectivist approach, and the interpretive approach category of Orlikowski and Baroudi is closely related to the subjectivist approach. The critical approach category of Orlikowski and Baroudi (1991) may favor longitudinal studies (Järvinen, 2001).

Action research (Kock, McQueen & Scott, 1997; Susman & Evered, 1978) has been used as the dominant research method in this dissertation. According to Järvinen (2001) action research applies the results of basic research to a certain objective, which commonly considers building an artifact. Hence action research can be considered as a constructive research method (*ibid.*). Constructive research can be considered as holding inherent objectivist and positivist assumptions, since constructing an artifact is commonly considered as a systematic, engineering-type of task. Action research itself is contradictory in this sense: some researchers claim that action research was developed to overcome the limitations of positivist research and is therefore opposite to it, whereas some consider it as residing in juxtaposition to positivist research (Kock, 1997). There are elements of subjectivist and interpretive approaches present in this study. They evolve from the action research method, and from the theory of

genres utilized in this study. The interpretive background influenced the techniques that were adopted in the methods developed in the course of the study.

In this research, there were two types of objectives present: research objectives and practical objectives posed by the collaborative organizations. Susman and Evered (1978) have characterized action research as a research that aims to "contribute both to the practical concerns of people in a problematic situation, and to the goals of science by joint collaboration" (ibid. p. 587). Action research can be characterized with six properties (Järvinen, 2001; Susman & Evered, 1978): 1) it is future oriented, 2) it is collaborative, 3) it implies system development, 4) it generates theory grounded in action, 5) it is agnostic, suggesting that researchers' theories and prescriptions should be re-examined when entering into a new context, and 6) it is situational, implying that the appropriate action should be based on knowing how particular actors define their current situation, instead of creating replicas of previously known situations.

Action research favors collaboration between researcher(s) and practitioners. For example, it applies small face-to-face groups of researchers and practitioners diagnosing and designing actions through which the problem situation may be changed. Ad hoc and permanent face-to-face groups are usually developed within a client system for conducting the research, and carrying out the action needed (Susman & Evered, 1978). The *action* in action research does not necessarily refer to drastic changes in organizations. According to Susman and Evered (1978), the intervention can be direct, and have an element of unexpectedness, as in a new way to solve a problem or in a way to conceptualize an old problem, presented by an action researcher. The nature of intervention is an act of communication between two or more self-reflecting subjects, with mutual understanding of the meaning of the acts, and of supposed consequences of them. "Even if the intended target of an intervention were changing the physical aspects of a system," .. "it still would be mediated by communication of such intentions to members of the social system and gaining the consent of at least its most influential members" (Susman & Evered, 1978, pp. 593-594). The intervention can also be indirect. The researcher can act as a catalyst or a facilitator for people of the organization. In an indirect intervention "acts of communication take the form of helping organizational members to articulate a desired future against which to compare present situation (ibid., p. 593)".

Action research can be viewed as a cyclical process with five phases: diagnosing, action planning, action taking, evaluating, and specifying learning (Susman & Evered, 1978). In these phases, especially in diagnosing and evaluating phases, other methods and techniques can be adopted. Usually all the five phases are needed and used, but action research projects may differ in the numbers of phases that are carried out in collaboration with the researcher and the client system (Susman & Evered, 1978). External rigor and generalizability of findings can be improved by carrying out multiple action research cycles, thus iterating the results of a previous research cycle on another (Kock, McQueen & Scott, 1997). Action research itself can be divided into multiple types, in which the role of the researcher and the extent to which the five phases are being applied and iterated can vary a lot (Baskerville & Wood-Harper,

1998). With respect to Baskerville and Wood-Harper's classification (1998) this research can be characterized as *canonical* action research.

A collaborative, agnostic and situational stance towards research, apparent in action research, surfaced also in the North-American school of genre researchers, partially due to the researchers' backgrounds in structuration theory, and partially via findings from previous research, which emphasized the reciprocal, intertwined relationship between genres and their community of users. Albeit action research is intended for solving practical problems together with research objectives, the philosophical background of action researchers is akin to that of genre researchers who adopt the North-American school of philosophy. It cannot be explicitly stated whether the interest on the genre theory influenced in the selection of action research for this study, or the other way around.

### 3.3 Research Process and Relationship to the Included Articles

The study consisted of two literature reviews and four action research studies that were carried out in four collaborative organizations. Figure 3 illustrates the studies with respect to the articles that are included in this dissertation.

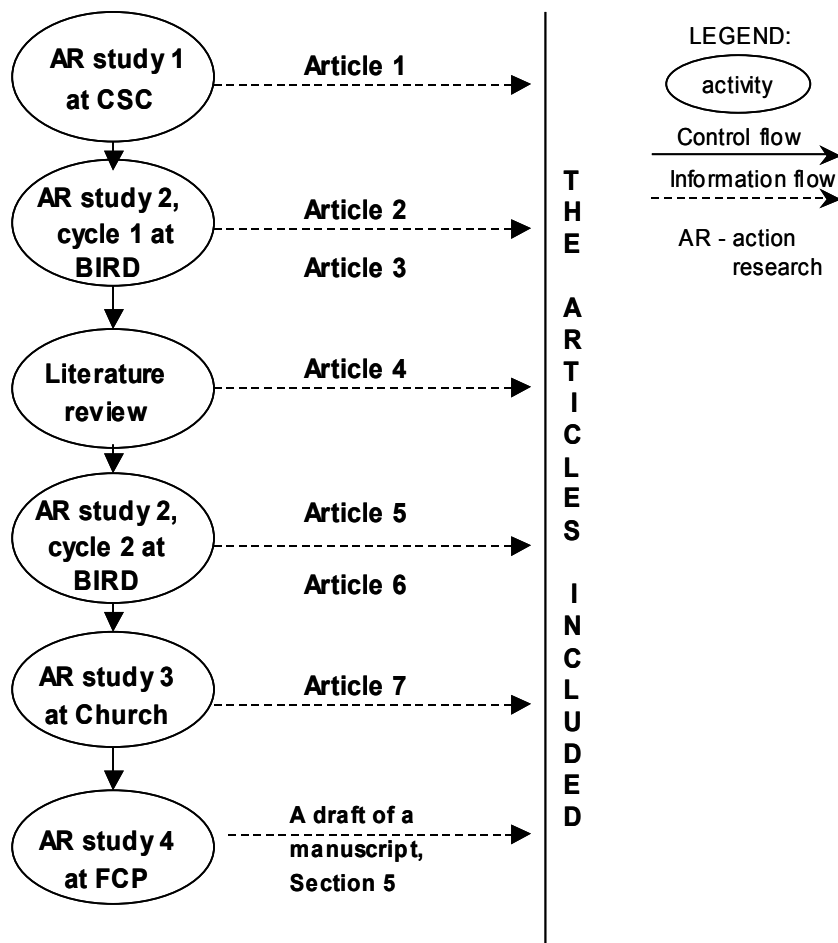


FIGURE 3 The Research Process with Respect to Included Articles

Figure 3 adopts a modified version of Information Control Nets (ICN) process model (Salminen, 2003a; Salminen, Lyytikäinen & Tiitinen, 2000; Ellis, 1979). In the figure, acronym AR refers to action research. The ovals depict research activities, and their ordering indicates the order in which these activities were started. Parts of the activities were conducted in parallel. The action research study carried out at FCP is not reported in an included article, since the manuscript considering the study was prepared concurrently with the text of the dissertation. The study carried out at FCP is hence presented as a draft manuscript, which is embedded to the dissertation as Section 5. At CSC, Church, and FCP the action research study consisted of one action research cycle, whereas at BIRD there were two action research cycles carried out.

As is the case with a research process in general, this one was started by a literature review as well. Since the reports of the review are not included in this dissertation, the review is not shown in Figure 3. In the review, the research on document management and definitions of document management systems available were studied (Doculabs, 1998; Staunton, Tougard, Richardson et al., 1998). As a result, a report defining a framework of document management features available or requested in contemporary systems was made (Karjalainen, 1999b), together with a report (Karjalainen, 1999a) that describes the standards related to document management. Other members of the METODI project research group discovered the vagueness of the concept of a document (Päivärinta & Tyrväinen, 1998), and the theory of genres was identified as a possible analytical tool (Yates & Orlikowski, 1992) that could be applied for document management (Tyrväinen & Päivärinta, 1999). A practical need of CSC provided the possibility for exploring the theory of genres. In the study, the contemporary document management as well as the requirements for a future system were studied by operationalizing the genre-based method for requirements analysis (Method 1). The method was explicated for the study and evaluated by the findings. In the study, the framework developed by the author was adapted for metadata categories for the genres identified, which allowed the requirements for a future system for CSC to be defined.

Articles 2 and 3 describe the first action research cycle that was carried out at BIRD, as a part of the inSGML project. BIRD was intended for investigating content reuse (which was later refined to content assembly), using the O&M manuals as the source, and the training documentation as the target. Article 2 describes the complexity of content production processes in organizations by giving an example of O&M manual production in BIRD. The example illustrates the multitude of contemporary technologies, such as SGML and XML, and software applications utilized at BIRD.

Article 3 was motivated by an unsuccessful trial of trying to analyze the content of training at BIRD using the RASKE methodology. The potential provided by the genre theory was re-diagnosed, and the first trial for utilizing it was instantiated. The article introduces the genre theory as a means for analyzing the content of training. The O&M manuals and the training content were identified as hard and soft genres (Schulze & Boland, 1997). The content of training, which consisted of a mixture of speech, documents, slide shows, and

multimedia was analyzed and defined in an explicit way by the trainers, thus making the content of training harder. Yet the content of training was not defined in enough detail for allowing the content reuse to be scrutinized.

Findings of the literature review are reported in Article 4. Implications and examples from previous research, illustrating how the theory of genres has been operationalized for content management were scrutinized. Article 4 discusses implications provided by the genre-based studies for researchers and practitioners. The literature review confirmed that the genre theory could be applied for content management in multiple ways. Implications for selecting techniques and methods for genre-based studies were not discussed in this article, and pointers for a great number of genre-based research focuses were also excluded. However, these were useful for planning further action for the study at BIRD, and were reported in Section 2 of this introduction. The objectives for the dissertation research were re-examined by the light of the findings. It was decided to focus on the theory of genres, and instead of focusing on content management the scope was narrowed to content analysis only. The field of contemporary technologies, details of use and production processes in this field, and countless software applications available were considered as too broad and complex for a thesis study.

The second action research cycle was carried out at BIRD (Karjalainen & Tyrväinen, 2001). Article 5 summarizes the findings of this action research study. It describes the genre-based method for content analysis (Method 2) that was explicated as a tool for studying requirements for content assembly in BIRD. It also introduces Method 2 and describes how it was utilized for analyzing the content of training, and studying requirements for content assembly at BIRD. Article 6 discusses Method 2 and the action research study carried out at BIRD from the eLearning viewpoint. It describes how the method can be applied for defining learning objects, which are the units of content used on eLearning applications and related specifications.

The action research studies carried out at Church (Article 7) and at FCP (Section 5 in this introduction) iterate and evaluate the genre-based method for content analysis (Method 2). For conclusions, the findings of the studies and literature reviews were analyzed with respect to this study as a whole. The discussion of findings for research themes was produced, and the findings were synthesized for the benefit of researchers in the genre area, and for researchers and practitioners in the field of content management.

In action research the researcher's role and intervention with the target system (Baskerville & Wood-Harper, 1998) can be quite different from one study to another. The role and type of intervention of the researchers in the studies conducted were as follows. In the study carried out at CSC (Article 1), the role of a group of the researchers (the author as one of them) defined the objectives for the study, and planned the action together with a document management development group at CSC. Action taking was observed in multiple workshop sessions, in which the researchers acted as facilitators and observers. The evaluation was carried out by observing the workshop sessions, and by interviews and discussions with the members of the document management

group in CSC. In the study carried out at BIRD (Articles 2,3,5 and 6), the researcher's (the author) role was similar to that of the researchers at CSC. In the study carried out at Church, there were two researchers. The first one had carried out a requirements analysis study in the church (Lyytikäinen, 2003), and found that there was a need for scrutinizing the content of a genre system in a more detailed and explicit way. A more detailed analysis of the genre system of interest was carried out with the author (Article 7). The researchers enhanced the genre-based method for content analysis (Method 2) for the purpose of the study. Action taking, which consisted of multiple workshops, was planned together with a contact person at Church. The role of the researchers, and data gathering for evaluation was similar to the previous studies. At FCP (Section 5 of this introduction), a consultant was asked to carry out a content analysis study as a part of a larger content development initiative. The consultant requested assistance from the researcher (the author). The researcher and the consultant diagnosed and drafted an instantiation of Method 2 for the study. The consultant planned the action together with the manager of the content development initiative at FCP. They carried out the action in FCP together as a combination of workshops, group work and individual tasks. The consultant acted as a facilitator on the workshops. The author took part in one of the workshops as an observer. She also gathered data for evaluation. The development initiative is still continuing, hence the findings discussed are preliminary.

The following section introduces, in more detail, the included articles with respect to research objectives and findings.

## 4 OVERVIEW OF THE INCLUDED ARTICLES

*"..there is a knife moving here. A very deadly one; an intellectual scalpel so swift and so sharp you sometimes don't see it moving. You get the illusion that all those parts are just there and are being named as they exist. But they can be named quite differently and organized quite differently depending on how the knife moves." (Pirsig, 1974, p. 66)*

This section discusses the included articles with respect to their role and contribution for the dissertation study. Each of the articles stands on its own in the sense that it is tied to a certain time and place: at one point of the research some of the objectives were considered more important than others. Some of the jointly authored articles reflect research objectives of the different authors. Some articles emphasize the themes of the tracks of the conferences to which they were submitted. Many interesting research questions and findings discussed in the articles but not central to this study are left out of this overview, which does not summarize the articles *per se*. Repetition of objectives and findings in successive articles has been minimized. Articles 1 and 3 were written under the author's maiden name, Karjalainen.

### 4.1 Article 1: "Genre-Based Metadata For Enterprise Document Management"

Karjalainen, A., Päivärinta, T., Tyrväinen, P. & Rajala, J. 2000. Genre-Based Metadata for Enterprise Document Management. In R. H. Sprague (Ed.) Proceedings of the 33rd Annual Hawaii International Conference on System Sciences (HICSS). Los Alamitos CA: IEEE Computer Society. CD-ROM .

#### 4.1.1 Research objectives and process

This research was carried out as action research at the Center for Scientific Computing (CSC). It had both practical and research objectives. Practical objectives included increasing the understanding of contemporary electronic management at CSC, and preparing a requirements definition for future electronic document management. Explicating a genre-based method for requirements analysis (Method 1) and evaluating its use was the major research objective.

For action taking, the researchers explicated a preliminary version of the genre-based method for requirements analysis. Multiple collaborative workshops were organized for identifying the genres in use at CSC. In the workshops, wall-diagram matrices (Saaren-Seppälä, 1997) were used for identifying the genres produced and used by the group of experts. After the workshop, each group filled the metadata values related to the genres in the form provided to them. The metadata categories were defined with respect to the framework of document management functionalities (Karjalainen, 1999b) provided by document management systems. The three main categories for metadata were: 1) users, producers, and the use frequency of the genres, 2) technologies and software used, and 3) processing needs such as versioning or archiving. Each of these categories contained subcategories also. The requirements definition was based on the metadata values gathered, which, for example, showed when instances of genres needed to be versioned, or published in multiple media. The findings were evaluated by interviewing participants, discussing them with the CSC document management group, and by analyzing the observations made in the workshops.

#### 4.1.2 Results

The genre-based method defined during the study for requirements analysis, proved to be successful. The document management group at CSC had listed 40 types of documents that were used at CSC at the time. There were 850 genres identified. The difference was quite drastic. One reason for the difference was that the genre-based analysis brought up soft (cf. Schulze & Boland, 1997) genres, which the members of the expert groups had not considered as documents. For example, the experts at CSC provided support for the use of scientific computing applications, and answered questions related to problems with Internet technologies or connections. Customer support was provided by phone and e-mail by a number of experts in multiple departments. There were various genres related to these activities, albeit a majority of them was not considered as documents nor shown in the list of documents. Identification of these soft genres brought up many ideas for further development, such as developing Frequently Asked Question (FAQ) service on the Internet. The analysis also revealed variations of genres. For example, there were many kinds of memos developed for official and unofficial, for internal and external use. The researchers estimated that there were approximately 150 main genres in use at CSC, the rest being variants of them. In the workshops, even local genres and their variants



had been identified, and the participants had named the genres themselves. An important finding was not explicated in this article: when the researchers offered the preliminary classification of the 150 main genres and their variants for CSC, the classification was rejected by the document management group with the explanation that they would have classified and categorized the genres differently. This gives support to the Swale's (1999) finding that people want to take control of the nomenclature and categorization of the genres they use.

There was clearly some room for improving the method. The researchers were surprised by the number of genres identified, but asked the groups in the organization to define the metadata values for all of the 850 genres. People in the expert groups considered the task time-consuming and laborious.

The finding implies that a hierarchy of genres should be defined by the people using them before gathering the genre-based metadata. Collaboration in workshops, as well as with researchers and among CSC groups was regarded as an essential factor for successfully obtaining comprehensive results. The diagonal matrices served as an illustrative tool for collaborative discussions in identifying and naming the genres. All participants adopted the diagonal matrix technique, even by those having no earlier experience in modeling information systems.

An explicit genre repertoire with metadata values provided a base for studying the social and organizational aspects of content management. With respect to the theory of genres, the study provides an example in which users define the genre repertoire of an organization. The findings of the study imply that many of the genres used in organizations are soft (cf. Schulze & Boland, 1997), and may not be recognized as documents. The study also shows that the analysis carried out in small groups revealed variants of local genres (cf. Berkenkotter & Huckin, 1995).

## **4.2 Article 2: "Possibilities and Constraints for Managing and Reusing Information Content of Structured Documents: The Case of Operation and Maintenance Manuals"**

Honkaranta, A. & Tyrväinen, P. 2001. Possibilities and Constraints for Managing and Reusing Information Content of Structured Documents: The Case of Operation and Maintenance Manuals. In A. Mørch, A. L. Opdahl, B. Solveig & R. Moe, E. (Eds.). Proceedings of Information Systems Research Seminar in Scandinavia (IRIS) 24 Conference. Bergen: University of Bergen, Department of Information Science, 381-396.

### **4.2.1 Research objectives and process**

The research was motivated by a practical need of the manufacturing company we called BIRD. BIRD aimed to explore content reuse using its O&M manuals as one source for producing training materials as the target. This article ex-

plores the aspects of a source documentation that should be considered when planning content reuse. It provides an example of the use of the SGML standard in content production in a manufacturing company. With respect to the study of the dissertation, the article illustrates a part of the first action research cycle carried out at BIRD. The objective for this cycle was to diagnose the practical needs of the company, and possible avenues for the research. In this part of the first action research cycle at BIRD, the O&M manuals and their production processes were analyzed by experimenting with the use of the RASKE methodology (Salminen, 2003a; 2000). The experimentation is not reported in the article in an explicit way. The general discussion about structured documents in this paper is more or less illustrative and vague, and has limited value for the research.

#### **4.2.2 Results**

The content production process, in which the O&M manual documentation was produced, was found to be complex. During the process, SGML editor software, three different DTDs, in-house tools, and transformation scripts defined by two transformation languages were combined for producing both an electronic and a print version of the O&M manual. The deployment of structured documents was found to simplify and automate tasks such as layout processing. However, the utilization of structured documents required technical expertise, managing and comprehending DTDs, and the use of editor software. For content analysts and designers, the example implies that the processes of content production with structured documents are complex. The document preparation process, document metadata, and management of DTDs and their versions need to be scrutinized for content assembly. The detailed design of a document assembly and content reuse (which was later re-defined as content assembly) is a complex and demanding task. The use of structured documents needs the support of organizational document management functionalities.

### **4.3 Article 3: "Bridging The Gap Between Hard And Soft Information Genres"**

Karjalainen, A. & Salminen, A. 2000. Bridging the Gap between Hard and Soft Information Genres. In M. Khosrowpour (Ed.) *Challenges of Information Technology Management in the 21st Century*. Proceedings of 2000 Information Resources Management Association International Conference. Hershey: Idea Group Publishing, 92-95.

#### **4.3.1 Research objectives and process**

This article also discusses the first action research cycle carried out at BIRD. A trial of adopting the RASKE methodology for defining the content of training at

BIRD had proved unsuccessful. This article explores the use of the genre theory, as proposed by Yates & Orlikowski (1992), as means for analyzing the content of training. In addition, the relationship between strictly structured source, i.e., the O&M manuals in SGML format and less strictly structured target content, i.e., various types of training materials were studied with respect to hard and soft genres (cf. Schulze & Boland, 1997).

The training genres were analyzed while making them harder. The hardening was based on the assumption that a genre is a product of a history of ongoing negotiations (Yates & Orlikowski, 1992). If the members of a community are given means for negotiating about, and defining the genres they use, the genres should become harder (cf. Schulze & Boland, 1997). Organizing a series of workshops carried out the hardening. The adoption of the workshop technique was a mutual decision by the researcher and the training manager, who collaborated closely in this study. The workshop technique utilized was later identified as consensus-creating workshop technique (Coughlan, Lycett & Macredie, 2003). In the workshops that were carried out, the trainers defined the topical content of genres of training. They were aided in their task by the use of a wall-diagram technique (Saaren-Seppälä, 1997). The consensus-creating workshops and group negotiations have been commonly used for these kinds of tasks, and shown to be successful in the area of requirements elicitation (Coughlan, Lycett & Macredie, 2003). The genre features were scrutinized by gathering metadata related to their substance and form, and genre rules were elaborated by using questionnaire forms. The use of forms is not being brought up in the article in an explicit way, and the organization of work in workshops was enhanced after the article was written.

#### 4.3.2 Results

The research implies that the genre theory can be operationalized for making the genres of a community harder. The genres of training at BIRD were found to be soft. The finding implied that it would be difficult to design computer support for enacting the genres. The content of training was produced individually by the trainers and varied considerably. The O&M manual documented in SGML format was, without a doubt, hard. Effective reuse of the content units of the O&M manual required finding more explicit descriptions for the training materials, i.e., hardening the training genres. Four training groups and four genres of training were identified. As the series of workshops continued, the trainers became familiar with the names used for the topical content of the training genres, and defined the content of their genres in a more coherent way with respect to others. In the study, consensus-creating workshops (Coughlan, Lycett & Macredie, 2003) and the wall-diagram technique (Saaren-Seppälä, 1997) were found successful for hardening the genres and for specifying the topical units of content of the training. The actions taken and techniques used were considered successful, and might be utilized in other settings, too. Hence a need to define a method was identified as this action research cycle was carried out at BIRD.

#### **4.4 Article 4: "Developing Document And Content Management In Enterprises Using A "Genre Lens"**

Honkaranta, A. 2003. Developing Document and Content Management in Enterprises Using a "Genre Lens". In O. Camp, J. Filipe, S. Hammoudi & M. Piatini (Eds.). Proceedings of the 5th International Conference on Enterprise Information Systems. Setúbal, Portugal: Escola Superior de Tecnologia do Instituto Politécnico de Setúbal, 334-340.

##### **4.4.1 Research objectives and process**

This article describes a literature review that was aimed at discovering ways in which the theory of genres can be used to support content management. The researcher thought that the genres of training at BIRD would need to be scrutinized further, and sought for answers from the previous findings from genre-based research. This article reports the findings that were considered as potentially interesting also for other researchers. Another objective for the article was to provide content management researchers and practitioners starting points to the genre-based research. With respect to the dissertation, the article diagnoses the potential of the genre theory for content analysis.

The research process proceeded as follows. At first, 80 potential papers in which the theory of genres was adopted were identified from the information systems field. The detailed analysis dealt with 68 of these papers. Possible categorizations for the papers were studied with respect to content management by applying inductive reasoning, and by formulating trial categorizations of the findings. The categorization approach was loosely adopted from a technique called Open Coding in the Grounded Theory (Strauss & Corbin, 1990). In Open Coding, the instances of data that are gathered are grouped into conceptual categories according to the researcher's inductive reasoning. The categorization selected was discussed and evaluated with two colleagues. A group of representative references were selected into each category by the individual judgment of the researcher. The findings of the paper in each category were analyzed, and described in the article.

##### **4.4.2 Results**

Three categories of articles were defined. The category named "Content analysis" contains references that illustrate the use of the "genre lens" for content analysis or a part of it. The category "Genres on the Web" contains references with implications relevant for the Web site and delivery media designers and related to content management. Finally, the category "Indirect implications" contains a selection of references with implications that are relevant for content management developers.

The findings related to content analysis imply that universal definitions of "documents" should not be used. Instead, organizational genres and genre sys-

tems should be systematically studied and revised together with domain experts (Tyrväinen and Päivärinta, 1999). The 5W1H aspects of the genres were operationalized as a metadata framework for studying genre systems (Orlikowski & Yates, 1998) and information coordination between genres. This provided opportunities for workflow design by applying the 5W1H aspects of genres (Yoshioka, Yates & Orlikowski, 2001; Yoshioka & Herman, 2000). The findings related to genres on the Web (e.g., Eriksen & Ihlstöm, 2000; Crowston & Williams, 1997; McAdams, 1995) and enactment of genres in new systems (e.g., Schulze & Boland, 1997; Yates & Sumner, 1997) imply that in new environments people tend to have increasing expectations of being able to use genres familiar to them. If a new system is to be implemented, users should be provided with an opportunity to use genres familiar to them. Another implication for application designers is that many times genres are enacted as bundles. Thus when users browse or search for content, a genre system instead of an individual genre should be provided to them. Findings of the literature review suggest that the theory of genres can provide a conceptual basis for developing content analysis in enterprises. The review provides examples and cues for developing genre-based content analysis methods in enterprises. The article does not explicate the types of genre research that was found to be lacking, nor does it report the findings that may affect the selection of data gathering techniques for genre-based research. For the researcher, the review did not provide direct implications of how to proceed. Instead, a portion of findings provided a rationale for reasoning about directions and techniques used for continuing the analysis of content and requirements at BIRD. This rationale has been exemplified as the “genre lens”, and introduced in the Section 2.3 of this introduction.

#### **4.5 Article 5: “Evaluating A ‘Genre Lens’ For Analysing Requirements For Content Assembly”**

Honkaranta, A. 2003. Evaluating a "Genre Lens" for Analysing Requirements for Content Assembly. In K. Siau, J. Krogstie & T. Halpin (Eds.). The 8th Caise/IFIP 8.1 International Workshop on Evaluation of Modeling Methods in Systems Analysis and Design (Emmsad03). Velden/Klagenfurt, Austria: IFIP 8.1 WG, 11 p.

##### **4.5.1 Research objectives and process**

This article reverts back to the practical problems at BIRD. It was discovered, that for content assembly the content of training needed to be defined in more detail, and that the units of content in O&M manuals needed to be identified for assembly. The article describes the results of the content analysis study in BIRD. The first objective of the article is to report Method 2 that was developed as a research tool, and illustrate how it can be applied for eliciting requirements for

content assembly. The second objective of the article is to demonstrate how the method was instantiated at BIRD.

The article describes Method 2 (The genre-based method for content analysis) and the techniques embedded in it. Workshop sessions, wall-diagrams and metadata definition forms were used for empowering the people of a community with tools to define and re-design the genres they use themselves. This paper also exemplifies the need to define the topical content within genres in order to study content assembly and the coordination of content between two different types of genres, i.e., the O&M manual, and the genres of training.

#### 4.5.2 Results

The genre-based method for content analysis was explicated (Method 2). As a result of carrying out the analysis, both topic definitions and topic templates defined the genres of interest. A hierarchy of genres and units of content for studying the content assembly was also produced. A metadata framework, which was enhanced from the 5W1H aspects, proved to be usable. Genre metadata definition provides information of the organizational context, in which the genres are enacted.

With respect to practical objectives at BIRD, the use of the method was considered successful. A description of content assembly process could be defined and the needs for seeding (i.e., adding locators to the source content) and other requirements could be studied. The topic templates, which describe the topics of training genre content, and locators for a source content from O&M manuals, can be used as a human-oriented definition of a genre for a target audience. They also provide a base for defining the templates for content assembly. The results seem to imply that the instantiation of the genre theory and previous findings from the genre-based research - the first, inherent version of the "genre lens" adopted - provided a conceptual background for the study, and gave indications about how to construct the method.

Workshops were considered useful in multiple ways. They allowed the analyst to familiarize herself with the content of the training and the language and acronyms used by the trainers (cf. Swales, 1999; Byrd, Cossick & Zmud, 1992). The trainers seemed to be able to reflect upon their work tasks and provide suggestions for further improvements in the workshop setting (cf. Mumford, 2000). Furthermore, the trainers who were not familiar with systems design diagrams were able to define the content of training for the target groups by using the wall-diagram technique (Saaren-Seppälä, 1997). If the trainers had been interviewed individually, the analysis of the results might have been more laborious and even contradictory, since the trainers would not have had a chance to negotiate current and future metadata features and topical content of genres amongst them. The manager of the training department pointed out that he would not have been able to define the genres of training himself. Experts from both the training and customer documentation departments were needed for studying the units of similar topical and substantial content in the source and target content.

## 4.6 Article 6: “Designing Training In Manufacturing Organizations Using The Genre-Based Method”

Honkaranta, A. & Tyrväinen, P. 2003. Designing Training in Manufacturing Organizations Using the Genre-Based Method. *Journal of Educational Technology and Society (JETS)*. Special Issue on Digital Contents for Educations, edited by Aedo, I. and Landoni, M. 6 (4), 73-85.

### 4.6.1 Research objectives and process

This paper discusses the field of eLearning. In this field, there is a myriad of technologies, different types of initiatives for learning technology standardization (e.g., LTSC, 2002; GLC, 2002; ADLI, 2002) and a multitude of different types of learning management systems available (e.g., Microsoft, 2002). In previous research the learning content has been studied, for example, with respect for enacting it in e-Books (Shiratuddin, Landoni, Gibb et al., 2003; Hillesund, 2002), or when defining the logical units of content for supporting various techniques which are used by the learners (Karjalainen, 1997a; Karjalainen, 1997b). Most of the learning content standards are aimed to define metadata about *learning objects* (units of content for learning), but the definition of a learning object is quite vague, and is often adopted in contradictory ways (Polsani, 2003).

A pioneer on distance learning (Holmberg, 1989) has proposed that the trainers have an inherent understanding about which topics of the content should be enacted as a unit of learning content. Hence he suggests the learning objects to be defined by their substantial, topical content. In contrast, the learning technology specifications base the definitions of learning objects for their enactment on learning management systems as units of storage. Learning objects used in these systems should be self-standing and reusable (Polsani, 2003). In practice, the concept of a learning object has been utilized in a variety of ways, and it seems that the concept as a whole is quite vague (*ibid.*). On the eLearning area, there is a lack of studies in which the content of learning is studied from the organizational perspective. For example, how the units of content such as a learning package, or a learning object should be defined from a learning content of an organization has not been addressed in previous research. This article proposes using Method 2 (The genre-based method for content analysis) for defining a hierarchy of training content as a genre hierarchy and illustrates the use of Method 2 for the purpose.

### 4.6.2 Results

The article illustrated the use of the genre-based content analysis method (Method 2) for defining a hierarchy of genres, which can be analyzed vis-à-vis the definitions of content units defined in the field of learning technology. The analysis of topics and subtopics proved to be useful from three viewpoints. First, it allowed the analysis of genre variants that needed to be defined. Sec-

ond, the variation due to machines that need to be trained could also be analyzed and harmonized using the participative techniques. The variation due to the individual features of each machine produced by BIRD was embedded to the definitions of the four main genres of training. Third, it enabled the collection of information sources for content on an appropriate level of content grain size.

At BIRD, the accurate grain size for a reusable content unit, i.e., the candidate for learning object unit as defined by Polsani (2003), turned out to be a topical unit of content within a genre. If the topics are marked up in XML documents, they are accessible for reuse and processing even though they are not stored as separate learning objects. This contrasts with the common idea used in the learning technology field about creating self-standing units for content manipulation and management in learning management systems. Another difference was observed, too. At BIRD, the learning objects could not be used as “produce once use many times” objects, as envisioned in the eLearning field. All in all, this article illustrates the need to define a hierarchy of genres starting from a system of genres down to subtopics. Only then the units of content that need to be defined in an application area, such as eLearning, could be defined at BIRD.

#### **4.7 Article 7: “Operationalizing A Genre-Based Method For Content Analysis: A Case Of A Church”**

Honkaranta, A. & Lyytikäinen, V. 2003. Operationalizing a Genre-Based Method for Content Analysis: A Case of a Church. In W. Abramowicz & G. Klein (Eds.). *The Proceedings of the 6th International Conference on Business Information Systems*, 4-6 June, Colorado Springs, U.S.A. Poznan, Poland: Department of Management Information Systems at Poznan University of Economics, 108-116.

##### **4.7.1 Research objectives and process**

The article is intended to define how the genre-based method for content analysis (Method 2) can be instantiated for a particular church (referred to as Church in the article). The article discusses on how the method was elaborated in the study, and discusses the findings with respect to the method, techniques used, and to the theory of genres. A requirements analysis that was previously carried out at Church implied that a system of genres would need to be further scrutinized for defining the coordination of units of content between the genres of the system. Two researchers, including the author, elaborated Method 2 and carried out an action research study in the church. In the study, the method was elaborated in the following way: the domain was already defined in a previous analysis process, so the domain definition also considered studying the genres in focus. It was decided to gather the metadata needed in one workshop de-



voted for the purpose instead of collecting a portion of metadata values in each workshop.

#### **4.7.2 Results**

The joint analysis and design carried out by the users, and the use of simple wall-diagram models and tables were found to be useful in multiple ways. The users were not acquainted with information systems design, but the use of these simple models helped them to express their requirements for the anticipated units of content in an explicit way. The users could also negotiate and decide on the topic names and their ordering, and generate ideas for future workflows. Thus a consensus for future improvements and content schemas seemed to exist, at least at some level. The system analysts and designers were able to learn about the domain and domain-specific terminology while acting as facilitators in the workshops that were carried out (cf. Swales, 1999; Byrd, Cossick & Zmud, 1992). Contradictory proposals for a future system could be negotiated, and agreed upon in workshops. There was a need for developing consensus about collaborative work with future shared content repositories. These aspects could be negotiated as the metadata about the genre system was defined. The researchers noticed that any fear about changes in work practices was reduced and positive expectations for future improvements were increased as the process continued.

This article does not bring out explicitly the finding that the users had difficulties in defining the substantial content of genres below the level of sub-topic. This observation is similar to that of Doll and Deng (2001): user participation is beneficial in systems analysis, but may not be useful for a more detailed design. In this study, Method 2 was used to complement a requirements analysis (Lyytikäinen, 2003) carried out in advance. The observations of the study suggest that Method 2 may be instantiated in small enterprises, despite the fact that it was developed for a large manufacturing company.

### **4.8 About the Joint Articles**

The author of this dissertation was the main author of articles 2, 3 and 6. She contributed by presenting the ideas for the research, and carried out the study discussed in the paper in a collaborative organization. The co-authors contributed both to the content, and especially to the language and organization of the articles. Articles 4 and 5 were written solely by the author of the dissertation. She also carried out the study discussed in the articles. The author got feedback and help on revising the language of the articles from her supervisors.

The research discussed in Article 1 was developed and carried out jointly in the METODI project research group. The author of this dissertation developed the article together with the two other researchers. She acted as the project manager for carrying out the research in a collaborative organization, took part

on the practical work, participated in the development of the preliminary method, and proposed applying the connotations of hard and soft genres for the research. She also took charge of developing the metadata categories from the document management framework prepared. Tero Päivärinta proposed a genre-based method to be developed, and explicated the method for the study. Jari Rajala was in charge of the action research project in the collaborative organization.

For article 7, both the co-authors are responsible for an equal share. Method 2 was adjusted, the study was carried out, and the article written collaboratively with Virpi Lyytikäinen. Steve Legrand has provided help for revising the language of articles 2 and 4-7.

## 5 THE CASE OF FCP

*“During this project I have learnt that document management is not so simple I thought it was”...“I understand now, that document management is a part of the new digital way of work” (comments from interviewees in FCP)*

This text is a part of a manuscript, the main part of which was prepared by the author of the dissertation (Anne Honkaranta). Since the study was partially carried out at the same time as this dissertation was produced, the manuscript has not yet been published, and hence could not be added into the included articles. The author of the dissertation prepared the text by studying the literary resources provided by the Finnish Centre for Pensions (FCP) and by a consulting company involved (Ahovaara, 2003; 2002; Eläketurva-keskus, 2003a; b; ETK, 2003; FCP, 2003; Peltola, 2003). Other means for data gathering were interviews and discussions with people working at FCP and with Tuomo Peltola in a consulting company. The also observed the proceedings of one workshop session that was carried out at FCP. In the following, we refer to her as the researcher. Tuomo Peltola has commented on the manuscript and rewritten parts of it. He works as a consultant in the area of document and content management for a consulting and software development company located in Finland. In the following, he is referred to as the consultant. The development and consulting work was carried out under the auspices of the FCP content management development initiative, primarily by the consultant.

### 5.1 The Finnish Centre for Pensions

Majority of the text in this subsection is based on the English Web site of the Centre for Finnish Pensions (FCP, 2003), and no further references to it are made within this subsection.

In Finland, there are two types of pension systems, which are regulated by the laws, norms, and administration. These are also known as pension schemes. The administration of the Finnish statutory earnings-related pension scheme is decentralized. Most employers and self-employed persons may choose their own pension institution, whether it be a pension insurance company, an industry-wide pension fund, or a company pension fund. Pension provision covers tasks such as collection of contributions, management of funds, and payment of pensions. However, a pension institute must operate according to some common principles, so that a person's pension provision is carried out within certain guidelines regardless of the institution involved. A person can work in many different jobs, and hence her or his pension can consist of several parts, which are affected by different types of pension schemes. The Finnish Centre for Pensions (FCP) acts as the central body of the private sector pension institutions. It functions under the supervision of the Ministry of Social Affairs and Health and the Insurance Supervision Authority. FCP can be characterized as an expert organization, which carries out multiple types of tasks (ETK, 2003; FCP, 2003; Eläketurvakeskus, 2003a). One of its functions is to safeguard earnings-related pension for each person. For the purpose, it maintains registers and databases that contain information on a person's work history, and delivers information upon requests. It also ensures that the pension institutions apply pension provisioning in a coherent way. FCP can negotiate the changes caused by a change of a law or norm with the institutions. It gives advice, and organizes courses. It also produces directives and circulars for pension institutions. A *directive* is used to issue ordinances and rules for pension provisioning at a time when a new law or norm affects the provisioning tasks that should be carried out. A *circular* can announce forthcoming changes in laws or norms, or international affairs or labor regulations that affect provisioning. Circulars have also been used for informing the pension institutions about revisions of content parts of published directives. A circular is less formal and less obligatory by its nature, and usually shorter than a directive.

FCP has faced challenges due to the changes in the environment: acts of the European Union can affect pension provisioning, and employees work abroad more often than before. Use of information technologies provides both opportunities and challenges. For example, the Internet has been employed more and more for providing information and advice about pension-related matters. It provides a relatively easy, fast, and reliable channel for distributing information both to individuals and pension institutions. Technologies provided could also be applied for negotiations (e.g., chat, newsgroups, and computer-supported meetings) and for creating Frequently-Asked-Questions type services.

## 5.2 The Content Management Development Initiative

FCP aimed for making use of the possibilities provided by new technologies and functionalities on the Internet, and for improving the management and content of directives and circulars. A content management development initia-

tive was founded for the purpose. The initiative was organized as a project that had multiple objectives (Ahovaara, 2002; Peltola, 2003):

- to create user-friendly and efficient web service for publishing the directives and circulars concerning the earnings-related pension provision and related issues,
- to improve production, maintenance, and publishing processes, as well as the content of directives and circulars by redesigning and standardizing them,
- to develop the quality of the substantial content of directives and circulars, by inspecting the requirements for user groups, and
- to elicit requirements for a content management system which would support content production and publishing on the Internet and on paper, and to implement such a system.

The development project was launched in 2002. In the beginning, the user requirements for the content of directives and circulars were inspected. For example, the pension institutions needed to have directives available faster than currently, and improvements for the preparation schedules were desired. The number of directives is large, and it can be difficult to locate a directive needed for a task from a paper-print folder. The producers of circulars and directives and the experts from the pension institutions both suggested improvements to the substantial content and organization of directives and circulars. For example, on the Internet directives and circulars could be grouped according to topical and pension-type categories (to be developed), and some common information, such as descriptions of a pension types, could be left out as they may be grouped to a set of introductory documents (Ahovaara, 2002).

The preparation processes of directives and circulars are quite complex. Sometimes it can be difficult to reason if a new directive is needed, or whether a circular should be produced instead. The process of producing a directive involves searching previous directives on the subject matter from archives. Locating directives and circulars from the archives can be laborious. A directive or a circular may not exist in an electronic form at all, the electronic version may not be easily usable with the existing word processing software, or the content in the electronic form may not be complete. The changes in laws and norms and transformation regulations also need to be studied. This study can involve searching internal and external information sources, such as laws, the content in law databases, and the content in law and norm in Internet services. Changes in rules and provisioning processes need to be negotiated with several expert groups. These groups consist of experts from in-house departments and external pension institutions. Scheduling the work so that all the statements and comments on a directive version are considered takes time. The preparation of directives consists of many phases, and the person responsible for the preparation at FCP is not able to control the process as a whole. Typically a directive consists of dozens of pages of paper printing, and its preparation process can take from a few months to close to one year (Ahovaara, 2002).

The content management development initiative was large. There were multiple internal and external actors involved. The steering group consisted of executives at FCP and the project manager from the FCP's Information Technology department. The project group consisted of the project manager, a pro-

ject assistant, and experts in substance areas from multiple departments of the CFP. A content management consultant guided the work on the content analysis project, and a project manager from the content management software company participated in carrying out the tailoring and configuring the content management software purchased according to FCP's requirements.

The development initiative consisted of three projects, in which multiple subprojects were (and one still is) carried out (Ahovaara, 2002). The three projects were as follows.

Project 1 was carried out during the first half of 2002. In this phase the needs for content and process development were scrutinized. The contemporary preparation processes of directives and circulars were inspected, and new standardized processes were designed. The phase involved selecting, purchasing, implementing and training the use of a process modeling software application (QPRSoftware, 2003). The members of the project group defined the new processes by using this tool and by negotiating with their colleagues.

Project 2 was carried out between June 2002 and June 2003. This phase considered eliciting requirements for, as well as testing, comparing, and selecting a content management system to be implemented (Hummingbird, 2003). The project group also elicited the requirements for directive and circular content, and the information needs in typical use cases by pension institutions. The content analysis carried out by the FCP project group and the consultant was a part of this project. For content analysis, multiple types of techniques were used, such as workshops, interviews, group, pair, and individual work. A mind map software (Mindjet, 2003) was utilized for documenting a part of the analysis results. The content analysis process carried out by the consultant is described in more detail in the following subsection.

Project 3 started in June 2003, and was expected to last until February 2004. The content analysis will be continued, as well as the implementation and evaluation of the content management system purchased. The Internet service will also be designed and launched. A portion of the directives and circulars have been left out from the implementation. During this project, the maintenance and development routines for content management will be defined.

### **5.3 The Content Analysis Process**

At first, the consultant diagnosed the needs for content development at FCP. He was familiar with the genre theory, and had taken part in the study in which the genre-based method, described in Article 1 was operationalized (Karjalainen, Päivärinta, Tyrväinen et al., 2000). However, the method did not fit the analysis needs. In the case of FCP, there was a need to re-define directives and circulars with respect to their substantial content and length. The preparation process of a broad substantial content, instantiated as long directive documents, took a long time. Most of the content of directives and circulars was to be published on Internet in some future date. The inspection of the user requirements had revealed

that the length of the directives was found to be unfit for current production and use, and it would not fit any better to future Internet publication either. It was also anticipated that if the content of directives could be chopped into smaller units, the preparation process could be quicker, smoother, and easier to handle. The consultant discussed the situation with the researcher. She had just explicated the genre-based method for content analysis (Method 2, Article 5 included), and found this method potentially feasible. The researcher explained the use of the method, and the possibilities to enhance it for the content analysis at FCP were examined together. Directions and techniques for carrying out the analysis were also discussed. The researcher and the consultant agreed that the former would not participate in the analysis process at FCP, nor interfere with the way the consultant applied the method. Instead, she would observe the instantiation of the method and the content management environment at FCP, and evaluate the findings for genre-based research and content analysis.

The consultant carried out the content analysis process at FCP in two phases. In *the first phase* the consultant acquainted himself with the development initiative. He and the project manager at FCP selected exemplar directives and circulars to be analyzed. A metadata framework for providing information about the context of the production and use environment of directives and circulars was explicated. It considered information about groups of users, typical tasks and information needs attached to them, and about the content management functions available or requested. The resulting metadata definition, i.e., metadata categories and values gathered together with the project group at FCP, was recorded into a spreadsheet. This metadata can be utilized, for example, for defining access rights and groups of users, and workflows to the content management system.

A series of workshops for redefining the substantial content and size of directives were conducted. The substantial content in directives varied according to the department that produced them. For example, the directives produced by the law department were different by their substantial content from the ones produced by the pension register department, or from those prepared by the department that took care of the matters related to insurances. Due to the different types of substantial content, the analysis was not carried out in a similar fashion for all directives. For example, the content of pension directives produced by the law department was analyzed and redefined in the following manner. The project group organized a workshop, in which the consultant acted as a facilitator. A mind map software (Mindjet, 2003) was utilized for producing topic definitions of the exemplar pension directives. The consultant used the mind map software and a LCD projector, which was used to project the mind map diagrams on the wall. The members were asked to reason about substantial topics of pension directives, and consider which of the topics might be self-standing enough to become documents in the future. For defining the topics of pension directives, the project group studied the indexes, tables of contents, and heading and section names, and discussed about themes and topics of the exemplar pension directives. By utilizing the mind map software, the proposed topics could be easily added to the topic definition diagram, which

the consultant constantly updated according to the findings. In the mind map software, the ordering of topics as content branches could be easily switched, and the topics could be removed or renamed in an iterative manner. Furthermore, subtopics could be easily attached to a topic, and notes and ideas could be easily recorded as text to be stored as an addition to the diagrams themselves. Figure 4 gives an illustrative example of topic definition of a pension directive, visualized as a mind map diagram.

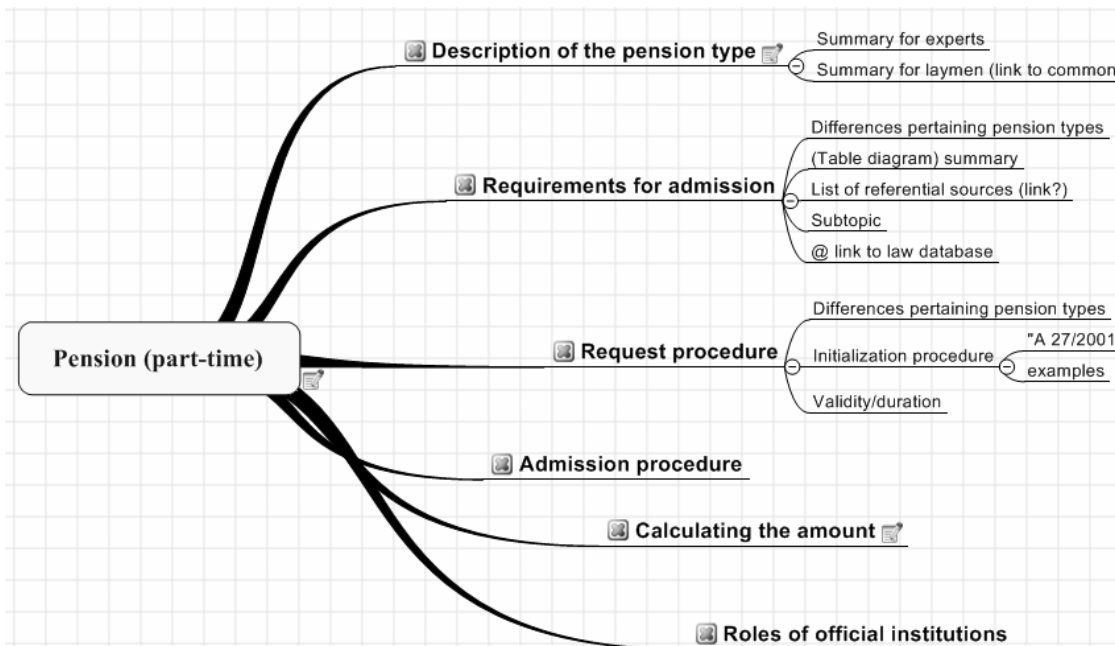


FIGURE 4 Exemplar Topic Definition of Pension Directive as a Mind Map Diagram

In the figure, the text box on the left illustrates the name or a type of the directive that was analyzed. It can be considered as the root of a content tree. The branches extend to the right. For example, a branch named as “Description of the pension type” illustrates a content topic with a name. There are also sub-branches, which illustrate subtopics and their names for a topic. In the figure, the topics, which could constitute a substantial content for a future document, are marked with a cross in front of the topic name. In pension directives, there were commonly around a dozen main sections, which were shown as heading names in the tables of content of exemplar documents. The sections were studied as mind map figures, which also showed some of the relevant subtopics pertaining to a topical section. As a result of the analysis, there were 12 new document types defined, with respect to the topical sections of the old pension directive document. These new documents were named as statements, and classified according to a statement type categorization produced. For example, a statement can be of type “Description of an X pension type”, “Requirements for admission” or “Request procedure”. Hence the name directive ceased to exist, at least on the part of the new documents to be produced. The directive type regarding insurances was, however, much more complex. A lawyer who was a member of the project group carried this analysis out with a group of experts of insurancing matters at FCP’s insurance department.



At this time, the topical content of directives related to insurancing was scattered across the directive sections in the exemplar documents. The group had to define the topics, subtopics, and sub-subtopics of exemplar documents from top to bottom, and then rebuild the topic definition of insurance directives from bottom-up as a whole. The new topic definition allowed defining most of the topics as new types of statement documents.

*The second phase* considered definition of metadata for content management system, preparing topic definitions about statements, and defining a new unit of substantial content, to be known as aggregation. These tasks were carried out as follows.

FCP had already selected a content management system that employs the Dublin Core (1999) scheme to attach metadata to documents. This metadata is used, for example, to support versioning, classifications, and search. The project group participated in a workshop, in which the consultant first explained the concept of metadata, and how it is utilized in content management systems. Then the participants studied the metadata scheme and designed modifications to it. For example, they decided to change some of the metadata category titles. They also defined the obligatory and optional metadata categories, and pre-defined values or lists of choices for some of them. Later on, the metadata definition for the statements was explored with the evaluation version of the document management system adopted.

A selected portion of new statement types were defined in more detail with respect to their substantial content by preparing topic definitions of them. The consultant explained the use of the mind map software (Mindjet, 2003) selected for the purpose and the concept of a topic to the project group at FCP. Each expert member, together with the other experts, was given a task to define a statement type used in their substance area. In the mind map software, the topics were defined as named mind map levels. The members could also define subtopics, and write a description of a topic with the software tool, which would then produce a textual report of a topical content of a statement type. The topic definitions were studied in the workshops, in which experts from multiple content substance areas introduced their topic definitions to others, and the definitions were negotiated and cross-examined. In workshops, the consultant acted as a facilitator. He used the software and an LCD projector for studying, negotiating, and re-defining, with the group, the definitions prepared as mind maps.

When the mind maps as topic definitions of statements were studied, it was discovered that the content residing elsewhere could be reused for some topics of statements. The reuse of topical content across statements and other content sources would prevent the production and versioning of the same topical content instance multiple times. For example, content for a topic "description of Employees' Pensions Act (TEL)" could be imported from a catalogue that introduces the earnings-related pension types. Some of the group members had come to a conclusion that a statement may include another statement as a topic. This type of definition couldn't be implemented in the content management software that only considered files and folders. A need for referencing

parts of law and norm collection texts was also identified. The reference could be realized, for example, as a hyperlink. A hyperlink would be attached either to a statement or a topic of it. The sources of reusable content parts, and needs for referencing a law or norm, or a part of it, were recorded into topic definitions, in as much detail as possible.

A concept of *aggregation* was adopted by the project group. An aggregation consists of a set of statements, which need to be enacted together in a typical user task, or of a set of statements with similar substance. As an example of an aggregation, consider the following. The statements in an Internet service are grouped in multiple ways for the end-users; by their types, or by their publication date, for example. This grouping only provides the latest versions of statements. To be able to reason about a pension provision for a person, an expert in a pension institute may need to study the older versions of some of the statement types, since she or he needs to decide about transferal rules, i.e., which law or norm should be applied. An aggregation that also contains older versions of statements is provided for the task. The experts at FCP create aggregations, and metadata for them can be provided in the content management system. The project group members anticipated that the aggregations might need to be versioned, but the request was dropped due to the lack of support by the system. With respect to the content management system, an aggregation is called a virtual folder. The consultant and the members of the project group at FCP envisioned and defined the types of aggregations to be implemented.

## 5.4 Evaluation

The researcher evaluated the use of the method and techniques as follows. The researcher was allowed to observe one workshop session, in which five people at FCP were interviewed, including the project manager. The interviews were carried out as a theme interviews face-to-face and by telephone, and as a form interview via email. The researcher also had informal discussions with the consultant and with the project manager about carrying out the content analysis. The interviews were focused on the opinions of the project group members about the way the units of content and related metadata were defined, about the techniques and tools used, and about the development initiative and its chances for success in general.

The project member interviews confirmed that the large size of the directive documents hindered the production, use, and search for content. The project members saw the process of redefining the directives and circulars to statements as important and necessary, and their expectations towards the new process and the use of the statements were positive. The members had some experience with the new content management software, but felt that they were not experienced enough to evaluate the success of the project as a whole. Those content producers who had explored the production of statements with the new system had positive expectations about the outcome. Improvements were ex-

pected on the content management activities and on the quality of the content itself. Multiple interviewees expressed, directly or indirectly, that the quality of the substantial content might have improved due to redesign. One of them explicated that since the content of a statement type is defined by its topics, a content producer is able to produce instances of statements that are more consistent by their substantial content than before, when produced by others.

All of the interviewees reported having identified technical problems in the test version of the system, while it was experimented with. The concept of content metadata, used for describing the documents for search and navigation in the content management system, was at first found confusing by several of the group members. There was a consensus that the metadata definition would need to be adjusted as more experience with the use of the system was gained. All interviewed project group members were quite confident and satisfied with the content management project as a whole. They had liked the combination of workshops, group work, and individual work. Workshops and group work had been used before, and the project group members stated that they would like to use these techniques in future development initiatives, too. They were satisfied with the process design tool (QPRSoftware, 2003), and especially with the mind map (Mindjet, 2003) software. The content management project was found to be longer and more comprehensive than an ordinary development project at FCP. Several of the interviewees reported that they felt that their knowledge about content production processes and content management, both in theory and practice, had increased during the project. More than one of the five interviewees had found the content and document management more complex than expected. Multiple of the interviewees provided negative feedback on the scarce resources allocated, and on the tight time schedules. One of them stated that (s)he had wanted to be able to focus more on the project work, and suspected that the project group, or members in his/her area of expertise had not fully understood the importance of the content management development. Another person brought the same idea up in a less direct way. One of the interviewees thought that the limited understanding of the technological capabilities was a barrier for effective work in the project.

The researcher observed that the use of concepts in the environment was complex. The project manager did not want the limitations of the technology to prevent the users from expressing their ideas. Efforts were made to prevent communicational breakdowns in the project, but concepts and terms invaded from multiple directions: the concepts adopted in the content management system, the substantial concepts utilized and developed at FCP, and the concepts used on the content management field. Table 2 on the following page illustrates some of the concepts in use during the project. In the table, the concepts are grouped into columns according to the source of the concept. The concepts appearing on the same row might not be related to each other.

The concept of a document ('dokumentti' in Finnish) was discussed during the study. In Finnish language there is another word signifying documents that are being used for administrative or official purposes (asiakirja). It was not quite clear if the statements at FCP should be referred to as "dokumentti" or "asiakirja". The content management system considered a document as a file.

TABLE 2 Some Concepts Used in the FCP Project and Their Sources

Some concepts or terms used in the FCP project			
<i>Content management system</i>	<i>FCP (substantial terms)</i>	<i>Content management field</i>	<i>Genre theory</i>
Metadata	Directive	Metadata	
File	Circular	Element	
Folder	Statement		
Virtual folder	Document/ official document	Aggregation	
Document (=file)	Topic	Content unit	Topic

The content management system utilized two hierarchical levels of content: files and folders. The files in the system were also referred to as documents. In the content analysis, there were two additional hierarchical levels used: a topic, a common term used in Finnish language (albeit adopted from genre terminology), and an aggregation, which is a special kind of a set of statements. The concept of aggregation is akin to a concept of content assembly, used elsewhere in this dissertation. In FCP's content management system an aggregation is enacted as a virtual folder. At the beginning, the concepts of content unit and element, stemming from the content management field were applied by the consultant. The use of concepts that refer to the substantial content side by side with concepts that refer to storage units of a system may, at some occasions, seem confusing. It seems that in the content management field, the practitioners are bound to struggle with concepts and hierarchies of content units with multiple types. In this project the management of concepts was quite essential, since even the names for documents were redefined.

## 5.5 Preliminary Findings and Implications

As a consequence of carrying out the content analysis, the members of the project group defined a new hierarchy of the content to replace directives and circulars. FCP now has aggregations, statements, statement types and topics. The directives were redefined into statements, and a classification of statements was produced. As the statements can be produced and published much quicker than before, the circulars are not needed any more. The statements were further defined by topic definitions, which describe the topical content of a statement type. Topic definitions allowed identification of reusable content units from external content sources. They were also utilized for defining the references to internal and external sources. A new concept of an aggregation was developed. An aggregation is a set of statements and their versions, which needs to be provided as a content bundle for a user. Hence, the content of directive and circular documents was restructured into a hierarchy of three types of content units: aggregations, statements, and topics.

With regard to content analysis, this study illustrates the complexity of content management as a whole. The units of content were redefined by their names and grain size. There were two types of content assemblies defined. An

aggregation is an assembly of statements. A statement itself can be considered as an assembly if the potential of reusing topical units of content across statements and external sources is to be realized in the future. Current combination of the content management system and statement files produced by a word processing system does not support assembly of statements. Future versions of the word processing software are expected to support the XML format, thus opening an avenue for further research on assembly of statements. A need for multiple types of cross-referencing was also identified. The management of multiple types of references across in-house and external sources can become quite complicated and laborious to maintain, and may also need to be scrutinized further in the future. Concepts from multiple aspects of content management were used, which may obfuscate the communication over the content to be analyzed. At FCP, the substantial names of the content, which were given by the project groups, worked the best.

To genre researchers the study provides an example of a deliberate evaluation of genres – carried out in the target domain (cf. Yates & Orlikowski, 1992) by the user community (cf. Berkenkotter & Huckin, 1995). Initially, the directive genre was very broad by its substantial content. It was suspected that one reason for this was the use of paper print deliveries. During the advent of the Internet publishing era the need was just the opposite, since up-to date information could be offered at all times. The evaluation was not driven by technologies and media only. The old document genres used were laborious to produce, and difficult to use for information content searching. A common solution might have been to span the content over multiple Web pages (Crowston & Williams, 1999). Instead, FCP chose to redevelop the content of its directives and circulars in a more thorough way. Similar to findings of genre evaluation on the Web (e.g., Shepherd & Watters, 1999; 1998), a future statement genre will be enhanced by its functionality, along the implementation of multiple types of references in and across in-house and external sources.

For the researcher and the consultant, the theory of genres, and findings from previous research provided tools for making sense of the situation. The concept of genre hierarchy (Crowston & Williams, 1997), genre system (Bazerman, 1994), genre (Yates & Orlikowski, 1992), genre variant (Crowston & Williams, 1997) and topic (cf. Swales, 1999; Bazerman, 1994), as well as the 5W1H (Yoshioka, Yates & Orlikowski, 2001) genre aspects as a genre-based metadata definition were utilized. The findings from previous research implied group-based, collaborative activities, which were instantiated by workshops and group work. To avoid further confusion over the concepts to be used, the concept of a genre was not used in the project meetings and discussions with the people at FCP. The consultant provided rationale for focusing on communication and substantial content, and, at the beginning of the content analysis project, discussed the basic principles provided by the theory of genres with the content management group. The concept of topic was introduced and used. It is a quite common concept in Finnish language and thus did not refer explicitly to the theory of genres in this case.

Multiple adjustments for the genre-based method for content analysis (Method 2; Article 5; Articles 6 & 7) were identified. The phases of the method were adjusted with a development plan of FCP. A wall-diagram technique (Saaren-Seppälä, 1997) was replaced by the use of mind map software (Mindjet, 2003). There were two types of metadata gathered: metadata about organizational context in which statements are used, and metadata for defining statement files for search and classification in the content management system. A combination of workshops, group work, and individual work was used. The researcher was puzzled by the enhancements to the method. She asked the consultant whether he considered the action taken as an instantiation of the method, or rather as a development of an in-house method. The consultant found the question somewhat strange. He stated that, in his opinion, the knowledge gained by the method drove the interpretations and actions carried out. He had felt easy with the use of the method, and saw no problems in applying it for other content analysis studies, too. The consultant suspected that the researcher was insinuating that he had not applied the method in an appropriate way. Table 3 summarizes some of the aspects by which the original and the evaluated method may be compared.

TABLE 3 Adjustments to Method 2 at FCP

Aspect	Method 2	Adjustments at FCP
Number of phases	4	2
Diagrams used for topic definitions	Wall-diagrams	Mind Map diagrams
Techniques used	Consensus-creating workshops, group work, forms for gathering metadata	Consensus-creating workshops, group and individual work, spreadsheet for recording metadata
Kinds of metadata defined	5W1H metadata and extensions of it; mainly metadata about organizational context	Metadata about organizational context (groups/roles of producers and users, typifications of use situations and content needs), metadata about documents; modified version of Dublin Core schema

While the researcher and the consultant reflected upon their findings and on what was learned from the study, they also discussed the general content analysis needs in organizations. According to the consultant, more than 80 % of the organizations would rather use and enhance a content management system off-the-shelf than design and build the whole application themselves. He also anticipated that more than half of the content management projects carried out in organizations are targeted towards eliciting requirements for, and selecting a content management system.

## 6 RESULTS, IMPLICATIONS AND SHORTCOMINGS

*“To see a world in a grain of sand  
And a heaven in a wild flower  
Hold infinity in the palm of your hand  
And eternity in an hour.”  
(Blake, 1863)*

Section 3 presented the dual objectives for the dissertation research. The potential uses of the genre theory for content analysis were to be explored. Another aim of the study was to find examples of types of content analysis that are needed in organizations. Five themes were defined for characterizing the objectives further. This section discusses the contribution to the study by the data collected from the four action research studies, and the literature review. The five themes, and the potential uses of the genre-based methods are discussed. Discussion of the limitations of the study concludes the section.

### 6.1 Research Themes Revisited

In the following, each of the themes posed for this research are discussed one at a time.

#### **What kinds of content analysis objectives there are in organizational development initiatives?**

Table 4 provides an overview of the four case organizations with respect to their development goals. The following paragraphs provide a more detailed discussion about the aims for content analysis in the four organizations.

TABLE 4 An Overview of the Four Case Organizations and Their Development Goals

Organization/ characterization	CSC	BIRD	CHURCH	FCP
Type of organization	Non-profit, highly technical	Manufacturer of production lines	Voluntary community (religious base)	Center body for pension institutions
Aim for development	To describe contemporary content management, elicitate requirements for future document management system	To develop production of training materials; to study requirements for content assembly between O&M manuals and training content	To analyze requirements for future system for informing about events on the Web	To develop content management, to implement an Internet service, to redesign the directives and circulars
Content to be analyzed	The content units of CSC (the repertoire of genres)	A set of training genres; variations according to products/training groups	Documents +other content related to informing about events (a genre system)	Two kinds of documents: directives and circulars

CSC aimed to describe its *contemporary content management*, and to study *requirements for future content management system*. The studies at BIRD and at FCP consisted of requirements elicitation. At FCP the requirements for, and definitions about documents to be managed were elicited. The *information needs of users* in pension institutions and in FCP were also studied. At BIRD, *the requirements for content assembly* between O&M manuals and other materials as the source, and training materials as the target were scrutinized. For the purpose, *the topical content* of training for four target audiences was defined. *The topics as units of reusable content*, and *their coordination across the genres*, and the internal and external sources were studied at BIRD, FCP, and Church. At FCP, two *types of content assemblies* were defined: an aggregation of a set of statements and their versions needed, and a statement itself, which can be produced as an assembly in the future.

The *content to be delivered in the Web*, and the *functionality features* needed in them were studied at FCP and at Church. The literature review (Article 4) reported other examples of requirements analysis in an organization (Tyrväinen & Päivärinta, 1999). Newspaper companies (e.g., Eriksen & Ihlström, 2000; McAdams, 1995) and academics have studied the *changes in document genres*, such as the *functionality* feature (e.g., Shepherd & Watters, 1999; 1998), and *effects of linking* (Crowston & Williams, 1999) as genres migrate to the Web.

Two *kinds of metadata* were defined for the analysis. The *metadata that defines the organizational context* in which the genres are enacted was derived from the 5W1H aspects of genres at BIRD and Church. This metadata describes the users, producers, typical use situations, and form and media or technology used for enacting the content. At FCP and CSC this metadata was enhanced by the content management functionality aspects available or requested. Another kind of *metadata* was needed for *defining the enactment of instances of genres in systems*.



The purpose of this metadata is to allow categorizations, search, and navigation of the content in a content management system. At FCP, the system selected had adopted the Dublin Core (1999) metadata schema.

One of the objectives for content analysis at CSC and at BIRD was to *identify and define the content to be managed*. At CSC, a great deal of content was not considered as documents. At BIRD, the content of training consisted of a mixture of speech, multimedia, documents, and slide shows. If application area-specific systems (Article 6, Section 5) are to be utilized, *the units of content need to be scrutinized with respect to the storage units or content units that are applied in an application-area specific system*.

### **What means could be used for helping to avoid the communication breakdowns, and helping the users to express their requirements in a more coherent and explicit way in content analysis?**

*Collaborative analysis carried out by the users themselves* (cf. Berkenkotter & Huckin, 1995) was found advantageous in multiple ways. The users were able to negotiate about the names and categorizations for genres (cf. Swales, 1999), requirements for a system, and to generate ideas for future. The researchers were offered an opportunity to observe a user community and learn about their content-specific language by the joint techniques utilized. Thus the communication breakdowns between the users, and between the users and the analysts could be avoided, at least partially.

At BIRD, CSC, and Church, *consensus-creating workshops* (Coughlan, Lycett & Macredie, 2003), *a wall-diagram technique* (Saaren-Seppälä, 1997), and *questionnaire forms filled in jointly* by the users were utilized. At FCP a combination of workshops, group and individual work was applied, and spreadsheets, a process design tool (QPRSoftware, 2003) and a *mind map system* (Mindjet, 2003) were used to support collaboration. The benefit of consensus-creating workshops is that user requirements can be negotiated before reporting them. Hence the ambiguity that may exist when individual users are interviewed can be reduced. A workshop setting was also observed to support the definition of the requirements on a quite abstract level. The techniques were found to be successful.

Preliminary findings from the study at FCP (Section 5) seem to support the necessity of using concepts and names that are familiar to users. When the content was defined at FCP, the concepts such as “content unit” or “virtual folder” hindered the orientation of the community members. Familiar names for communicative units, such as “statement”, were easy to utilize. In a similar fashion, the trainers at BIRD used substance-oriented names of their content, such as “machine 1 training to key personnel” rather than “learning module”, for example. At CSC, the users resisted the categorization of their genres provided by the researchers. These findings imply that the *language of the community* (cf. Swales, 1999), *including the substantial names, categorizations and hierarchies of content units, should be applied in analysis*. Concepts and names for units of content that refer to storage units in systems or to an application area, which is not familiar to users should be avoided.

The analysts observed that the adoption of domain-oriented language required efforts from them. *Glossaries of domain-oriented acronyms and concepts* were utilized for making sense of the domain-oriented language. It is commonly acknowledged that content analysis in organizations is complex, due to multiple systems and technologies already used in the domain. Therefore, the analysts' reflections about the complex domains were left out from research articles. Additional factors for complexity became apparent as the research was analyzed as a whole. The need to analyze and define the content on multiple levels, such as assemblies, aggregations, documents, and topics within them added the amount of analyzable content, and increased the complexity of the content to be addressed. The soft content at BIRD was also found quite complex to deal with.

### **Which aspects of the genres, and in what detail, should be scrutinized in content and requirements analysis?**

The aspects and details of analysis depend naturally on the objectives. As pointed out by some document researchers (e.g., Salminen, 2003a; Murphy 2001; Salminen, 2000), and genre researchers (e.g., Berkenkotter & Huckin, 1995; Yates & Orlikowski, 1992) the content should be analyzed in the organizational context. Hence the *genre-based metadata* that provides information about this context is needed. Examples of genre-based metadata and frameworks that can be utilized were given in research theme 1.

The concepts of *genre hierarchy* (cf. Crowston & Williams, 1997; Figure 2), *genre repertoire* (e.g., Swales, 1999; Orlikowski & Yates, 1994; Miller, 1994) and *genre system* (Bazerman, 1994; Orlikowski & Yates, 1994) were found to be useful conceptual tools for defining the focus of the analysis. At CSC, the repertoire of genres was defined. As local expert groups carried out the analysis, *variants* of genres were revealed. At Church, the system of genres was inspected. Some of the genres were defined in more detail by their topical content for the purpose of studying content assembly. At BIRD (Articles 3, 5, and 6), a *hierarchy* of selected training genres was defined with respect to training *genres*, *genre variants*, and their *topics*. The *hierarchy of training genres* was also utilized for *reasoning about the units of content for a system or for a specific application area*. Depending on the objectives for content management, there is a *tradeoff between the coverage and depth of the analysis, and the reasonable allocation of resources*. The coverage considers whether a whole repertoire or genres, or only a few of them are inspected. The depth considers if the genres of interest are studied only with respect to their names, or if the metadata values and the topical content of the genres of interest are scrutinized in detail. For example, if a contemporary content management system consists only of files and folders, it might not be feasible to define the topics within the repertoire of genres. Similarly, the *amount of metadata categories utilized* and the *number of values to be gathered can be adjusted* with regard to objectives. For example, at CSC, the metadata values would have been gathered for the main genres only, instead of gathering metadata about all variants, too.

At CSC, the identification of genres as units of content was sufficient for eliciting the requirements for a content management system. In studies at BIRD,

Church, and FCP, identification of topics and subtopics within genres was considered necessary and useful. The observation seems to imply that *in content and requirements analysis there is a need to define units of content that are smaller by their grain size than those of contemporary document management systems.*

At Church it was observed that the *users were unable to define the content of their genres below the level of subtopics.* They had difficulties when they were asked to reason about a content item within a topic or subtopic. The finding implies the following: *the line between the analysis carried out by the users and the design carried out by the analysts and designers is drawn along the topics and subtopics of genres.* This finding supports the finding of Doll and Deng (2001) that user-based, collaborative analysis can be feasible, while design carried out by the users may not. The observation made at Church implies that *the collaborative analysis carried out by users can be useful up to the point where the grain size of topics and subtopics of genres is reached.* User participation on a more detailed design may not provide further advantages.

### **What are the findings, if any, for the use of structured documents?**

The findings from previous research (Article 4) implied that *genres that are soft can be difficult to enact upon in computer systems.* The study of BIRD provides support for this finding. On the other hand, it was suspected that overly detailed enactment of contemporary genre rules and features in a system may block the evaluation and birth of genres. The previous research also implied that when a system or media is being implemented, the genres and rules for enacting them that are familiar for the user community should be adopted (e.g., Schulze & Boland, 1997; Yates, Orlikowski & Rennecker, 1997; Yates & Sumner, 1997).

At BIRD it was found that *enactment of genres in the SGML format was complex:* there were multiple DTDs and tools, and the process in which the O&M manuals were produced was complex. *Utilization of the SGML format, and especially the content-oriented DTD employed provided advantages for BIRD.* However, also *structured documents need to be supported by content management functionality.* Contemporary content management systems have so far managed documents as files (e.g., Boiko, 2002; Doculabs, 1998). The management of document content in structured documents has been proved feasible, but this capability has not yet been incorporated in commonly utilized office and content management tools. Support for XML is provided for the latest releases of content management (e.g., Documentum, 2003) and text processing systems (Cover, 2003) or expected to be provided for their future versions.

At Church (Article 7) the *collaborative analysis* carried out by the users *was utilized for an XML schema design.* Topic definitions allowed the researchers to design XML DTDs for the genre instances. The task was found to require expertise on XML schema design. At BIRD it was envisioned that genre templates could act as human-crafted definitions for designing XSLT (Clark, 1999) transformation templates, providing a possibility to automate a part of the content assembly process. An example of operationalizing an XSLT template was given. The findings seem to suggest that the user-based design may be adopted for

structured document schema design. At FCP it was noticed, that the hierarchy of genres defined was not adoptable to a content management system *per se*. At BIRD, the hierarchy of training genres was utilized as a base for defining conceptual units on eLearning field. The content schema based on genres was found to be comprehensive and timeless, and may be utilized in multiple contemporary technologies or systems.

### **How the genre theory and related research could support content analysis?**

Yates and Orlikowski (1992) suggested that genres as units of communication are independent from technologies and media. Therefore genres may be harnessed as analytical lens for studying the technologies and media in organizations (*ibid.*). The studies carried out in the four organizations and the findings of the literature review indicate that the theory of genres may also be adopted for analyzing the content in organizations. Genres as units of communication may be identified even if the content to be analyzed is vague by its features. The contemporary concepts adopted, such as documents, content blocks, or learning objects, or terms used for systems, such as virtual folders, or files may hinder the ability of people to express their requirements, or define their content with respect to these concepts. The findings of the studies indicate that the theory of genres may be utilized as a mediator between the world of the users and that of systems. The genres may be applied as an analytical lens for studying the language and the content used in an organization, and for comparing them with concepts utilized in varying application areas, and units of storage components in contemporary systems.

The findings of previous genre-based research provided important implications. The most influential of them were reported as the “genre lens” in Section 2 of this introduction. The findings seemed in favor of adopting techniques and tools for empowering the users for designing their genres. As a consequence, consensus-creating workshops, the wall-diagram technique, and simple, non-technical diagrams, such as tables and mind maps were utilized. The findings at BIRD, Church, and FCP implied that the use of these techniques might provide advantages. The communication breakdowns were avoided, at least partially.

The theory of genres provided useful conceptual tools for the analysis carried out in the study organizations. In addition to the concept of genre, the concepts of genre hierarchy, genre system, genre variant and genre topic, as well as the aspects of genres as genre-based metadata, such as the 5W1H, were operationalized. The concepts of soft and hard, and a local or common genre helped on adjusting the scope and focus for inquiries. In the studies carried out, most of the genres inspected were local. They were inspected in quite a detail, by topics and by subtopics.

The literature review revealed examples of other instantiations of the genre theory for multiple purposes in the content management area. For example, the review came up with studies in which requirements for content management systems were elicited, and with studies in which the information and content coordination between genres for designing document content work-

flows were located. The findings of the literature review confirmed that the theory of genres could be applied for studying content enacted in organizations. The findings of this study imply that the genre theory as an analytical tool provides both means and implications that may be advantageous, and help in overcoming some of the communicational obstacles in content analysis.

## 6.2 The Genre-based Methods

During this study two genre-based methods were explicated. Findings from the action research studies suggest that these may be considered as outcomes of the study itself. March and Smith (1995) have proposed that the basic question to be asked of an artifact or a method developed is whether it works. The genre-based method for requirements analysis (Method 1) was successfully instantiated at CSC. It allowed the definition of the contemporary content management at CSC. It also provided genre-based metadata values for eliciting requirements for a content management system. The genre-based method for content analysis (Method 2) was instantiated three times. The method was utilized to study requirements for content assembly, and to define the hierarchy of training content at BIRD. The hierarchy also provided a base for defining conceptual units of content for eLearning, and for defining a hierarchy of storage units for a system. At Church, the researchers were able to enrich the requirements definition prepared in the preceding study. At FCP, the content management system is adjustable for content enactment.

Tolvanen (1998, p. 212) has suggested a more thorough examination when evaluating a method. Four aspects of a method should be studied: 1) the conformity of the steps taken during the action taking stage with regard to the prior planning, 2) the problem solving capabilities of the method with regard to the problem setting, 3) the modeling capabilities of the method with regard to the problem setting, and 4) the rationale for choosing the method in the first place. The steps of Method 1 were carried out in an orthodox way, and its problem solving capabilities were considered successful. The rationale for the method was to study the adoption of the genre theory for a problematic setting in an organization. For Method 2, the steps taken with respect to the steps planned were in line with the enhanced method in the study of the church. At FCP the consultant enhanced and adjusted the method. The content analysis was carried in line with the enhanced plan. The objectives for the analysis were fulfilled. Both methods have poor modeling abilities with regard to detailed content and workflow design. It is anticipated, that experienced system analysts will be able to import the genre-based definitions of the domain into the modeling methods of their choice. Method 2 was developed as a consequence of a failed trial of applying a document analysis method for analyzing the content the features of which were found to be vague. Tolvanen (1998, p. 14) has been quite skeptical towards the applicability and usability of methods in common. He observed that numerous methods exist, yet companies create their own in-

house variants. He doubts whether any universal agreement about methods being useful in information systems development exist at all. The action carried out at FCP lends support to Tolvanen's observation. The consultant developed an in-house variant of Method 2. The method, and the implications provided by the "genre lens" defined, were feasible as tools for transferring knowledge that allowed the problematic task at hand to be solved.

The genre-based methods could be compared with document analysis or more common information systems analysis methods. Päivärinta (2001) has compared the genre-based method for information systems planning, which was enhanced from Method 1, with several methods for developing electronic document management in organizations. His comparison includes the RASKE methodology (Salminen, 2003a; 2000), a method of Sutton (1996), a method of Koulopoulos and Frappaolo (1995), a method of Bielawski and Boyle (e.g., Bielawski & Boyle, 1997), a method of Schäfer, Hirscheim, Harper et al. (e.g., 1998), and a method for Business Systems Planning by IBM (1984). According to Päivärinta (2001) the genre-based method differs from the others by its emphasis on user participation, hence the role of an analyst is less dominant than on the other methods. The genre-based method may be located in between the consultancy-based methods and the research-based methods with regard to the number and complexity of its basic concepts and modeling techniques (ibid.). Päivärinta (2001) also suggests that the genre-based method may complement the other methods for electronic document management in organizations.

To illustrate the range and capabilities of the genre-based methods a bit more, they are briefly discussed with regard to the RASKE (Salminen, 2003a; 2000) methodology for document analysis. It is quite evident that the genre-based methods developed are far more modest and prototypical by their nature than the RASKE methodology for example. The RASKE methodology uses a variety of modelling methods in multiple phases for describing the contemporary document management, and analyzing a future one. For example, it applies parts of OOA (Shlaer & Mellor, 1992) and office information systems modeling (Ellis, 1983; 1979) to the task at hand. In the methodology, the organizational context is studied with respect to user roles, and processes in which the information is produced and used, and with respect to technologies, software, and applications implemented. User interviews, literary sources, and exemplary documents are mainly utilized. For the analysis and design of structured document schemas, the methodology utilizes the method of Maler and ElAndaloussi (1996). Their method makes use of tables and graphical "Elm"-models for visualizing document schemas. The RASKE methodology and the method of Maler and ElAndaloussi were developed for the application areas in which the documents and document classes are quite stable and easily identifiable.

The potential of the genre-based methods with respect to these more developed methods is the ability to handle content that is more vague. The theory of genres allows identification and inspection of the content regardless of whether the content is considered as documents or not. In addition, the use of workshops and collaborative design enhances the abilities to overcome communication breakdowns. Users can negotiate about their requirements by using familiar,

domain-oriented language. The requirements can be expressed by utilizing simple tools, such as forms and wall-diagrams, which make them more explicit.

The findings of the studies, and the comparison carried out by Päivärinta (2001) indicate that the genre-based methods may be adapted for eliciting requirements for systems off-the-shelf and for analyzing the content of an organization for implementing one. Our findings imply that they may allow requirements for content assembly to be elicited. The genre-based methods may be used as sense-making tools, for example for diagnosing content and its enactment in domains that appear to be complex. One potential use for these methods is to merge them with other methods. For example, Lyytikäinen (2003) reports a requirements analysis study, in which she embedded a genre-based method for information systems planning (Päivärinta, Halttunen & Tyrväinen, 2001), developed from Method 1, in the RASKE methodology (Lyytikäinen, 2003). She pointed out that the workshop technique helped the users to focus on the domain in question, and also revealed that even in such a small organisation, there was no single person who could know all the details related to the content of the domain. The requirements analysis indicated that a genre system at Church needs to be scrutinized even in more detail. The genre-based method for content analysis (Method 2) was successfully operationalized for the purpose (Article 7).

### 6.3 Implications

Previous subsections studied the findings based on the data that was gathered from the four action research studies, and the literature review. This section takes a portion of the findings a bit further. The findings presented with respect to the research themes 2-5 in the previous subsections can be summarized as follows:

- The content analysis can be started from, and based on the units of organizational communication, i.e., on genres. The normative scope and the coverage and depth of analysis should be adjusted in an appropriate way (cf. research theme 3). The normative scope defines the community for analysis, and the breadth and width define the portion of hierarchy (Figure 2 on Section 2.1) of genre concepts utilized.
- The genres of an organization may be inspected by using the domain-oriented language of the users, including the names and categorizations of the genres given by users for hindering communication breakdowns. The user community should define the genres (cf. research theme 2).
- Users may define their communicational content, and the genre-based metadata successfully (cf. research themes 2 and 3).
- Content analysis may contain a mapping between the hierarchy of genres or a portion of it defined, and the communicational content of an organization expressed by a domain-oriented language. Another mapping may take place between the hierarchy of genres or a defined portion of it, and content management units adopted by a system or an application area to be implemented. Analysts or designers should carry out the mappings (cf. research themes 3, 4 and 5).

Altogether, these findings imply that the communicative content of an organization may be successfully captured in terms familiar for the user community.

The theory of genres and the hierarchy of genre-related concepts (cf. Figure 2, Sect. 2.1) may act as a go-between bridging the concepts of a user community and concepts of an application area or a system to be implemented (cf. research theme 5). Figure 5 (on the following page) illustrates the role of such a mediator, which the theory of genres and related research may adopt. It visualizes the key activities and resources related to content analysis.

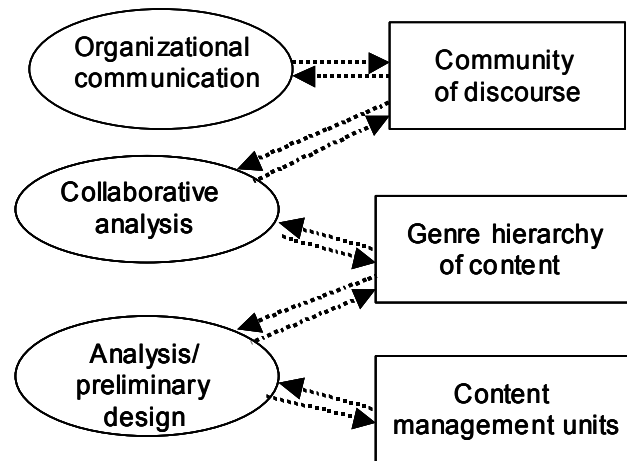


FIGURE 5 Key Activities And Resources Related to Content Analysis

In the figure circles depict activities. The resources are shown as rectangles. An arrow with a dashed line represents an information flow. The figure presents the key activities related to content analysis process in which the theory of genres is utilized. As illustrated, the focus of content analysis in organizations is on capturing the units of organizational communication of interest, with regard to the content management development focus in question. Organizational communication is managed by and evolves with a community of discourse. An organization can constitute one or several communities of discourse, depending on the scope of the analysis and communicative activities. Communicative units of content of an organization or a part of it can be identified and analyzed by using collaborative analysis. In the analysis the members of a community define their current and anticipated future communication needs with their own language, which may include domain-oriented concepts, acronyms, and terms. Appropriate techniques supporting collaboration are utilized. The analyst may learn about, and adopts the acronyms, terms and concepts, as well as the names used of the units of content in an organization while participating on the collaborative analysis, possibly as a facilitator. She or he applies the hierarchy of genre-related concepts (cf. Figure 2, Sect. 2.1) as a conceptual structure to capture the units of communicative content of an organization, or a relevant portion of it. The content captured can be analyzed with respect to units of content that are commonly used in system, such as files and folders. Another mapping may be needed if an off-the-shelf system is to be implemented, or if the content needs to be defined with terms and conceptual units of content adopted by a specific application area, such as eLearning. An analyst of designer can apply the hierarchy of genres or a portion



of communicational units of content defined as a base for studying the mappings needed between concepts and units of content for the new environment. In content management systems and in specific application areas there can be, at least three kinds of content management units:

- conceptual units used in an application area or in a system, such as “learning object”, “learning package”, or “document”,
- units that refer to a storage unit in a system, such as “file”, or “folder”, or “record”, and
- units that refer to a processing unit, such as “element of an XML document”, “attribute”, or “metadata field”.

The hierarchy of genre-related concepts can provide multiple advantages for the analysis process. A hierarchy of genres, or a portion of it, can be mapped against multiple types of content unit categories adopted by contemporary systems or by application areas. The use of genre-related concepts provides a conceptual structure for an analyst, allowing people in the communities use domain-oriented language for defining the content. Technical or system-specific concepts may be left unrevealed by the users. For example, the users may not be able to define what are “virtual folders” or “learning objects” in their domain. Yet they are capable of defining a set of statements (cf. the study at FCP), or topics of training session content (cf. the study at BIRD). The hierarchy of genre concepts provides a conceptual structure, which also allows relationships across units of content be inspected. It could turn out that a system or a technology adopted does not support the constellation of content units of an organization, and adjustments may be needed. Awareness of a gap between the communicational content and how it could be enacted in a system can reveal avenues for further development and cues for end-user training.

Findings for the research themes 1 and 5 indicate that the theory of genres may provide support for a variety of content analysis aims that were identified in the four collaborative organizations. The organizational environments were found to be quite complex, and the complexity may even grow further. The content assemblies for target audiences or for the Web require the units of content within genres to be defined. This may increase the number of units of content to be analyzed and managed in the organizations. The content is often delivered in multiple media in parallel, causing a part of digital content to be duplicated. Technologies and media continue to evolve. It is quite likely that the variety of content forms and formats, and mixtures of them continue to grow. Even though the theory of genres provides means for content analysis and for avoiding the communicational breakdowns between analysts and people, it is not capable for solving the myriads of challenges for content analysis in organizations.

## 6.4 Shortcomings of the Study

There were both research and practical objectives posed for this study. It considered many action research studies in different organizations. The research

was reported as a series of articles. The content of articles was modified to fit for multiple purposes. Some joint articles reflect multiple research orientations. Some articles reflect a theme of focus given by an organizer of the conference in which the article was published. Some articles promote the genre-based methods as an outcome, since the implications for content analysis considered only one domain. Some articles were targeted for forums that allowed fast publication, or did not require long or expensive conference trip to be made. Due to all these limitations, the included articles might not seem to constitute a coherent whole.

The action research method that was utilized may have influenced the selection of techniques for the methods. Action research (Kock, McQueen & Scott, 1997; Susman & Evered, 1978) typically emphasizes group work settings, since many of its phases are usually carried out with small groups including researchers and practitioners. The selection of the research method may have affected the selection of techniques, and put more emphasis on implications for collaborative design.

The genre theory and related concepts were found to be essential assets for analyzing the content of the four organizations described in the study. The included articles may reflect the fact that the task of content analysis in organizations is not a simple one. Organizations and their environments are complex. They may not be easy to get acquainted with, and adoption of the domain-oriented language may require a great deal of effort. Articles 2 and 3 may reflect the complexity of the domain at BIRD, and effort taken in order to make sense of the domain. At CSC the number of genres and metadata was large, and software tools may need to be developed to assist in recording and processing the data about the content analyzed. Regardless of the benefits the genre theory may offer for content analysis, the domains still remain complex. The amount and diversity of content, as well as the range of technologies and media applied for managing it continues to grow. Further research on content analysis in organizations, and development of tools to support it are needed.

There were many aspects of the genre theory that were left unutilized. For example, the intertwined connections between genres, i.e., the genre systems and their effect on analysis could have been studied in a more thorough manner. Content assembly has been previously studied in the genre area from the viewpoint of information coordination (Yoshioka, Yates & Orlikowski, 2001; Yoshioka & Herman, 2000). In these studies (*ibid.*), it was pointed out that genres, and genre systems especially, can share content with each other. The information coordination theory defines types of information coordination between the source and the target content. For example, a document instance can act as a source with respect to another, or two document instances can share the same document as their source. In these studies (Yoshioka, Yates & Orlikowski, 2001; Yoshioka & Herman, 2000), the information coordination types were adopted, and embedded to the 5W1H metadata framework. The coordination of information across the documents was made explicit for investigation by gathering metadata values about documents. The studies (Yoshioka, Yates & Orlikowski, 2001; Yoshioka & Herman, 2000) do not discuss how the units of content within

documents for information coordination could be identified. Yet the examples given by Yoshioka, Yates and Orlikowski (2001) indicate that they do not consider a document instance as a resource or target for another as a whole. Our findings of genre topics for content assembly may support the findings of Yoshioka, Yates and Orlikowski (2001) and Yoshioka and Herman (2000). On the other hand, the information coordination types discussed in these studies could have enhanced our studies of content assembly, and especially the metadata categories that were utilized.

The research carried out in the dissertation was motivated by the findings and views provided by the researchers of the North-American school of genres (Freedman & Medway, 1994). Among other issues, these researchers have emphasized the constant evaluation and change on genres. This aspect has not been operationalized on the content analysis studies included. For one thing, the changes in organizational communication are quite difficult to foresee while genres are analyzed. The evaluation may only be witnessed if a historical view on a genre repertoire, or about the genres of interest is available. It may be possible to solicit ideas about future changes from the users, as was done in the studies. Use and enablement of genre variants may provide, at least to some point, flexibility for changes and adjustments on the management of content in systems. Content analyst may add a requirement that system implementation should not restrict the enactment of new types of units of content, and means of communications. However, there is no way to ensure that the changes that remain unforeseen can be fully taken into consideration. It may be possible that the analysis process itself, and implementation of the system off-the-shelf, may restrict the natural development of communications, and hinder the enactment and evaluation of soft genres that emerge.

## 7 CONCLUSIONS AND AVENUES FOR FURTHER RESEARCH

*“.. my mama used to say, that  
life is like a box of chocolate  
– you never know what you’re going to get”  
Forrest Gump*

This Section concludes the dissertation by summarizing it. Avenues for further research are proposed.

### 7.1 Concluding Summary

This research was positioned on content management field, in which various types of content – such as documents and content on the Web are considered. The research was focused on the analysis phase of content management development projects in organizations. . The content analysis as a research area is relatively new; hence one aim of the study was to explore the kinds of focuses that exist for content analysis in the organizations. Previous research on requirements analysis provided findings of communication breakdowns that may hinder the analysis (e.g., Mumford, 2000; Byrd, Cossick & Zmud, 1992). For the study, a method for content analysis was needed as a research tool. The methods developed for document analysis, or more common methods for information systems analysis could have been applied (Päivärinta, 2000). Due to the communicative nature of some of the problems identified in requirements analysis, this study followed the direction provided by the theory of genres of organizational communication (Yates & Orlikowski, 1992). The other aim for this research was to explore the potential uses of the genre theory for content analysis.

The study that was carried out consisted of the literature review and four action research (Susman & Evered, 1978) studies. The studies were carried out

in four collaborative organizations of different types. One of them was a small, non-profit expert organization, one was a manufacturing company, one was a church, and one was an expert organization on the eGovernment field. The author was involved with action taking in three of the studies. A consultant conducted the study in the eGovernment organization. The results of the studies were reported in seven articles, which are included in this dissertation.

The findings of the four studies were discussed with respect to five research themes, which characterized the aims of the research further. Implications were also presented. In the collaborative organizations content analysis was needed for eliciting requirements for content management system, and for defining the organizations' content for implementation. Requirements for content assembly, and for migrating a part of the content to Web were also studied. It was observed that a great deal of content in one collaborative organization was not in the form of documents. In another study, the content of training was found as too vague to be analyzed with a document analysis method. Other findings imply that domain-oriented language and concepts should be emphasized in the analysis, and that the users should be allowed to define the names and categories for their content themselves. The focus of content analysis may require covering units of content the grain size of which is smaller than, for example, that for a document. The ability of users to define their content may demarcate the field of content analysis from that of content design. It was found that users were not able to define content units smaller than topics and subtopics within genres they used.

During the study, two genre-based methods were developed. These may be applied as sense-making tools in complex environments. They may also be embedded to other methods. The "genre lens" that was explicated harnesses a portion of previous genre-based research and the implications provided. It can provide support for adapting the genre-based methods. The "genre lens" implied for collaborative analysis of content, carried out by the users themselves. Hence consensus-creating workshops (Coughlan, Lycett & Macredie, 2003), wall-diagram technique (Saaren-Seppälä, 1997), mind maps (Mindjet, 2003), and jointly filled questionnaire forms were utilized in the genre-based methods. The use of the techniques was observed to reduce the possibilities for communication breakdowns, at least in some cases. This also allowed the users to negotiate, reach consensus, and express their requirements in an explicit way.

The theory of genres and related concepts can be adopted by an analyst as conceptual structure. This structure may act as mediator between the content and concepts used by the user community and the units of content and concepts used in systems available, or in certain application areas. For genre researchers, the studies provide examples of local genres, which are defined by their users. This kind of genre research was seen to be lacking (Berkenkotter & Huckin, 1995; Yates & Orlikowski, 1992). Findings of the study confirm the suggestion of Berkenkotter & Huckin (1995); the local genres were allowed to be inspected in a great detail, including the topics and subtopics of genres. Yates and Orlikowski (1992) proposed that the genres are independent of technologies and media, and can therefore provide an analytical tool for studying organizational

communication. This findings of this research indicated that the genre theory may be harnessed for content analysis, too. It may also be applied for analyzing the requirements for the use of technologies, systems and media. The genre-based research findings implied for collaborative analysis of content, which was found to help in avoiding the communication breakdowns between people of the user community and the analyst, at least partially. The content analysis in organizations is complex and the complexity may even grow in the future. The theory of genres cannot provide solutions for the myriads of challenges of content analysis in organizations, but it may help to overcome a portion of them.

## 7.2 Avenues for Further Research

Some of the limitations of this study imply for further research. The findings of this study could be enhanced and tested in further studies. In these studies, the concept of the genre system may be studied more thoroughly. The genre-based methods may also be enhanced. Elicitation of requirements for, and the analysis of content assembly could be enhanced with the information coordination aspects as suggested by Yoshioka and Herman (2000) and Yoshioka, Yates and Orlikowski (2001). Tools and software support for the data gathering and processing on content analysis in general, and on the two methods might be developed. Further studies on the content analysis in organizations are needed. Challenges posed, and means developed for overcoming them are needed on the field of content analysis in general. Future research might also entail different kinds of organizations and their content management endeavors, as data available for comparison of different kinds of domains is still meager.

Some of the studies carried out could be continued with another analysis. For example, the genre repertoire at CSC could be re-analyzed with the genre-based method (Method 1), and the names and numbers of genres, as well as the genre-based metadata might be compared with the ones gathered on the previous study. The impacts of document management system implementation might also be scrutinized; especially from the viewpoint of genre evaluation. An evaluation could be made to inspect, whether a system implementation has restricted or enforced genre evaluation, and emergence of soft genres.

Another potential avenue for further genre-based studies might be to juxtapose the theory of genres with the research on communities of practice. In the studies carried out in this research the normative scope was quite narrow. The communities of discourse as groups of users were akin to communities of practice, which are studied on the area of organizational learning and knowledge management (e.g., Wenger, 1998; Brown & Duguid, 1996; Brown & Gray, 1995). A *Community of practice* can be characterized by three dimensions: it shares the *repertoire of routines*, concepts and symbols in a *joint enterprise*, which is supported by a *mutual engagement* (Wenger, 1998). For example, Wenger (1998) studied the normative rules of a community of practice as re-enforcers of the community, or as triggers for launching the members into separate trajectories.

The study bears an inherent connection to the Giddens's (1984, e.g., Rose & Scheepers, 2001) structuration theory, on which the North-American school of genres (Freedman & Medway, 1994) is anchored. The studies about communities of practice as knowledge-creating communities have brought into daylight the role of documents in this process, as instantiations of communities' norms, and capturers of knowledge (Brown & Duguid, 1996). Murphy (2001) explicated the functions of documents in communities of practice as being three-fold. Documents are animated community-builders, border-patrollers, and coordinators of work. A word-game, in which the contemporary concept of a document would be replaced with that of a genre, may lead to interesting consequences. For example, the features of the genres as instantiations of community rules and norms could be more thoroughly examined by using the genre aspects. The knowledge possessed by, and the ways for expressing it in a community of practice could be investigated by studying topical schemes of genre instances. For example, the implications for background knowledge can imply what is considered as the "known" or "state of affairs" in a community (Giltrow, 1994). The boundary objects refined as boundary genres may allow the features of boundary objects to be investigated, with a potential for implications for the design of interorganizational or inter-community content exchange (cf. Tyrväinen & Päivärinta, 1998). For theory of genres, the concept of community of practice may open up new ways for reasoning about a normative scope for inquiries, at least for the part of local genres. The roles of genres with respect to a community of users could be studied by adopting the three dimensions proposed by Murphy (2001).

Further research might also entail embedding the genre-based content analysis method (Method 2) into the method of Maler and ElAndaloussi (1996) for SGML/XML (Goldfarb, 1990; Bray, Paoli, Sperberg-McQueen et al., 2000) DTD design. The convergence of these methods might provide schema designers a way for importing user-crafted definitions of the content. For example, a topic definition of a genre may be adopted for a document component hierarchy. A component hierarchy is a base for further detailed design of structured document elements and attributes.

## REFERENCES

- ADLI. 2002. Advanced Distributed Learning Initiative. Advanced Distributed Learning Initiative. <http://www.adlnet.org/>. [20 Aug. 2002]
- Ahovaara, R. 2002. Yleiskirjehankkeen 1. Osaprojekti 10.01.-07.06.2002. Helsinki: Eläketurvakeskus.
- Ahovaara, R. 2003. Väliraportti: Yleiskirjehanke, 2.Vaiheen Osaprojekti: Palvelun Sisällön Määrittely Ja Suunnittelu. Versio 1.1. Helsinki: Eläketurvakeskus.
- Askehave, I. 1999. Communicative Purpose as Genre Determinant. *Hermes - Journal of Linguistics* 23, 13-23.
- Askehave, I. & Swales, J.M. 2001. Genre Identification and Communicative Purpose: A Problem and a Possible Solution. *Applied Linguistics* 22 (2), 195-212.
- Barry, R.E. 1993. Electronic Document and Records Management Systems: Towards a Methodology for Requirements Definition. In *Document & Management 93, Proceedings of the Document Management Conference*. London.
- Baskerville, R. & Wood-Harper, A.T. 1998. Diversity in Information Systems Action Research Methods. *European Journal of Information Systems* 7, 90-107.
- Bazerman, C. 1994. Systems of Genres and the Enactment of Social Intentions. In A. Freedman & P. Medway (Eds.). *Genre and the New Rhetoric*. London: Taylor & Francis, London, U.K., 79-101.
- Berkenkotter, C. & Huckin, T.N. 1995. *Genre Knowledge in Disciplinary Communication: Cognition/Culture/Power*. Hillsdale, New Jersey, Hove, UK: Lawrence Erlbaum Associates, Inc.
- Bhatia, V.K. 1993. *Analysing Genre: Language Use in Professional Settings*. London: Longman.
- Bielawski, L. & Boyle, J. 1997. *Electronic Document Management Systems: A User Centered Approach for Creating, Distribting and Managing Online Publications*. Upper Saddle River NJ: Prentice Hall.
- Blair, D., C. 2002a. The Challenge of Commercial Document Retrieval, Part I: Major Issues, and a Framework Based on Search Exhaustivity, Determinacy of Representation and Document Collection Size. *Information Processing & Management* 38 (3), 273-291.
- Blair, D., C. 2002b. Exemplary Documents: A Foundation for Information Retrieval Design. *Information Processing & Management* 38 (3), 363-379.
- Blake, W. 1863. *Auguries of Innocence (Excerpt)*. RPO Editors, Department of English, and University of Toronto Press 1994-2002. <http://eir.library.utoronto.ca/rpo/display/poem161.html>. [30 Oct. 2003]
- Blokdijk, A. & Blokdijk, P. 1987. *Planning and Design of Information Systems*. London: Academic Press.



- Boiko, B. 2002. Content Management Bible. New York, U.S.A: Hungry Minds, Inc.
- Bray, T., Paoli, J., Sperberg-McQueen, C.M. & Maler, E.E. 2000. Extensible Markup Language (XML) 1.0. (2nd Edition). W3c Recommendation. 1st Version 10 Feb. 1998, Revised 6 Oct. 2000. W3C Consortium. <http://www.w3.org/TR/2000/REC-xml-20001006>. [1 April 2003]
- Brown, J.S. & Gray, E.S. 1995. The People Are the Company. Fast Company, Com, The Magazine. <http://www.fastcompany.com/online/01/people.html>. [04 Feb. 2002]
- Brown, J.S. & Duguid, P. 1996. The Social Life of Documents. First Monday 1 (1).
- Burrell, G. & Morgan, G. 1979. Sociological Paradigms and Organizational Analysis. London: Heinemann.
- Byrd, T.A., Cossick, K.L. & Zmud, R.W. 1992. A Synthesis of Research on Requirements Analysis and Knowledge Acquisition Techniques. Management Information Systems Quarterly 16 (1), 117-138.
- Clark, J., Ed. 1999. XSL Transformations (XSLT) Version 1.0. W3C Recommendation 16 Nov. 1999. W3C Consortium. <http://www.w3.org/TR/xslt>. [2 April 2003]
- Coughlan, J., Lycett, M. & Macredie, R.D. 2003. Communication Issues in Requirements Elicitation: A Content Analysis of Stakeholder Experiences. Information and Software Technology 45 (1), 525-537.
- Cover, R. 2003. Technology Reports: Microsoft Office 11 and Infopath [Xdocs]. OASIS. <http://xml.coverpages.org/microsoftXDocs.html>. [15 Aug., 2003]
- Crowston, K. & Williams, M. 1997. Reproduced and Emergent Genres of Communication on the World-Wide Web. In R. H. Sprague (Ed.) Proceedings of the 30th Annual Hawaii International Conference on System Sciences. IEEE Computer Society Press, Los Alamitos, CA, U.S.A., 30-39.
- Crowston, K. & Williams, M. 1999. The Effects of Linking on Genres of Web Documents. In R. H. Sprague (Ed.) Proceedings of the 32nd Annual Hawaii International Conference on System Sciences. Genre in Digital Documents-Minitrack. IEEE Computer Society Press, Los Alamitos, CA, HICSS Digital Library. p. 2006, 10 pp. <http://www.computer.org/proceedings/hicss/0001/00012/00012006abs.htm> [15 Aug. 2003]
- CSC. 2003a. Csc Tieteen Tieteen Tietotekniikan Keskus. Espoo: CSC Tieteen Tietotekniikan keskus. <http://www.csc.fi/suomi/info/yhteystiedot.phtml.fi>. [15 Aug. 2003]
- CSC. 2003b. CSCn Vuosikertomus 2002. Annual Report 2002. Espoo: CSC Tieteen Tieteen Tietotekniikan keskus. [http://www.csc.fi/suomi/info/vuosikertomus\\_2002.pdf](http://www.csc.fi/suomi/info/vuosikertomus_2002.pdf). [15 Aug. 2003]
- Curtis, K., Foster, P.W. & Stentiford, F. 1999. Metadata - the Key to Content Management Services. In Proceedings of the Meta-Data '99. Third IEEE Meta-Data Conference. Bethesda, Maryland: National Institutes of Health. <http://www.computer.org/proceedings/meta/1999/paper-index.html>. [15 Aug. 2003]

- Dervin, B. 1992. From the Mind's Eye of the User: The Sense-Making Qualitative-Quantitative Methodology. In J. D. Glazier & R. R. Powell (Eds.). *Qualitative Research in Information Management*. Englewood (CO): Libraries Unlimited, Inc., 61-84.
- Doculabs, I. 1998. *Special Report on Document Management Products*. Second Edition (2.1). Chicago, Illinois: Doculabs, Inc.
- Documentum. 2003. Documentum Web Site. Documentum, Inc. <http://www.documentum.com/products/editions/dam/index.htm>. [15 Aug. 2003]
- Doll, W.J. & Deng, X. 2001. The Collaborative Use of Information Technology: End-User Participation and System Success. *Information Resources Management Journal* 14 (2), 6-16.
- Dublincore. 1999. Dublin Core Metadata Element Set, Version 1,1. Dublin Core Metadata Initiative. <http://purl.org/DC/documents/rec-dces-19990702.htm>. [17 Feb. 2000]
- Ellis, C.A. 1979. Information Control Nets: A Mathematical Model of Office Information Flow. *Proceedings of the Conference on Simulation, Measurement and Modeling of Computer Systems* (special issue). ACM SIGMETRICS Performance Evaluation Review 8 (3), 225-238.
- Ellis, C.A. 1983. Formal and Informal Models of Office Activity. In R. E. A. Mason (Ed.) *Proceedings of the IFIP 9th World Computer Congress, Paris, France, September 19-23*. New York: North-Holland.
- Eläketurvakeskus. 2003a. Eläketurvan Keskus. (The Finnish Center for Pensions) Esite/Brochure. Helsinki: Eläketurvakeskus.
- Eläketurvakeskus. 2003b. Vuosikertomus 2002. (The Finnish Center for Pensions, Annual Report) Helsinki: Eläketurvakeskus. <http://www.etk.fi/suomi/pdf/Vuosikertomus%202002.pdf>. [15 Aug. 2003]
- Eriksen, L.B. & Ihlström, C. 2000. Evolution of the Web News Genre - the Slow Move Beyond the Print Metaphor. In R. H. Sprague (Ed.) *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences (HICSS)*. Los Alamitos CA: IEEE Computer Society, p. 3014, 10 pp. HICSS Digital Library <http://computer.org/proceedings/hicss/0493/04933/0493toc.htm>. [15 Aug. 2003]
- ETK. 2003. Eläketurvakeskuksen Www-Sivusto. (The Finnish Center for Pensions, WWW site) Eläketurvakeskus. <http://www.etk.fi>. [20 May 2003]
- Fallside, D.C.E. 2001. XML Schema Part 0: Primer. W3c Recommendation, 2 May 2001. W3C Consortium. <http://www.w3.org/TR/xmlschema-0/>. [16 June, 2001]
- FCP. 2003. The Central Pension Security Institute: The Finnish Centre for Pensions. The Finnish Centre for Pensions. <http://www.etk.fi/english/etusivu.asp>. [25 Aug 2003, 2003]
- Freedman, A. & Medway, P. 1994. Locating Genre Studies: Antecedents and Prospects. In A. Freedman & P. Medway (Eds.). *Genre and the New Rhetoric*. London, U.K.: Taylor & Francis, London, U.K., 1-22.

- Galliers, R.D. 1992. Choosing Information Systems Research Approaches. In R. D. Galliers (Ed.) *Information Systems Research*. Oxford: Blackwell, 144-162.
- Giddens, A. 1984. *The Constitution of Society. Outline of the Theory of Structuration*. Berkley: University of California Press.
- Giltrow, J. 1994. Genre and the Pragmatic Concept of Background Knowledge. In A. Freedman & P. Medway (Eds.). *Genre and the New Rhetoric*. London: Taylor & Francis, London, U.K., 155-180.
- GLC. 2002. Ims Global Learning Consortium, Inc. IMS Global Learning Consortium, Inc. <http://www.imsproject.org/aboutims.cfm>. [20 Aug. 2002]
- Goldfarb, C.F. 1990. *The SSML Handbook*. Oxford, UK: Oxford University Press.
- Gordon, M.D. 1997. It's 10 A.M. Do You Know Where Your Documents Are? The Nature and Scope of Information Retrieval Problems in Business. *Information Processing & Management* 33 (1), 107-212.
- Heikkinen, B. 2000. *Generalization of Document Structures and Document Assembly*. Helsinki: University of Helsinki. Ph.D. Thesis.
- Hertzum, M. & Pejtersen, A.M. 2000. The Information-Seeking Practices of Engineers: Searching for Documents as Well as for People. *Information Processing & Management* 36 (5), 761-778.
- Hillesund, T. 2002. Many Outputs - Many Inputs: XML for Publishers and E-Book Designers. *Journal of Digital Information* 3 (1), 22 p. <http://jodi.ecs.soton.ac.uk/Articles/v03/i01/Hillesund/> [20 Aug. 2002]
- Hirschheim, R., Klein, H., K. & Lyytinen, K. 1996. Exploring the Intellectual Structures of Information Systems Development: A Social Action Theoretical Analysis. *Accounting, Management and Information Technologies* 6 (1/2), 1-64.
- Holmberg, B. 1989. *Theory and Practice of Distance Education*. London, New York: Routledge.
- Honkaranta, A. 2003. Developing Document and Content Management in Enterprises Using a "Genre Lens". In O. Camp, J. Filipe, S. Hammoudi & M. Piattini (Eds.). *Proceedings of the 5th International Conference on Enterprise Information Systems*. Setúbal, Portugal: Escola Superior de Tecnologia do Instituto Politécnico de Setúbal, 334-340.
- Hummingbird, 2003. WWW Homepage. <http://www.hummingbird.com>. [15 Aug. 2003]
- IBM. 1984. *Business Systems Planning: Information Systems Planning Guide (Ge20-0527-4)*. Atlanta: International Business Machines Corp.
- Ingwersen, P. 1992. *Information Retrieval Interaction*. London: Taylor Graham Publishing.
- Järvinen, P. 2001. *On Research Methods*. Tampere, Finland: Opinajan kirja.
- Karjalainen, A. 1997a. *Etäoppimateriaalin Rakenteistaminen*. Computer Science and Information Systems. Jyväskylä: University of Jyväskylä. Master's thesis.
- Karjalainen, A. 1997b. *Structured Hypercourseware for Distance Learning*. In *The Proceedings of Enable 97. Enabling Network-Based Learning Confer-*

- ence. Espoo: Espoo-Vantaa Institute of Technology and the Learning Info-  
bahn Project, WWW-publication. [http://www.enable.evitech.fi/  
enable97/submissions/anne.karjalainen/paper.html](http://www.enable.evitech.fi/enable97/submissions/anne.karjalainen/paper.html) [15 Aug. 2003]
- Karjalainen, A. 1999a. Dokumenttien Hallintaan Liittyvät Standardit. Metodi-  
Projektin Raportti. (Standards Related to Document Management. Report,  
Metodi project.) Jyväskylä: Jyväskylän yliopisto, Tietotekniikan tutkimus-  
instituutti, Elektronisten dokumenttien hallinnan tutkimusalue.
- Karjalainen, A. 1999b. Dokumenttien Hallintasovellukset: Arviointiviitekehys ja  
Sovelluskartoitus. Metodi Projektin Raportti. (Framework for Document  
Management Systems. Report, Metodi project.) Jyväskylä: Jyväskylän yli-  
opisto, Tietotekniikan tutkimusinstituutti, Elektronisten dokumenttien  
hallinnan tutkimusalue.
- Karjalainen, A., Päivärinta, T., Tyrväinen, P. & Rajala, J. 2000. Genre-Based  
Metadata for Enterprise Document Management. In R. H. Sprague (Ed.)  
Proceedings of the 33rd Annual Hawaii International Conference on Sys-  
tem Sciences (HICSS). Los Alamitos CA: IEEE Computer Society.
- Karjalainen, A. & Tyrväinen, P. 2001. Defining Genres and Their Features for  
Studying Information Reuse: Preliminary Findings of a Case Study of  
Training Materials. In M. Khosrowpour (Ed.) Managing Information Tech-  
nology in a Global Economy. Proceedings of Information Resources Man-  
agement Association 2001 Conference (21-23.5.2001 Toronto, Canada).  
Toronto: Idea Group Publishing, Hershey, U.S.A, 346-348.
- Kock, N.F., Jr. 1997. Myths in Organizational Action Research: Reflections on a  
Study of Computer-Supported Process Design Groups. *Organizations &  
Science* 4 (9), 65-91.
- Kock, N.F., Jr., McQueen, R.J. & Scott, J.L. 1997. Can Action Research Be Made  
More Rigorous in a Positivist Sense? The Contribution of an Iterative Ap-  
proach. *Journal of Systems & Information Technology* 1 (1), 1-24.
- Koulopoulos, T., M. & Frappaolo, C. 1995. Electronic Document Management  
Systems. A Portable Consultant. New York: McGraw-Hill, Inc., U.S.A.
- Krippendorff, K. 1980. Content Analysis: An Introduction to Its Methodology.  
Beverly Hills, CA: Sage.
- Lassila, O. & Swick, R.R. 1999. Resource Description Framework (RDF) Model  
and Syntax Specification. W3C Recommendation 22 February 1999.  
<http://www.w3.org/TR/REC-rdf-syntax/>. [8 March 2001]
- Lindén, G. 1997. Structured Document Transformations. Helsinki: Department  
of Computer Science, University of Helsinki. Ph.D. Thesis.
- LTSC. 2002. IEEE Learning Technology Standards Committee (LTSC). IEEE  
Computer Society. <http://grouper.ieee.org/groups/ltsc/index.html>. [20  
August, 2002]
- Lyytikäinen, V. 2003. Analysing Requirements for Content Management. In O.  
Camp, J. Filipe, S. Hammoudi & M. Piattini (Eds.). Proceedings of the 5th  
International Conference on Enterprise Information Systems, Angers,  
France, April 23-26, 2003. Portugal: Escola Superior de Tecnologia do Insti-  
tuto Politécnico de Setúbal, 104-111.

- Maler, E. & El Andaloussi, J. 1996. Developing SGML DTDs. From Text to Model to Markup. Upper Saddle River (NJ): Prentice Hall.
- March, S.T. & Smith, G.F. 1995. Design and Natural Science Research on Information Technology. *Decision Support Systems* 15, 251-266.
- Mcadams, M. 1995. Inventing an Online Newspaper. *Interpersonal Computing and Technology: An Electronic Journal for the 21st Century*. Published by Georgetown University, Washington, U.S.A., Center for Teaching and Technology. 3 (3), 64-90. <http://www.sentex.net/~mmcadams/invent.html> [10 Aug 2002]
- Maingueneau, D. 2002. Analysis of an Academic Genre. *Discourse Studies* 4 (3), 319-342.
- Mcintosh, M. 2000. Content Management Using the Rational Unified Process®. Rational Software White Paper. Cupertino, CA: Rational Software Corporation. <http://www.rational.com/media/products/rup/TP164.pdf>. [15 Aug 2003]
- Medina, R., Meyers, S., Bragg, J. & Klima, C. 2002. Doculabs Evaluates Document Management for Enterprise Content Management. [http://www.transformmag.com/db\\_area/archs/2002/02/tfm0202f1.shtml?contentmanagement](http://www.transformmag.com/db_area/archs/2002/02/tfm0202f1.shtml?contentmanagement). [2 April, 2003]
- Meier, J. & Sprague, R.H. 1996. Towards a Better Understanding of Electronic Document Management. In R. H. Sprague (Ed.) *Proceedings of the 29th Annual Hawaii International Conference on System Sciences. Digital Documents Track*. Maui (HA): IEEE Computer Society Press, 53-61.
- Microsoft. 2002. Microsoft LRN 3.0 Toolkit (Learning Resource Interchange). Microsoft, Inc. <http://www.microsoft.com/elearn/support.asp>. [20 Aug. 2002]
- Miller, C., R. 1994. Genre as Social Action. In A. Freedman & P. Medway (Eds.). *Genre and the New Rhetoric*. London: Taylor & Francis, London, U.K., 23-42.
- Mindjet. 2003. Mindmanager Software. WWW site of Mindjet, Inc. <http://www.mindjet.com/eu/>. [15 Aug. 2003]
- Mumford, E. 2000. A Socio-Technical Approach to Systems Design. *Requirements Engineering* 5 (2), 125-133.
- Murphy, L.D. 1998. Digital Document Metadata in Organizations: Roles, Analytical Approaches, and Future Research Directions. In R. H. J. Sprague (Ed.) *Proceedings of the 31st Annual Hawaii International Conference on System Sciences (HICSS)*. Los Alamitos, CA: IEEE Computer Society, 267-276.
- Murphy, L.D. 2001. Digital Documents in Organizational Communities of Practice: A First Look. In R. H. Sprague (Ed.) *Proceedings of the Thirty-Fourth Hawaii International Conference on System Sciences*. Los Alamitos, CA: IEEE Computer Society, CD ROM.
- Orlikowski, W.J. & Baroudi, J.J. 1991. Studying Information Technology in Organizations: Research Approaches and Assumptions. *Information Systems Research* 2 (1), 1-28.

- Orlikowski, W.J. & Yates, J. 1994. Genre Repertoire: The Structuring of Communicative Practices in Organizations. *Administrative Science Quarterly* 39 (4), 541 - 574.
- Orlikowski, W.J. & Yates, J. 1998. Genre Systems: Structuring Interaction through Communicative Norms. MIT Sloan School of Management. <http://ccs.mit.edu/papers/CCSWP205/>. [20 Aug. 2002]
- Overmyer, S.P. 2000. What's Different About Requirements Engineering for Web Sites? *Requirements Engineering* 5 (1), 62-65.
- Peltola, T. 2003. Yleiskirjehankkeen Sisällön Määrittelyosio-Väiliraportti. Jyväskylä, Helsinki: Done Information Oy.
- Phillips, J. & Greenfield, A. 2002. Content Management Demystified. Platts, Inc. The McGraw-Hill Company, New York. [http://www.platts.com/businessstech/issues/0210/0210ebt\\_contentman.shtml](http://www.platts.com/businessstech/issues/0210/0210ebt_contentman.shtml). [2 April 2003]
- Pirsig, R., M. 1974. *Zen and the Art of Motorcycle Maintenance*. NY, USA: Bantam New Age.
- Polsani, P., R. 2003. Use and Abuse of Reusable Learning Objects. *Journal of Digital Information* 3 (4), 10. <http://jodi.ecs.soton.ac.uk/Articles/v03/i04/Polsani/> [15 Aug. 2003]
- Päivärinta, T. & Tyrväinen, P. 1998. Documents in Information Management: Diverging Connotations of "a Document" in Digital Era. In M. Khosrowpour (Ed.) *Proceedings of the 9th Information Resource Management Association International Conference*. Boston, M.A.: Idea Group Publishing, Hershey, PA, U.S.A., 163-173.
- Päivärinta, T. 1999. A Genre Approach to Applying Critical Social Theory to Information Systems Development. In C. H. J. Gilson, Grugulis, I., Willmot, H. (Ed.) *Proceedings of the 1st Critical Management Studies Conference; Information Technology and Critical Theory - Stream*. Manchester: WWW-resource, 24 p. <http://www.mngt.waikato.ac.nz/ejrot/cmsconference/documents/Information%20Tech/A%20Genre%20Approach%20to%20Applying%20Critical%20Social%20Theory.pdf>. [15 Aug. 2003]
- Päivärinta, T. 2000. *Towards a Genre-Based, Critical Approach to Enterprise Document Management*. Department of Computer Science and Information Systems. Jyväskylä: University of Jyväskylä. Lic. Thesis.
- Päivärinta, T. 2001. *A Genre-Based Approach to Developing Electronic Document Management in the Organization*. Jyväskylä: University of Jyväskylä. Ph.D. Thesis.
- Päivärinta, T., Halttunen, V. & Tyrväinen, P. 2001. A Genre-Based Method for Information Systems Planning. In M. Rossi & K. Siau (Eds.). *Information Modelling in the New Millennium*. Hershey, PA: Idea Group Publishing, Inc., 70-93.
- QPRsoftware. 2003. Qpr Process Guide. QPR Software. <http://www.qpr.com/Products/QPRProcessGuide/index.html>. [15 Aug. 2003]
- Raggett, D., Le Hors, A. & Jacobs, I. 1999. HTML 4.01 Specification. W3c Recommendation 24 Dec 1999. World Wide Web Consortium. <http://www.w3.org/TR/html4/>. [10 Jan. 2003]

- Rehm, G. 2002. Towards Automatic Web Genre Identification. A Corpus-Based Approach in the Domain of Academia by Example of the Academic's Personal Homepage. In R.H. Sprague, Jr. (Ed.) Proceedings of the 35th Annual Hawaii International Conference on System Sciences. Los Alamitos, CA: IEEE Computer Society Press, 101, 10 pp. HICSS Digital Library. <http://csdl.computer.org/comp/proceedings/hicss/2002/1435/04/14350101abs.htm>. [15 Aug. 2003]
- Rose, J. & Scheepers, R. 2001. Structuration Theory and Information Systems Development - Frameworks for Practice. In S. Smithson & S. Avgerinou (Eds.). European Conference on Information Systems. Bled, Slovenia. <http://www.cs.auc.dk/~jeremy/Publications.htm>. [15 Aug. 2003]
- Saarela, J. 1999. The Role of Metadata in Electronic Publishing. Acta Polytechnica Scandinavica, Mathematics and Computing Series No102. Espoo, Finland: Finnish Academy of Technology. Ph.D. Thesis.
- Saaren-Seppälä, K. 1997. Seinätekniikka Prosessien Kehittämisessä. (Using the Wall-Chart Technique in Process Development). Kari Saaren-Seppälä Ltd., Finland.
- Salminen, A. 2000. Methodology for Document Analysis. In A. Kent (Ed.) Encyclopedia of Library and Information Science. New York: Marcel Dekker, Inc., 299-320.
- Salminen, A., Lyytikäinen, V. & Tiitinen, P. 2000. Putting Documents into Their Work Context in Document Analysis. Information Processing & Management 36 (4), 623-641.
- Salminen, A. 2003a. Document Analysis Methods. In C. L. Bernie (Ed.) Encyclopedia of Library and Information Science. NY, USA: Marcel Dekker, Inc., 916-927.
- Salminen, A. 2003b. Document Management Research. Description of Research Projects at the University of Jyväskylä. <http://www.cs.jyu.fi/~airi/docman.html>. [15 Aug. 2003]
- Savolainen, R. 1993. The Sense-Making Theory: Reviewing the Interests of a User-Centered Approach to Information Seeking and Use. Information Processing & Management 29 (1), 13-28.
- Savolainen, R. 1999. The Role of the Internet in Information Seeking. Putting the Networked Services in Context. Information Processing & Management 35 (6), 765-782.
- Schamber, L. 1996. What Is a Document? Rethinking the Concept of Uneasy Times. Journal of the American Society for Information Science 47 (9), 669-671.
- Schulze, U. & Boland, R., Jr. 1997. Hard and Soft Information Genres: An Analysis of Two Notes Databases. In R. H. Sprague (Ed.) Proceedings of the 30th Annual Hawaii International Conference on System Sciences. Digital Documents Track. Maui (HA): IEEE Computer Society Press, 40 - 49.
- Schäfer, G., Hirscheim, R., Harper, M., Hansjee, R., Domke, M. & Bjørn-Andersen, N. 1998. Functional Analysis of Office Requirements: A Multi-perspective Approach. Chichester: Wiley.

- Shepherd, M. & Watters, C. 1998. The Evolution of Cybergenres. In R. H. Sprague (Ed.) Proceedings of the 31th Annual Hawaii International Conference on System Sciences. Digital Documents Track. Maui (HA): IEEE Computer Society Press, 97-109.
- Shepherd, M. & Watters, C. 1999. The Functionality Attribute of Cybergenres. In R. H. Sprague (Ed.) Proceedings of the 32th Annual Hawaii International Conference on System Sciences. Digital Documents Track. Maui (HA): IEEE Computer Society Press, p.2007. HICSS Digital Library. <http://www.computer.org/proceedings/hicss/0001/00012/00012007abs.htm?SMSESSION=NO> [15 Aug. 2003]
- Shlaer, S. & Mellor, S., J. 1992. Object Lifecycles: Modeling the World in States. Englewood Cliffs (NJ): Yourdon Press.
- Silverman, D. 1993. Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction. London: Sage Publications.
- Sprague, R.H. 1995. Electronic Document Management: Challenges and Opportunities for Information Systems Manager. MIS Quarterly 19 (1), 29-49.
- Star, S.L. & Greisner, J. 1989. Institutional Ecology, "Translations", and Coherence: Amateurs and Professionals in Berkley's Museum of Verterbrate Zoology, 1907-1939. In Social Studies of Science. London: Sage Publications, 387-420.
- Staunton, R., Tougard, J. -F., Richardson, J. & Carnelley, P. 1998. Guide to Understanding and Selecting Document Management Systems. Ovum Evaluates Document Management. London: Ovum Ltd.
- Strauss, A. & Corbin, J. 1990. Basics of Qualitative Research. Grounded Theory Procedures and Techniques. Newbury Park: Sage Publications, Inc. California, U.S.A.
- Susman, G.I. & Evered, R.D. 1978. An Assessment of the Scientific Merits of Action Research. Administrative Science Quarterly 23 (4), 582-603.
- Sutton, M.J.D. 1996. Document Management for the Enterprise: Principles, Techniques and Applications. Wiley and Sons, Inc. New York, U.S.A.
- Swales, J.M. 1999. Genre Analysis. English in Academic and Research Settings. Cambridge: Cambridge University Press.
- Tiitinen, P., Lyytikäinen, V., Päivärinta, T. & Salminen, A. 2000. User Needs for Electronic Document Management in Public Administration: A Study of Two Cases. In H. R. Hansen, M. Bichler & H. Mahler (Eds.). Proceedings of the 8th European Conference on Information Systems. Wien: Wirtschaftsuniversität Wien, 1144-1151.
- Tolvanen, J-P. 1998. Incremental Method Engineering with Modelling Tools. Computer Science and Information Systems. Jyväskylä: University of Jyväskylä. Ph.D. Dissertation.
- Tozer, G. 1999. Metadata Management for Information Control and Business Success. Norwood, MA: Artech House, Inc.
- Tyrväinen, P. & Päivärinta, T. 1998. Analysing Document Genres for Smoothening It Integration in Global Organization Networks. In E. M. Roche, K. Kangas & R. Suomi (Eds.). Proceedings of the IFIP Wg 8.7 Working Conference on Informatics in Multinational Enterprises. "MIS - Shaped of Mis-



- shaped? Emerging Issues in the New Global Arena". TUCS General Publication No 12. Turku, Finland: Turku Centre for Computer Science, 100-113.
- Tyrväinen, P. & Päivärinta, T. 1999. On Rethinking Organizational Document Genres for Electronic Document Management. In R. H. Sprague (Ed.) Proceedings of the 33th Annual Hawaii International Conference on System Sciences. Maui (HA): IEEE Computer Society Press, 2011-2020. HICSS Digital Library. 10 pp. <http://computer.org/proceedings/hicss/0001/00012/0001toc.htm>
- Tyrväinen, P. & Päivärinta, T. 2003. How Digital Is Communication in Your Organization? - a Metrics and an Analysis Method. In O. Camp, J. Filipe, S. Hammoudi & M. Piattini (Eds.). The Proceedings of the 5th Information Conference on Enterprise Information Systems, ICEIS. Setúbal, Portugal: Escola Superior de Tecnologia do Instituto Politécnico de Setúbal, 156-166.
- UML-1.1. 1997. UML Notation Guide. Version 1.1. <http://www.rational.com/uml/resources/documentation/notation/index.jtmdl>. [March 31, 1999]
- Vakkari, P. 1999. Task Complexity, Problem Structure and Information Actions. Integrating Studies on Information Seeking and Retrieval. *Information Processing & Management* 35 (6), 819-837.
- Wenger, E. 1998. *Communities of Practice. Learning, Meaning, and Identity*. Cambridge, U.K.: Cambridge University Press.
- West, K., Brook, F. & Klima, C. 2002. Understanding the ECM Market. *AIIM E-DOC Magazine*. [http://www.edocmagazine.com/vault\\_articles.asp?ID=25375&header=e\\_features\\_he](http://www.edocmagazine.com/vault_articles.asp?ID=25375&header=e_features_he). [15 Aug., 2003]
- Yates, J. 1989. *Control through Communication. The Rise of System in American Management*. Baltimore, MA: Johns Hopkins University Press.
- Yates, J. & Orlikowski, W.J. 1992. Genres of Organizational Communication: A Structural Approach to Studying Communication and Media. *Academy of Management Review* 17 (2), 299-326.
- Yates, J., Orlikowski, W.J. & Rennecker, J. 1997. Collaborative Genres for Collaboration: Genre Systems in Digital Media. In R. H. Sprague (Ed.) Proceedings of the 30th Annual Hawaii International Conference on System Sciences. Digital Documents Track. Maui (HA): IEEE Computer Society Press, 50 - 59.
- Yates, S.J. & Sumner, T.R. 1997. Digital Genres and the New Burden of Fixity. In R. H. Sprague (Ed.) Proceedings of the 30th Annual Hawaii International Conference on System Sciences. Digital Documents Track. Los Alamitos: IEEE Computer Society Press, 3-12.
- Yoshioka, T. & Herman, G. 2000. *Coordinating Information Using Genres*. Cambridge, MA: Massachusetts Institute of Technology, Sloan School of Management, Center for Coordination Science. <http://ccs.mit.edu/papers/pdf/wp214.pdf>. [10 Oct. 2002].
- Yoshioka, T., Yates, J. & Orlikowski, W.J. 2001. Genre Taxonomy: A Knowledge Repository of Communicative Actions. *ACM Transactions on Information Systems* 19 (4), 431-456.

## YHTEENVETO (FINNISH SUMMARY)

Nykypäivän organisaatioissa tietoa tallennetaan eri muodoissa ja eri kokoisina koostein - kuten tekstinkäsittelytiedostoina, Internet-sivuina, multimedia-tiedostoina, sähköpostiviesteinä ja tietokantojen kenttäsisältöinä. *Sisällön hallinnassa* pyritään kartoittamaan hallittavat sisältöyksiköt ja määrittelemään toimintatavat tuleville sisällön käsittelyprosesseille. Sisällön hallinnan kehittäminen alkaa useimmiten sisällön analysoinnilla. Organisaatioissa kehittämistoimien kohteena olevan sisällön analysointi voi olla vaikeaa: tietosisältö voidaan tuottaa eri tallennusformaatteihin, sen jakeluun voi liittyä useita medioita tai osa sisällöstä voi olla tallennettuna useisiin eri järjestelmiin. Käytettävien teknologioiden ja medioiden kirjosta johtuen jopa sisältöyksiköiden määrittely ja nimeäminen voi olla hankalaa.

Tämän väitöskirjan tavoitteena on tutkia, kuinka organisaation kommunikoinnin analysoinnissa hyödynnettyä *genre*-teoriaa voi hyödyntää sisällön analysointiin organisaatioissa. Koska sisällön analysointi on tutkimusalueena verraten uusi, väitöskirjatutkimuksen toisena tavoitteena on kerätä tietoa sisällön analysoimisen tavoitteista ja toteutustavoista organisaatioissa. Väitöskirjassa kuvataan kirjallisuuskatsaus ja neljä toimintatutkimusta neljässä eri organisaatiossa.

Kohdeorganisaatioissa sisällön analysoimisen tavoitteita olivat vaatimusmäärittelyn laatiminen tulevalle sisällön hallintajärjestelmälle, sisältökoosteidien määrittely eri käyttäjäryhmille ja eri käyttötilanteisiin, sekä Internetissä jaettavan sisällön yksiköiden, koosteiden ja sisältöryhmitysten laatiminen. Vaatimusmäärittelyyn liittyvissä tutkimuksissa on aiemmin havaittu, että käyttäjien ja järjestelmänanalyttikoiden välille saattaa syntyä kommunikaatiokatkoksia. On myös huomattu, että käyttäjät voivat kokea vaatimusmäärittelyn ongelmalliseksi, koska he eivät tiedä millaisia toimintoja uudet teknologiat tai mediat voisivat mahdollistaa. Toimintatutkimuksissa tehdyt havainnot vahvistivat näitä löydöksiä. Lisäksi havaittiin, että järjestelmälähtöisten tai käyttäjille vieraiden sovellusalueiden käsitteistön ja termien käyttäminen saattaa häiritä käyttäjiä sisältöä, sen ryhmittelyä ja tulevalle järjestelmälle asetettavia vaateita määrittäessä.

Väitöskirjatutkimuksessa kehitettiin kaksi genre-pohjaista sisällön analysointimenetelmää. Genre-teorian avulla tunnistetut kommunikaatiokokonaisuudet ja niiden muodostama hierarkia hyödynnettiin teknologia- ja mediariippumattomana kommunikaatorakenteen mallina. Aiempi genre-tutkimus osoitti, että käyttäjien tulisi antaa itse määritellä kommunikoinnissa käytettävä tietosisältö ja sisältöyksiköiden väliset suhteet. Tapaustutkimuksissa käyttäjät kuvasivat oman kommunikaationsa sisältöjä neuvotteluperustaisia ryhmätyöistuntoja, seinätaulutekniikkaa, käsittekarttoja, ja yhteisesti täytettäviä lomakkeita käyttäen. Näiden genre-menetelmiin valittujen tekniikoiden havaittiin vähentävän kommunikaatiokatkokosten mahdollisuutta, ja tukevan käyttäjien järjestelmävaateiden sekä sisältöyksiköiden määrittelyssä.

- 1 ROPPONEN, JANNE, Software risk management - foundations, principles and empirical findings. 273 p. Yhteenveto 1 p. 1999.
- 2 KUZMIN, DMITRI, Numerical simulation of reactive bubbly flows. 110 p. Yhteenveto 1 p. 1999.
- 3 KARSTEN, HELENA, Weaving tapestry: collaborative information technology and organisational change. 266 p. Yhteenveto 3 p. 2000.
- 4 KOSKINEN, JUSSI, Automated transient hypertext support for software maintenance. 98 p. (250 p.) Yhteenveto 1 p. 2000.
- 5 RISTANIEMI, TAPANI, Synchronization and blind signal processing in CDMA systems. - Synkronointi ja sokea signaalinkäsittely CDMA järjestelmässä. 112 p. Yhteenveto 1 p. 2000.
- 6 LAITINEN, MIKA, Mathematical modelling of conductive-radiative heat transfer. 20 p. (108 p.) Yhteenveto 1 p. 2000.
- 7 KOSKINEN, MINNA, Process metamodelling. Conceptual foundations and application. 213 p. Yhteenveto 1 p. 2000.
- 8 SMOLIANSKI, ANTON, Numerical modeling of two-fluid interfacial flows. 109 p. Yhteenveto 1 p. 2001.
- 9 NAHAR, NAZMUN, Information technology supported technology transfer process. A multi-site case study of high-tech enterprises. 377 p. Yhteenveto 3 p. 2001.
- 10 FOMIN, VLADISLAV V., The process of standard making. The case of cellular mobile telephony. - Standardin kehittämisen prosessi. Tapaustutkimus solukoverkkoon perustuvasta matkapuhelintekniikasta. 107 p. (208 p.) Yhteenveto 1 p. 2001.
- 11 PÄIVÄRINTA, TERO, A genre-based approach to developing electronic document management in the organization. 190 p. Yhteenveto 1 p. 2001.
- 12 HÄKKINEN, ERKKI, Design, implementation and evaluation of neural data analysis environment. 229 p. Yhteenveto 1 p. 2001.
- 13 HIRVONEN, KULLERVO, Towards Better Employment Using Adaptive Control of Labour Costs of an Enterprise. 118 p. Yhteenveto 4 p. 2001.
- 14 MAJAVA, KIRSI, Optimization-based techniques for image restoration. 27 p. (142 p.) Yhteenveto 1 p. 2001.
- 15 SAARINEN, KARI, Near infra-red measurement based control system for thermo-mechanical refiners. 84 p. (186 p.) Yhteenveto 1 p. 2001.
- 16 FORSELL, MARKO, Improving Component Reuse in Software Development. 169 p. Yhteenveto 1 p. 2002.
- 17 VIRTANEN, PAULI, Neuro-fuzzy expert systems in financial and control engineering. 245 p. Yhteenveto 1 p. 2002.
- 18 KOVALAINEN, MIKKO, Computer mediated organizational memory for process control. Moving CSCW research from an idea to a product. 57 p. (146 p.) Yhteenveto 4 p. 2002.
- 19 HÄMÄLÄINEN, TIMO, Broadband network quality of service and pricing. 140 p. Yhteenveto 1 p. 2002.
- 20 MARTIKAINEN, JANNE, Efficient solvers for discretized elliptic vector-valued problems. 25 p. (109 p.) Yhteenveto 1 p. 2002.
- 21 MURSU, ANJA, Information systems development in developing countries. Risk management and sustainability analysis in Nigerian software companies. 296 p. Yhteenveto 3 p. 2002.
- 22 SELEZNYOV, ALEXANDR, An anomaly intrusion detection system based on intelligent user recognition. 186 p. Yhteenveto 3 p. 2002.
- 23 LENSU, ANSSI, Computationally intelligent methods for qualitative data analysis. 57 p. (180 p.) Yhteenveto 1 p. 2002.
- 24 RYABOV, VLADIMIR, Handling imperfect temporal relations. 75 p. (145 p.) Yhteenveto 2 p. 2002.
- 25 TSYMBAL, ALEXEY, Dynamic integration of data mining methods in knowledge discovery systems. 69 p. (170 p.) Yhteenveto 2 p. 2002.
- 26 AKIMOV, VLADIMIR, Domain Decomposition Methods for the Problems with Boundary Layers. 30 p. (84 p.) Yhteenveto 1 p. 2002.
- 27 SEYUKOVA-RIVKIND, LUDMILA, Mathematical and Numerical Analysis of Boundary Value Problems for Fluid Flow. 30 p. (126 p.) Yhteenveto 1 p. 2002.
- 28 HÄMÄLÄINEN, SEPPO, WCDMA Radio Network Performance. 235 p. Yhteenveto 2 p. 2003.
- 29 PEKKOLA, SAMULI, Multiple media in group work. Emphasising individual users in distributed and real-time CSCW systems. 210 p. Yhteenveto 2 p. 2003.
- 30 MARKKULA, JOUNI, Geographic personal data, its privacy protection and prospects in a location-based service environment. 109 p. Yhteenveto 2 p. 2003.
- 31 HONKARANTA, ANNE, From genres to content analysis. Experiences from four case organizations. 90 p. (154 p.) Yhteenveto 1 p. 2003.