IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITP4473

Bridging the Gap Between Hard and Soft Information Genres

Anne Karjalainen and Airi Salminen
University of Jyväskylä, Computer Science and IS, PO Box 35 (MaE), FIN 40351, Jyväskylä, Finland
Tel: +358 14 2603 041; Fax: +358 14 2603 068; ankarjal@cc.jyu.fi, airi@cs.jyu.fi

ABSTRACT

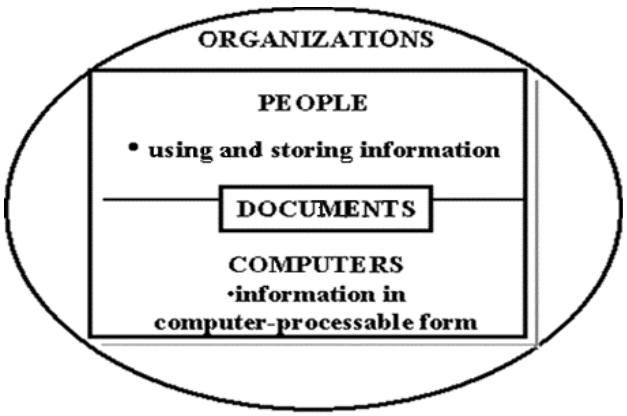
The genre theory has been earlier introduced as a means for analyzing organizational communication. The terms hard and soft information genre have been employed to refer to the extent and explicity of the rules that the genres are enacted by in a particular communication community. People have widely learned to use genres without explicit rules for their purpose, structure, and other features. For computer-mediated information management, techniques like SGML/XML have been developed to allow formal specification of logical information structures for computer processing. The deployment of SGML and XML in organizations creates huge structured information resources. In the paper we discuss the relationship of information genres and structured information resources, and demonstrate by a case that the genre theory offers a framework for analyzing information needs of people and for identifying the pieces which might be accessible from structured information resources. In the case study, the need to harden a set of genres was identified and the hardening was performed.

1. INTRODUCTION

Documents are essential information resources for organizations. They serve as a means for communication, organizational memory, and as a vehicle in work processes (Meier & Sprague, 1996). In a *document*, a set of information items pertaining to a topic is recorded as a unit for human consumption. Successful document management in a computerized community - be it an organization, a group of organizations, or a group of people in an organization - requires, on the one hand, capabilities by which people of the community are able to use documents in their activities effectively, on the other hand, capabilities by which computers in the same community can effectively process the digital form. The dual requirements for documents in computerized communities are illustrated in Figure 1.

To support computer-mediated information management, the information units and structures important to people of a community have to be presented in a computer-processable form. SGML and its subset XML are metalanguages used in many communities to define a format and a language for the computer-mediated information of the community (Goldfarb, 1990; W3C:XML1.0, 1998). For example, HTML is an SGML-based language developed to allow communication between people using different browsers in the Internet community, and AECMA 1000D is an SGML-based language developed for the technical documentation of the aerospace industry (AECMA, 1998). In SGML-based languages, the logical structure of documents can be defined by a *Document Type Definition* (DTD). In document instances, the structure is indicated by markup which SGML-compliant applications are able to recognize.

Figure 1. The dual context of documents in computerized organizations.



Communication *genres*, i.e., socially recognized types of communicative actions of organizational communication (Yates & Orlikowski, 1992) can be enacted as document types. The genre theory has been extensively used as a means for analyzing information classification and use, especially from the human perspective. Information genres useful for people have, to a great extent, been quite "soft": people have learned to use them without explicit rules. In using SGML, formal rules are needed to specify document types and thus information in SGML format could be classified as "hard". The notions of hard and soft information genres were introduced by Schulze & Boland (1997). The deployment of SGML and XML in organizations creates huge structured information resources. In the paper we discuss the relationship of information genres and structured information resources, and demonstrate by a case that the genre theory offers a framework for analyzing information needs of people and for identifying the pieces which might be accessible from structured information resources. In the case study, the need to harden a set of genres was identified and the hardening was performed.

The rest of the paper is organized as follows. Section 2 introduces the central notions of the genre theory. In particular, the relationship of communication genres and SGML technology will be discussed. Section 3 introduces the paper machine factory as a case environment and the use of SGML for the operation and maintenance manuals there. Section 4 describes a collaborative project for improving the usability of the information in operation and maintenance manuals for training in the case environment. In the project, the need to harden a set of genres was identified and the hardening was performed. Section 5 concludes the paper.

2. GENRES IN DOCUMENT MANAGEMENT ENVIRONMENTS

Yates and Orlikowski (1992) introduced genres for the investigation of organizational communication. Organizational communication takes place in a community such as an organization, a group of organizations, or a group of people in an organization. Since documents are used for organizational communication, genre

studies often concern documents (see e.g. Tyrväinen & Päivärinta, 1999; Päivärinta, 1999). Examples of document genres in business communities are memos, user manuals, purchase orders, and contracts. A genre is characterized by having a similar *substance* and *form* in the communication of a community (Orlikowski & Yates, 1994). Substance refers to the *purpose* and *content* of the communication. In case of documents, the purpose is closely related to the intended *target group*. Form refers to the observable features of the communication including its *structure*, communication *medium*, and *language*.

Genres are enacted through *genre rules*. The rules concern the substance as well as the form used in recurrent situations but the rules need not to be explicit. Neither do they always concern all of the attributes listed above. For example, sometimes medium is an important characteristics of a genre (e.g. email message) whereas sometimes it has no importance in the communication. The terms *hard genre* and *soft genre* refer to the extent and explicity of genre features and the rules in the communication community that the genre is enacted by (Schulze and Boland, 1997). The set of genres routinely enacted by a particular community is called a *genre repertoire* (Orlikowski & Yates, 1994). The genres can also form complex systems, genre webs of interrelated genres (Bazerman, 1994).

In respect to communication genres, SGML and XML can be regarded either as enabling or enforcing technologies. Like other technologies, they may also be irrelevant for communicating communities. As an enabling technology, SGML/XML provides a way to define the logical structure of documents by a DTD and rules for representing it. The content structure is expressed by markup in a way that both people and computers can understand it. The structure defined by a DTD is intended to be related to the subject matter of documents. In some communication contexts, the layout of documents is an essential subject matter. For example, HTML as the specification of the layout of Web pages has enabled the Web communication in the Internet community. Communication in the Web has evaluated some genres and facilitated the creation of new kinds of genres for Web, like home pages (Crowston and Williams, 1997; Shepherd & Watters, 1999). XML offers a technique to define logical content structures for specific purposes of the Internet communication.

In a communication community, SGML may also be regarded as enforcing technology. The deployment of SGML requires modifications and probably major changes in the established communication and communication genres. Case studies have shown that the implementation of SGML causes changes in logical content structures, work processes, and in the technology used (see e.g. Fahrenholz-Mann, 1999).

SGML as a metalanguage allows the specification of rules related to the substance as well as to the form of documents. Where very detailed rules are specified and agreed by the members of a communicative community, the corresponding documents are instances of hard genres. As discussed above, the use of SGML in communicative contexts varies greatly. The softness and hardness of the genres depends on the extend and explicity of genre features among the communicative community that enacts them.

One of the motivations for the use of SGML is its support for flexible reuse of information for different purposes and for different media. The benefit of the structured form in SGML documents brings also problems. The information in the structured form may be used in very different ways and for very different communicative purposes. People designing the formal rules for the information, or people authoring the "hard" structured content are rarely sufficiently well aware of the needs of different kinds of users and communicative situations. The established information genres of users may be extremely soft compared to the new structured form. The information needed from a document instance for a specific purpose may be a small subset of the whole content. Furthermore, the subset may be scattered amongst different parts of the document. Effective reuse of the information content stored in structured form requires capabilities to specify the subset of information needed by people in their work tasks. The genre theory offers a framework for identifying the information genres used by people, but as will be demonstrated in the following case description, hardening of genres may be required to achieve improvements in communicative actions.

3. THE CASE ENVIRONMENT

The case organization is a leading manufacturer of paper machines in the world. A paper machine is a huge construction. A complete paper producing line can be 10 meters wide, over 100 meters long, and 10 meters high. It can produce over 1800 meters of paper in a minute. The machine can be in a production use for over 20 years.

Since the machine is a one-of-a-kind product designed to meet the individual needs of the customer, the documentation and the training for the machine must also be tailored. And since the paper machines are becoming more and more sophisticated and technologically complex, the management of the information related to the machines becomes increasingly important. Currently the paper machine business is global, meaning that the communication related to the machines takes place in many different languages. When creating the documentation and training, the manufacturer should keep in mind the users of the documentation in the paper factory. People working with a paper machine have specialized tasks, such as maintenance, operation, design, automation, management and administration. They have different perspectives on the use of the machine and also different information needs concerning the machine and its documentation.

The case organization uses SGML as production and delivery format of the operating and maintenance (O&M) manuals of a paper machine. The O&M manual consists of "books". The books are divided into "interleaves" that contain either overall information or specific information needed mostly by one user group of a machine. Thus the O&M manual is actually a collection of genres, (i.e. genre repertoire) which carry a collection of information targeted for different machine user groups and pertaining information topics of several machine section parts. The printed manual can consist of over 40 folders of filed paper.

The O&M manual provides a computer-accessable, comprehensive information asset. Since the case organization is looking for improvements in the use of the information, a collaborative project with the University of Jyväskylä was started in 1998 (inSGML Project, 1998). In the project, the reuse of the O&M manual content in the training shall be explored. The research is carried out as action research (Susman & Evered, 1978). The case described below forms the first action research cycle of the research and is therefore organized according to action research phases of diagnosing, action planning, action taking, evaluating, and specifying learning.

4. HARDENING OF GENRES

4.1 Diagnosing

The preliminary study concerning the O&M manual and the training genres showed that the O&M manual genre in the organization is hard. There is an explicit definition of it's content and the people using it share a clear and unified understanding of it's fea-

tures. From the viewpoint of the manufacturing organization both the O&M manual and the training are genres. From the viewpoint of an individual trainer, content provider or reader, they both are, in fact, complex systems made up of numerous genres. They contain numerous identifiable content parts pertaining to specific purpose and specific information topics considering specific paper machine sections for specific target audience groups.

The smallest training units identified and used in the organization were the individual training sessions. A training session is lectured by an individual trainer using a combination of media: speech, audiovisual representations, and training material. It covers the trainable content of one individual machine section. Each training session or a training module is designed for one of the four target groups: *key personnel*, *operators*, *maintenance* and *automation***.** The training modules for different machine sections and for the four target groups result in nearly 80 different training modules as genres.

In the organization, approximately 100 experts participate in the training activity for one paper machine. Approximately 100-200 customer's employees are trained to use and maintain it. The coordination and design of the training had been organized as a separate unit only recently. The purpose was to find more consistent ways for training and producing the training material, and coordinating the training activities. So far, the trainers had varying individual ways for producing, organizing and archiving the training content and material, and they did not utilize the SGML data. The features of the training modules were not clear nor uniformly understood. Most of the trainers still come from separate departments. Their preliminary task is to design the paper machine section that they are specialized on. Therefore, the recognized training genres were not hard enough for specifying the reusable content parts from the O&M manual or providing the designers and trainers a unified understanding of the training genres for designing and carrying out the training.

4.2 Action Planning

Orlikowski and Yates (1994) have noticed that when a community is formed its members come to some understanding about the set of genres they will use collectively. This set of genres is often based on the communicative experiences of the members. Ongoing interaction among the members of the community tends to drag on and reinforce the genres established within the community (Orlikowski and Yates, 1994). The more people participate in defining the rules and norms applicable to a genre, the harder the genregets (Schulze & Boland, 1997).

Since the individuals of the communication community usually do the negotiating and refining of genres thus making them harder, it was decided to organize workshops for allowing people to negotiate the training genres' features. One workshop would gather experts of one section of a machine from separate departments to negotiate and define the training modules attached to the section. They would be asked to define the training modules for the separate target audiences of the machine section they are experts of.

The information in the workshops was collected in two ways. The wall-diagram technique was used for defining the training modules' content, i.e. the topics and subtopics that were to be covered. The content and the other genre features (such as media, the way of addressing the audience, and purpose) were then negotiated and defined considering the specific training module or a topic of it. The wall-diagram technique was originally developed for collaborative design of processes, information flow diagrams and structures (Saaren-Seppälä, 1997). It has been successfully used for defining genres and their metadata in a collaborative way (Karjalainen & al., 2000).

4.3 Action Taking

Collaboration in the workshops was carried out in the following way. First the objectives, schedule and assignments were introduced. The use of the wall diagram technique was explained to the participants. Then the participants were organized into workgroups. Two assignments were given to each workgroup. The first assignment was to define the main topics (see Table 1 below) of the training module onto the wall diagram. The second assignment consisted of defining the main topics further, and answering a set of questions considering other genre features of the training module to be defined. After completing the first and the second assignment the workgroups would gather together to discuss the results with other workgroup members and for negotiating the terms that were used for defining the training modules' content topics and other features.

Most of the workshops lasted for a day. The last workshop that was held could be carried out as a half-day session. Some of the participants were by now familiar with the technique. The workgroups could also use the results of the previous workshops as their starting point. Therefore, the participants discussed the main contents only shortly and focused on discussing other features of the training module. Among other subjects, the workgroups' discussion concerned following aspects:

- The target training group. Who the trainees actually are, what kind of previous knowledge they usually have, how different training target groups should be addressed?
- What is the appropriate level of abstraction of presentation? How much detailed information the training module should include?
- Which is the preferable topic organization, and can the topics be organized in a similar way for the other target training audiences, too?
- The use of the training techniques and media for certain topics.
- What are the preferable terms to be used?
- Do the training modules form a coherent and unified information content for the target training audience – and is the content relevant for them?

4.4 Evaluating

The objective of the workshops was to harden the training

Table 1. A part of a content definition of a training module.

Training module: Headbox section training for key personnel	
Main topics	**Definition**
Introduction	-Introduction of the trainer and schedule -Introduction of the content and objectives of the training module
Structure	-The main parts/composition of the structural section for the training
Functions	-List of main functions that the structural section includes -How are these functions related to the overall functions of the machine and the process of paper production?
Maintenance	-List of the procedures to be performed on the regular basis -List of the procedures to be performed when needed
Summary	-Summary of the things the trainees should know

genres in the manufacturing organization. The genre is hard if its substance (purpose, topics, target group) and form (visible structure, medium and used language style) are identified and clear to most of the people that use it. In the case organization the hardening of the training genres was observable both during the workshops and after them. In the workshops the trainers and the experts performing the majority of the training defined the content of training modules using the wall-diagram technique and discussed and negotiated the other features of training modules, too. During the latter workshops the focus of the discussion changed to other genre features away from content topics, thus implying that the participants had reached a mutual, unified understanding of the content topics and thus focused their negotiation efforts to other important features.

After the workshops the descriptions created facilitated the planning of a guidebook for trainers. The intent is also to use the knowledge provided in the workshops for developing techniques by which information available in the O&M manuals can be accessed for training purposes.

4.5 Specifying Learning

The chosen approach and used techniques (collaborative workshops and wall-diagram technique) worked well on the case domain and can thus be tested in other domains for hardening soft genres and especially for defining the logical information content of genres. In the future, the method could be used to support the planning of SGML/XML DTDs.

In the case, the genre theory offered a framework to analyze the information needs of people and to identify the pieces which might be accessible from structured information resources. In the case, the resulting definitions showed that there were differences in the target audience groups. For the mechanical maintenance group most of the content can be picked from O&M manual parts quite easily. Specification of the O&M manual parts useful for other target groups is not so easy, since the reusable parts are smaller, not so easy to identify, and they should be re-organized for training. There are some training modules that need other information content than that of the O&M manual.

5. CONCLUSION

The research implies that the genre theory offers a framework for improving enterprise document management. In this paper we discussed hard and soft genres and gave an example of hardening the training genres in a manufacturing company. The hardening of the genres was accomplished by accelerating the process by which the members of the community negotiate them. The hardening was regarded as an important requirement for enabling the definition of the reusable information resources stored in the O&M manual in SGML form in the case environment. The case study was accomplished successfully implying that the used approach can be tried out in other domains, too. Avenues for further research are the analysis of the hardened genres' features with respect to content reuse between genres, and the analysis of genre features and SGML/XML document management environment components and DTD design.

REFERENCES

AECMA Specification 1000D (1998). http://www.aecma.org/1000d.htm [01.10.1999]

Bazerman, C. (1994). Systems of genres and the enactment of social intentions. In A. Freedman and P. Medway (Eds.), *Genre and the New Rhetoric*, London: Taylor & Francis, pp. 79-101.

Crowston, K., Williams, M. (1997). Reproduced and emergent genres of communication on the World-Wide Web. In *The Proceedings of the 30th Annual Hawaii International Conference on System Sciences.* Los Alamitos, CA: IEEE Computer Society Press. [Electronic copy with the permission of IEEE in location: http://florin.syr.edu/~crowston/papers/webgenres.html]

Goldfarb. C. (1990). Rubinsky, Y. (Ed.). *The SGML handbook*. New York: Oxford University Press Inc.

Fahrenholz-Mann, S. (1999). SGML for electronic publishing at a technical society: Expectation meets reality. *Markup Languages. Theory and Practice*. Massachusetts Institute of Technology.

inSGML Project (1998). http://www.cs.jyu.fi/~airi/docman#inSGML.html.

Karjalainen, A., Päivärinta, T., Tyrväinen P. and Rajala, J. (2000). Genre-based metadata for enterprise document management. To be published in *the 33th Hawaii International Conference on Systems Sciences, Genre on Digital Documents Track*, Maui, U.S.A. Los Alamitos, CA: IEEE.

Meier, J., Sprague, R. (1996). Towards a better understanding of electronic document management In P*roceedings of the29th Annual Hawaii International Conference on Systems Sciences: Digital Documents.* Los Alamitos, CA: IEEE Computer Society Press. 53 - 61.

Orlikowski, W.J., Yates, J. (1994). Genre repertoire: The structuring of communicative practices in organizations. *Administrative Science Quarterly* 39 (4). 541 - 574

Päivärinta, T. (1999). A genre approach to applying critical social theory to information systems development. In C. Gilson (Ed.), *Proceedings of the 1st Critical Management Studies Conference, Information Technology and Critical Theory –stream* Manchester, England. [http://www.mngt.waikato.ac.nz/ejrot/cmsconference/cmsdefault.htm.]

Saaren-Seppälä, K. (1997). *Seinätekniikka prosessien kehittämisessä*. Finland: Kari Saaren-Seppälä Ltd.

Schultze, U., Boland, R, Jr. (1997). Hard and soft information genres: An analysis of two notes databases. In *Proceedings of the 30th Hawaii International Conference on System Sciences. Digital Documents Track*. Los Alamitos, CA: IEEE Computer Society Press.

Shepherd, M., Watters, C. (1999). The functionality attribute of cybergenres. In *Proceedings of the 32nd Annual Hawaii International Conference of System Sciences: Digital Documents.* Los Alamitos CA: IEEE Computer Society Press.

Susman, G.I., Evered, R.D. (1978). An assessment of the scientific merits of action research. *Administrative Science Quarterly* 23 (4). 582 - 603.

Tyrväinen, P., Päivärinta, T. (1999). On rethinking organizational document genres for electronic document management. In *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences,* Los Alamitos CA: IEEE Computer Society Press.

W3C:XML1.0 (1998). The W3C Recommendation of XML language version 1.0, 1998.
http://www.w3.org/TR/1998/REC-xml-19980210 [01.10.1999]

Yates, J., Orlikowski, W.J. (1992). Genres of organizational communication: A structurational approach to studying communication and media. *Academy of Management Review* 17 (2). 299 - 326.