# Defining Genres and Their Features for Studying Information Reuse: Preliminary Findings

Anne Karjalainen, Pasi Tyrväinen
University of Jyväskylä, Department of Computer Science and Information Systems
P.O. BOX 35 (Ag C516.3), 40351 Jyväskylä, Finland, Fax. +358 14 2603 068, ankarjal@cc.jyu.fi, pasi.tyrvainen@jyu.fi

**ABSTRACT**

*This paper discusses the use of genre theory for analysing information content of structured documents and mixed media training material. We describe an analysis phase of an ongoing research aimed at information reuse from Operation and Maintenance (O&M) manuals in training, consisting of spoken, multimedia and textual content. The content of training was given the form of templates. We used them successfully for locating reusable information contents available from O&M manuals and other information sources. These templates can also serve as a basis for designing XML document type definitions (DTD's) for training content, or as draft versions of XSLT templates for defining reuse transformations from source to target documentation. While defining and hardening genres of training, we collected metainformation and knowledge the trainers possessed. We then formalised this as another set of genres for training. Based on our findings we suggest that genre theory can be used as 1) a framework for defining content within genres, 2) for revealing metainformation needed for enacting genres and 3) for locating reusable information contents from structured source documents. We also observed potential problems for reuse in scheduling of production processes and a need for demonstrating a single point of access to all training knowledge and material available.*

## 1. INTRODUCTION

SGML (Goldfarb, 1990) or XML (W3C:XML 1998) information resource can be valuable also for purposes other than it was originally meant. In the case described here, an SGML-formatted O&M manual was found to contain valuable information content needed for training. This paper discusses the applicability of genre theory for analysing information reuse potential in a case study.

The case organization produces paper machines. A paper machine with the life cycle exceeding 20 years is a large, complex, and unique construction always designed to meet individual customer needs. The O&M manual of the paper machine in our case is defined by a DTD called *papmch.dtd*. This content definition has close to 200 elements and over 120 attributes and enables filtering and reorganizing small and precise content parts from a manual to training material.

The employees of a customer organization are trained to operate and maintain the machine. The four principal target groups - key personnel, operation personnel, mechanical maintenance personnel and automation personnel - each need their own particular training content. Usually training consists of sessions lasting from one to three days. One training session pertains to one section of a paper making line for one target group. A process of training a customer's employees can involve 100 trainers, 100-200 trainees and nearly 80 different training sessions.

## 2. THE ANALYSIS PHASE OF THE STUDY

The analysis phase consisted of two action research (AR) cycles (Kock & al., 1997), which are summarized in tables one and two.

**TABLE 1. Summary of the first AR cycle.**

| | ACTION RESEARCH CYCLE 1 |
|---|---|
| **Phase** | **Description** |
| **Diagnosing** | Development goals:<br><br>*need to harmonize existing training materials*<br><br>*need to support training content analysis* |
| **Action planning** | Decided to use genre theory for hardening the training genres, i.e. to make the features and content of training genres explicit. |
| **Action taking** | Organizing 6 workshops for groups of trainers to harmonize the training genres' features. |
| **Evaluating** | Preliminary study of training templates defined. The templates allowed comparison of content definitions made for target groups and the comparison of training and O&M manuals content. |
| **Specifying learning** | The genre theory provided a suitable framework also for defining the content of genres. |

Genre theory offered a basis for this part of the study. Genre is a typified communicational form in a discourse community. Genres can be characterized by similar substance and form taken in response to recurrent situation. The substance of a genre refers to the social motives, themes and topics of the communication of organization, whereas its form refers to the observable physical and linguistic features of the communication (Yates & Orlikowski, 1992). Genre theory has earlier been used also for detecting and analyzing documents in organizational settings (Tyrväinen & Päivärinta, 1999, Karjalainen & al., 2000).

We were interested in both the contents and features of training genres. Researchers have observed content or topic organization or repeated patterns within a genre, although there is a lack of research considering the content within genres (Karjalainen & Salminen, 2000). As the training substance and form was tailored for the defined four training target groups we first found four "main genres" in training. If genres can be commonly identified and their features are explicit, Schulze & Boland (1997) consider them

as "hard" as opposite to "soft". We organized workshops for trainers for negotiating features and content of training and thus hardening the genres (see: Karjalainen & Salminen, 2000). As a result, we had 16 templates each describing topics and subtopics of one training session with up to three levels of subtopics in them. We also recorded the workshop discussions in memos.

**TABLE 2. Summary of the second AR cycle.**

| | ACTION RESEARCH CYCLE 2 |
|---|---|
| *Phase* | *Description* |
| *Diagnosing* | Development goals:<br>• unified terminology used for training templates<br>• content organization of templates analyzed with respect to target groups and paper machine sections trained<br>• potential reusable parts from O&M manuals mapped and documented<br>• other possible sources for reusable information located |
| *Action planning* | Get assistance from the case organization experts. Schedule co-operative sessions and participants. Study the production process of O&M manuals. |
| *Action taking* | Working sessions with domain experts from O&M manual and training departments. Studying and modelling the production process, content and structure, and roles of O&M manuals. |
| *Evaluating* | Trainers' knowledge on enacting the training should be made explicit and available for sharing. Trainers' guide was designed. Reusable content and DTD parts in O&M manuals could be mapped with the training template's topics and subtopics. Other relevant information sources for training also found. Temporal constraints for reuse. |
| *Specifying learning* | Genre theory provided an approach for defining training content as templates. The reusable content parts from O&M manual could be located with help from domain experts. The production processes of source and target documentation need to be synchronized. |

### 3. EVALUATION OF THE FINDINGS

We reviewed the memos of the workshops and categorized their content to create a set of guidelines. In this process we actually documented another set of genres consisting of metainformation needed for enacting the training. We decided to arrange the genres as documents and call the collection Trainers' guide. It consists of the following documents:

· Short session definitions give a brief overview of each training session's duration, purpose, timing and overall content. It defines the knowledge and skills a trainee should gain and the target group for the training session. (For trainers, training designers and customers.)

· Training target group guidelines define the features of training target groups. (For non-professional trainers and customers.)

· Multimedia and controlled language guide explains what presentations are available and where and how the presentations are to be used. Controlled language guide discusses the rules for enacting different communicational content within training. (For training designers and trainers.)

· Terminology guide defines the terms used for machine components and other domain specific terminology.

· Templates for training sessions consist of 25 different training templates that have been defined so far. (For training designers and trainers.)

For locating the reusable content parts from O&M manuals we designed a generalized training template for each training target group and compared it with the content of the O & M Manual. We then documented reusable content and DTD elements for extracting it from O&M manuals (see Figure 1 for an example). Other relevant information was found from sources including the sales department (layout drawings in color), the multimedia group (multimedia presentations), and the technical writing and document translations group that had produced a detailed term repository.



Figure 1. Example of possible training topics and subtopics with information content sources defined.

Preparing O&M manuals is a rather complicated process with three different DTD's used. The information content is defined by *papmch.dtd*. Another DTD defines the relationships of *papmch.dtd* documents with the content sections of O&M manual and also the files concatenated to each section file. The concatenation script also transforms the content into a layout-oriented structure defined by a third DTD that is used for presenting digital and printed versions of a manual. In one phase of the production process, the content describing a machine section is represented by a single file, and in another, by several files produced by experts in several sites. The schedule of production process depends on the schedules of machine design and manual delivery but is independent of training schedule. The trainers reported that sometimes the final O&M content was not available on time for training. Furthermore, locating relevant files and studying them proved difficult because of the large number of files and lack of SGML compatible tools. We observed that these production processes need to be synchronized: the constraints this might create for information reuse have to be analysed in further research. In addition, a technical design and implementation of a demonstrator would be needed for studying the possibilities and drawbacks of XML and related languages for information reuse. The training templates seem to provide a draft version of DTD's main branches for training documentation. The element definitions needed for representing the structure and media content below these branches could be reused from other existing DTD's of the case organization. Alternatively, instead of using a DTD for representing training content, we could evaluate the use of training templates as XSLT (W3C:XSLT, 1999) templates for filtering and re-organizing needed information parts from an O&M manual.

### 4. CONCLUSIONS

In this research, we observed the positive impact of hardening genres on organizational information reuse, and on the knowledge management of the case organization. The explicit definition of the training genres also enabled the explicit definition of the associated training knowledge. This knowledge was captured in a set of (meta-)genres enabling sharing corporate knowledge and best practices among the trainers. People working in similar activities can use the shared information space containing both mixed media material and shared knowledge. It will help them in their daily tasks and facilitate sharing and reusing the materials, scripts and

applications they need to use as well as forming a base-line for further knowledge externalization and process development – these are additional benefits not envisioned in the original reuse initiative.

## REFERENCES

Goldfarb, C., F.,edited by Rubinsky, Y. (1990). The SGML Handbook. Oxford University Press Inc., New York.

Karjalainen, A., Päivärinta, T., Tyrväinen, P., Rajala, J. (2000). Genre-Based Metadata for Enterprise Document Management. *Proceedings of the 33th Hawaii International Conference on Systems Sciences*. IEEE Computer Society Press, Los Alamitos, CA. CD-ROM. Available at: http://www.jyu.fi/~ankarjal/ HICSS2000.pdf .

Karjalainen, A., Salminen, A. (2000). "Bridging the Gap Between Hard and Soft Information Genres". In: Khosrowpour, M. (ed*.). Proceedings of 2000 IRMA International Conference*. Idea Group Publishing, Hershey, U.S.A, 92-95.

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 Approach". *Journal of Systems & Information Technology*. Vol. 1(1), 1 – 24.

Schultze, U., Boland, R. Jr. "Hard and Soft Information Genres: An Analysis of two Notes Databases". (1997). *Proceedings of the 30th Hawaii International Conference on System Sciences*. IEEE Computer Society Press, Los Alamitos, CA. CD-ROM.

Tyrväinen, P., Päivärinta, T. (1999). "On Rethinking Organizational Document Genres for Electronic Document Management". *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences*. IEEE Computer Society Press, Los Alamitos, CA. CD-ROM.

[W3C:XSLT]. Clark, J. (ed.). (1999). " XSL Transformations (XSLT) Version 1.0. W3C Rec-16-November 1999". http:// www.w3.org/TR/1999/REC-xslt-19991116 [20.11.2000].

[W3C:XML]. Bray,T., Paoli, J., Sperberg-McQueen, C.M. (eds). (1998). "Extensible Markup Language (XML) 1.0". W3C Rec-10-Feb-1998. http://www.w3.org/TR/REC-xml [15.05.2000].

Yates, J., Orlikowski, W.J. (1992). "Genres of Organizational Communication: A Structurational Approach to Studying Communication and Media". *Academy of Management Review*. Vol. 17(2), 299 - 326.

## Related Content

Supply Chain Resources and Economic Security Based on Artificial Intelligence and Blockchain Multi-Channel Technology
Dong Wangand Ao Yu (2023). *International Journal of Information Technologies and Systems Approach (pp. 1-15).*
www.irma-international.org/article/supply-chain-resources-and-economic-security-based-on-artificial-intelligence-and-blockchain-multi-channel-technology/322385

On Inter-Method and Intra-Method Object-Oriented Class Cohesion
Frank Tsui, Orlando Karam, Sheryl Dugginsand Challa Bonja (2009). *International Journal of Information Technologies and Systems Approach (pp. 15-32).*
www.irma-international.org/article/inter-method-intra-method-object/2544

Lack of Characteristics Management Causing Biggest Projects Failure
Loredana Arana (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 5650-5659).*
www.irma-international.org/chapter/lack-of-characteristics-management-causing-biggest-projects-failure/184265

Factors Influencing Nursing Professionals' Computer-Based Information Systems (CBIS) Use Behavior
Princely Ifinedo (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 3332-3343).*
www.irma-international.org/chapter/factors-influencing-nursing-professionals-computer-based-information-systems-cbis-use-behavior/112764

Data Mining and Knowledge Discovery in Databases
Ana Azevedo (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 1713-1722).*
www.irma-international.org/chapter/data-mining-and-knowledge-discovery-in-databases/112576