Designing Training in Manufacturing Organizations Using the Genre-based Method

Anne Honkaranta and Pasi Tyrväinen
University of Jyväskylä, Department of Computer Science and Information Systems
P.O. Box 35 (Agora), 40014 University of Jyväskylä, Finland
anne.honkaranta@cc.jyu.fi
pasi.tyrvainen@cc.jyu.fi

Abstract

This paper discusses the analysis and design of training material in manufacturing organizations. Delivery of complex, large machines, such as airplanes, paper machines, or cabin cruisers, requires massive customer training, which must be tailored according to the specific training groups and features of the individual machine delivered. Hundreds of trainers and trainees can be involved. Thus the analysis and design of training is important. A small amount of automation or process improvement can lead to remarkable resource savings and free trainers' resources for educational design. We view the content of training as modular and hierarchical. It is proposed here that the trainers should be allowed to define the content they produce. We present the way we operationalized the theory of genres of organizational communication for a participatory genre-based method for analyzing variances of topical content within genres of training. We also discuss the potential for enhancing the production of content for training by content reuse using XML transformation techniques. Due to variation in customized content we propose reusing raw content sources and topic structures, rather than re-using learning objects as such for the assembly of raw learning content.

Keywords

Learning objects, Training genres, Content analysis and reuse, XML

Introduction

Recent initiatives for learning technology standardization (e.g. IEEE-LTSC, 2002; IMS-GLC, 2002; ADLI, 2002) have provided the essential specifications for defining learning objectives and Learning Objects (e.g. IMS-LRMDI, 2000) or Learning Object Packages (e.g. IMS-CPIM, 2001; SCORM, 2001), and definitions how to bind these specifications to XML (Bray et al., 2000) representations (e.g. IMS-CPXB, 2001). Multiple software vendors use these specifications or have defined combinations of them (e.g. Microsoft, 2002), and there is a multitude of different types of learning management systems available. However, these definitions developed for e-Learning have mainly focused on metadata necessary for describing learning intentions and defining units of learning content known as *Learning Objects*. It has remained somewhat vague what a learning object actually is (Polsani, 2003), and what the units of content within training are, and, more appropriately, who defines the units of content within learning, and how. All in all, there exist few methods for analyzing and designing content of training and combinations of content for various target audiences intended.

The purpose of this paper is to illustrate the complex and massive content of training in manufacturing organizations, and show how the units of content of training in manufacturing organizations can be analyzed. We discuss a study in which the broad and complex content of training of a manufacturing organization was designed for specified target training groups. For the purpose, we used the theory of genres as an analytical lens for studying organizational communication as suggested by Yates & Orlikowski (1992). The genre theory has been used e.g. for developing content and document management in organizations (Honkaranta, 2003a), and a genre-based method (Honkaranta, 2003b) has been explicated for analyzing content and studying requirements for content assembly. In the paper, the use of genre analysis is combined with studies of learning content, where the content and its objects are considered. We provide an example of a hierarchy of content units within a massive, complex content of industrial training, and provide an example on using the genre-based templates along with XML and XSLT languages (Clark, 1999) for fetching source content units and transforming them into raw training content intended for customer training of a large manufacturing organization.
The rest of the paper is organized as follows. In section two we discuss the content and its units in training materials. Section three discusses the theory of genres. Section four describes the manufacturing organization referred to as BIRD, and illustrates how the genre-based method was operationalized for studying the units of content within training for four training groups. Section five provides an example of using a genre-based

template together with XML and XSLT for assembling content for training. Sections six and seven discuss the implications and conclude the paper.

Content of Training Revisited

The nature of human information processing has been studied in multiple disciplines from philosophy and education to information systems design. For example, a set of theories related to constructive view of learning (Korhonen & Väliharju, 1995) see learning as an activity, in which the learner has an existing repository of information content, and by metacognitive knowledge is able to seek for new pieces of information. Recognized, suitable new piece of information is added into the learner's knowledge repository to accomplish a learning task. Information content is commonly represented by using concept maps (Mcknight et al., 1991) and abstract and concrete concept hierarchies (Rasmussen, 1985). Topic maps (Pepper, 1999) are kinds of concept maps that provide metadata over a set of content units called topics, as well as locators for content instances.

According to Holmberg, (1989) the learning content should be appended by elements of guided didactic conversation, such as communicating the learning objectives and motivating the learner. The guided didactic conversation is similar by nature to learning object definitions included in the recent e-Learning standards such as Learning Object Metadata (IMS-LRMDI, 2000) and Sharable Content Objects (SCORM). This content can hence be considered as metadata of the learning content. Partially due to this didactic conversation, a large portion of student communication can consist of other communication than the actual content provided by the instructor (Tyrväinen et al., 2003).

Holmberg (1989) has pointed out that a natural, logical model of content organization can be found for most learning areas, and the content should be clearly organized into content topics. The model of content is evident for experts in the area. The learnable content itself is a mix of several types of information, including examples, references, discussion, and assignments. The learning content should be divided into suitable units for learning: an amount of topical content that can be learned in one session should be enacted as a learning unit, or module. It should be easily navigable, from bottom-up to top-down and by general browsing, since the learners tend to use either inductive or reductive reading strategies. The content should also contain indexes, introductions and concluding remarks for quick browsing and top-down approach (Holmberg, 1989). Holmberg has also envisioned that if the content of learning were organized as a collection of independent content modules, the learners could construct the appropriate learning content themselves.

Polsani (2003) provides a well-founded discussion over the fuzzy nature of Learning Objects (LOs). The concept of an LO as an identified container of learning content is used in multiple standards for learning metadata, such as LOM and SCORM. Granularity of content units can vary a lot according to the domain or the definition used, which can confuse learning material content designers (Polsani, 2003). Polsani considers this to be the caveat of many definitions of LOs, limiting their successful reuse. Grain size of an LO is often based on learning time, and assets are often subjective and arbitrary. The grain size should be re-defined by asking the question: how many ideas about a topic can stand on their own, and can be reused in different contexts? Thus Polsani seems to indicate that an LO grain size should be defined by reasoning about *topical* content of training; an LO would preferably consist of one or a few content topics.

Furthermore, e-Learning LOs, should be accessed and reused independently from the delivery media or content/learning management system (LMS) used. An LO cannot become reusable if its creation is not separated from the instructional method used, therefore high level of abstraction for LO content should be emphasized. The content of LO itself consists of *elements*, i.e. smaller content parts such as text, video, animation, glossary, assessment, and multimedia. Polsani (2003) also points out that a content of an LO should be a combination of multiple types of elements. As XML is the standard for all future applications, it could facilitate creation of learning elements within LOs enabling cross-platform interoperability of LOs consisting of one or more XML elements. The inherent separation of structure, content and presentation in XML also allows the flexibility required for LO creation, deployment and manipulation. (Polsani, 2003)

As Polsan i (2003) and Holmberg (1989) have provided descriptions of learning content, we can ask what are the LOs, topics, and their inherent ordering in some domains, and how they can be defined? Polsani emphasizes the abstract, self-standing nature of the LO, Holmberg refers to domain-oriented nature of content expertise that is needed for organizing the content and units of it in a certain domain. Given that existing definitions of LOs differ so much, is the reason for differences actually caused by the differing learning content domains? It seems that studies of learning content units in different types of learning domains are needed for comparison. There also

seems to be a lack of methods for defining the hierarchy of units of the content of training, and for identifying potential LOs.

Genres as Analytical "Lens" for Content Analysis

Content can be comprehended and analyzed from multiple perspectives. Documents, texts and speech used for communicating the content in organizations can be analyzed by focusing on the communication itself - hence emphasizing the human-oriented, non-technical aspect that provides a base by which contemporary technologies and media can be scrutinized. Following Yates & Orlikowski (1992) we propose using the *genre theory* as an analytical "lens" for scrutinizing communication within organizations. A g*enre* can be defined as a typified communicative action, characterized by a similar substance and form, and enacted as a response to recurrent situations (Yates & Orlikowski, 1992). Genre s*ubstance* defines the social motives, themes and topics, whereas *form* refers to the physical and linguistic features of the communication. A *topic* defines a name for a unit of content within a genre. Examples of commonly known genres include a memo, a business offer, and a meeting agenda. Recurrent situations, such as work tasks that occur repeatedly, can be typified. *Genre rules* associate appropriate elements of form and substance with a certain type of recurrent situation within a community of discourse (Yates & Orlikowski, 1992). *Community of discourse* can be considered as a type of a special interest group (SIG), that shares understanding of the communicative situations, and genre features that are to be applied (Swales, 1999). A community can develop a specific language with jargon and acronyms for both uniting itself and preventing outsiders for entering (Swales, 1999).

Yates & Orlikowski (1992) have pointed out that genres should be studied in their context, i.e. in the community using them, since genre rules and features are governed and evaluated by the community using them. Berkenkotter & Huckin (1995) noticed that genres and rules for enacting them can be strikingly different from one community to another, even if both communities refer to the genre with the same name. Genre studies, in which the users are allowed to define the genres they use themselves, are therefore needed albeit rare.

For the purpose of an analysis it is useful to group and categorize genres according to their features. For example, properties of genres in regard to substance and form can be thought of as genre-based metadata values, and substance and form can be thought of as metadata category names for the genres inspected. The *5W1H* (Yoshioka et al., 2001) aspects of genres can be thought of as genre-based metadata framework, which can be operationalized for studying genres and their use with regard to others. The 5W1H considers the **why** (socially recognized purpose of communication), **what** (expected content), **how** (media, type of language), **who(m)** (who communicates, to who or whom), **when** (e.g. time schedules or deadlines, duration) and **where** (physical or virtual places; such as company building or URL) aspects of genre features. The framework can also be enhanced by developing categories according to the needs of the analysis, or by expanding the 5W1H aspects. For example, the "how" can consider technology or software used for enacting genres in addition to media and language used.

Multiple genres and combinations of them are commonly used in organizations (see, e.g. Karjalainen et al., 2000; Päivärinta et al., 2001). Genres that are used habitually in a community form a *genre repertoire* (Orlikowski & Yates, 1994), a collection of genres of a community, whereas genres that have a generalization (variant) relationship with respect to each other are considered as *genre variants* (Crowston & Williams, 1997). The theory of genres hence provides us with means for scrutinizing genres as communicational units of content, and also with means for expanding our inspection to the whole repertoire of units of content in an organization - from a genre repertoire as a broad content collection to a subtopic of a genre, which is a unit of content with a quite small grain size. Figure 1 visualizes the genre-related concepts and content units that can be operationalized for scrutinizing the communication and relationships between the communicational units of content. The figure uses UML class diagram (UML, 1997) notation.

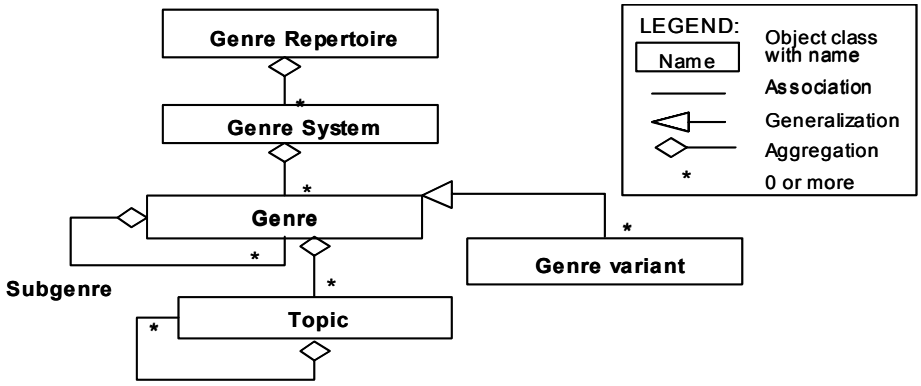

Figure 1. Genre-related concepts and their relationships

The differences between a genre and its instance, and a genre and media used for delivering an instance of it should be clarified: a memo genre can be thought of as a model for numerous memo instances, which can be produced, manipulated and delivered as documents, views, web sites, files, mail bodies or databases. In a way, a genre can also be considered as a human-crafted model for a document type schema, which (e.g. DTD; Maler and El Andaloussi, 1996 or XML schema: Fallside, 2001) is a more exact definition prepared by using a formal notation, and meant both for human and computer comprehension.

Training Content in Manufacturing Organizations: The Case of BIRD

Context of the Study

The target organization of this study - i.e. BIRD - is a large, multinational enterprise producing and delivering production lines consisting of machines, which are customized for individual factories. There is a need to train the people working in customer factories for operating and maintaining the machine purchased. Training is organized as a project that starts with a training plan and an offer made to a customer. Course design considers the requirements set by the machine ordered by the customer and knowledge obtained about the employees to be trained. The actual training includes carrying out the training in various locations, possibly using multiple languages, and using a variety of training techniques. Duration of a training project can exceed a year and involve intensive communication between people at the customer organization and at the product design, marketing and customer documentation departments. In each training project, up to 100 experts of BIRD train 100-200 employees of a customer organization engaged in different types of work with a production line. BIRD has defined four separate training groups according to the types of work in factories: key personnel (such as managers and office workers), operators using the machines of the production line, mechanical maintenance staff, and automation staff maintaining software and ventilation systems.

The training domain is divided along two main dimensions: the four training groups and the maximum of 20 machines of the production line to be trained. Each of the about 80 training sessions pertains to one machine for one training group. Further, training customized to the features of customized machines adds the third dimension. Sessions are carried out in a certain order or in tandem. A training session can be carried out as a combination of lecturing and group work, factory visits, or group work with multimedia-enriched training materials. A session can last from two hours up to three days. The amount of training material to be managed can be high: from 20-60 pages of paper pertaining a session to 160-480 pages of content for each of the possibly 100 to 200 trainees.

A large portion of training content could potentially be reused from the Operation and Maintenance (O&M) manual. An action research study (Susman & Evered, 1978) was established to collect requirements for content assembly from O&M documents to training documents. An analyst who was assigned to the research of reuse requirements tried to apply the structured document content analysis methods – such as the Maler & ElAndaloussi (1996) method for the task, since the content of O&M manuals was defined by a SGML (predecessor and superset of XML) language. She studied the content of training materials and their production, but was unable to apply the method because of the heterogeneous nature of training materials. The existing content and material production processes varied from trainer to trainer in the recently established training department. One trainer producing the training for his training session reused the text and figures from O&M

documents with a copy/paste function producing a layout similar to the O&M documentation. Another trainer produced slide shows containing bullet lists and figures of a machine according to another layout definition. Trainers used the term training materials of a variety of content including a slide show presentations, spoken content, and multimedia presentations. The amount of training material and its variations seemed overwhelming. Even the manager of the training department stated that he could not define all the variations of a training session contents and requested for methods for making it explicit and harmonizing it.

The Genre-based Method

A preliminary method for the study was defined based on the theory and findings related to genres. Following Yates & Orlikowski (1992) the genres were studied in the domain they are used. As there is a lack for studies in which the domain experts design the genres they use themselves (Berkenkotter & Huckin, 1995), user participation was emphasized rather than utilization of existing studies of training content. The benefit of participatory collaborative design is that it ensures that the terms, definitions and domain-oriented language are all commonly shared in the community of training, and also transferred to the content analysts and designers. The findings of Tyrväinen & Päivärinta (1999) supported the selection of participatory analysis also: the authors point out that a single expert of an organization can usually define only a fraction of the genres used.

The method consists of four main phases, each containing multiple tasks. Figure 2 visualizes the phases. Three techniques are used within the method. Workshops (Coughlan et al., 2003) are used to allow trainers to reach consensus over the genres and the ways they should be enacted within the community. The metadata values are gathered using questionnaire forms filled up collaboratively. A wall-diagram technique (Saaren-Seppälä, 1997) is used for defining topics and subtopics of genres. In the following sections we first describe the phase in a general level followed by a description on how it was actually implemented at BIRD.

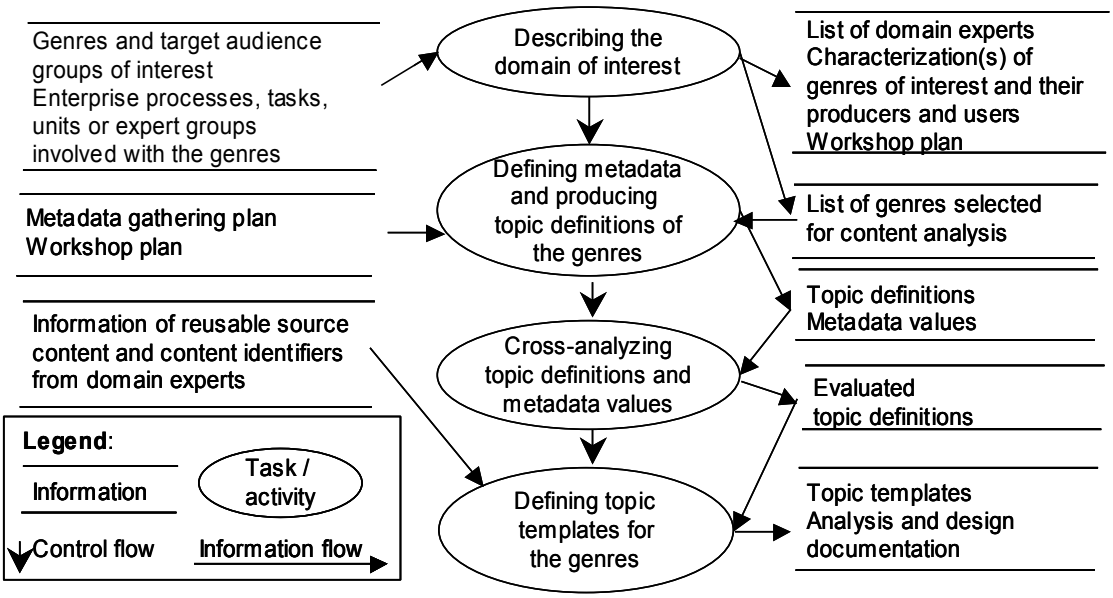

Figure 2. The method used for analyzing training content.

Describing the Domain of Interest

In Phase 1, expert groups for administrating and carrying out the development initiative, and for defining the genres of interest are defined. The 5W1H aspects for genres can be used as preliminary metadata categories for genre-based metadata definition.

In BIRD the analysis group established consisted of the manager of the training department, a development engineer from the customer documentation department, and a content analyst. The group listed and grouped domain experts available from BIRD and subcontractor organizations.

There were multiple perspectives from which to define the initial genres of training. A choice was made to use the training sessions for the four training groups as the genre candidates for further inspection. As some of the machines, due to their special characteristics, could require specialized training, there was a possibility that separate genres might be needed to be defined for some or all of the machines. Product structure of a production line seemed to form the common ontology of the domain, according to which the trainers seemed to organize their content (some machine designers worked as part-time trainers, too). The analysis group decided to scrutinize the training sessions pertaining machines as variant candidates for the four training group specific genres. Customer-specific variation was considered as an instantiation of the genre variants. Altogether, the final number of the genres that could be defined by the results of analyzing the candidate genre variants was expected to be between 4 and 80.

The group also studied the processes of content production, the relationships between the candidate and variant genres, and the machines. Reports of previous studies, organizational charts, and examples of training content and O&M documents were collected and analyzed. A workshop plan and preliminary metadata categories originating from the 5W1H framework were defined.

Defining Metadata and Producing Topic Definitions of the Genres

In Phase 2 the genres and metadata values pertaining them are scrutinized in workshops using questionnaire forms and the wall-diagram technique (Saaren-Seppälä, 1997). The metadata values are defined as the first assignment, and the wall-diagrams of topical content unit names within the genres as another. Results are presented and unified in joint sessions in between and after the assignments. Terms and ordering of topics are cross-examined across the groups and alternative definitions are evaluated and harmonized, when needed.

In BIRD a pilot workshop, six actual workshops, and two smaller working sessions were carried out. The genres were defined by producing wall-diagrams of topics and subtopics for a total of 21 genre variants. Most of the workshops involved training one machine for the four target groups. Thus there were four parallel expert groups each working on one variant of the four initial genres. Metadata values were collected collaboratively by filling in a questionnaire. Topic definitions for each genre variant consisted of 6-8 topics and 4-9 subtopics each. There were also sub-subtopics defined for a few genre variants.

Figure 3 represents an example of a topic definition of a genre variant. The top row defines the genre and the variant by the names of the target group and the machine, respectively. The topics and subtopics pertaining to the genre variant are recorded on the two columns on the left while notes and ideas are recorded on the right.

Training target group: key personnel, production line component: presser		
Topic	**Subtopic**	**List of items, notes**
Introduction		▪ trainer, schedule ▪ intro of content and training objectives
Structure	Steering equipment	▪ list of equipment part names and picture of equipment location ▪ a multimedia presentation of steering is available in URI (....)
Maintenance	Mechanical part maint.	▪ list of maintenance prodedures of our factory ▪ list of subcontractors and their maintenance qualifications

Figure 3. An example of a topic definition of a genre variant

The time needed for negotiations in joint sessions decreased in later workshops, and the consistency of definitions increased. This implies that the people gained commonly shared and more explicit understanding of the genres and their use on the domain. For the analyst, the workshops provided a possibility to discuss with experts in an informal way and to learn about the domain and language used.

Cross-analyzing Topic Definitions and Metadata Values

Phase 3 is carried out with a selected group of domain experts in one or several collaborative workshops. It consists of three steps:

1. cross-examining and harmonizing the topic definitions of the genres
2. producing a metadata definition for the genres by gathering, analyzing and harmonizing the metadata values collected in workshops
3. scrutinizing the hierarchy of genres and topics in the domain by analyzing and possibly re-defining the genres, and defining the logical units of content for computerized manipulation.

In BIRD the analysis group studied the workshop memos, metadata values and topic definitions. A coherent metadata definition was prepared, and the topic definitions were re-defined for consistency again. The initial definition of the four candidate genres, one for each training group session, was shown to be feasible. Variations based on machines were recorded into topic definitions according to topics and subtopics, resulting in unified topic definitions for each of the four genres, including information about the variants. The division and grain size of the genres were also considered as appropriate unit for computerized processing in the case of BIRD, whereas processing needs pertaining to topics was studied on the next phase.

Defining Topic Templates

Phase 4 aims at locating potential source content units and their identifiers from existing content repositories for content reuse and assembly. The expert group carrying out this phase should include trainers and experts from all source content domains. Potential source content can be identified using tables of content, document schemas, as well as database schemas of source content repositories. The content is studied by scrutinizing definitions and example contents from content source repositories for each genre and topic and subtopic on the genre topic definitions

In BIRD the analysis group studied the topic definitions with respect to the O&M manuals as a potential source of training content. The printed form of the manuals can consist of more than 40 folders of paper, and the content-oriented SGML DTD of the manual consists of about 200 elements and 130 attribute names. The customer documentation department had prepared a document, which defined the “sections, headings, subheadings and sub-subheadings” of the topical content of the manuals. Since the content for a genre could not be reused to form a unit of content in a O&M manual directly, the analysis group matched each topic and subtopic in the topic definitions against these listings for potential source content and its identifiers. For example, a topic “safety” on the training topic definition was studied with regard to O&M content by studying the O&M content definition document, example document instances, and possible element names and attribute values of the DTD definition. Properly identified source content heading or subheading names, DTD elements and attribute values for source identification were recorded in topic templates.

The documentation was published as “A trainer’s guide” in BIRD’s Intranet. The trainer’s guide consists of a metadata definition, topic definitions and topic templates with notes and examples. Metadata definition describes the four training groups, and provides other information on the timing and use of the training genres. A *topic template* is similar to a topic definition extended with the rightmost column for information of the topical (content headings, DTD elements) source content identifiers. Figure 4 provides an imaginary example of a topic template.

The topic templates resulting from the analysis process were made use of with manual copy-and-paste operations to assemble a couple of genre instances according to the templates. As this proved to be feasible, the analyst drafted a process of content assembly based on the use of XML and XSLT transformations.

Training target group: key personnel, production line component: presser			
Topic	**Subtopic**	**List of items, notes**	**Source**
Introduc-tion		▪ trainer, schedule ▪ content and training objectives	
Structure		▪ list of equipment part names and picture of equipment locations	Content: Section5/Cookpit DTD: section (id="sect5")/ heading (partid="cookpit") ▪ Multimedia presentation, query from database "MM", table "components" in which compnent name is "steering/construction"
	Steering equip-ment	▪ remember to introduce following new components: ▪…	Content: Section5/Cookpit/Steering DTD: section (id="sect5")/ heading (partid="cookpit")/ header(partid="steering")

Figure 4. An example of a topic template of a genre variant

Genre-based Templates in Action: an Example of Using XSLT Transformations

In XML, content parts are *elements* that are defined by using start and end tags as delimiting mark-up. *Attributes* are used for defining characteristics of elements. Allowed names for elements and attributes can be defined by a DTD or an XML schema, which also contains the rules for element and attribute organization with respect to each other. XML documents can be *transformed* by using a language such as XSLT. Within transformation process, content parts from the source document can be located by element names or attribute values. Located source content can be re-arranged and mixed with static content parts defined within a transformation script when producing an output document. A document assembly system can use a transformation script, such as an XSLT template, for producing different versions of the document content for separate target audience groups, or for delivery via multiple media, such as the Web or e-Books. By using the XML language the content can be separated from the layout, and we can produce multiple outputs from a single content source, or combine multiple sources for producing a single output.

Heikkinen (2000) has envisioned a large repository of SGML documents, from which users could search, browse and pick up part content for producing tailored training materials. Personalization for target audiences could also be based on using specifying attribute values that define the content for each target audience. However, Hillesund (2002) has claimed that using XML for producing multiple outputs from a single source, or a single output form multiple sources is not feasible for producing e-books for learning. She claims that the use of XML is too complex for trainers and training content designers, and that the separation of content, layout and structure by using XML is, in practice, impossible.

In BIRD, the material to be reused includes heterogeneous content sources from multimedia presentations to SGML documents. Most of the source content can be located from the O&M manuals, which are structured in accordance with a single, content-oriented DTD. Each production line machine is documented as an instance of this DTD. As each machine and production line is a one-of-a-kind product, a new version of an O&M manual according to the variance of the individual machine is created for each. One cannot predict the parts whose content will remain the same as in previous O&M manual constructions. Neither can one predict the topics of training content that could be reused as such. Instead, one can define the O&M manual schema structures that can be used to point out the reusable units of content from O&M manuals as the raw source material for the topics of training material. We defined these in the topic templates by using the genre-based method. The next phase is to turn them into a machine-readable form called genre-based templates, which enable (semi-)automation of content assembly of raw training material. This enables support for assembly operations by programs that locate the potential source topics, re-organize them, and add static content for raw training material output. If we do not have the means for analyzing and unifying the multiplicity of training session content combinations, we can end up with the need to maintain tens of scripts or a large program, which cannot be done by the trainers. In the following text we envision the use of the genre-based templates that provides us with means for managing this complexity and for documenting the training content and source content topics in

a unified form that both IT and non-IT people can understand. In the following, we leave out most of the complexity and phases of a content assembly process, which have been discussed in more detail by, for example, Heikkinen (2000) and Honkaranta (2003b). We also assume that the source content has already been transformed to XML.

A *genre-based template* (GBT) is an XML document that defines the mix of source content units to be collected and static content to be added for producing a raw training content for a genre. It uses its own DTD, which describes the rules for copying the source DTD elements and attributes to raw training material output. By using the GBTs, the trainers can write and modify one XML definition document related to each training group. For example, a "Keypersonnel.xml" could consist of the GBT definition as a XML document for training the key personnel. Editing this XML document will change the content for the training group. Instead of writing a transformation script for assembling raw training material content for each training target group defined, we can use a single transformation script. The XSLT transformation script would read the content of a GBT document (e.g. "Keypersonnel.xml") and produce an output document that consists of content topics defined for the key personnel.

Documents of raw training content for each genre variant (i.e. machine) can be produced by applying the same XSLT transformation script and the GBT of the training group to the O&M manual of the machine. Further, the variations of customized machines will produce new variants of the source XML documents. They all can be processed with the same transformation script and the same four GBTs for the training groups as they all use the same O&M manual schema.

Figure 5 illustrates the process for the example of the topic template visualized in Figure 4. In Figure 5 the source document (originating from an O&M manual) and the GBT in the form of a XML document are illustrated as document trees consisting of nodes (depicted by circles in the figure). Each node represents one XML element. The text written beside the document nodes refer to document element names and attribute values (written as attribute name="value"). The rectangle on the right illustrates the output document. The "TRANSFORMER" depicted by a rectangle below denotes an XSLT transformation template document that produces the raw output for training material document.

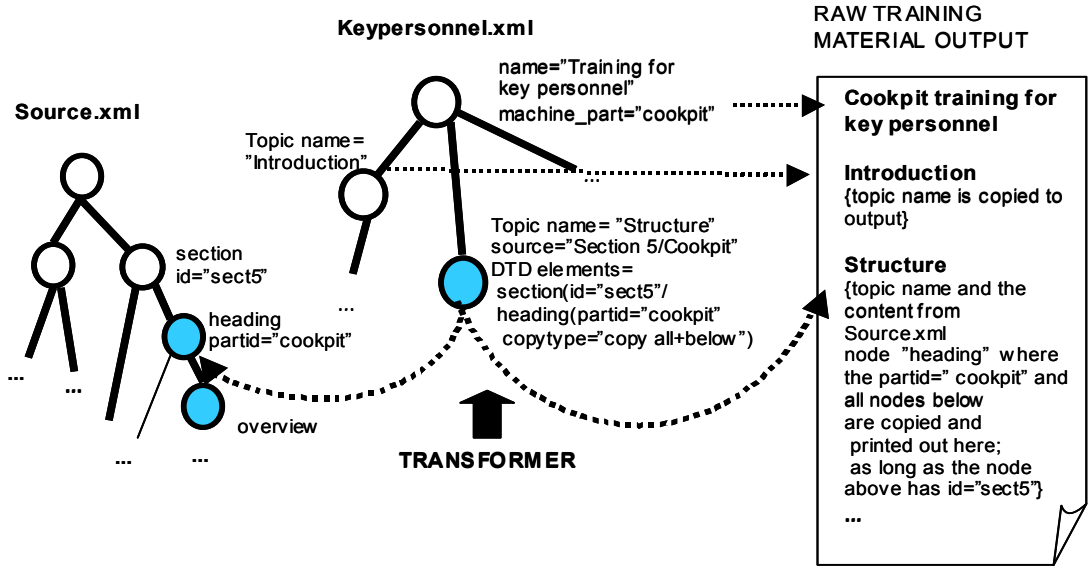

Figure 5. Example of content assembly using genre-based template as XML document

An XML parser and XSLT transformation template process the documents as tree-type content schemas as illustrated in Figure 5. An XSLT template can copy the content of a node as a whole, including all the elements that belong to the nodes subtree. As illustrated in Figure 5, the output document can be produced as a combination of content presented in the "Keypersonnel.xml" file (such as the text "Introduction" and "Structure") and content copied from the source document. Changing the GBT document that the transformer reads (e.g. instead of "Keypersonnel.xml" the transformer may use "Operators.xml") we can produce another

kind of training content output document.

Discussion

The topic templates can be used as a human-oriented definition of genres and their variants for a training group, as well as models for defining the genre-based templates needed for assembling the raw content for training. The topic templates define the content units needed for producing indexes and navigating the content, and producing specialized versions of the content for training groups. Naturally, the raw content assembled from O&M manual content needs to be edited for training purposes. The O&M manuals and their content as a source provide trainers with a possibility to acquire knowledge from designers about each machine that is tailored for a customer.

The use of topics and subtopics proved to be useful from two viewpoints. First, it enabled analyzing the feasibility to use the four genres by providing the means by which to extract the variation from the genres. It was now possible to separate the variations inherent to the information content specific to each machine from the variations caused by individual preferences of the trainers. The variation due to preferences could be harmonized using the participative techniques. The machine-specific variation could later on be attached to the four unified main genres. Secondly, it enabled collecting information about additional sources of source content, which was made available by the experts of the related units participating in the workshops.

If the company in the future wishes to deliver content also as e-Learning materials via the Web, the XML-based training content can make the content production smoother. There is also a possibility for extracting XML-marked content from training documents, and adding this information into e-Learning content metadata definitions, such as SCORM and LOM, or content packaging manifestos (e.g. IMS-CPIM, 2001) of the content by using XML-related languages for automated content processing.

In BIRD, the accurate grain size for a reusable content unit, i.e. a candidate for LO unit as defined by Polsani (2003), turned out to be a topic within a genre. As the topics can be marked-up within XML documents they are accessible for reuse and processing even though they are not stored as separate learning objects. For e-Learning, a training content for each of the training groups would be a suitable unit of content to be defined by using e-Learning metadata standards. It would consist of a package of training sessions. It seems that the needs for defining the units of content in BIRD differ from the perspective of e-Learning community. A great deal of source content is available as SGML, and the use of XML and content-oriented mark-up schemas reduce the need for defining LOs as physical, reusable content objects.

In BIRD, the reuse of content would also mean reusing structure of content rather than reusing content objects, as envisioned by the e-Learning community. For example, Polsani (2003) refers to the reuse of LO instances for creating learning content. In BIRD, each machine and production line is one-of-a-kind product, and changes to learning content are common. When the learning content reuse is designed, one can reuse the content structure and define the location of a reusable content for a topic by referring to a schema of O&M manual content. When a new version of the manual is produced a (semi-)automated content assembly script should be able to pick up the right instance.

From a methodology viewpoint the use of the genre “lens” provided a conceptual background for the study, and insights for analyzing the massive content of training. The genre “lens” focused the emphasis on user participation and collaboration thus providing support for Holmberg’s (1989) notion that the logical content of training can be defined by domain experts. Also the “5W1H” metadata framework (Yoshioka et al., 2001) proved to be usable for designing the metadata categories. The use of the workshop technique (Coughlan et al., 2003) was successful and in line with the findings of previous research promoting participatory design (Honkaranta & Lyytikäinen, 2003; Berkenkotter & Huckin, 1995) of a “specific interest group” (Swales, 1999) hindering the possible communication breakdowns between an analyst and the training community (Byrd et al., 1992). The methods used were also adopted by another unit of BIRD, the implication being that these methods were considered as feasible and usable.

Conclusions and Further Research

In this paper we discussed customer training in a manufacturing organization dubbed BIRD as a complex domain in which the genre “lens” and related findings from previous studies were extended with topic analysis

for training content. In BIRD the variations of training genres were analyzed based on their metadata and topic structure. The machine-based variation of training content structures was captured in topic definitions, and potential content sources for the topics were identified in topic templates. The content of training in BIRD turned out to be massive and complex with multiple levels of content units. The content of training consisted of four content collections with respect to the four training groups. The content for each training group contains variations of content structures and content for separate machines of a production line - these were enacted as training sessions in the domain of BIRD. Use of the genre-based method allowed the definition of the main content units and gathering of contextual metadata, while extending it with the topic hierarchies enabled definition of the content unit hierarchy and study of potential source content units for content reuse. By using the method, the topics and subtopics of training genres were negotiated and defined amongst training experts. In the end, we described the use of genre-based templates for automating parts of content reuse from existing source content. Rather than reusing learning objects as such, we apply content assembly according to static transformations for producing raw learning material.

The results seem to encourage the use of the method and provide avenues for further research. Due to the lack of platform in which to test the assembly in BIRD, more detailed system development and testing was left to be carried out in a further research. For example, the content assembly was considered to be embedded into a product data management system at first, but later the company has considered purchasing a Learning Management System (LMS). Also the requirements for learning content types and elements, such as content indexes, examples, references, didactic conversation, assignments (Holmberg, 1989) and multimedia objects and glossaries (Polsani, 2003) need to reflected in a learning content. How these could be included in the existing methods or evaluated for learning-oriented schema design methods remains an interesting and challenging avenue for further research. Other interesting questions still open include:

- Should the O&M manual structure (i.e. the source DTD) be optimized for aiding production of LOs?
- How could the transformation approach presented here be combined with the use of existing, more abstract LOs in the best possible way?
- Could we embed the method to schema design methods, such as the Maler & El Andaloussi (1996) method, and how the learning content types and elements could be embedded in the design process?

References

ADLI (2002). *Advanced Distributed Learning Initiative,* http://www.adlnet.org/.

Berkenkotter, C. & Huckin, T. (1995). *Genre knowledge in disciplinary communication: Cognition/Culture/Power*, Hillsdale, New Jersey, Hove, UK: Lawrence Erlbaum Associates, Inc.

Bray, T., Paoli, J., Sperberg-McQueen, C. & Maler, E. (2000). *Extensible Markup Language (XML) 1.0. (2nd Edition),* http://www.w3.org/TR/2000/REC-xml-20001006 .

Byrd, T., Cossick, K. & Zmud, R. (1992). A synthesis of research on requirements analysis and knowledge acquisition techniques. *Management Information Systems Quarterly* , 16 (1), 117-138

Clark, J. (1999). *XSL Transformations (XSLT) version 1.0,* http://www.w3.org/TR/xslt.

Coughlan, J., Lycett, M. & Macredie, R. (2003). Communication issues in requirements elicitation: a content analysis of stakeholder experiences. *Information and Software Technology,* 45 (1), 525-537.

Crowston, K. & Williams, M. (1997). Reproduced and emergent genres of communication on the World-Wide Web. In Sprague, R. (Ed.) *Proceedings of the 30th Annual Hawaii International Conference on System Sciences,* Los Alamitos, CA: IEEE Computer Society Press, CD-ROM.

Fallside, D. (2001). *XML Schema part 0: primer,* http://www.w3.org/TR/xmlschema-0/.

Heikkinen, B. (2000). *Generalization of document structures and document assembly*. Ph D. Thesis. Helsinki, Finland: University of Helsinki, Department of Computer Science.

Hillesund, T. (2002). Many outputs - many inputs: XML for publishers and e-book designers. *Journal of Digital Information* , 3 (1), 22 p. Available: http://jodi.ecs.soton.ac.uk/Articles/v03/i01/Hillesund/

Holmberg, B. (1989). *Theory and practice of distance education*, London, New York: Routledge.

Honkaranta, A. (2003a). Developing document and content management in enterprises using a "Genre Lens". In Camp, O., et al. (Eds.) *Proceedings of the 5th International Conference on Enterprise Information Systems,* Portugal: Escola Superior de Tecnologia de Setubal, (Area 3) 334-340.

Honkaranta, A. (2003b). Evaluating a "Genre Lens" for analysing requirements for content assembly. In Siau, K., Krogstie, J. & Halpin, T. (Eds.) *Proceedings of the 8th CAiSE/IFIP8.1EMMSAD Workshop,* USA: CAiSE/IFIP8.1 WG, 95-105.

Honkaranta, A. & Lyytikäinen, V. (2003). Operationalizing a genre-based method for content analysis: A case of a church. In Abramowicz, W. & Klein, G. (Eds.) *Proceedings of the 6th International BIS Conference.* Poland: Poznan University of Economics, 108-116.

IEEE-LTSC (2002). *IEEE Learning Technology Standards Comittee (LTSC),* http://grouper.ieee.org/groups/ltsc/index.html.

IMS-CPIM (2001). *IMS Content Packaging Information Model. Version 1.1.2 final specification,* http://www.imsproject.org/content/packaging/cpv1p1p2/imscp_infov1p1p2.html.

IMS-CPXB (2001). *IMS Content Packaging XML Binding. Version 1.1.2 final specification,* http://www.imsproject.org/content/packaging/cpv1p1p2/imscp_bindv1p1p2.html.

IMS-GLC (2002). *IMS Global Learning Consortium, Inc,* http://www.imsproject.org/aboutims.cfm.

IMS-LRMDI (2000). *IMS Learning Resource Meta-data Information. Version 1.2.1 - final specification,* http://www.imsproject.org/metadata/mdinfov1p1.html.

Karjalainen, A., Päivärinta, T., Tyrväinen, P. & Rajala, J. (2000). Genre-based metadata for enterprise document management. In Sprague, R. (Ed.) *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences,* Los Alamitos, CA: IEEE Computer Society, CD-ROM.

Korhonen, V. & Väliharju, T. (1995). Learning, learning environments and hypermedia. In Panzar, E., et al.(Eds.) *Theoretical foundations and applications of modern learning environments,* Tampere, Finland: University of Tampere, Computer Centre/Hypermedia laboratory, 49-69.

Maler, E. & El Andaloussi, J. (1996). *Developing SGML DTDs. From text to model to markup,* Upper Saddle River (NJ): Prentice Hall.

Mcknight, C., Dillon, A. & Richardson, J. (1991). *Hypertext in contex,* Cambridge: Cambridge University Press.

Microsoft (2002). *Microsoft LRN 3.0 Toolkit,* http://www.microsoft.com/elearn/support.asp.

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

Pepper, S. (1999). Navigating haystacks and discovering needles. *Markup Languages: Theory and Practice,* 1 (4), 41-68.

Polsani, P. (2003). Use and abuse of reusable learning objects. *Journal of Digital Information,* 3 (4), 10 p. Available: http://jodi.ecs.soton.ac.uk/Articles/v03/i04/Polsani/.

Päivärinta, T., Halttunen, V. & Tyrväinen, P. (2001). A genre-based method for information systems planning. In Rossi, M. & Siau, K. (Eds.) *Information Modelling in the New Millenium,* Hershey, PA: Idea Group Publishing, Inc., 70-93.

Rasmussen, J. (1985). The role of hierarchical knowledge representation. *Decisionmaking and System Management. IEEE Transactions on Systems, Man, and Cybernetics,* 15 (2), 234-243.

Saaren-Seppälä, K. (1997). *Seinätekniikka prosessien kehittämisessä. (Using the wall-chart technique in process development),* Finland:Kari Saaren-Seppälä Ltd.

SCORM (2001). *Sharable Content Object Reference Model (SCORM) Version 1.2. The SCORM overview. October 1, 2001,* http://www.adlnet.org/index.cfm?fuseaction=scormabt.

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

Swales, J. (1999). *Genre analysis. English in academic and research settings*, Cambridge: Cambridge University Press.

Tyrväinen, P., Järvenpää, M. and Sievänen, A. (2003). On estimating the amount of learning materials - A case study. In Camp, O., et al. (Eds.) *Proceedings of the 5th International Conference on Enterprise Information Systems,* Portugal: Escola Superior de Tecnologia de Setubal, (Area4) 127-135.

Tyrväinen, P. & Päivärinta, T. (1999). On rethinking organizational document genres for electronic document management. In Sprague, R. (Ed.) *Proceedings of the 33th Annual Hawaii International Conference on System Sciences,* Los Alamitos, USA: IEEE Computer Society Press, CD-ROM.

UML (1997). *UML notation guide. Version 1.1,* http://www.rational.com/uml/resources/documentation/notation/index.jtmpl.

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

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.