Using StoryML for Content Representation

Petri Gerdt[1], Jarkko Suhonen[1], and Erkki Sutinen[1]

Dept. of Computer Science, University of Joensuu
{pgerdt,jsuhon,sutinen}@cs.joensuu.fi

Abstract. The Woven Stories is a tool for writing stories collaboratively. In order to represent the written stories adaptively, we need to construct description about the content of the stories and the relative information between different stories. The formal presentation of Woven Stories is called StoryML and it can be used for adaptive content representation of the stories.

1 Introduction

Adaptive IR tools in web-based education are essential to represent the content differently according to the specific needs of the students. For example the learning methods and styles between students with holistic-style of learning and students with serialistic-style of learning differs tremendously. It has been shown for example that the holistic-style of learners have flexible and easier way of solving problems than serialistic-style of learners. This means that there is a clear need for representing the learning materials according to nature of the students learning [4, ?]

The *Woven Stories* is a tool where several authors' efforts is managed in a shared writing space, where the authors may write story sections and link them together. A woven story is thus a hyperdocument or alternatively hyperspace, which consists of an arbitrary set of story sections and links between them. The Woven Stories is intended to be used to support group activities in educational settings, such as different distance education situations. The users of the tool can construct, manage and share information with it. Thus the environment works as a backbone of group work for a group of students. A prototype system for the environment labeled Woven Stories has been described in [2]. In this paper we will present our ideas that we are developing for the next implementation of the environment.

The Woven Stories can be defined as hypermedia in following way. The Woven Stories visualizes the content of a hyperspace as a graph, whose nodes are content pages, passages of text which may include hyperlinks. The arcs of the graph represent links between the content pages. Each content page may be written by a single user or several users. The Woven Stories uses an XML-based data structure to store information about the hyperspace. [3]

One of the basic requirements for the Woven Stories is the need for adapting the representation of the content for example between holistic-style and serialistic-style of learner. Hence, the important feature of the Story ML is to capture the episodic nature of stories. However, in our opinion such adaptation techniques are not available [1].

A node in a concept map is often related only to its neighboring concepts. One might say that a concept map represents only the local information of a particular concept. On the contrary stories typically represent episodes that consist of a several sub-sequent story sections connected by links. The analysis of episodes could be useful to make assumptions about the students' reasoning during the learning process. Moreover, information on different episode patterns could be stored by the StoryML specification, and actual story lines corresponding to a given pattern could be retrieved on demand. An example of episode patterns is a sequence A1-B1-A1 which represents a three step dialog, between students A and B.

Furthermore, the users of the Woven Stories can customize their views of the collaboratively constructed hyperspace. A member of the group may need a specialized view of the hyperspace to support a specific activity. If the students co-author an essay one of the students may be an editor who needs to all material produced by the group. On the other hand an author constructing the introduction of the essay will not want to see everything when she is concentrating on her own contribution.

To support the user specific customization of the view to the hyperspace the environment collects user model for each user. The user model includes data about the users' view preferences, personal settings (such as the user name and the color representing the user) and the browsing history of the user. The environment extracts much of the data for the user model from the XML-based description of the hyperspace and its contents, but it collects some of the data by examining user actions (see section 3).

There are no limits for the size of the hyperspace in the environment. Thus the users may write a single content page together, or write several, perhaps interlinked, pages. The hyperspace can become very complex if it contains many pages written by several authors. In this case a user may not find what she wants or she may not notice some input that she should. To help the users to cope with in the hyperspace the environment provides searching and information filtering tools, which are outlined in section 4.

A challenge for the environment is to adaptively support very different views as the users customize their personal views of the hyperspace and use the information filtering and search tools that the environment provides (see section 5).

2 Adapting Content Representation in Woven Stories

In educational setting the adaptive content representation in Woven Stories means that the same topic or story can be presented in multiple ways to different users. Hence the representation of the stories can be varied according to the learning style of a particular learner.

For example there are evidence that concept mapping fits best to the natural study style of the holistic [4]. It proved to be difficult to make serialistic students benefit from the concept mapping approach. Here we expect to find woven stories as a beneficial compensation for the problems holistic students have in chronological reports. In summary: woven stories and concept mapping have the natural tendency to be mutually supportive. In global terms we may expect that the serialistic students will benefit more from concept mapping, while the holistic students get a typical compensation from the woven stories approach.

In order to make the transformation between different types of representation we need a StoryML description. Figure 1 presents the general idea of transforming a particular topic to various modes of presentation. In Figure 1 CM indicates concept map and WS woven stories. Topic *t* can be for example a physical phenomenon or course contents. The change from Woven Stories representation to concept map form is straight operation. The StoryML includes all the necessary information. The transformation from concept maps to Woven Stories is a little bit elaborated, because some information needs to be collected from the authors before the transformation can be done. For example the order of the concepts does not necessarily appear in the concept-like representation.

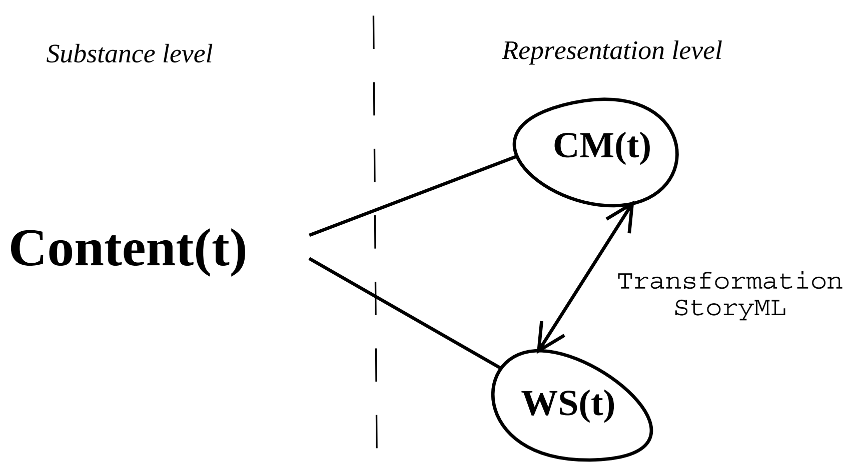

Fig. 1. An Example of Different Representations for a Given Topic t

3 Data Management And The User Model

We have defined a XML-based data structure, called StoryML, to describe a hyperspace and its contents in the Woven Stories environment [3]. The StoryML is divided into five main categories which are summarized in table 1. Each manifestation of StoryML represents a given hyperspace.

The elements presented in Table 1 can be stored in three different layers: locally, episodically and globally. The *local* metadata description deals with the single story node and its relationship to the other nodes. The *episodic* level includes two or more nodes that form a distinctive episodic structure. The episodes can be defined by the authors of the stories or by the Woven Stories environment. The *global* level includes all the single story nodes and episodes in one hyperspace of stories.

Naturally the content of a particular StoryML element can differ when the story node is treated locally, episodically or globally. Furthermore the content of the element may change when the structure of the stories is altered. For example the union of two stories probably creates a new set of metadata elements. The metadata description of different levels can be done in couple of ways. In first approach the metadata description of every story node holds the data about the local, episodic or global information. Second approach creates metadata descriptions to different layers. Hence, every episodic structure has its own metadata description.

A user model for a specific user can partly be extracted from the Author and Content elements. The Author element stores information that the user has inputted about herself: name, affiliation, color, email and preferences. The content element on the other hand has knowledge about what the users have been doing in the environment. This is a case of collaborative user modeling: the user inputs data about herself and her preferences and the environment builds a part of the user model by examining what the user has done in the environment.

To complete the user model the environment records data about the users behavior: the environment collects the browsing history of the user, and data about any searches and filtering operations that she has done.

4 Adaptive Operations Using StoryML

We plan to include various tools in our environment in order to let the users customize their views and find relevant data.

Table 1. The main categories of the StoryML core element set.

Element	**Description**
General	General information about the Woven Stories.
Visualization	Determines the visual appearance of a graph-like Woven Story.
Content	Descriptive information about the content (e.g. keywords).
Author	Data about the authors who have been active in the hyperspace.
Episodes	Links and time factors between story nodes.

4.1 Information Filters

Information filters provide a subset of the hyperspace. They are not intended to be used to find a specific piece of content, but to restrict the hyperspace by some constraints to make it easier to handle.

Subsets of the hyperspace defined by a subset of the authors. The most intuitive use of filtering the hyperspace by a subset of the authors would be a user checking her own contribution. By broadening the filter to include other authors the user can for example examine how the contributions of others' are related to her own ones. A teacher can examine what each student member of a group collaborating in the environment has done.

Subsets defined by creation date. The environment may present a subset of the hyperspace, which is limited by the creation or modification dates of the content. This filter may be implemented as a dynamic view, where the user can alter the subset with a slider component, which controls the date parameters. We call this filter type the time line filter, as the user can follow a time line along which the hyperspace is constructed visualizing the creation process. This filter is useful when a user wants to see what has changed to the hyperspace since she last time logged into the environment.

4.2 Search Mechanisms

Searches to the hyperspace are made to find specific content. The search can be made to a filtered subset of the hyperspace. An user can for example search for a keyword among content produced by a specific user.

Keyword search. The authors may attach keywords to the content pages. These keywords can then be used by the search engine to find certain topics.

Searching the written part of the content pages. Unfortunately keywords can be interpreted differently by different people, or they may be missing altogether. In this case free form searches are needed. The environment lets the users to search the content pages with a arbitrary string, looking for matches among all written content.

Searching among the relationships of the content pages. The relationships, or links, between the content pages can be searched. The user may for example look for all content pages linked to a specific page.

Other search mechanisms. Different context analysis heuristics can be used to analyze the content of the text. We can extract such information as subject of the matter, the intention of the text or even the emotional "tone" of the text by using Natural Language Processing (NLP) and Machine Learning (ML) [6]. With this functionality the environment can search for content pages and links based on the deep understanding of the actual content of the hyperdocument. These analysis heuristics may be used to create filters too.

5 Adapting The Results

The user model maintained by the environment makes it possible to present the results provided by searches and filters adapted to the situation of the user performing the operations.

One example of adaptation in the environment is the highlighting of such content pages in a result of a filtering or search that the user has not visited. This adaptation of results use the browsing history of the user who makes the filtering or search.

Another approach for the environment to adapt the results is to use the browsing history of other users. The environment can suggest such results of searches to the user that other users have found useful. Furthermore the environment can use the user models of others together with a specific user model to show the user possible searches or filtering options that other users in her situation have used.

A more complex way to adapt the view of a user would be the environment presenting content pages written by others to an author in order to motivate or help him. For example if a user A often collaborates with user B, then the environment can notify A about B:s behavior and writing, or the it can weight search result that include contributions from B more than other users contributions.

6 Conclusion

In this short paper we have presented our ideas for a collaborative hypermedia authoring environment that uses a XML-based representation for a hyperspace, provides different search and filtering tools and adapts the results of these tools to support its users. We are currently working on the implementation of the environment.

References

1. Brusilovsky, P. Adaptive hypermedia. User Modeling and User Adapted Interaction, Ten Year Anniversary Issue (Alfred Kobsa, ed.) **11** (1/2), 2001, 87 – 110.
2. Gerdt, P., Kommers, P., Looi, C., Sutinen, E.: Woven Stories as a Cognitive Tool. **Cognitive Technology 2001**, LNAI 2117, Springer-Verlag, 2001, 233 – 247.
3. Gerdt, P., Kommers, P., Suhonen, J., Sutinen, E.:**StoryML: A XML Extension for Woven Stories**. Accepted to be presented in ITS 2002.
4. Huai, H.; Cognitive Style and Memory Capacity: **Effects of Concept Mapping as a Learning Method**. Doctoral Thesis, University of Twente, 2000.
5. Huai, H., Kommers, P.: **Concept Mapping as a Learning Strategy for Autonomous Students with a Serialistic Cognitive Style**. In International Journal of Continuing Engineering Education and Life-Long Learning, 11 (1/2), 135 – 151, 2001.
6. Paik, W. Automatic Generation of Educational Metadata Elements to Enable Digital Libraries. **Proceedings of the 9th International Conference on Computers in Education** (ICCE/SchoolNet2001), Seoul, Korea, November 12-15, 2001, 1406 – 1413.