Partial Generation of Contextualized Metadata in a Collaborative Modeling Environment

Niels Pinkwart, Marc Jansen, Maria Oelinger, Lena Korchounova, Ulrich Hoppe

University of Duisburg-Essen, Faculty of Engineering,
Institute for Computer Science and Interactive Systems
Lotharstr. 63, 47048 Duisburg, Germany
Phone: +49 / 203 / 379-1403, Fax: +49 / 203 / 379-3557
{pinkwart,jansen,oelinger,korchounova,hoppe}@collide.info

Abstract. In asynchronous collaboration scenarios, document metadata play an important role for indexing and retrieving documents in jointly used archives. However, the manual input of metadata is usually an unpleasant and error prone task. This paper describes an approach that allows the partially automatic generation of metadata in a collaborative modeling environment. It illustrates some usage scenarios for the metadata within the modelling framework – including concepts for document based social navigation and ideas for tool embedded archive queries based on the current state of the user's work.

1 Exploiting Task Context in a Modeling Environment

Document metadata play an important role for indexing, retrieving and thus re-using material in collaborative learning scenarios. Yet, one critical problem concerning metadata is the reliability of user provided data. Users tend to minimize their efforts describing their documents. So incompleteness and inconsistency of metadata are frequent; incompleteness results from partially skipping of information, different understandings by various users cause inconsistency. Using standards may help, but this demands additional effort from the users, namely to learn the terminology. Yet there is a need for annotated, classified material with a rich context, which supports retrieval on a higher level than a just content based text search. An automation of metadata generation would disburden the user from these problems. A general solution seems out of reach. Our approach is based on using task (esp. tool) information available in a multifunctional collaborative modeling framework. Beyond metadata generation it also supports certain forms of content based awareness in a user community.

Cool Modes (COllaborative Open Learning and MODEling System) is a collaborative tool framework designed to support discussions and synchronous cooperative modelling in various domains. Like some other environments [e.g., 1], Cool Modes supports synchronous cooperation by a shared workspace environment with coupled objects. In Cool Modes, these objects can be defined externally, which offers the option to develop domain dependent plug-ins. These consist of five different elements (see [2] for details):

- Primitive element types (atomic elements of the constructed representations)
- Connection types that can be used to link different primitive elements
- Usage rules to restrict the syntactic structures a user can create
- Interpretation patterns (like rules and algorithms) that can be applied on constructed graphs in a workspace – the means to e.g. realize simulations
- A UI that offers the elements for direct manipulation and may contain specific control elements for the plug-ins (e.g., start buttons for simulations)

These system extensions can differ considerably with respect to the underlying formal semantics (e.g., Petri nets vs. handwritten annotations), but all can be used synchronously in an integrated way, mixing and combining different conceptual representations (see [3] for details).

2 Archiving and Retrieval Functions in Cool Modes

Some of the metadata slots we use have been inspired by existing standards like Dublin Core [4]. The basic standard conform set of metadata is enriched by adding tool dependent metadata like the used plug-ins which indicate a domain context of the activity, e.g. the use of the stochastics plug-in would indicate a probability theory activity in a math course. Thus the usage within an educational context is possible as well as in the specific group of Cool Modes users. Several of the educational scenarios for the system design originate from the European project COLDEX[1] which enables scientific modeling and simulation in distributed collaborative settings. The learning domain here is the study of various phenomena from a scientific and a subjective experiential perspective. Tools like Cool Modes can contribute to forming integrated (a)synchronous access to a "group memory" on different levels. In this context, the development of scenarios for the automation system has been motivated by pedagogical scenarios of COLDEX, based on the following use cases: A teacher organizes and supervises learning groups, i.e. he creates new projects, or groups. He assigns projects, students, and groups. A student collaborates with his group using Cool Modes. After registration, any user can retrieve documents using a query form.

Based on the standards and use cases as identified in the previous section and specific features available by the use of Cool Modes as a collaborative and customizable framework, we have defined the metadata types to be used. Some of these are independent of the tool that generated the data, some are specific to Cool Modes. Another dimension to categorize the entries is in how far they can be generated automatically. Some possible levels are “completely and deterministically”, “inferable under certain conditions” or “not at all”. The following table systematically classifies the entries with their attributes.

It is observable that most of the entries are independent of the concrete tool that is used and that for all the metadata entries, at least a heuristic or a calculation based on previous document versions is possible. Furthermore, the fields which build the core for the querying scenarios can be generated automatically.

---

[1] Collaborative Learning and Distributed Experimentation, EU IST project No 2001-32327, http://www.coldex.info.

Table 1. Properties of metadata information types

Entry	Dependency	Possible generation type
Name	independent	proposal inferable for higher versions of documents
Description	independent	proposal inferable for higher versions of documents
Keywords	independent	heuristics based on keyword comparisons with the archive possible; proposal inferable for higher versions of documents
Write flag	independent	manual input needed, yet “false” is a reasonable default
Creation date	independent	fully automatic
Version	independent	fully automatic (1.0 for initial versions, increment for follow-ups); manual conflict resolution may be needed
Language	independent	heuristics based on user profile and previous versions of documents
Used plug-ins	customizable tools	fully automatic
Authors	independent	fully automatic
Cooperators	collaborative tools	fully automatic
Project	independent	proposal inferable for higher versions of documents
Activity	independent	proposal inferable for higher versions of documents; heuristics based on project and user information

Based on this metadata structure, we have implemented a number of search and retrieval mechanisms that work on the archive. The guiding principles and central features can be summarized as follows:

The core consists of *generic archive queries*. These can be formulated in an SQL style syntax and allow the flexible access of the archive in an expressive language. However, for most users (considering the intended educational context), the manual formulation of queries is definitely too complicated. Thus, some of our developments aimed at realizing an *associative lookup strategy* (cf. [5]) that is visualized in figure 1: The user can take any document that has metadata information attached to it and *generate a query from this document* in one simple generalization step – the system creates a query template in which the attribute values are taken from the source document and all the attributes are marked as “relevant”. The user can then modify the content of slots to modify the search query, “free” specific slots (to make them return values rather than search parameters) or label specific slots as “not relevant”, which hides these slots in the result. The modified query is then run against the community database and returns a set of matching document metadata. These can be used to form new queries, or the documents can be opened in Cool Modes. As the slots that were used for the search process represent information that holds semantics and is related to the educational ontology, this allows for navigation and retrieval which considers educational and domain related contexts.

Based on this retrieval function, we have added tool embedded support for *task contextualized queries*: A user working within the Cool Modes environment has the option of asking “is there something in the archive similar to what I am working on” without providing additional manual information. The needed system side knowledge about the domain and task context the user is currently in is generated by the approaches described above. Especially under the condition that a user is working with material that (at least in parts) originates from the archive, semantically rich queries that take into account domain and task context are enabled.

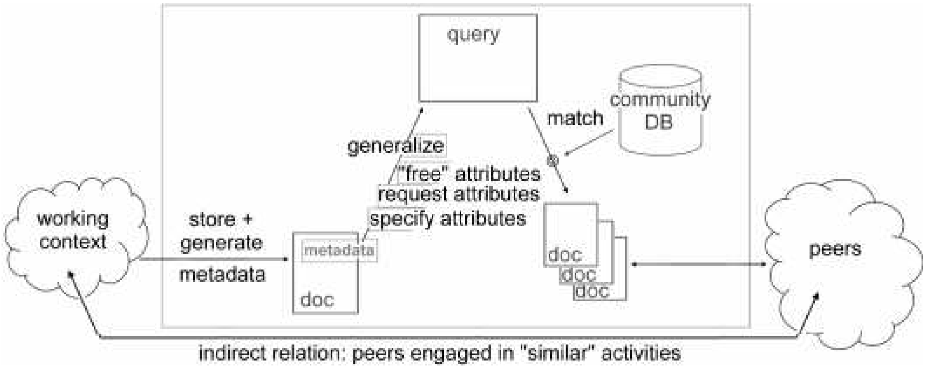

Fig. 1. The use of documents as search query templates

With the retrieval strategies described above, it is possible to allow users to take their current document as a query template, search for related content in the archive and display to the user not just the found documents themselves but a ranked list of other users that created these documents. Under the assumption that the automatic generation of metadata works sufficiently, this approach enables finding peers or groups of similar interest and, iterating this process, to realize *social navigation* [6] based on the document archive (see illustration in right part of figure 1).

Furthermore, the retrieval strategies as presented build a good base to personalize the system by a *what's interesting* feature, either on a web page (this is the way we realized it) or tool embedded. This feature performs some of the queries as presented above automatically, considering additional constraints like language, time, and organizational proximity of persons and is thus able to propose documents of potential interest to the user that log into the system.

3 Architecture and Implementation

Cool Modes enables the upload of documents to a central database which contains not only the documents themselves, but also metadata information about them. Additionally, metadata are stored within the document itself. This approach allows the metadata to be used without the document itself, and the user can also work offline. The retrieval of the documents is either possible via a web interface or directly from Cool Modes. A WebService is used to evaluate queries against the database.

A high degree of flexibility is gained by using an XML structure to describe general "query patterns". A query pattern can contain two types of slots: slots with values which are used as filters, slots without values which represent "requested" attributes. Irrelevant attributes are not included. The example in figure 2 shows such a pattern which abstracts from the author but fixes the "plug-in" feature as an indicator of task semantics. A possible result is shown on the right side. After sending the request the user has the choice of either downloading one of the presented documents or to refine her query based on the results she received, e.g. she may want to refine her query if the result does not fit her needs.

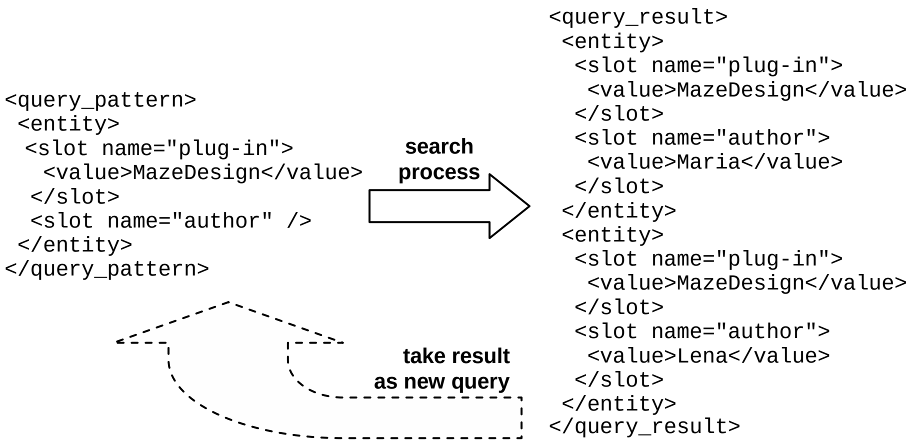

Fig. 2. XML representations of retrieval queries and answers

4 Outlook

To consider the balance of automation and human control, the metadata mechanism will be further enriched. The possibility to change the automatically generated metadata suggested by the system is one of the factors for this balance, another is the adjustment of the automation grade. User profiling will be a means to adjust these parameters.

References

1. Suthers, D., Connelly, J., Lesgold, A., Paolucci, M., Toth, E., Toth, J., & Weiner, A. (2001). Representational and Advisory Guidance for Students Learning Scientific Inquiry. In D. Forbus & P. Feltovich, P. (eds.): Smart Machines in Education, pp. 7-35. AAAI Press, Menlo Park
2. Pinkwart, N. (2003). A Plug-In Architecture for Graph Based Collaborative Modeling Systems. In U. Hoppe, F. Verdejo & J. Kay (eds.): Shaping the Future of Learning through Intelligent Technologies. Proceedings of the 11th Conference on Artificial Intelligence in Education, pp. 535-536. Amsterdam, IOS Press.
3. Pinkwart, N., Hoppe, H.U., Bollen, L. & Fuhlrott, M. (2002). Group-Oriented Modelling Tools with Heterogeneous Semantics. In Cerri et al (eds). Proceedings of the International Conference on Intelligent Tutoring Systems. pp. 21-30. Springer.
4. Dublin Core Metadata Initiative [Homepage] 2003. http://www.dublincore.org
5. Hoppe, H., Ammersbach, K., Lutes-Schaab, B., Zinßmeister, G. (1990). EXPRESS: An Experimental Interface for Factual Information Retrieval. In J.-L. Vidick (ed.), Proceedings of the 13th annual international ACM SIGIR conference on Research and development in information retrieval (ACM SIGIR '91). pp 63 - 81. New York, USA, ACM Press.
6. Munro, A. J., Höök, K. & Benyon, D. (1999) (eds). Social Navigation of Information Space. London et al., Springer.