Hendrix, M., and Cristea, A.I., UM'2007, A3H International Workshop
http://www.iit.demokritos.gr/um2007/UM2007_WS1_A3H.pdf

Evaluating Adaptive authoring of Adaptive Hypermedia

Maurice Hendrix and Alexandra Cristea

Department of Computer Science, The University of Warwick, Coventry CV4 7AL, United Kingdom
{maurice, acristea}@dcs.warwick.ac.uk

Abstract. Now-a-days a large amount of research exists into the design and implementation of adaptive systems, but still, a lot less targets the complex task of authoring of such systems, or their evaluation. In order to tackle these problems, we have looked into the causes of the complexity. Manual annotation has proven to be a bottleneck of authoring of adaptive hypermedia. Therefore, all means for supporting this authoring process by reusing automatically generated metadata should be sought. Previously, we have proposed the integration of the generic Adaptive Hypermedia authoring environment, MOT, into a semantic desktop environment, indexed by Beagle++. A prototype was built based upon this integration approach. This paper describes an evaluation of both the approach in general, and the prototype in particular.

Keywords: Authoring; Adaptive Educational Hypermedia; CAF; Evaluation; Metadata; RDF; Semantic Desktop; Semi-automatic adding

1 Introduction

Authoring of adaptive hypermedia is a notoriously difficult endeavour [4] although its results can be extremely valuable, generating, e.g., in the educational context, personalized (learning) experiences [3]. A solution to this problem is using as much as possible automatically generated authoring instead of authoring by hand. There is some research into how to automatize authoring in different ways [9]. A good basis of this is to use already annotated resources, which can be automatically retrieved when necessary, as dictated by the authoring process. A rich source of information that can be exploited in this sense is the semantic desktop [6,13]. In the semantic desktop, resources can be categorized by rich ontologies, and semantic links express various kinds of semantic relationships between these resources. For a document, for example, the semantic desktop stores not only a filename, but also information about where this paper was published, when and by whom, which of the colleagues sent it, and how often and in what context it was accessed. All these metadata are generated automatically, by the appropriate applications, and stored in an application independent way as RDF metadata, in the user's personal data store. This rich set of metadata clearly makes it easier for the user or applications to semi-automatically retrieve appropriate material for different contexts: for example, when an author

wants to select fitting materials for a lecture. Of course, in this context, the author still has to create some basic lesson material, serving as a retrieval framework.
Previously, in [10], we described the interaction and exchange of data between the Beagle++ environment [1,5], which is an advanced search and indexing engine for the semantic desktop, generating and utilizing metadata information, and the adaptive hypermedia authoring environment MOT [7,11], a sophisticated system for authoring personalized e-courses. In this paper we evaluate our approach in general, and the prototype in particular. This paper focuses only on the evaluation of the sesame2MOT Enricher (converter) application that guides the interaction process between the two systems above and performs the automatic authoring and population of the authoring environment with new resources (here, papers) from the semantic desktop.
The remainder of this paper is organized as follows. Section 2 shortly introduces our approach to Adaptive Authoring of Adaptive Hypermedia, as well as the system setup, which implements a prototype of this approach. In Section 3 we introduce the evaluation setup. In section 4 the results of the evaluation are discussed. Finally, in section 5, we discuss what these results mean for our approach in general, and prototype in particular.

2 The approach & system setup

In this section we introduce our method and system setup. As can be seen in Figure 1, Beagle++, the Semantic Desktop Environment used in our prototype, stores all metadata in the Sesame RDF database [12]. All Beagle++ components that *generate* metadata (for example, the email, publication, web cache and file metadata generators) add the metadata to this database. All Beagle++ components which *use* metadata (for example, the search and indexing module, the ranking module or the browsing modules) retrieve their data from this repository, and, in some cases, write back new data (such as the PageRank value for documents or other resources).

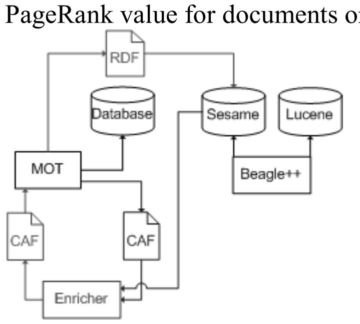

Figure 1. System overview

It is easy to accommodate additional modules in this environment by writing appropriate interface components, which read and write from this repository. This is what we have done for MOT [7, 11]. We have focused on the semi-automatic addition of articles stored on the user's desktop to a MOT lesson [7]). This represents an instantiation of the concept of adaptive authoring: authoring that adapts to the author's needs. In MOT, the addition is done to an *existing* lesson. Based on pedagogic goals, the author can then process the data, by adding more information on the article after the conversion. These additions can then be fed back into the RDF store, if necessary. We use CAF [8], a system-independent XML exchange format, to simplify the transformation process from RDF to the MOT MySQL storage format.

2.1 Enrichment of the lesson and domain model

As MOT is mainly a tool for authoring educational (adaptive) material, the internal information structures are based on strict hierarchies. When enriching the domain-models and lessons, we therefore aim at getting the right information in the right place in this hierarchy. To achieve this, the program first queries the Sesame database, using as search terms *title* and *keywords* of each domain concept found in the current existing lesson. The RDF query in the SeRQL [2] language looks as follows:

```
select x from x {p} y where y like "*keyword" ignore case
```

Some alternative retrieval methods have been studied and implemented, as follows.

Concept-oriented versus Article-oriented

For computing the mutual relevance between a given article and a given concept, in order to decide the appropriate place of articles in the concept hierarchy, we have developed two slightly different theoretical alternatives, as follows.

Concept-oriented relevance ranking method

The first relevance ranking method is computing relevance according to Equation 1:

Equation 1. Concept oriented ranking

$$rank(a,c) = \frac{|k(c) \cap k(a)|}{|k(a)|}$$, where:

rank(a,c) is the rank of article a with respect to the current domain concept c;
k(c) is the set of keywords belonging to the current domain concept c;
k(a) is the set of keywords belonging to the current article a;
|S| = the cardinality of the set S, for a given set S.

This formula is *concept-oriented*, in the sense that articles 'battle' for the same concept: a given article is placed in the appropriate place in the hierarchy by it.
The actual implementation of the concept-oriented formula is not trivial for the RDF database that we have, as there is no keyword property available for retrieved objects (articles). Therefore, these keywords had to be extracted from the title property of the object. This means that the computation of the cardinality of the keywords, |k(a)|, can be interpreted in many ways. The easiest way is to set it to the number of words in the title property. Another option is to compute the number of relevant keywords (e.g., eliminating words such as 'the', 'a', etc.) by connecting them to a given ontology or dictionary, such as WordNet. Both options were implemented; however, the current WordNet connection is very slow and thus inefficient.

Article-oriented relevance ranking method

A second implementation of the rank is given by Equation 2.

Equation 2. Article oriented ranking

$$rank(a,c) = \frac{|k(c) \cap k(a)|}{|k(c)|}$$

The equation shows how many of the keywords (shared by the article and the concept) are present in the concept. E.g., if a concept has less keywords than another one, but the keywords shared between article and concept are the same, the former concept will have a higher rank and 'win' the article. This formula shows thus to which of the concepts the article should be attributed. Therefore, this formula is article-oriented, in the sense that concepts 'battle' for the same article.

Sets versus multisets

Next, once the formula is chosen, there is another possible distinction to be made: we have chosen to implement two different versions for computing the cardinality of the intersection, one *set*-based (with intersection operation on sets, as defined above), and one with *multisets* or bags (and the respective intersection operation on bags).

In set theory, a *bag* can be formally defined as a pair *(A, m)* where A is some set and $m : A \rightarrow N$ is a function from A to the set N = {1,2,3,...} of (positive) natural numbers. The set A is called the underlying set of elements. For each $a \in A$ the multiplicity (i.e., number of occurrences) of a is the number *m(a)*.

The intersection can then be defined as: $(A \cap B, f(A \cap B))$, where:

$f(x) = \min\{m(x \in A), m(x \in B)\}$. For example: $\{(a,2),(b,1)\} \cap \{(a,3),(x,1)\} = \{(a,2)\}$

The reason to use bags instead of sets is that the number of times keywords appear in certain texts can be relevant in itself (not just *which* keywords). A text containing a greater number of occurrences of a specific keyword could be a better match for that keyword than a text with only one occurrence of the respective keyword.

Allowing duplicates or not

The same resource may be relevant in more then one place within the hierarchy. In that case, the resource will be added to the place where it has the highest relevance, by default. If there exist more places in the hierarchy with a value equal to the highest relevance, currently, the one with the higher position in the tree wins. If, however, there are siblings with the same position in the tree, and with the (same) highest relevance, a decision has to be made: either to allow duplicates, or to select randomly one of the candidate sibling concepts and allocate the resource to it. The '*allow duplicates*' option in the Enricher program controls this particular option.

Adding meta-data as separate concepts or as attributes

The retrieved metadata also has a structure. For example, a retrieved paper might have a location it was presented at and a year it was presented in. This metadata can be added either as attributes of the new *article* concept in MOT, or as a set of new sub-concepts, with their own attributes. With the '*Add meta-data as separate concepts*' option, the author can select between these two possibilities in the Enricher program.

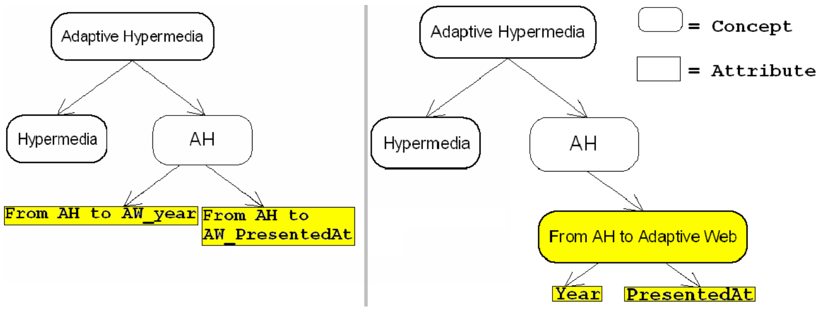

Figure 2. left: result *without*; right, result *with* the '*add medatada as separate concepts*' option.

3 Evaluation

The evaluation of the conversion process and the Enricher application program performing it has taken place in two steps so far: the first step was a small-scale qualitative experiment with about 4 PhD students of the IMPDET course organized by the Joensuu university, Finland, based on the *think-aloud* method. The second step is the evaluation described in the current paper, which was of a larger scale, and contains therefore a larger amount of quantitative evaluation results, although qualitative information was also sought.

3.1 General evaluation setup

This second evaluation was conducted at the Politehnica University of Bucharest in January of 2007. It took place within an intensive two-week course on Adaptive Hypermedia and The Semantic Web, which was delivered as an alternative track to the regular Intelligent Systems course. The students were 4th year undergraduates in Engineering studies and 2nd year Master students in Computer Science, from the English-language stream. Firstly, basic knowledge on Adaptive Hypermedia and Semantic Web was addressed. The first course week was dedicated to theory, and finished with a theoretical exam. Out of the initial 61 students, only the students with a satisfying knowledge of the theory where selected to continue with the practical part of the alternative track. The rest returned to the main stream course. The 33 students that passed the theory worked with the actual systems described in section 2. After these experiments, they where requested to submit a number of questionnaires, to answer both generic and specific issues regarding the automatic generation of adaptivity and personalization via the methods described in section 2.

3.2 Hypotheses

We based our evaluation firstly on a number of generic, high level hypotheses, as follows:

1. The respondents enjoyed working as authors in the system.
2. The respondents understood the system.
3. The respondents considered that theory and practice match.
4. The respondents considered the general idea of Adaptive Authoring useful

We refined these into more specific, lower granularity hypotheses (see Table 1), which ultimately generated our questions for the questionnaires. To explain the construction of the sub-hypotheses, let's take, for instance, hypothesis 3. There, we check the matching between theory and practice, i.e., between theory and the implementation. For the Enricher application, from a theoretical point of view, we have defined different ranking methods and other options, such as allowing duplicates or not between the imported articles, etc. These have been implemented as options for the user to select, and therefore, in this particular case, matching theory and practice means that these methods render different results, firstly, and secondly, that these different results should look like the theory has predicted1. Therefore, sub-hypothesis 3.4, and its own sub-hypotheses, 3.4. 1, 3.4.2 and 3.4.3 emerged. As said, the hypotheses and sub-hypotheses feed into and determine the questions.

4 Evaluation Results and Discussion

Due to lack of space, here we present and discuss only the numerical results from one of our questionnaires that focussed on the Enricher application (sesame2MOT). For testing our hypotheses we have used, where possible, numerical averages, and tested their significances with the help of Student's T-test. We assumed a confidence of 95%. Thus, a hypothesis can be confirmed, if the mean is M>0 and the probability is P <0.05. In order to obtain numerical averages, we mapped the multiple-choice answers follows: '*Yes*' was mapped to 1, '*no*' to -1 and '*mostly*' to 0. Hence the average was always 0 and the T-test was applied by comparing against the neutral result of 0. Below we present a table with each hypothesis, T-test results (T value, degrees of freedom Df, Mean M, probability P) and whether the results show that it was confirmed or not. The main hypotheses are shown in bold. Their result is obtained by combining the results of the sub-hypotheses.

Table 1 Sesame2MOT Conversion hypotheses results

Nr.	Hypotheses	T	Df	M	P	Confirmed (M>0; P<0.05)
1	**The respondents enjoyed working as authors in the system**	2.709	31	0.438	0.011	Confirmed

[1] This is the reason why it was important to have students understand the theory first; otherwise such questions would not be easy to answer by the layman.

2	**The respondents understood the system.**					**Confirmed**
2	Respondents understood how it works					Confirmed
2.1	Concept oriented	4.458	31	0.625	0.000	Confirmed
2.2	Article oriented	3.788	31	0.563	0.001	Confirmed
2.3	Allow duplicates	10.063	31	0.875	0.000	Confirmed
2.4	Compute resources as set	5.271	31	0.688	0.000	Confirmed
2.5	Add meta-data as separate concepts	6.313	31	0.750	0.000	Confirmed
3	**The respondents considered that theory and practice match.**					**Confirmed**
3.1	The two ranking methods (concept-, article-oriented) do deliver different results.					Confirmed
3.1.1	Concept Oriented	6.313	31	0.750	0.508	Confirmed
3.1.2	Article Oriented	6.313	31	0.750	0.508	Confirmed
3.2	Ranking methods (concept-, article-oriented) in line with the theory.		31			Confirmed
3.2.1	Concept Oriented	2.252	31	0.375	0.032	Confirmed
3.2.2	Article Oriented	2.709	31	0.438	0.011	Confirmed
3.3	The different options influence the result		31			Confirmed
3.3.1	Allow duplicates	7.760	31	0.813	0.000	Confirmed
3.3.2	Compute resources as set	3.215	31	0.500	0.032	Confirmed
3.3.3	Add meta-data as separate concepts	6.313	31	0.750	0.000	Confirmed
3.4	The results of the conversion are in line with the theory					Confirmed
3.4.1	The two ranking methods	see 3.3.1.2	see 3.3.1.2	See 3.3.1.2	see 3.3.1.2	Confirmed (see 5.3.1.2)
3.4.2	Allow duplicates	7.760	31	0.813	0.000	Confirmed
3.4.3	Compute resources as set	2.252	31	0.375	0.032	Confirmed
3.4.4	Add meta-data as separate concepts	4.458	31	0.625	0.000	Confirmed
4	**General idea useful**	**15.000**	**31**	**0.938**	**0.000**	**Confirmed**

As we have seen, most hypotheses have been confirmed based on the current data. The Sesame2MOT conversion is indeed considered useful and in line with the theory. Its options are understood. Respondents agreed strongly with most of our hypotheses, with all means above zero. Looking at the ones with lower scores, such as concept-oriented and article-oriented method, as well as computation of resource as set, they were less sure in their statements. This is probably due to the fact that they did not work with these options enough. This shows that more targeted evaluations may be necessary to establish without a doubt the acceptance rate of these features.

5 Conclusions

In this paper we have briefly reviewed an authoring environment for personalized courses, as well as an Enricher mechanism and prototype based on Semantic Desktop technology. The paper describes the theoretical considerations for the implementation of the Enricher, and then, in parallel, the evaluation of these considerations as well as

of the prototype. From the two evaluation steps performed so far, the general result is that, to the extent it was understood, the theoretical concept of Adaptive Authoring of Adaptive Hypermedia was perceived as useful. We have also gained some important feedback into possible improvements to the Enricher application itself. Respondents in our experiments pointed out that the integration is currently not optimal and the user interface has to be improved. We plan to integrate the Enricher further into MOT by making a web based version and enhance the usability of the selection options.
Beside the hypotheses analysed here, we are also looking into the correlation between the students' responses and their comprehension of the theory on adaptive hypermedia and authoring thereof. We are still in the process of analyzing SUS questionnaires on MOT and the Enricher, and will perform comparisons of the students' preferences.

Acknowledgments. We would like to thank the Socrates mobilities exchange program at The University of Warwick, Politehnica University of Bucharest and Prof. Dr. Paul Cristea for giving us the opportunity to conduct our user study.

References

1. Beagle++ website, viewed 31 January 2007, http://beagle.kbs.uni-hannover.de/
2. Broekstra, J., Kampman, A., 2004 SeRQL: An RDF Query and Transformation, viewed 15 December 2006, http://wwwis.win.tue.nl/~jbroekst/papers/SeRQL.pdf
3. Brusilovsky, P., Adaptive hypermedia, User Modelling and User Adapted Interaction, Ten Year Anniversary Issue, 2001 (Alfred Kobsa, ed.) 11 (1/2), 87-110.
4. Celik, I., Stewart, C., Ashman, H. Interoperability as an Aid to Authoring: Accessing User Models in Multiple AEH Systems, Proceedings of A3H: 1st International Workshop on Authoring of Adaptive and Adaptable Hypermedia, 2006
5. Chirita, P.-A., Costache, S., Nejdl, W., and Paiu, 2006, R. Beagle++: Semantically Enhanced Searching and Ranking on the Desktop. Proceedings of the 3rd European Semantic Web Conference, Budva, Montenegro.
6. Chirita, P.-A., Gavriloaie, R., Ghita, S., Nejdl, W., and Paiu, R., 2005, Activity-Based Metadata for Semantic Desktop Search. Proceedings of the 2nd European Semantic Web Conference, Heraklion, Crete.
7. Cristea, A. I., De Mooij, A., 2003, Adaptive Course Authoring: My Online Teacher. Proceedings of ICT'03, Papeete, French Polynesia.
8. Cristea, A.I., Smits, D., and De Bra, P., 2005, Writing MOT, Reading AHA! - converting between an authoring and a delivery system for adaptive educational hypermedia -, A3EH Workshop, AIED'05, Amsterdam, The Netherlands
9. Cristea, A., Stewart, C., Automatic Authoring of Adaptive Educational Hypermedia, book chapter in "Web-Based Intelligent e-Learning Systems: Technologies and Applications", Zongmin Ma. IDEA Publishing group 2005
10. Hendrix, M., Cristea, A.I., Nejdl, W., Authoring Adaptive Learning Material on the Semantic Desktop, 1st International Workshop on Authoring of Adaptive and Adaptable Hypermedia, Adaptive Hypermedia Dublin Ireland, 2006
11. MOT homepage, viewed 31 January 2007, http://prolearn.dcs.warwick.ac.uk/mot.html
12. Schlieder, T., Naumann, F., Approximate tree embedding for querying cml data, ACM SIGIR Workshop on CML and Information Retrieval.
13. Semantic Desktop, viewed 31 January 2007, http://www.semanticdesktop.org/