[LRSC8-2] L..sse L..hmann, Christoph Rensing, Ralf Steinmetz; **Capture of Lifecycle Information to Support Personal Information Management**; In: Pierre Dillenbourg, Marcus Specht: Times of Convergence: Technologies Across Learning Contexts, Third European Conference on Technology Enhanced Learning, EC-TEL 2008,Springer,September 2008. Seite 216 - 221

Capture of Lifecycle Information to Support Personal Information Management

Lasse Lehmann, Christoph Rensing, Ralf Steinmetz

KOM - Multimedia Communications Lab
Technische Universität Darmstadt
Merckstrasse 25
64283 Darmstadt

{Lasse Lehmann, Christoph Rensing, Ralf Steinmetz}@kom.tu-darmstadt.de

Abstract. Re-use is a key aspect of today's Learning Resource creation. Authors often re-use objects, which they had originally created. Additionally organization of documents has become a complex task and users tend to have more and more problems to manage documents stored on their local computers. With our approach we combine these two aspects by supporting users in their Personal Information Management with information, emerging from re-use processes. We propose a framework capable of capture, management and utilization of this so called lifecycle information and present our implementation for PowerPoint presentations. A first evaluation shows promising results and demonstrates the feasibility and validity of our approach.

Keywords: Re-Use, Metadata Generation, Lifecycle Information, PIM

1 Introduction and Motivation

Nowadays, re-use is a key aspect of the creation of Learning Resources as well as knowledge documents. Processes like authoring, re-authoring and re-use of images, texts, slides or other parts of these resources provide for the emergence of multiple types of information. This information – if captured and processed – can help to support retrieval, authoring or management of the documents involved.

It is a known fact that users tend to have problems organizing documents stored on their local computers [12]. Modern PIM (Personal Information Management) tools try to support users here. With the above mentioned information, PIM applications like desktop search tools, semantic desktops or dedicated information management tools can be enhanced.

Authors of Learning Resources and knowledge documents usually do not want to create metadata or additional information for their documents. Therefore we propose to acquire this so called *lifecycle information* without explicit user interaction but by monitoring actions users take anyway when working on their resources. In [7] we have presented a framework for capture, utilization and management of lifecycle information (LIS.KOM). In this paper we present the application of our approach for the capture of lifecycle information in PowerPoint. Among others the captured information can be used to support management of the documents involved. To increase the readability we refer to Learning Resources, when meaning both Learning Resources and knowledge documents. In the following we present our notion of lifecycle information and discuss how lifecycle information can be used to support PIM (Section 2). After a discussion of related work (Section 3) we present the overall architecture of the LIS.KOM framework (Section 4). Finally we discuss first evaluation results (Section 5), summarize and give an outlook on future work (Section 6).

2 Lifecycle Information for PIM

Lifecycle information is a special kind of metadata. However, in contrast to the common notion of metadata it is not related to a specific object but emerges from a certain process. Therefore it is only available while the corresponding processes take place. That means that lifecycle information has to be captured during these processes - or else is lost. In order to identify information emerging from processes, the processes themselves must be identified. A detailed description of the analysis of a Learning Resource's or knowledge document's lifecycle and a thorough view on lifecycle information and the corresponding processes can be found in [7] and [8]. In the following we describe how lifecycle information can be used to support PIM.

Boardman [1] defines PIM as "Management of personal information", where "personal information" means information a user possesses and not information *about* a user. PIM is not only related to e-mail and bookmarks, though these concepts are strongly associated with it, but to all kinds of information (or codified knowledge) in a user's possession. That includes various information and document types, like pictures, videos, audio files, text files or presentations. PIM tools are often designed to support one specific type of information only (like management tools for images or audio files). Nevertheless there are applications like Semantic Desktops, which aim to cover several information types.

Lifecycle information can constitute valuable input for those applications. When content is re-used, the source document and the document the content has been re-used in (target document) are often very closely related. In most cases a relation that connects both documents on a semantic level can be assumed. Thus the possibility is high that the source document might again be interesting if the target document is re-opened, e.g. for a revision. These relations between documents can be captured as lifecycle information.

There are two main possibilities to use lifecycle information to support PIM presented in the following. The first possibility is to provide information about source and target documents for his currently opened document to the user. The related documents can then be made accessible directly from the working context of the user.

The second possibility to utilize lifecycle information for PIM is an external application for the retrieval, browsing or search of documents. To achieve this, document management systems or desktop search engines like Beagle++ [4], could be extended. Lifecycle information like relations between documents would add nicely to the already featured full-text search, enrichment with contextual information and social recommendations Beagle++ provides. Search results could be extended with links to re-used or otherwise related resources or a visualization of document relations could be rendered. Of course it is also possible to utilize the lifecycle information in an independent application, for example a relation browser or file system explorer.

3 Related Work

There are quite some interesting approaches in current research that are related to our work. With the *Ecological Approach* McCalla (et al.) laid the foundation for our approach [9], [2]. *Contextualised Attention Metadata* [3] is used to store the attention a user pays to different Learning Resources in different applications. It is e.g. utilized for ranking and recommendations of Learning Resources [13], learner modelling [11] or even for knowledge management [15]. The main difference to our approach is that the information is collected user-centric instead of document-centric. *Semantic Desktops* are tools which, among other things, aim to improve the PIM on a greater scale. Tools like Gnowsis [14] or Haystack [6] try to provide a holistic solution for Personal Information Management, often mixed with social aspects. *Desktop search engines* like Google Desktop or Beagle++ [4] try to make information stored on a local computer search- and retrievable. In *TeNDaX* [5], a system for the collaborative creation and editing of documents, user actions are stored as transactions in a database. Thus it is possible to track copy and paste relations between documents. However, other kinds of lifecycle information are not considered. Mueller proposes in his approach a system for "*consistent management of change*", i.e. for improved versioning of documents [10]. The approach takes relations both, within and between documents into account and tries to provide versioning functions on a semantic level. However, re-use or lifecycle information is only marginally considered.

4 LIS.KOM Framework

As stated in [7] a system is needed which allows the collected lifecycle information to cross system borders. That means that information gathered in one system has to be transported to a different system in order to be fully utilizable. The LIS.KOM framework (Figure 1) provides these features. The main component of the framework is the LIS.KOM Server. Here, the captured lifecycle information is stored, processed and provided for utilization. Local components can connect to the central server via a web service API to either send the lifecycle information captured or obtain processed information. The LIS.KOM Client, located on a user's computer, is responsible for the handling of the connection to the server and the synchronization of the locally cached lifecycle information. Due to the local storage the LIS.KOM Client works even in an offline case. The Client provides an API for add-ins and applications that utilize the lifecycle information as well as for those that capture information. Capture is done by the ReCap.KOM add-ins. They are plugged into the different applications where information should be captured, e.g. authoring and office applications, repositories or

Learning Management Systems. The utilization can be either done with a standalone application (LIS.KOM Utilization Tool) or in turn by means of add-ins (ProCap.KOM) for existing applications like office tools, repositories, desktop search tools or Semantic Desktops.

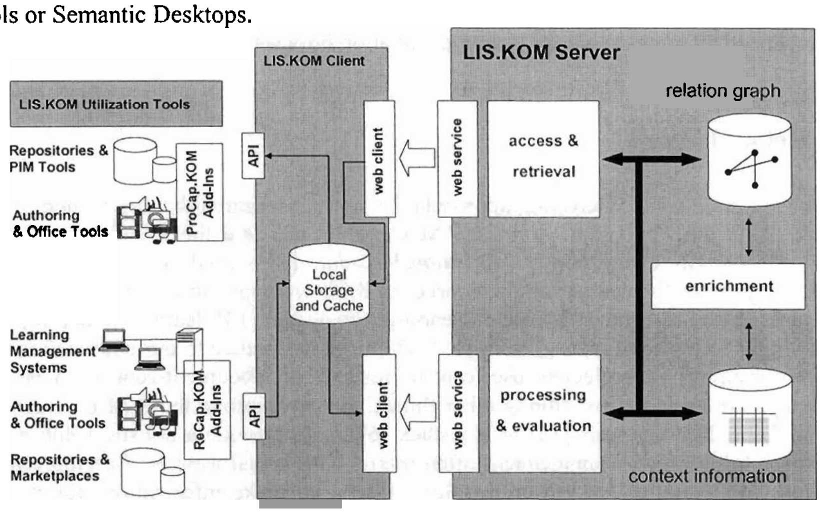

Figure 1: LIS.KOM framework

We have implemented the LIS.KOM client as well as a ReCap.KOM add-in for PowerPoint and are currently working on the implementation of the LIS.KOM server.

5 Evaluation

The goal of this evaluation was to prove the validity of lifecycle information captured with our framework. We deployed a ReCap.KOM capture module for PowerPoint on the computers of 4 test persons. We focused on the capture of relations emerging during the creation, re-use and editing of PPT presentations. We captured *provision relations* when elements were re-used within PowerPoint, *asset relations* when an external asset, e.g. an image, was (re-)used in a PowerPoint presentation and *variant relations* each time a presentation was saved under a different name. The main purpose of this evaluation was to test if the tool works with respect to the *validity* of relations captured. To achieve this, the source and target document were examined by an expert. A relation was found valid if it was traceable by the expert. It was not the goal of this evaluation to judge the *significance* or *importance* of relations captured.
The evaluation was done for 6 weeks in a realistic usage scenario, i.e. the test persons used PowerPoint as they would have without being test persons. Since the capture of information happened completely in the background it can be assumed that the test persons have not been influenced by it in any way. Because of the naturally different amounts of working time, working styles and re-use behaviour the amount and types of relations captured were quite different. Table 1 shows the results of the evaluation with the types of relations captured, their respective quantity, the distribution among

the test persons and their validity. It is remarkable that one test person opened and created significantly more presentations than the others. Altogether there were 58 provision (i.e. re-use) relations for 29 different documents collected. This shows that there is actually a high amount of reuse happening when PowerPoint presentations are created.
Relations were captured on slide level. About 75% of the provision relations were valid, 19% invalid and 6% inconclusive. Relations were marked as inconclusive when the validity of a relation could not be determined. This was the case if a target slide (a slide a relation pointed to) did not exist anymore, due to the fact that the test persons were not forced to keep all versions and revisions of their presentations.

Table 1: Number and Distribution of Captured Relations and their Validity

Relation Type	Total	Distribution (Testperson 1-4)	Valid	Invalid	Inconcl.
Provision (total)	58	3 | 5 | 0 | 50	43	11	4
Prov. - Slides	38	3 | 0 | 0 | 35	31	7	0
Prov. - Text	12	0 | 5 | 0 | 7	9	3	0
Prov. - Shapes	8	0 | 0 | 0 | 8	7	1	0
Variant	16	2 | 4 | 1 | 9	16	0	0
Asset	23	18 | 0 | 5 | 0	23	0	0

We identified three reasons for invalidity of relations:

1. The evaluation scenario was not closed. I.e. there was the possibility that other users without ReCap.KOM adds-in changed the presentations leading to invalid relations. This problem does not occur in a closed evaluation scenario where all users have the mandatory add-ins installed.
2. Some of the invalid relations were caused by a minor event handling issue that we have solved meanwhile. We estimate that about 25% of the invalid relations were invalid due to this error.
3. Lastly, invalid relations were captured when slides or shapes were re-used for structural or formal reasons only. This constitutes the biggest challenge. To solve this we need to analyse the content of related slides to judge if the relation is valid. A similar problem occurs if a slide is re-used and then the contents of the slide are deleted successively. Here, a measurement to judge when a relation is not valid anymore is needed.

Asset and variant relations were captured with a reliability of 100%. The overall validity of captured relations was around 85%.

6 Conclusion and Future Work

In this paper we have shown that it is feasible to capture lifecycle information for Learning Resources which are created with standard office applications. We have proposed and developed different possibilities to utilize this information. The evaluation has shown that there is a significant amount of re-use when presentations are created and that the resulting relations can be captured with high reliability. The next step, beside the improvements of the capture in PowerPoint, will be the connection of the LIS.KOM Client to the LIS.KOM Server. Thus we can change from a personal to a community environment, where lifecycle information und thus their value can be shared with other users. More evaluations will be conducted to determine the significance of relations and to test our approach in a community scenario. Due to the modular nature of the LIS.KOM framework it is easily possible to implement further add-ins for other document types, like e.g. MS Word.

References

1. Boardman, R. (2004), 'Improving Tool Support for Personal Information Management', PhD thesis, Imperial College London
2. Brooks, C. & McCalla, G. (2006), Towards Flexible Learning Object Metadata, in 'In Proceedings of Int. J. Cont. Engineering Education and Lifelong Learning, Vol16, Nos 1/2'.
3. CAMs (2007), Conceptual Base Scheme, Online (http://ariadne.cs.kuleuven.ac.be/empirical/attentionCAM%20schema_Document_v1.5.pdf).
4. Chirita, P.; Gavriloaie, R.; Ghita, S.; Nejdl, W. & Paiu, R. (2005), Activity Based Metadata for Semantic Desktop Search, in 'Proceedings of the 2nd European Semantic Web Conference'.
5. Hodel, T.; Hacmac, R. & Dittrich, K.R. (2005), Using Text Editing Creation Time Meta Data for Document Management, in 'Proceedings of Conference on Advanced Information Systems Engineering'.
6. Karger, D.; Bakshi, K.; Huynh, D.; Quan, D. & Sinha, V. (2005), Haystack: A General Purpose Information Management Tool for End Users of Semistructured Data, in 'Proceedings of CIDR'.
7. Lehmann, L.; Hildebrandt, T.; Rensing, C. & Steinmetz, R. (2007), Capturing, Management and Utilization of Lifecycle Information for Learning Resources, in 'Proceedings of EC-TEL 2007 - Second European Conference on Technology Enhanced Learning'
8. Lehmann, L.; Hildebrandt, T.; Rensing, C. & Steinmetz, R. (2007), Utilizing Lifecycle Information for Knowledge Document Management and Retrieval, in 'In Proceedings of the 7th International Conference on Knowledge Management'.
9. McCalla, G. (2004), 'The Ecological Approach to the Design of E-Learning Environments: Purpose-based Capture and Use of Information About Learners', Journal of Interactive Media in Education 7.
10. Mueller, N. (2006), An Ontology-Driven Management of Change, in 'In Wissens- und Erfahrungsmanagement, LWA (Lernen, Wissensentdeckung, Adaptivitaet) conference proceedings'.
11. Najjar, J.; Wolpers, M. & Duval, E. (July 2006), Towards Effective Usage-Based Learning Applications: Track and Learn from Users Experience(s), in 'Proceedings of the IEEE ICALT 2006'.
12. Nejdl, W.; Paiu, R. (2005), 'I know I stored it somewhere - Contextual Information and Ranking on our Desktop'.
13. Ochoa, X. & Duval, E. (2006), Use of Contextualized Attention Metadata for Ranking and Recommending Learning Objects, in 'Proceedings of the CAMA 2006'.
14. Sauermann, L. (2005), The Gnowsis Semantic Desktop for Information Integration, in 'Proceedings of the IOA Workshop of the WM2005 Conference'.
15. Wolpers, M.; Martin, G.; Najjar, J. & Duval, E. (2006), Attention Metadata in Knowledge and Learning Management, in 'Proceedings of I-Know 2006'

Application for an ETA for Australia

Your payment has been authorised and your application for an ETA has been approved

Your ETA has been recorded on our database. It will be available for checking by your airline at check-in and for Australian immigration authorities on your arrival in Australia. You do not need a visa label in your passport.

Your ETA has been recorded with the following data and conditions:

Expiry Date	21 OCT 2009
Period of Stay	3 Months
Number of Entries	Multiple

- The Expiry Date of the ETA is the latest date on which you may enter Australia with this ETA.
- The Period of Stay is the longest time you may spend in Australia on any one visit.
- The Number of Entries is the number of times you may enter Australia with this ETA.

Your credit card has been charged by IVisa Services for **AUS$20.00**

Please retain the following Reference Number for making enquiries through this site. If you do not supply the Reference Number when making an enquiry, we will not be able to tell you the status of your ETA. To protect your privacy, this unique reference number cannot be provided at any other time.

Reference Number: **8967669**

Print this screen

Would you like to submit another application? Yes No

This document is not your authority to enter Australia and cannot be used as such. It is not required for airport check-in nor on arrival in Australia.