Lifecycle Information Management for Learning Resources and Knowledge Documents

Lasse Lehmann
Christoph Rensing, Ralf Steinmetz (Supervisors)

KOM Multimedia Communications Lab
Technische Universität Darmstadt
Merckstr. 25, 64283 Darmstadt, Germany
{lasse.lehmann, christoph.rensing, ralf.steinmetz}@kom.tu-darmstadt.de

Abstract. Lifecycle information can be of great help in the retrieval, authoring and usage of both, Learning Resources and Knowledge Documents. In my PhD thesis I want to show how the different types of information can be captured, managed and utilized so that the persons involved in the lifecycle of these documents benefit from it in various ways. This article shows my overall strategy and my progresses in the different segments.

Keywords: Lifecycle Information, Document Management, Retrieval

1 Background and Motivation

Both, Learning Resources and Knowledge Documents, usually take part in various processes, where they are adapted, modularized, aggregated, re-purposed, generally re-used, updated or just plainly used. Commonly this is done by different persons. These processes generate a lot of information, which persons involved in the processes might take advantage of. If we consider a realistic scenario, where a lecturer re-uses some slides of several other professors, adapts them and - given he got the allowance to do so - puts them in his new presentation, and we assume that the slides he re-used were produced by the other professors in a similar way, we can conduct that all the presentations, the lecturer re-used slides from, are connected somehow with the presentations, the other professors got slides from. In such a scenario authors involved may want to know:

- Who is reusing my slides?
- Who has changed the slides in which way?
- In which context are the slides re-used?
- How often are the slides re-used?
- How popular are my slides?

On the other hand, users of the slides, or learners respectively, might want to know:

- Are there presentations related to the one I am currently viewing?
- Are there slides covering similar or related topics?

- Which slides are most popular?
- How can I retrieve the proper documents efficiently?

2 Identification of Research Questions

Unfortunately, with today's methods, we are not able to answer most of the aforementioned questions in a satisfying way. Even if every person, involved in the processes, would have used one and the same state of the art version control system - like CVS or SubVersion - we would not be able to reproduce all the information, generated during the processes described above. Therefore all the information needs to be captured at the moment it emerges, or otherwise is lost. This is one of the central issues of my PhD thesis. Among others, the following questions are of interest:

- What kinds of information emerge during the lifecycle of a document?
- Which information is important and can be helpful in later on processes on a document?
- How can the lifecycle of a document be modelled properly?
- How can the desired information be captured / collected?
- Which phases of a document's lifecycle provide which kinds of information?
- How can the information be stored and traversed beyond system borders?
- In which ways can the collected information be utilized?
- Is there a critical mass of information (relations) or documents for this approach to be effective

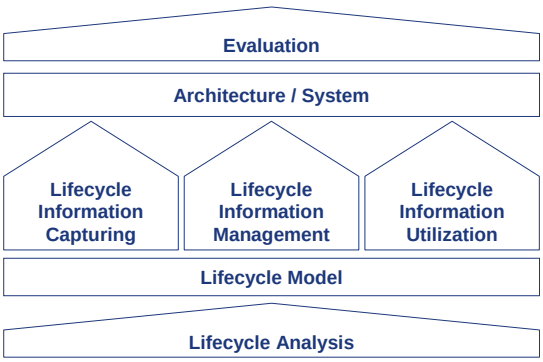

Figure 1: Concept of the Overall Approach

3 Overall Approach

To cope with the aforementioned challenges, we divided the overall problem into three segments: *Capturing*, *management* and *utilization*. Additionally, we start from a

rather specialized domain - the one of learning resources - and try to transfer the results to the more general domain of what we call knowledge documents. We chose an analytical approach that we will prove with mostly user driven evaluations. Therefore we start from analyzing the lifecycle and the information generated herein. The results from this analysis are brought into a lifecycle model. Then, based on that, we cover the three above mentioned segments and derive a conceptual architecture. After the implementation work is done, the approach will be evaluated. This whole process is depicted in Figure 1.

4 Results, Solution Approaches and Work in Progress

At first, the results and progress in the domain of Learning Resources are described before we depict our approaches in the field of knowledge documents.

4.1 Learning Resources

For the field of Learning Resources the analysis of the lifecycle, the systems taking part in it and the information generated herein has been done. We differentiate two types of information that have to be handled, processed, as well as captured differently: **Context** and **relation** information. We modelled the lifecycle of Learning Resources and developed a schema to store this information [7]. Our first proof of concept for the overall approach was conducted in our Authoring by Aggregation environment ResourceCenter [5]. In [7] we have shown that the collection, management and utilization of lifecycle information for Learning Resources is possible and can support the users of the ResourceCenter in various ways. In [6] a more general approach, which is not dependent on one specific system, is described. We implemented a generic framework that is easily extendable and provides the infrastructure for arbitrary applications. Capture Components do the **capturing** of lifecycle information within the applications in use, like repositories, market places or authoring, re-authoring or authoring by aggregation tools. The Lifecycle Information System (LIS) is the central instance that manages, processes and distributes the collected information - in short: it is responsible for the **management** of lifecycle information. At last, the Accessing Components allow for the **utilization** of the processed information within the systems in use. Nevertheless, the evaluation of the approach is still missing. The components are described in detail in [6].

4.2 Knowledge Documents

In our understanding the knowledge documents do not have a strict definition. It is rather the case that we discovered a group of documents with certain attributes and tried to find a name for it. Our definition of knowledge follows the definition of Probst et al. in [14]. Here, characters become data if syntax is added, data becomes information when context is added and information becomes knowledge when the linking of information is given. Documents are in that case digital text-based

documents, which can have various formats and contain different types of media like images, audio, animation etc. Starting there from, we define knowledge documents as documents with the following attributes:

- Documents containing (codified) knowledge
- Digital, text-based
- High potential and probability of re-use
- Often created in collaborative processes

Typical knowledge documents are slides, reports, papers or articles, in contrast to bills, receipts, invoices or telephone books, since these contain information at most. We found that in this domain lifecycle information can be of even more value than for Learning Resources, since existing Document Management Systems do not take into account the huge amounts of information generated during the documents' lifecycle.

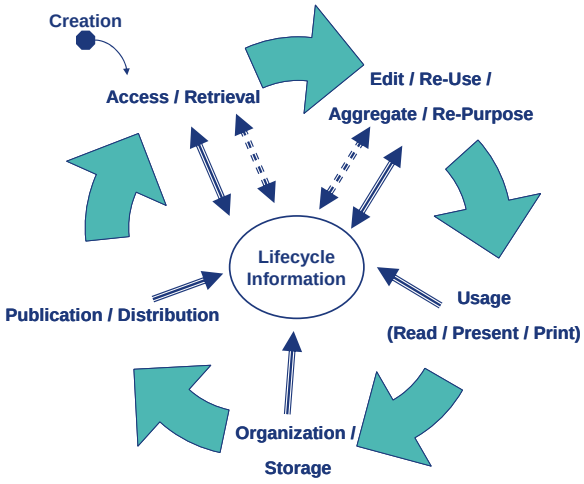

Figure 2: Schematic Knowledge Document Lifecycle

So far, we have developed a schematic model for their lifecycle and we are currently analyzing the different types of lifecycle information emerging in the different phases, in order to put them into a proper schema. We decided to stay with the division in context and relation information. Figure 2 shows the schematic lifecycle for knowledge documents, we propose. Here, the arrows show, where information emerges (arrows to the middle) and where information is utilized (arrows from the middle). The dotted arrows represent relation information while the other ones depict context information. In order to have this lifecycle properly analyzed, we have to generate views from different perspectives, including a system view, considering the applications involved, and an instance view, where the proceeding of one instance of a knowledge document through the lifecycle is shown. In addition, we have transferred the existing architecture concept to cope with the new situation. Here, we adapted the framework we designed for Learning Resources to the new usage context [8].

5 Related Work

Several approaches within this domain show, that the collection, management and utilization of lifecycle information are important issues in ongoing research. The ecological approach [2], [9] states that there should not only be a dedicated labelling phase for learning objects in the beginning of their lifecycle, but metadata should be created during the whole lifecycle of a learning object, which motivates the capturing of lifecycle information. On top of that, the capturing of usage information is proposed in order to improve user or learner modelling.

The research works around the Attention Metadata framework [12] is actually dealing with a special sort of contextual usage information to help in retrieving - e.g. ranking and recommending - learning objects [13], learner modelling [11] or in semantic desktop search [3]. However relation information seems to be not captured at all and there is no publication to be found where this kind of information is taken into account. Thus a lot of potential value is dismissed here.

Systems and approaches that take relations between documents into account are rather sparse. In [10] a concept for a management of change for collaborative authoring of structured and unstructured documents is proposed. This is somehow build upon relations but mostly relies on semantic, intra-document relations that are inevitable for a consistent management of change, but do not make use of typed relations that connect different instances of documents like we do.

In the HyLOS system [4] relations between instances of learning objects are used to provide additional links to learners in order to support constructivist learning. However, these relations are generated at most semi-automatically and are not caught while they occur. Nevertheless the HyLOS system uses algebraic rules to enrich existing relation information sets.

Additionally there are several frameworks, systems or approaches where this work might fit in nicely. Semantic Desktops [15] for instance, deal with the retrieval of documents in desktop environments; they already make some use of basic lifecycle information, like e-mail context or information gathered from the file system. Nevertheless, the consequent capturing and utilization of lifecycle information might make these tools even more powerful.

6 Conclusions and Future Work

The attention in the research community shows that the capturing, management and utilization of lifecycle information is an important and interesting research task. We have shown that in the domain of learning resources - under the given conditions - our approach is reasonable and applicable [6], [7]. The next steps will have to prove, if our approach can be generalized and transferred to the domain of knowledge documents, where it could serve an even bigger target group and likely have more impact. They include:

- Evaluation of the implementation in the ResourceCenter
- Refinement of the lifecycle model and the lifecycle information schemas for knowledge documents

- Refinement and implementation of the above concepts
- User driven evaluation of the developed system

References

1. Berndt, O.; Biffar, J. & Zoeller, B. (2005), DMS – vom elektronischen Archiv zum Enterprise-Content-Management aus VOI: Code of Practice - Dokumenten-Management, VOI, chapter 1.
2. Brooks, C. & McCalla, G. (2006), Towards Flexible Learning Object Metadata, in 'Proceedings of Int. J. Cont. Engineering Education and Lifelong Learning, Vol16, Nos 1/2'.
3. 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'.
4. Engelhardt, M.; Hildebrand, A.; Lange, D. & Schmidt, T.C. (2005),A Semantic Approach to Automated Content Augmentation for eLearning Objects, in Michael Auer & Ursula Auer, ed.,'Proceedings of the ICL 2005', Kassel University Press, Villach (Austria).
5. Hoermann, S.; Hildebrandt, T.; Rensing, C. & Steinmetz, R. (2005), ResourceCenter - A Digital Learning Object Repository with an Integrated Authoring Tool, in 'Proceedings of the Edmedia 2005'.
6. 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'.
7. Lehmann, L.; Rensing, C. & Steinmetz, R. (2007), Lifecycle Information Management and Utilization in an Authoring by Aggregation Environment, in 'Proceedings of the Edmedia 2007'.
8. Lehmann, L.; Hildebrandt, T.; Rensing, C. & Steinmetz, R. (2007), Utilizing Lifecycle Information for Knowledge Document Management and Retrieval, In proceedings of I-KNOW 2007 - 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. Najjar, J.; Wolpers, M. & Duval, E. (2006), Attention Metadata: Collection and Management, in 'Proceedings of the WWW2006 workshop on Logging Traces of Web Activity: The Mechanics of Data Collection, Edinburgh, Scotland'.
13. Ochoa, X. & Duval, E. (2006), Use of Contextualized Attention Metadata for Ranking and Recommending Learning Objects, *in* 'Proceedings of the CAMA 2006'.
14. Probst,G./Raub, S./Romhardt, K.: Wissen managen. Wie Unternehmen ihre wertvolle Ressource optimal nutzen. 4. überarb. Aufl. Wiesbaden 2003
15. Sauermann, L.; Bernardi, A. & Dengel, A. (2005), Overview and Outlook on the Semantic Desktop, *in* 'Proceedings of the 1st Workshop on The Semantic Desktop at the ISWC 2005 Conference'.