Attention Metadata in Knowledge and Learning Management

Martin Wolpers
(L3S Research Center, Germany
wolpers@l3s.de)

Gunnar Martin
(DFKI, Germany,
gunnar.martin@iwi.dfki.de)

Jehad Najjar
(K.U.Leuven, Belgium
jehad.najjar@cs.kuleuven.be)

Erik Duval
(K.U.Leuven, Belgium
erik.duval@cs.kuleuven.be)

Abstract: The paper outlines how attention metadata enables a tight integration between organisational knowledge stores and human resource management on the one hand and learning management system in corporate contexts on the other hand. The approach relies on an extension of AttentionXML, a metadata standard to capture the attention a user spends on digital content. We suggest relying on attention metadata to unobtrusively capture the attention of employees at the workplace (e.g. through workflow systems, knowledge management systems, human resource management systems, etc.) and use the captured information to enable a targeted steering of learning process of the employee in companies in accordance with company's and employee's aims and goals.

Keywords: Attention Metadata; AttentionXML, Knowledge Management, Learning Management, Learning Object Metadata, Organisational Intelligence, Organisational Learning, Workflow Management

1 Introduction

The importance of the companies´ knowledge, its development and usage for the realisation of market successes is discussed in (research-) theory and (industrial-) practice since almost 40 years. Here, one has to refer to the terms of Organisational Intelligence [Wilensky, 67] and Organisational Learning [Argyris, 78] as starting points for establishing knowledge as a manageable resource and crucial business factor for companies. Beside those – more or less – organisation theory-driven concepts a lot of different IT-solutions and -tools has been developed and implemented into the business IT-infrastructure landscape in order to ensure the ongoing identification, retrieval, creation and provision of relevant knowledge to the point of business process execution as well as its usages, storage and controlling [Romhardt, 98]. Currently, many different types of information, learning and knowledge supporting systems that support a broad spectrum of functionalities are existent. For example, information retrieval techniques, data-warehouse technology as well as document-, content- and learning management environments are used to support the core functions of an “intelligent organisation” [Pinchot, 93]. But, in reality there are no real interoperable linkages between learning and knowledge management systems and other relevant enterprise-wide information systems, that enable the dynamic accumulation of working contexts with relevant content and/or up-to-date information in an organizational-, individual- and application specific way to satisfy heterogeneous learning and knowledge needs.

As an example, KM-systems organize the location of explicit knowledge, provide retrieval functionalities and – in case of ontology-based organisational memories – pinpoint existing relations between the stored information and working contexts based on predefined patterns and stereotypes. These systems work quite well on capturing the knowledge which is available through the companies´ databases but they cannot provide any information concerning the single user, his role within the organizational structure, his involvement in several business processes and his (implicit) knowledge and skills. To identify the users´ working context, data from business process automation tools and workflow management systems (WFMS) can be analysed. Following a detailed description of a business process, including organisational units, employees, applications etc., WFMS electronically hand over objects (e.g. documents) to be processed from one work place to the next. Additionally, WFMS collect information on the processing status, execution times and the context of the single users involved. These data provide a good quantitative basis for the process performance measurement and the cost/time controlling, but WFMS does not contain any information about the usage of knowledge by the individual employee and his competencies. The learner-centered data – such as personal skills, competencies and acquired knowledge – is often imparted in learning history databases of learning management systems (LMS). LMS accumulate the data for the definition of prerequisites for course bookings or sequencing and navigation constrains within closed course structures. So far, LMS do not make this data available for other, related information systems.

Obviously, none of the existing information systems is neither able to derive precise specifications about the usage of knowledge in certain working conditions, its allocation or to show the correlation of knowledge to the individual user nor the definition of requirements for the development of learning and knowledge supporting actions. But this is needed to improve the employee’s and the company’s knowledge base continuously. Thus, the employees’ qualification as well as the constant enhancement of their (individual) knowledge constitutes an important precondition for an effective and efficient business process execution, the implementation of change management strategies and the realization of competitive advantages.

To address this problem, the attention metadata approach aims to improve the collection of recent user data across the various IT-system boundaries and applications used in different working surroundings. Thus, it represents information about each session a user attends in IT-environment. By unobtrusively monitoring and analyzing the users´ activities and habits, a more holistic information model about the single user can be created. Furthermore, using attention metadata facilitate the usage of competencies in human resource management (HRM) systems by identifying the knowledge and by observing the user’s activities. This information will be mapped into competency models and enables a tight integration of HRM-systems and LMS.

This paper describes the notion of attention metadata for learning and knowledge management environments. The understanding of attention metadata is defined in section 2. Section 3 outlines how attention metadata can be used in learning scenarios. The development of a framework to capture attention metadata is briefly imparted in section 4. Section 5 concludes the paper.

2 Attention Metadata

From our point of view, attention metadata lies directly at the intersection of data about the user and data about the content and applications. It represents data about the user, e.g. as represented in user profiles like IMS LIP [LIP] and PAPI [PAPI] profiles, while at the same time represents the information on activities the user has carried out with which digital content. We are thus able to conclude on the interests and knowledge of the user by analyzing the content he deals with, what he does with the content and how he handles the content. This includes the information on the processes the user is involved in thus allows the analysis of the underlying workflow and knowledge management processes.

The recently introduced AttentionXML [Attention] standard sets out to capture attention of users in terms of attention metadata. The attention metadata term was introduced to the field of information technology by Steve Gillmore, the president of AttentionTrust (http://www.attentiontrust.org/) and co-author of AttentionXML standard of tracking attention of users.

As an open specification, AttentionXML provides applications with a schema to describe data on tracking, prioritizing and sharing how people use digitally provided information, e.g. what people are reading, looking at or listening to. It is used to capture and share information on the attention that users spend on web pages, news feeds and blogs. Table 1 lists the elements that the AttentionXML schema provides. Please see [Attention] for an explanation of each element.

Post/Item/Page	Blog/Feed/Site
- Title	- Title
- GUID/identifier	- url
- Type (mime type)	- Alt url
	- Etga
- Etag	- Last updated
- Last updated	- Date added
- Last read	- Date removed
- Duration	- Last Read
- Followed links	- Read times
- Rel/vote link	- User feed title
- Tags	- Rel/XFN
	- Rel/vote link
	- Tags

Table 1: AttentionXML schema

It is apparent from Table 1 that AttentionXML provides basic elements to describe the attention metadata. Nevertheless, it is missing elements to capture the wealth of information provided in logs (Apache webserver logs or from standards as used in digital libraries [Gonçalves, 03] and their analysis [Jones, 00]), on the user or captured through monitoring how the user works with learning objects within systems. For example, the schema does not allow capturing the information about users' activities such as downloading, viewing or editing learning objects and applications and contexts where objects were used. The following section 3 motivates the usage of attention metadata in a learning scenario in more detail and briefly summarizes our extension of the attention XML schema.

3 Attention Metadata for Learning

As mentioned in the previous section, attention metadata provides us with detailed information how users handle specific learning objects. For example, it captures information on the context in which objects were used, how long users spent with them and how users located those objects. It also indicates which specific user is interested in which objects, based on their attention given to certain learning objects. This attention metadata is not covered by current learning metadata specifications (i.e. IEEE LOM [Duval, 03]) or by user information models (i.e. IEEE PAPI). Attention metadata is especially useful in the context of learning and knowledge management to improve the technologies that enable learning and working experiences with multiple contemporary technologies. The following examples can be given:

- Support administration decisions: the analysis of attention metadata provides decision makers in organizations with supportive information on the development of their employees, their status in knowledge acquisition and the location of knowledge within the organisation. Furthermore, by analyzing attention metadata, systems will be able to provide insights into the user's learning and working behaviour thus enabling management to recognize needs and potential savings (e.g. cheaper learning material) earlier.
- Enrich metadata about digital objects: (manually/semi-automatic) generating content metadata is one of the challenges content providers are facing. Attention metadata can provide rich metadata about digital content based on the attention it receives. Users use objects in a certain context and they provide tags and descriptions for objects they pay attention to. This data forms a rich source of information to be used in workflow and document management systems. In addition, it can be used to update user profiles with recent information about users' interest, attention and information needs. Users seldom provide accurate data about themselves or their interests, thus automatically derived attention metadata can be used to identify knowledge gaps and provide suitable learning offers automatically.
- Detail user models: most of the current user models draw and store conclusion about users rather statically. Those models might suggest, for instance, that a German should receive digital content in German, while they do not include information if that given user ever really spent attention on German content or

if he prefers English content. Attention metadata improves this situation by describing the real user behaviour. In our example, attention metadata captures the fact that the user has not read any German content but instead queried the knowledge management system for English information. Such metadata at the intersection of learner and content enables to build successful and more detailed learner models.

- Support recommendation systems: attention metadata can be a perfect source of information for recommender systems because it describes the past user experience with digital content. For example, when searching for suitable learning objects, the user gets a ranked results list presented with the ranking basing on the user's attention so far. Another example is a recommender system that helps the user to identify the most suitable learning object for his/her context, e.g. an automatic suggestion of related learning objects following the style of Amazon's "users who have bought this book also bought". In other words, the process of matching relevant content is transformed from searching through keywords in huge repositories to the "information finding".

Privacy in terms of data capturing, storing and usage within systems is one of the main issues that needs to be addressed and tackled properly when working with attention metadata. As this aspect is not in the scope of this paper; sources like http://www.attentiontrust.org/ provide more information on attention privacy.

The examples mentioned make clear that the attentionXML schema needs to be extended to suite these advanced usages. Therefore, we proposed the extension in [Najjar, 06] and will only briefly summarize it here to support readability. CAMs extends attentionXML with a number of additional elements required to capture user activities in various systems at the item element level. See http://ariadne.cs.kuleuven.ac.be/empirical/attention.php for all elements. Our extension focuses on three major elements:

1. The **Action** element provides information on the *action type* that the object was involved in (e.g. if it was downloaded, inserted, viewed, etc.). In case of a query, this element also stores information on the query terms used to locate learning objects. In case of an insertion of an object, it also holds information on the metadata schema used to index the learning object.

2. The **Application** element groups information related to the tool used by the user to carry out actions, e.g. search, use or integrate learning objects. Information hold includes the name of the tool, its URI and the type of the tool, e.g. LOR or LMS. Information collected in this element allows the identification and combination of attention information collected from different applications.

3. The **Session** element holds the information that is needed to identify the different working sessions that the user engages within the systems. However, this element is not intended to hold the session of a user across multiple system boundaries. Instead, it only captures session information per system.

The following section will give a brief overview of the framework that employs this extended schema for capturing and merging attention metadata.

4 The CAM Framework

Users usually interact with a wide variety of tools while working with digital content in one way or another. We provide here a none-exhaustive list that exemplifies the large number of tools and possible interactions within. The list illustrates the sources from which attention metadata can be obtained:

- Knowledge Management Systems: Attention metadata can easily obtained by observing which employee inserts and uses which knowledge.
- Workflow Systems: users work with documents that either store information or enable conclusions on the user's knowledge by comparing the real document status with the theoretically necessary document status at each station in the workflow.
- Human Resource Management Systems: These systems provide information on the user's knowledge, his learning status and the context the employee works in (e.g. job description, etc.).
- Learning Object Repositories (LOR): the main activities of users interacting with LOR (such as MERLOT, EdNa, ARIADNE and SMETE) usually include the search, download and/or upload of relevant learning objects.
- Learning Management systems (LMS): In addition to LOR, LMS (such as Clix, Blackboard, WebCT and Moodle) include activities that are related to the aggregation and management of learning objects. Furthermore, LMS provide functions related to the management of courses, students and teachers so that a first correlation between users and their activities with learning objects can be established.
- Internet Browsers: Users view and download relevant learning objects from sources on the internet, e.g. found through Google or within a LOR or LMS.
- Authoring tools: Examples of such tools are word processing suites such as MS-Word, OpenOffice and MS-PowerPoint that are used to create new learning resources by reaggregating existing ones, by authoring new contents and by modifying existing ones.

The above list clearly illustrates that attention metadata, generated from all these sources needs to be combined to provide a more complete set of information on the user. Figure 1 presents the contextualized annotation metadata (CAM) framework that we are developing to exchange and manage attention metadata regarding the users and their interaction with learning objects.

Our framework is intended to publish attention metadata related to each tool in a separate stream using the CAM schema. Afterwards, the set of those Attention streams are merged into one extended attention metadata stream. As outlined in Figure 1, each tool the user interacts with stores attention metadata locally (currently in XML format and valid against CAM schema) and sends a separate attention stream to the attention repository (XML DB) where the streams are merged into the repository. The merge process bases on our CAM schema in which we attempt to capture information about the user in the various systems and environments.

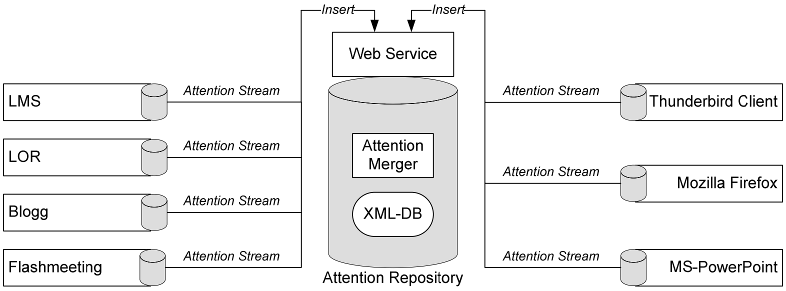

Figure 1: The CAM framework

Applications connect to the attention repository to retrieve the necessary data that enables the above examples and more. Using a central attention metadata store enables us to provide a large base of data on which statistical algorithms can perform the necessary analytics, e.g. provide the HR department with an overview of the recently acquired information of one unit or deducing the skills of employees based on their advances in learning. A recommender system could use the attention store to conclude on recommending which learning material to which employee by basing the decision on their learning styles and learning advances.

Tracking the user over system boundaries requires a strong identity management. In this respect, we rely on research projects to provide solutions to the identity issue (e.g. liberty alliance; http://www.projectliberty.org/).

Our framework provides the merged extended AttentionXML streams to participating KM, HMRS, LOR, LMS recommender and adaptation systems to enable advanced and personalized services. Thus, we propose the Contextualized Attention Metadata schema (CAM) and framework (CAMf) to facilitate the collection and management of user rich attention metadata from a variety of learning tools.

5 Conclusions and Future Work

The paper describes the theoretical bases to bridge the gap between the estimated and the real knowledge in companies. Attention metadata forms the basis on which observations about real knowledge in companies can be represented and system overarchingly collected. Thus, attention metadata will enable a tight integration between the whole notion of knowledge management and related systems on the one hand and corporate learning management systems in general. The approach presented here allows us to use organisational systems as additional sources for attention metadata thus providing us with the ability to capture the user behaviour in learning and working contexts. The analysis of this data allows the targeted steering of the learning process of the single user for the company while, at the same time, provides the learner with an improved learning experience that is better suited to his requirements.

So far, we have developed first tools, e.g. a plugin for MS PowerPoint and a transformer of Slogger attention metadata to CAM (see [Najjar 06] for more details) to capture attention metadata in learning contexts. We intend to extend this approach to corporate environments in which we envision to build large attention metadata bases and develop respective analytical methods.

Acknowledgements

We like to thank Katrien Verbert, Riina Vuorikari, Xavier Ochoa and Julien Broisin for the fruitful discussions that supported this work. This work has been carried out in the context of the EC/IST supported PROLEARN NoE project.

References

[Argyris, 78] Argyris, C., Schön, D. A.: “Organizational learning: A theory of action perspective”. Reading et al. (Addison-Wesley), 1978.

[AttentionXML] AttentionXML, "AttentionXML specifications", http://developers.technorati.com/wiki/attentionxml.

[Duval, 03] Duval, E., Hodgins, W., (2003). "A LOM Research Agenda". In: Hencsey, G., White, B., Chen, Y., Kovacs, L., Lawrence, S. (Eds.): Proceedings of the twelfth international conference on World Wide Web, pp. 1-9, 2003.

[Gonçalves, 03] Gonçalves, M., Panchanathan, G., Ravindranathan, U., Krowne, A., Fox, A. E., Jagodzinski, F., Cassel, L.: "The XML Log Standard for Digital Libraries: Analysis, Evolution, and Deployment". In: Proceedings of the third Joint Conference in Digital Libraries, Houston, Texas, May, 2003.

[Jones, 00] Jones, S., Cunningham, J. S., McNab, J. R., Boddie, J. S.: "A transaction log analysis of a digital library". International Journal on Digital Libraries 3(2): 152-169 (2000).

[LIP] LMS Learner Information Package Specification. (LIP), http://www.imsglobal.org/profiles/index.html.

[Najjar, 06] Najjar, J., Wolpers, M., Duval, E.: "Attention Metadata: Collection and Management". In: WWW2006 workshop on Logging Traces of Web Activity: The Mechanics of Data Collection, Edinburgh, Scotland, May 23, 2006.

[PAPI] IEEE P1484.2/D8, "Public and Private Information (PAPI) for Learners", http://edutool.com/papi/.

[Pinchot, 93] Pinchot, G., Pinchot, E.: “The end of bureaucracy and the rise of the intelligent organization”. San Francisco (Berrett-Koehler), 1993.

[Romhardt, 98] Romhardt, K.: “Die Organisation aus der Wissensperspektive – Möglichkeiten und Grenzen der Intervention”. Wiesbaden (Gabler) 1998.

[Wilensky, 67] Wilensky, H. L.: “Organizational Intelligence: Knowledge and Policy in Government and Industry”. New York, London (Basic Books), 1967.