LM-DTM: An Environment for XML-Based, LIP/PAPI-Compliant Deployment, Transformation and Matching of Learner Models

Mohamed Amine Chatti, Ralf Klamma, Christoph Quix, David Kensche
RWTH Aachen University, Informatik V
{chatti\klamma\quix\kensche}@cs.rwth-aachen.de

Abstract

Our shared belief is that learning, like other human activities, cannot and will not be confined within rigidly defined course systems or learning repositories, inclosing learning resources which cannot be tailored to the different learner´s needs, skills, interests, preferences, goals, etc. Therefore, a learning environment, beside supporting communication between knowledge providers and consumers, has to be organized in a flexible manner based on different learner profiles. Learner modeling has become a highly challenging task to provide personalized, adaptive and context-based learning. The work presented in this paper address this issue by providing a meta-level solution for description, transformation and matching of learner models, based on standards such as IMS LIP, IEEE PAPI, XML to foster the reuse and exchange of learner models between learning platforms, both by universities and corporations.

1. Introduction

In the past few years, personalized adaptive learning has become increasingly important in the e-Learning domain since learners have a clear need to retrieve information that fits to their preferences, goals and cultural backgrounds in a world facing information overload in general. However, the majority of the learning platforms do not take into account those heterogeneous needs of users, provide the same learning material and process to students in different contexts. In addition, they are still lacking powerful and reliable search methods and techniques that enable the localization of relevant learning resources which are tailored to the learner needs. A solution to this problem can be best achieved through the use of learner models, the combination of standards with current and emerging technologies to offer better information retrieval strategies that take into account the characteristics of the learner performing the search; thus, achieving **personalized adaptive learning**. This paper is concerned with **learner modeling** and its role in achieving personalized adaptive learning, as well as the investigation and implementation of standard based methods and techniques for definition, manipulation, storage, matching and retrieval of learner models. We present a system, LM-DTM (Learner Models Deployment, Transformation and Matching), that supports these tasks. The rest of the paper is structured as follows: Section 2 deals with the personalization issue which becomes the key element of e-Learning. Section 3 gives an overview of the main issues and requirements related to personalized adaptive learning and highlights possible solutions to provide personalization of the learning process by means of metadata, standards, learner models, as well as several techniques and algorithms from both the model management and information retrieval research areas. Section 4 introduces the most important standards dealing with learner modeling. A brief description and comparison of those follows. Section 5 covers the standard based LM-DTM environment including an IMS LIP-compliant learner profile metadata editor, XML-based learner model tranformation facilities and storage possibilities, as well as powerful matching and retrieval algorithms. Finally, Section 6 gives a summary of the paper and outlines perspectives for the future.

2. Personalized adaptive learning

Learners are different: they have different needs, learning goals, cultural backgrounds, domain expertise and a wide range of skills, abilities, preferences, interests, processing speeds, etc. Besides, learning content adapted to a learner's need has been shown to produce more effective learning, for learners respond better to such personalized content. Therefore, in recent years, personalization has become the key element in the learning process in order to find appropriate learning content which satisfy a learner's need and support the accomplishment of a specific learning objective, thus enhancing the effectiveness and efficiency of the learning activity and achieving higher learner satisfaction. The need for personalization and adaptation has even accompanied the evolution of the learning object definition from any entity - digital or non-digital - that may be used for learning, education or training [1] to a collection of information objects assembled using metadata to match the personality and needs of the individual learner [2].

Recognizing its importance, personalization has been investigated in various different areas and systems in the learning context. *Intelligent tutoring systems* adapt their explanations and teaching strategies to the individual needs of users in terms of their knowledge level and learning progress [3]. *Adaptive hypermedia systems* support personalized presentation and navigation [4]. Other systems dealing with personalization issues are *information filtering and recommender systems*. These systems aim at adapting the resource content based on information about the learner which is maintained in a central or virtually integrated repository [3] [5]. However, the majority of these systems either makes information about a learner available as predefined stereotypes or try to generate a learner model that is not based on learner modeling standards, which are fundamental requirements for achieving personalized adaptive learning. Moreover, this system dependent learner model representation does not allow the reuse or the exchange of learner information between different learning environments.

We are working on personalization issues in the context of the EU Network of Excellence PROLEARN for professional training. The main aim of this network is to bridge the currently existing gap between research and education at universities and similar organizations and corporate and life long learning that is provided for and within companies [6]. Personalized adaptive learning is one of the key research areas in PROLEARN that focuses on improving the efficiency and cost effectiveness of learning, for individuals and organisations, independent of time, place, and pace and try to provide solutions to achieve personalized and effective learning.

3. State of the art in personalized adaptive learning

As mentioned in the last section, personalization is an important issue that still has to be resolved. We believe that metadata, standards and learner modeling are a fundamental requirement and an important step towards achieving personalized learning.

Personalized adaptive learning requires the use of metadata for describing learner profiles and learning resources as well as adopted, common, open and accredited standards. Metadata is traditionally defined as descriptive, meta information about information and, as defined in [2], is the full and rich set of information needed in order to find, filter, select, and combine the information. In e-Learning, metadata becomes increasingly important and is required to annotate learning resources and describe learner profiles in order to support and facilitate the search and retrieval of appropriate learning resources or learning paths relevant for a specifific learner or a group of similar learners. However, metadata used in annotation of learning objects and in the description of learner profiles should be based on a common metadata schema to reduce variability in models used in the learning service and avoid the large number of complex and time consuming mappings that might be required in case many learning environments need to communicate with each other. Additionally, the use of standards yields the following advantages: reusability, accessibility, durability of learning resources and learner profiles, and interoperability across heterogeneous learning platforms. Many e-Learning standards have existed as basis for personalized services. Besides, XML, and its related standards, have provided some elegant potential solutions to the content personalization issue. These standards will be discussed in details in sections 4 and 5.

Learner modeling is the cornerstone of the personalization process. The learner model reflects information that is specific to each individual learner or a group of resembling learners such as current knowledge level, performance, progress, learning objectives, personal interests and preferences as well as the topics from the supported learning domain that the student has already covered. Furthermore, the degree of adaptivity of provided learning contents is proportional to the information available about the learner´s characteristics and features. The more a learning system knows about a learner, the greater is the chance to deliver learning content that best suits to the learner´s needs. Therefore, a learning system has to have access to the learner information and handle learner profiles to determine which content is the most appropriate and provide the learner with learning resources or complete learning paths tailored to his/her needs.

The goal of model management is to develop mechanisms and techniques to transform, match, and merge models. The most important operators are Match, Diff, Merge, and ModelGen/Transform (to generate a new model from an existing model). In the context of e-Learning, especially the transformation and matching of models is relevant. The application of the Match operator to two models should return a mapping between the models, relating elements with the same or similar semantics [7]. Similarity Flooding [8] is an algorithm that works very well for schemas with similar structure and without a high similarity of the labels. However, it has been shown that the current matching algorithms are not efficient for matching large schemas [9], or if a schema should be matched against a huge number of models (as it is the case for e-Learning where, for instance, a learner model should be matched against learner models available in a repository). Thus, more sophisticated mechanisms are required in this case. One way to solve the problem is to apply a filter to the set of models, to limit the number of candidates for which a detailed match operation can be applied [10][11].

Information retrieval is also of great importance in the context of personalized adaptive learning in order to find learning resources relevant to a learner´s interests or to determine similarities between models such as learner models. In the vector space model [12], similarity between vectors (e.g., document and query) is represented by the mathematical similarity of their corresponding term vectors.

4. Learner modeling standards

A commonly agreed representation of learner profiles is still missing. In recent years, there have been some efforts to standardize learner profiles. Numbers of standards have emerged for learner modeling. The two most important specifications in this area are the IEEE Personal and Private Information (PAPI) [13] and IMS Learner Information Package (LIP) [14]. Standards for bindings of the abstract data model such as the XML binding have also been produced for both specifications. The aim of those standards is to define a standard exchange format for learner metadata and make information on learner profiles available in a flexible way that permits storage, manipulation, access and more advanced searching. There are significant differences in the metadata defined in PAPI and LIP standards. The focus of PAPI is to provide a minimum of an information set that would enable tracking a learner's performance during his/her study. Accordingly, this standard specifies categories such as performance, portfolio and relations to other learners. Unlike PAPI, LIP offers richer structures and aspects, takes into account important learner´s features such as goal and interest and describes learner´s characteristics for the purpose of personalization of content.

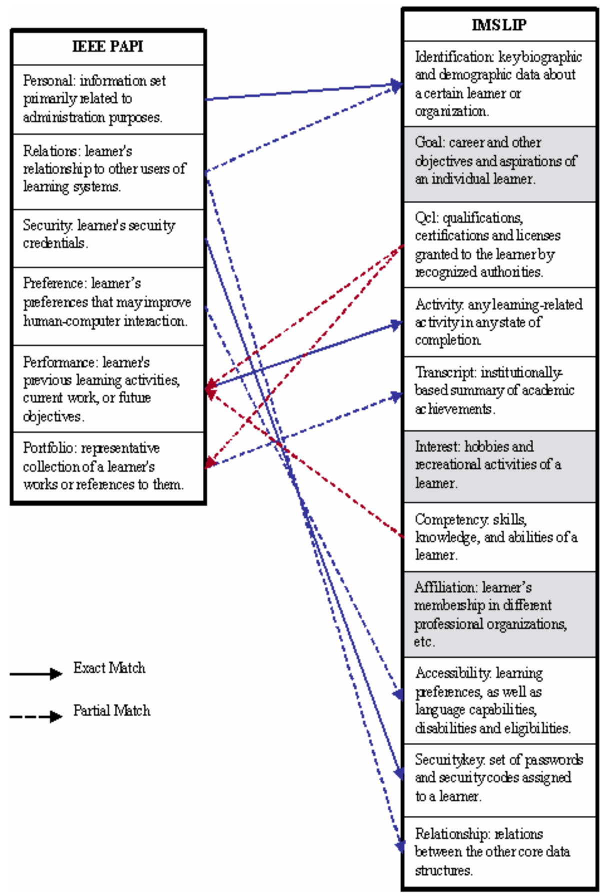

Figure 1: Mapping between PAPI and LIP categories

As shown in figure 1, using LIP as analysis base, all PAPI features can be expressed via IMS features explicitly (e.g. *Personal*) or implicitly (e.g. the *Relations* category in PAPI can be represented in LIP via relationships between different records of the *Identification* category). Besides, PAPI does not cover important learner´s features such as interest and goal, though information about learners' goals can be useful when applying recommendation and filtering techniques in adaptive systems [15]. Furthermore, Unlike PAPI which is limited regarding implementation examples and readability, LIP is extensible, has an implementation from the IMS, provides best practice guides, and is easier to use and read. On this account, we opt for the LIP solution, and we use it as a baseline for learner modeling in LM-DTM. Since PAPI is being used as learner modeling standard from several learning environments, a transform operator has also been implemented to generate a PAPI-compliant learner model from a LIP-based one, thus ensuring more flexibility and interoperability.

5. LM-DTM structure and implementation

As mentioned in section 3, a personalized adaptive learning environment should be based upon learner profiles and inclose learner model based information retrieval techniques to locate learning resources that fit to the learner needs. The success of the retrieval results in this case depends largely on the learner model retrieval performance and how these models are built, stored and mapped. For achieving this, a first, standard based software prototype for deployment, transformation, matching of learner models has been implemented and will be described in the following.

In LM-DTM, the learner or the system user is asked to define features describing the learner profile. The new learner model is then constructed and shown in the left frame of the LM-DTM graphical user interface, depicted in figure 2, either as an XML DOM tree view or as XML code text.

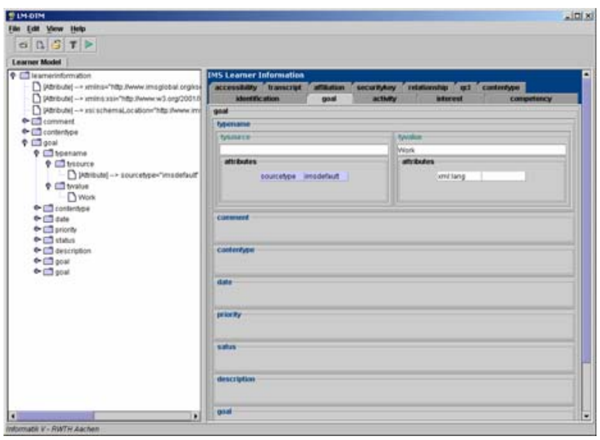

Figure 2. LM-DTM graphical user interface

In order to enhance interoperability among learning platforms, a learner profile can also be imported to the LM-DTM environment as an XML document. The composed learner model is based on features inspired by the LIP standard and its XML representation is conform to the LIP XML binding. However, learning environments, though based on learning modeling standards, cannot be expected to use the same representation scheme. Therefore, some mechanism is needed to transform at least between the the two most important learner modeling formats namely LIP and PAPI. Based on a detailed comparison of these two standards (section 4), a set of XSLT files are maintained which represent mappings from LIP to PAPI core features. These mappings are expressed as 'select-match' expressions, which are used to select the current element from the LIP format to match it to a particular, semantically similar target element from the PAPI representation.

After building the learner models, the next issue to consider is their storage and retrieval. In order to effectively store and manage learner profiles, we have chosen a native XML database: Apache Xindice [16]. Contrarily to XML-enabled relational databases or object-oriented databases, native XML databases provide more natural, sophisticated and reliable solutions for storing, managing, searching and retrieving of complex XML data. Moreover, multiple XML documents can be contained in a single collection and thus be queried as a whole. Xindice supports a variety of W3C standards, including the XPath query language and the XUpdate language for database updates. In Xindice, queries are executed at the collection level. In addition, Xindice allows indexing at the element and attribute levels of XML documents; thus the queries' run-time performance could be further improved.

The defined learner model in the deployment step can be matched to other models which are stored in the XML learner modeling repository. For this step, we have designed a vector model-based matching and rating algorithm which is currently being implemented. The output of this algorithm are identifiers of the results found and a ranking as a degree of similarity and relevance of each result. The algorithm takes into account the personalization domain (i.e. learning environment). We have delimited two learning environments: learning in the educational environment (educational learning) and learning in the context of work where learning is an integrated part of working (corporate learning or learning on demand). In the educational context, it is important to consider aspects such as learner performance, qualification and relationships to other learning environment users. In the corporate learning or learning on demand context, the most important information used in describing learners are goal, interest, experience (level of expertise), and preference. Based on the personalization domain, different learner features are assigned different weights in the matching and rating algorithm according to their importance and relevance. Additionally, the Xindice indexing system is used to define indexes on those features in order to speed the queries´ run-time performance.

The matching and rating algorithm performs as follows: Using XPath, queries are constructed on the basis of a set of core learner features (i.e. goal, interest, accessibility, identification, qcl, activity, competency, transcript, affiliation, securitykey and relationship) that the system user has defined in the deployment step. The queries are then sent to the XML learner model repository to get relevant sections and feature values from the learner models already stored in the repository and to compare them with the user´s defined values in the deployment step. During this comparison, similarity and rating heuristics have to be applied. Similarity by features with values in a predefined value set, such as language preference, can be easily exploited. However, for more complex features with, for instance, sub-features or textual ranges, additional heuristics and powerful statistical methods have to be applied. On this account, a similarity sim_i is calculated for each defined core feature f_i using the vector model. Thereafter, we calculate the global similarity as follows:

$sim_{global} = \sum sim_i.w_i$. The weights w_i with $\sum w_i = 1$ represent the relative importance of the core feature f_i and are set according to the learning context as discussed above.

6. Conclusion and future work

The focus of this paper was e-Learning personalization. We have shown that metadata, standards, learner modeling and powerful information retrieval techniques are crucial factors for achieving personalized and adaptive learning. The investigation of existing learner modeling standards revealed that the IMS LIP specification is quite complete. A standard based environement for deployment, transformation and matching of learner models has been presented. From the technological point of view, common standards of the XML language family such as XSLT and XPath have been deployed. We have designed an algorithm for matching and ranking of learner models using the vector model for similarity calculation at the core feature level. However, an experimental evaluation of the retrieval quality of our matching algorithm is still missing. In further work, we plan to integrate the LM-DTM environment in a complete, standard based learning environment using LOM as learning resource annotation standard and including learner model based learning object retrieval techniques. Furthermore, we will evaluate schema matching algorithms and other model management techniques. Besides, methods that support XML-oriented information retrieval such as the extented vector model [17] will be investigated to enhance the retrieval performance.

7. References

[1] 1484.12.1: IEEE Standard for Learning Object Metadata.

[2] H. Wayne Hodgins, Into the Future. Available at http://www.learnativity.com/download/MP7.PDF.

[3] G. Kappel, B. Proll, W. Retschitzegger, and W. Schwinger. Customization for Ubiquitous Web Applications: a Comparison of Approaches. International Journal of Web Engineering and Technology, 11, January 2003.

[4] P. Brusilovsky. Adaptive Hypermedia. User Modeling and User-Adapted Interaction, 11(1-2):87–100, 2001.

[5] A. Kobsa: "Generic User Modeling Systems", User Modeling and User-Adapted Interaction, Vol. 11, No. 1-2, Ten Year Anniversary Issue, pp. 49-63, 2001.

[6] M. Wolpers, G. Grohmann: PROLEARN: technology-enhanced learning and knowledge distribution for the coroporate world. Int. J. Knowledge and Learning, Vol. 1, Nos. 1/2, 2005.

[7] E. Rahm, and P. A. Bernstein: A Survey of Approaches to Automatic Schema Matching, VLDB Journal 10, 4 (Dec. 2001), pp. 334-350.

[8] S. Melnik, H. Garcia-Molina, E. Rahm. Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching, 18th ICDE Conf., 2002.

[9] P. A. Bernstein, S. Melnik, C. Quix, M. Petropoulos: Industrial-Strength Schema Matching, ACM SIGMOD Record 33(4), December 2004, pp. 38-43.

[10] E. Rahm, H. Do, S. Maßmann: Matching Large XML Schemas, SIGMOD Record 33(4), December 2004.

[11] K. Kailing, H.-P. Kriegel, S. Schönauer, T. Seidl: Efficient Similarity Search for Hierarchical Data in Large Databases, Proc. 9th Int. Conf. on Extending Data Base Technology (EDBT'04).

[12] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. Comm. ACM 18(11), 613-620, 1975.

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

[14] IMS learner information package specification. Available at: http://www.imsglobal.org/profiles/index.html.

[15] P. Dolog and W. Nejdl: Challenges and Benefits of the Semantic Web for User Modelling. In Proc. of AH2003 workshop, Budapest, Hungary, May 2003.

[16] Apache Xindice native XML database. Available at http://xml.apache.org/xindice/index.html.

[17] E. A. Fox. Extending the Boolean and vector space models of information retrieval with p-norm queries and multiple concept types. Ph.D. Dissertation, 1983.