Modeling Learner information within an Integrated Model on standard-based representations

Mario Chacón-Rivas[1,*], Olga C. Santos[2], Jesus G. Boticario[2]

[1] TEC Digital, Instituto Tecnológico de Costa Rica, Cartago, Costa Rica
machacon@itcr.ac.cr
[2] aDeNu Research Group, Artificial Intelligence Departament, Computer Science School, UNED C/ Juan del Rosal, 16. Madrid 28040. Spain
{ocsantos,jgb}@dia.uned.es

Abstract. Learner modelling is a process consisting of collecting information explicitly from users and inferring some data from the learner activity. This information is basic for recommending resources as well as to predict performance. There are open issues when it comes to integrate in standards-based user models that information, which covers learning styles, competences, affective states, interaction needs, context information and other learner´s characteristics. In particular, there are standards that can be used to cover several of the subjects to be integrated into those models, such as IMS-LIP, IMS-RDCEO, IMS-AFA. This paper presents a work on implementing a user model that aims at providing a holistic UM perspective, which is able to hold and collects all relevant information, thus supporting its real-life usage. This is expected to facilitate interoperability and sustainability while we are progressing on filling the gaps, where representation and management is required.

Keywords: User modelling, IMS standards, Interoperability of user models, Lifelong Learning User Modelling

1 Introduction

User Models (UM) have been considered as a representation of information on individual users, which is essential for building applications of adaptive systems, intelligent interfaces, intelligent information retrieval and expert systems, among others [1]. Also UM are being used for over the last two decades on implementing personal learning environments, adaptive learning environments and intelligent tutoring systems [2]. Information about UM is usually categorized in terms of personal, affective and cognitive information [2]–[4].

Nowadays there is an increasing interest in taking advantage of new interaction data which cater from learner affection thus requiring integrating into UM affective state indicators [4] [5] [6]. These indicators provide valuable pedagogical pointers, which affect the cognitive process. Actually, learners´ affective modelling is impacting positively on adaptive systems, recommender applications and personalized learning environments [6].

In order to cope with both existing UM information and providing a real life standards-based application this paper introduces existing challenges in terms of the information to be integrated into the model (ie., competences, learning styles, socio-economical data, among others) and the available standards to cope with (e.g., IMS-RDCEO, IMS-LIP, IMS-AFA). The rest of the paper consists of section 2, where UM components and the identification of variability levels are presented, section 3, which summarizes the IMS family international specifications to be integrated into the UM, and last but not least, section 4 where some lines of work in progress are introduced.

2 Identifying UM Components and Information Levels

UM components have been specified in terms of categories or data to be captured to model the learner. Those components could be specified explicitly asking information to the learner or could be specified inferring from the learner interaction with the e-learning platform.

Independently the way to capture the information of the learner, as commented by Brusilovsky and Millán in [2], the interest of information to be modelled in learning environments must allow to identify the user as an individual, thus supporting a feedback process which can be managed by providing recommendations oriented to the meet learners´ needs.

In the context of this research, we identify the UM components and classify them in terms of variability. The variability term is oriented to classify the information depending on the frequency of change, because it will influence any process of recommendation. The UM attributes identified are generated by a methodological process that integrates several sources of information and stakeholders. During the methodology application, the identification of stakeholders designs preliminary security roles.

2.1 UM Components

Several authors defined the UM components oriented to personal information, knowledge and interest. In [5] Bull and Kay enumerate as cognitive, affective and social attributes. Baldiris et. al. [7] present the user model in terms of learning styles, competences and access device preferences, also it includes knowledge level based on six levels of knowledge defined by Bloom´s taxonomy. This proposal also includes a collaboration level based on indicators obtained from learners´interaction in the learning management system.

Based on aforementioned and related literature we are currently defining the UM components information in terms on the following information categories: *personal*, *provenance*, *academic record*, *socio-economical*, *accessibility* or special needs, *psychological*, *learning styles*, *competences, knowledge level and collaborative level*.

Those components can be modelled and organized in terms of standards, such as the IMS family specifications. The use of these international specifications is aimed to support collaboration and systems interoperability and have the advantage of being specifications that are already integrated into dotLRN [7] [8] [9]. IMS-LIP is a collection of information about the learner, which supports data exchange between applications, agents, server and other services concerned about the learners` characteristics [10] IMS-RDCEO is a concise and flexible structure to represent competencies, furthermore this specification is extensible to any competence model [7] [11].

To provide the required standards-based modelling support at TEC, we are following these previous approaches while extending them and filling information gaps when needed. Based on IMS-LIP categories, the element *identification,* is loaded with *personal, provenance, academic, socio-economical* information. The element *Competency* is loaded with competences information. The TEC competences model is based on CEAB model [13] and this competencies information is represented using IMS-RDCEO. *Accessibility* is represented using IMS-AFA. Additional details on IMS family specifications are provided in the section 3.

The UM in our context, called td-um means TEC Digital-User Model. **td-um** is an integrated model because it is gathering together learner information from applications, databases and some indicators collected from learner interaction with the e-learning platform.

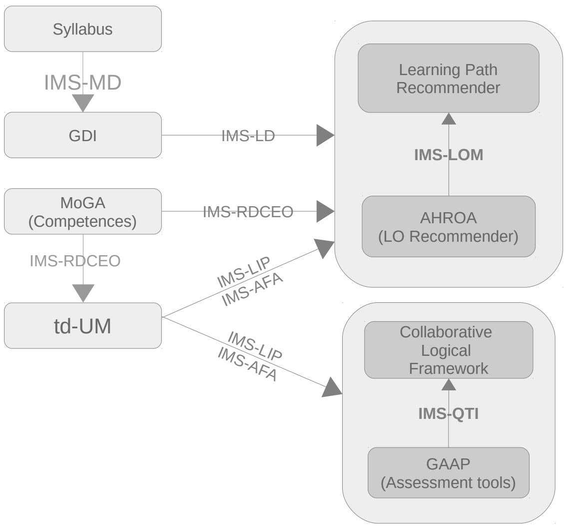

Figure 1: td-UM Integrated model

From that integrated approach and after studying available information from literature, we have detected some gaps in the information to be modelled, these are the following:

- Learner knowledge level: it is been modelled based on specific background of knowledge from each discipline studied by the learner. In case of computing students, the knowledge level is modelled using knowledge areas presented by ACM in [14]. The gap to be resolved is based on several bodies of knowledge from different disciplines. The other disciplines to be modelled in TEC are: Industrial engineering, Electronic engineering, Materials engineering, Electromecanic engineering, Construction engineering, Agricultural engineering, Industrial Maintenance engineering, Occupational Safety and Environmental Hygiene Engineering. This other disciplines have a different body of knowledge, therefore the structures used to model should be sufficiently flexible.

- Academic record attributes: these cover information reflecting the progress in program courses in terms of final grades or qualifications. It is frequently confused with the knowledge level. This is mainly required to preserve historical information.
- Competences attribute: it contains a set of competences that requires reflecting the level of domain of each competence, and evidences used to assess each competence, among others. The work in progress is designing the model which we are adapting to cope with TEC competence model.
- Variability of information: an important issue is to track the progress in competence domain, as well in academic records. This progress tracking represent some level of variability of information that could impact the recommendation and adaptivity of platform.
- Categories and attributes privacy levels: the privacy level of some attributes or for the whole category are not clear in the specifications. For example, in socio-academic attribute, the sub-attributes: *level of sociability*, *esteem*, *motivation*, *coping strategies* contains private information accessible only for department of Psychology and the learner. In this work is needed to define and to model the privacy level by category and attribute based on privacy modelling [15]. The model is contemplating the user roles definition and the integration with dotLRN.

Currently at TEC Digital we have implemented several applications that are using partial learner models, Figure 1 shows the integration architecture. The immediate work is being focused on adapting td-UM, to be used as source of integrated learners´ information, which will be able to support recommendations and assessments. Those applications are:

- Adaptive Learning Paths, which use the learning design of a course and the students´ performance information to recommend learning resources. The recommendation is based on bayesian networks [16].
- Hybrid Agent Recommender of Learning Objects (AHROA), based on the learning design and syllabus of the course, recommends learning objects to learners. The recommendation is prepared using TF-IDF[1] to work the terms relevancy, also uses cosine similarity. The UM will provide information about learner needs to AHROA in order to identify the impact on the quality of recommendations.
- Collaborative Logical Framework, (CLF) implemented by aDeNu [17], it uses collaborative indicators to assess the learners collaboration. An important activity on the CLF is the identification of each group moderator during

[1] More details in http://nlp.stanford.edu/IR-book/html/htmledition/tf-idf-weighting-1.html

the consensus stage, this activity is currently based on the learner interaction in the platform. The UM will impact the CLF integrating specific attributes concerning to the leadership and entrepreneurship. The TEC Digital adapted CLF to improve the indicators information [18] and to analyse the impact of the UM in CLF assessment.

- Learning activities editor application (GAAP) implements learning styles test based on Felder&Soloman theory [19].
- Several test to determine personality and character, leadership, entrepreneurship, communications competences. These tests are defined by the Psychology department and by the team responsible to design the competence model. Some tests are in processes to be patented by TEC. The variability and tracking progress of these competences are very important to be considered in the CLF, GAAP and AHROA.

2.2 UM Variability Information Levels

In order to take advantage of the information being modelled in real-life situations we are particularly interested in taking into consideration the "variability" factor. The levels of information *variability* reflect the frequency of variability or changes on the values of UM components. Authors as Sosnovsky and Dicheva in [4], defined this variability as *long term* and *short term* variability. In our research we are defining these variability levels as *low*, *medium* and *high* as explained bellow:

Low Variability: Some UM components hardly ever vary during the learning period and are used either for managing personal information (e.g., *name*, *birthdate*, *provenance*, *native language*) or academic processes, such as birth date, which can be used to calculate the learner age, the native language and other language domain features that may have an impact on the learning process.

Medium Variability: UM components seldom vary on a daily basis but the change more frequently than those being described as *low*, including periods of relatively stable values; usually are characteristics that are modified during the learner progress in the curricula. It could be presented with the competences category. Improvements on competences are been registered and assessment once at year. Some examples of competencies are *communication skills*, *team work, problem analysis, knowledge base of engineering, ethics and equality,* among others. The impact on medium-level variability changes in the learning process is very important for learners because the progress on some of those components affected by them has a direct and immediate influence in the learning performance.

High Variability: Those components vary almost continuously and hardly ever remain stable, some even could vary daily.

3 Representation based in standards

The use of international specifications, such as the IMS family, is aimed to support integration and interoperability. The LMS used in TEC is based on dotLRN, which supports IMS-LD and IMS-QTI. The model td-um is based on IMS specifications adapting to the specific needs mentioned above as gaps in the models.

Table 1 shows the categories and standards we are using along with the main attributes of the UM.

Table 1: Standards and information category in td-UM

Standard	**Category**	**Attributes**[2]
IMS-LIP	personal	name, birthdate, address, phone number, email address, native language, affective, socio-academic needs{ study conditions, study habits, metacognitive study strategies, development study strategies, organizational study strategies, level of sociability, esteem, motivation, coping strategies}
	provenance	place of provenance (based to recognize the social development index)
	socio-economical	scholarship, loan financing
	academic record	based on progress record on each course
	learning style	based on Felder and Solomon learning style test
	knowledge level	based on the body of knowledge of specific disciplines
	collaborative level	based on indicators tracked from the interaction with the platform.
IMS-AFA	accessibility	visual adaptation, hearing adaptation, cognitive adaptation, learning needs {reading-writing, understanding, speaking, math, attention, depression, anxiety, difficulties with peers, family problems, difficulties with teachers}
IMS-RDCEO	competences	knowledge bases of engineering, problem analysis, investigation, design, use of engineering tools, communication skills, professionalism, impact of engineering on environment and society, ethics and equity, economics and project management, lifelong learning, resource utilization

IMS-LIP categories: *Academic record* is being adapted to support historical information. *Knowledge level* is being adapted to cover the body of knowledge of several disciplines. Also the *knowledge level* should model the knowledge area or topic with a level reflecting the expertise in the given domain. This level is going to be described in terms of the Bloom´s Taxonomy, following previous approaches [7].
Personal category is been adapted to model *socio-academic needs*{ study conditions, study habits, metacognitive study strategies, development study strategies, organiza-

[2] This column enumerates only principal attributes.

tional study strategies, level of sociability, esteem, motivation, coping strategies} all captured using a test of 40 questions.
IMS-AFA category of *Accessibility* is being adapted to model *learning needs* {reading-writing, understanding, speaking, math, attention, depression, anxiety, difficulties with peers, family problems, difficulties with teachers}. These learning needs are captured using a test of 66 questions.
IMS-RDCEO category of competences {*Knowledge base of engineering, Problem analysis, Investigation, Design, Use of engineering tools, Individual and team work, Communication skills, Professionalism, Impact of engineering on environment and society, Ethics and equity, Economics and project management, Lifelong learning, Resources utilization*}. Adaptation to be provided require to support the representation of the domain level of each competence, the evidences used to assess each competence and the authority. Another modelling issue is to support the integration and matching of the competence model with those required for an international accreditation process. Accreditation processes are oriented to model the program of courses or careers in universities, while learner models reflects personal and individual information. The competence model for international accreditation is based on statistical samples [20].

4 Works in progress towards an integrated learner model

Once the aforementioned issues are designed, structured and integrated to cope with the information to be represented, the next decision will be the specific way to represent information on each competences, learning styles and affective indicators. The interoperability of those indicators will require an ontological representation that allows to deal with the information to be used in each foreseeable situation.

As an example of the decision to be done, we are planning to apply a test of temperament to identify (1) extroverted – introverted *temperaments*, (2) ways to *capture information*: by intuition-by senses, (3) *ways of making decisions*: by thought- by feeling, (4)*ways to organize time*: judicious-mandatory. This test has 70 questions, the interpretation of the results will give a value for each element to identify. A learner could have a value for the way to capture information of 6-4 (ie, 6: by thought, 4: by feeling). The internal decision about the representation of those values could affect the adaptive process, if the UM stores the 6 value or the pair 6-4. A recommender system could take several considerations concerning the type of resources to recommend. The tests used to capture information are being used by TEC since 2002 and they are bases in [21].

Finally, this research is aimed to fill the gaps beyond current usage of UM in adaptive learning systems thus making it really extensible, sustainable and applicable in any situation.

We are currently progressing on the first stage of this research, which covers: (1) understanding the dimensions of UM and the results of experimental research in learning scenarios, (2) focusing on reviewing the state of the art in UM and its components, (3) identifying a methodological approach to gather the UM attributes in a real learning environment, (4) identifying possible gaps that may come up when integrating UM into real dimensions of learners characteristics that have impact on the learning process.

This contributions of our research are focused in (1) defining a methodology to identify attributes of UM in real learning environment that support personalised and inclusive e-learning scenarios, (2) identify the UM attributes that really impact in recommendation processes using AHROA and CLF, (3) validating if the standards are enough to cope modelling real learning environments supporting relevant recommendations.

The work in progress is done in the context of a PhD thesis research with aDeNu group. This is implemented in the Instituto Tecnológico de Costa Rica (TEC). The implementation is based on dotLRN platform, instantiated by TEC Digital [22]. This research is aimed to provide a model containing learner information to be used in adaptive and recommendation processes, based on interaction indicators computed from large scale setting which corresponds to official courses run in TEC.

5 Acknowledgements

Authors would like to thank the Spanish Ministry of Economy and Competence (MINECO) for funding BIG-AFF project (TIN2014-59641-C2-2-P), where this research is partially supported. Authors would also like to thank the Department of "Orientación y Psicología" (DOP) at TEC, specially to Alejandra Alfaro.

References

1. A. Kobsa, "User modeling: Recent work, prospects and hazards," *Hum. Factors Inf. Technol.*, vol. 10, pp. 111–111, 1993.
2. P. Brusilovsky and E. Millán, "User Models for Adaptive Hypermedia and Adaptive Educational Systems," in *The Adaptive Web*, Berlin, Heidelberg: Springer-Verlag, 2007, pp. 3–53.
3. S. Bull and J. Kay, "Metacognition and open learner models," in *The 3rd Workshop on Meta-Cognition and Self-Regulated Learning in Educational Technologies, at ITS2008*, 2008, pp. 7–20.
4. S. Sosnovsky and D. Dicheva, "Ontological technologies for user modelling," *Int. J. Metadata Semant. Ontol.*, vol. 5, no. 1, pp. 32–71, 2010.
5. S. Bull and J. Kay, "Open learner models," in *Advances in Intelligent Tutoring Systems*, Springer, 2010, pp. 301–322.
6. C. Conati and H. Maclaren, "Empirically building and evaluating a probabilistic model of user affect," *User Model. User-Adapt. Interact.*, vol. 19, no. 3, pp. 267–303, 2009.
7. S. Baldiris, O. C. Santos, C. Barrera, J. Boticario, J. Velez, and R. Fabregat, "Integration of Educational Specifications and Standards to Support Adaptive

Learning Scenarios in ADAPTAPlan.," *Int. J. Comput. Appl.*, vol. 5, no. 1, pp. 88–107, 2008.

8. O. C. Santos and J. G. Boticario, "Requirements for Semantic Educational Recommender Systems in Formal E-Learning Scenarios," *Algorithms*, vol. 4, no. 2, p. 154, 2011.
9. J. Boticario, A. Rodriguez-Ascaso, O. C. Santos, E. Raffenne, L. Montandon, D. Roldán, and F. Buendía, "Accessible Lifelong Learning at Higher Education: Outcomes and Lessons Learned at two Different Pilot Sites in the EU4ALL Project.," *J UCS*, vol. 18, no. 1, pp. 62–85, 2012.
10. "IMS GLC: Learner Information Package Specification." [Online]. Available: http://www.imsglobal.org/profiles/index.html. [Accessed: 14-Apr-2015].
11. "IMS GLC: RDCEO Specification." [Online]. Available: http://www.imsglobal.org/competencies/. [Accessed: 17-Apr-2015].
13. "Accreditation Resources | Engineers Canada." [Online]. Available: https://www.engineerscanada.ca/accreditation-resources. [Accessed: 07-Apr-2015].
14. "Computing Curricula 2005: The Overview Report." [Online]. Available: http://www.acm.org/education/education/curric_vols/CC2005-March06Final.pdf. [Accessed: 28-Jul-2014].
15. Y. Wang and A. Kobsa, "A PLA-based privacy-enhancing user modeling framework and its evaluation," *User Model. User-Adapt. Interact.*, vol. 23, no. 1, pp. 41–82, Mar. 2013.
16. I. Gámez, C. Garita, and M. Chacón-Rivas, "Generación de Sugerencias de Rutas de Aprendizaje Adaptativas en Entornos de e-learning," presented at the Conferencia Latinoamericana en Informática - CLEI 2012, Medellín, Colombia, 2012, pp. 1–10.
17. O. C. Santos and J. G. Boticario, "Involving Users to Improve the Collaborative Logical Framework," *Sci. World J.*, vol. 2014, pp. 1–15, 2014.
18. M. Chacón-Rivas, O. C. Santos, and J. G. Boticario, "Collaborative Logical Framework adapted to instructors and learners," in *Artificial Intelligence in Education-Interactive Events*, Madrid, Spain, 2015.
19. R. M. Felder and B. A. Soloman, "Learning styles and strategies," *N. C. State Univ. Httpwww Ncsu Edufelder-PublicILSdirstyles Htm*, 2000.
20. E. Raffenne, "MIRLO: una ontología para dar soporte a un modelo de estudiante abierto," UNED, Madrid, Spain, 2013.
21. A. Alfaro, "Demandas académicas y afrontamiento en estudiantes con adecuaciones curriculares | Alfaro Barquero | Actualidades en Psicología." [Online]. Available: http://www.revistas.ucr.ac.cr/index.php/actualidades/article/view/36. [Accessed: 04-May-2015].
22. M. Chacon-Rivas and C. Garita, "A Successful OSS Adaptation and Integration in an e-Learning Platform: TEC Digital," in *Open Source Software: Mobile Open Source Technologies*, Springer, 2014, pp. 143–146.