Using Ontologies for eLearning Personalization

Paulo Gomes, Bruno Antunes, Luis Rodrigues
AILab, Centro de Informática e Sistemas da Universidade de Coimbra
DEI – Pólo II
Universidade de Coimbra, Portugal
pgomes@dei.uc.pt

Arnaldo Santos, Jacinto Barbeira, Rafael Carvalho
PT Inovação SA
Aveiro, Portugal
arnaldo@ptinovacao.pt

ABSTRACT

Learning is a cognitive activity that differs from student to student. Most of the eLearning systems do not take into account individual aspects of students, ignoring the different needs that are specific to existing cognitive profiles. In this paper, we present an approach to eLearning personalization based on an ontology. We developed a student model that is integrated with an ontology, enabling the personalization system to guide the student's learning process. The developed model monitors the student's progress so that it can update the concepts known by the student and decides which concepts s/he should learn next.

Keywords

Ontologies, eLearning, Personalization, Student Modeling, SCORM, LMS.

1. INTRODUCTION

Personalization is the next step in the evolution of eLearning systems. Students can have several cognitive styles [1], which makes the efficiency and efficacy of an eLearning system different with distinct students. In this paper we address the problem of personalizing an eLearning system. We have developed an approach based on a student model and an ontology. The system responds differently according to the student characteristics and performance, depending also on the concepts that the student knows. Another important aspect of our approach is the use of SCORM (Sharable Content Object Reference Model, [2]) as a guideline format for content development, and to implement our student model.

Our project is named as PERSONA, and it is developed in collaboration between PT Inovação and the Artificial Intelligence Lab of Coimbra University. PT Inovação (PT In) is the Portuguese company responsible for the technological learning in the Portuguese telecommunications group – Portugal Telecom – and has developed the eLearning platform – FORMARE. As a company that follows innovative product development, PT In is pioneer in the area of eLearning, for which project PERSONA is a clear bet in the development of new features for its platform FORMARE.

The development team of FORMARE follows a pedagogical methodology that is in constant evolution, and that descends from the distant learning theory developed by Moore [3], using Keegan characterization [4].

SCORM [2] comprises several technical specifications and guidelines for developing learning objects. It was created by the Advanced Distributed Learning (ADL) initiative to meet the Department of Defense (DoD) learning needs in terms of web-based learning contents. SCORM is also an attempt to unify the interests and goals of different groups and organizations that work in the eLearning area.

Figure 1 shows the different SCORM components. The specifications that make part of SCORM are organized in two major groups: the Content Aggregation Model, and the Run-time Environment. The Content Aggregation Model provides the specifications for content development, according to the main creation guidelines for learning objects: accessibility, interoperability, reusability and durability. The Run-time Environment defines the mechanisms for establishing the communication between the Learning Management System (LMS) and the learning objects. SCORM is a leading standard for eLearning content development, and a valuable asset for any LMS that wants to win customers.

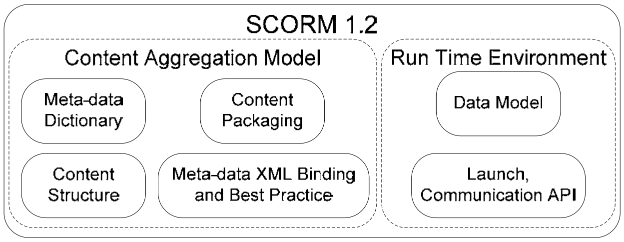

Figure 1. The components of SCORM 1.2.

PERSONA integrates personalization with the standard SCORM 1.2 in FORMARE. In the remaining of this paper, we will present our approach. Section 3 describes the student model and the personalization models developed for PERSONA. Section 4 shows how the ontology is integrated in our system and how it is used. Then section 5 illustrates our approach with an example. Section 6 presents the related work, and we conclude with some final remarks and the future work.

2. OUR APPROACH

The approach that we present in this paper is based in a student model developed from the literature of different areas: authoring systems, user modeling, adaptive web-based educational systems, adaptive educational hypermedia, adaptive web-based tutoring system, intelligent tutoring systems, and semantic web for user modeling. The student model defines what can be known about the student by FORMARE. This model is built incrementally by the system, and it uses several data sources: from student enrollment form, data from the student-system interaction, teacher, or from the system administrator. We also define which parts of the student model are available for inspection from the point of view of the different system actors: students, teachers and content developers. The next subsection describes in greater detail the student model.

3. STUDENT MODEL

The student model comprises two kinds of data: static data, which are not altered during the student-system interaction; and dynamic data, that changes according to the student learning progress and with the system interaction. The static data is referred to as the static model, and the other kind of data as the dynamic model.

3.1 Static Model

The static model comprises five different parts: personal, personality, cognitive, pedagogical and preference data. Each one is an aggregation of student characteristics, which are not usually changed during an eLearning session.

Personal data comprises the biographical information about the student, and can be easily obtained from the course enrolment form. This information is:

- Student name;
- Special accessibility needs to course materials that the student must have;
- Affiliation;
- Student's professional activities;
- List of degrees and qualifications;
- Information of student security and access credentials.

The personality data models the student characteristics that represent the type of person the student is. These characteristics can be inferred from personality tests as the Myers-Briggs test [5]. The attributes of personality data are:

- Personality type;
- Concentration skills, based on the average time spent in the learning contents;
- Collaborative work skills based on the participation in group works;
- Relational skills based in the interactions with students and teacher.

The cognitive data models the student characteristics that represent the type of cognition the student possesses. These characteristics can be inferred from cognitive tests as the Ross and Witkin tests [1]. They try to formalize the type of information processing and reasoning the student uses. These are properties used in the user modelling area, so that contents can be tailored to the student needs. This information is:

- Cognitive style;
- Level of experience the student possesses in using the eLearning system;
- Student experience in using computers.

The pedagogical data defines the student characteristics that deal directly with the learning activity. This data intends to model the student's behaviour in learning situations, comprising two strictly personal properties:

- Learning style;
- Learning approach.

The pedagogical data includes three more operational properties:

- Course objectives: list of concepts that the student must learn in the session course;
- Course evaluation: defines if the student is taking an evaluated course or not;
- Course navigational control: defines what type of control is being used in content navigation.

The preference data stores a set of student preferences regarding the system customization. Most of the preferences are gathered from the student, but some of them are defined by the system administrator. The attributes of preference data are:

- Preferred presentation format;
- Preferred language for content display;
- Web-design personalization;
- Command personalization;
- Personal notebook;
- Sound volume;
- Video speed;
- Subtitles.

3.2 Dynamic Model

The dynamic model comprises two sets of data: the performance data, and the student knowledge data. The performance data (level of motivation and confidence for learning, ability to formalize and comprehend course concepts, global performance level shown by the student in the current course, level of effort spent by the student in the course, and the portfolio that stores all the results obtained by the student in the current course) gathers information about the student's current performance in the course sessions. Data is constantly being gathered in order to keep an updated data model. This data is obtained from the student-system interaction.

The student knowledge data (domain ontology comprising all the concepts referenced in the course, message information that enables to infer the student's active collaboration, and the student's progress regarding the knowledge concepts and competences referenced by the course) describes the knowledge concepts and competences relevant for the current course that the student possesses and must possess until the end of the course. This set of data also gathers information about the student's progress during the course sessions. As the previous data set, all of this data is gathered from the student-system interaction.

The student model is the basis for the personalization models developed in PERSONA. There are two personalization models developed, which are described in the next subsection.

3.3 Personalization Architecture

Two personalization models are proposed for using the student model: on-line personalization and off-line personalization.

The on-line personalization (see Figure 2) monitors the student-system interaction continuously, in real-time, trying to adapt the contents and navigation paths according to the student model. The system uses a reasoning engine to decide the adaptations to be made. These adaptations are then sent to the adaptation mechanisms to be executed.

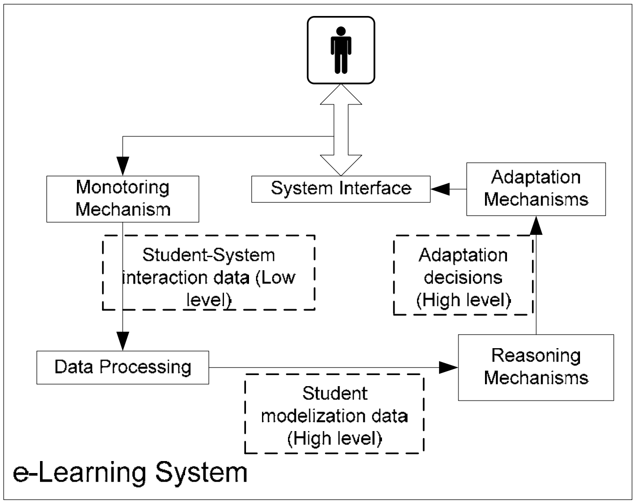

Figure 2. The on-line personalization model.

The off-line personalization (see Figure 3) gathers student-system data, and then analyzes this data to recommend content developers changes in the course contents. This analysis is performed using data mining tools, resulting in changes that are suggested using an authoring tool.

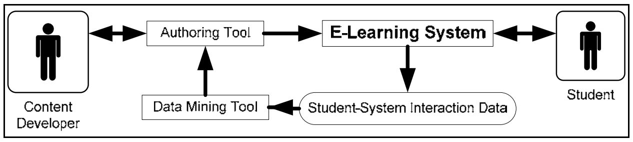

Figure 3. The off-line personalization model.

The implementation of PERSONA was made through a DLL library coded in C# that comprises the specified student model, enabling the FORMARE platform to store and access the student model. The library enables the LMS platform to store the student data, which can then be analyzed generating different values for student attributes. These attributes enable the LMS to get those values, and use them.

3.4 Teacher and Developer Templates

One of the important aspects in any eLearning platform is the correct development of contents. One factor that influences the development of effective contents is the communication between the content providers (usually the Teachers) and the content development team. Miscommunication of ideas between these two system actors can result in poorly developed contents, which can jeopardize the eLearning course success. This communication has been enhanced through the use of templates based on the student model, which enable an easy and more precise communication between the Teacher and the rest of the development team.

These templates have also taken into account the SCORM standard, associating to each student model attribute a set of SCORM data model variables. There are two correspondent templates for each attribute data set, one for the Teacher another for the development team. Table 1 presents the Teacher's template for the student's performance data; notice that the template's first column is used for the teacher to select which attributes s/he wants implemented.

Table 1. The Teacher's template for the student's performance data.

	Characteristic	Description
☑	Motivation / self esteem	Student's level of motivation and confidence for learning.
☒	Cognitive development	Student ability to formalize and comprehend course concepts.
☑	Performance	Global performance level shown by the student in the current course.
☑	Study effort	Level of effort spent by the student in the course.
☒	Portfolio	Stores all the results obtained by the student in the current course.

The correspondent template for Table 1 is presented in Table 2. Only three fields are presented, since in this example only these three were selected.

Table 2. The develoment team template for the student's performance data, according to template in Table 1.

Characteristic	Data Source	SCORM variables to be used
Motivation / self esteem	FORMARE	cmi.objectives.n.score.raw; cmi.objectives.n.score.max; cmi.objectives.n.score.min; cmi.objectives.n.status; cmi.student_data.mastery_score; cmi.core.score.raw; cmi.core.score.max; cmi.core.score.min; cmi.interactions.n.result
Performance	FORMARE	cmi.objectives.n.score.raw; cmi.objectives.n.score.max; cmi.objectives.n.score.min; cmi.objectives.n.status; cmi.student_data.mastery_score; cmi.core.score.raw; cmi.core.score.max; cmi.core.score.min; cmi.interactions.n.result; cmi.interactions.n.latency
Study Effort	FORMARE	cmi.core.session_time

4. ONTOLOGY USAGE

The student dynamic model makes reference to course concepts, which are then used for making decisions about what contents should be shown to the student. These concepts are organized in an ontology [6], which represents the domain knowledge. Basically the ontology is a formalization of the domain concepts, in which concepts are represented by classes. There can be relations between classes and class attributes. In our model we use classes and generalization relations between classes, forming a taxonomic structure.

The course itself, as defined in SCORM standard, comprises several modules implemented as SCO's (Sharable Content Objects), which can have a set of associated objectives and interactions. The objectives represent goals that a student must achieve for a specific module. The interactions represent points of interaction between the student and the system, normally used to attest what have been learned by the student.

Concepts are associated to course modules, objectives and interactions. These associations are traced during sessions to determine the student's knowledge about the related concepts. This enables the system to compute the learning progress of a student, and to known which concepts are well understood and learned by her/him.

Each concept present in the ontology has four associated attributes: correct answers, wrong answers, completed SCO's and uncompleted SCO's. These attributes derive from interactions, objectives and SCO's data, and are explained below:

- **Correct Answers (CA):** Contains the number of correct answers associated with the concept, obtained by adding the number of passed objectives and interactions in the course, which are associated with this concept.
- **Wrong Answers (WA):** Contains the number of wrong answers associated with the concept, obtained by adding the number of objectives and interactions not passed in the course, which are associated with this concept.
- **Completed SCO's (CSCO):** Contains the number of passed SCO's associated with the concept.
- **Uncompleted SCO's (USCO):** Contains the number of SCO's not passed associated with the concept.

Beside these four attributes, each concept in the ontology has an associated state, which can take four values: known, well learned, learned and unknown. This state reflects the values of the four parameters referred initially, as explained below:

- **Known:** The student already knows the concept, based on the historical information about the student;
- **Well Learned:** The student performed tests that assure his understanding of the concept. To consider that the student learned well the concept we verify if there are correct answers or wrong answers associated with the concept, and if the percentage of correct answers is more than 50%, as described in the formula below:

$$Well\ Learned \Leftarrow (CA > 0 \vee WA > 0) \wedge \frac{CA}{CA + WA} > 0{,}5 \qquad \textbf{(1)}$$

- **Learned:** The student went through contents that represent the concept. To consider that the student learned the concept we verify if there is completed or uncompleted SCO's associated with the concept and if the percentage of passed SCO's is more than 50%, as described in the formula below:

$$Learned \Leftarrow (CSCO > 0 \vee USCO > 0) \wedge \frac{CSCO}{CSCO + USCO} > 0{,}5 \qquad \textbf{(2)}$$

- **Unknown:** The concept doesn't fit in any of the previous conditions.

With the student progress data being updated in the domain ontology, we can take a global overview about his knowledge on the specified area at any point in time.

5. EXAMPLE

To exemplify our approach, we will use an example course entitled "Ethernet". We defined an ontology (see Figure 4 for an example of part of the ontology), that represent the course domain. Associated with each course's interaction, objective and SCO, there is a set of concepts in the ontology.

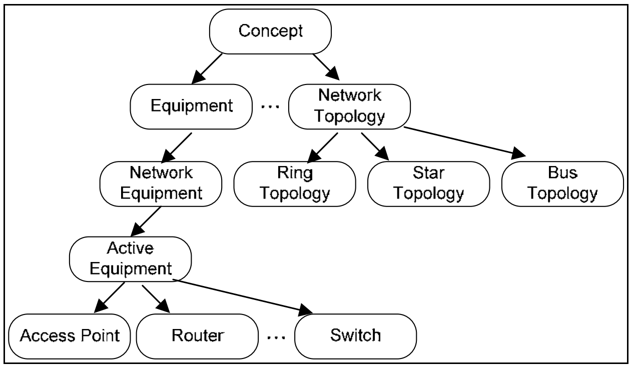

Figure 4. Part of the ontology that represent our example course domain.

For example purposes, let's imagine that our student completed the modules defined in Table 3. As we can see, there are several concepts associated to each module. Once the student completed these modules, the number of completed SCO's of each concept, associated to these modules, increases.

Table 3. Some modules, and associated concepts, from our example course.

Module (SCO)	Associated Concepts	SCO State
LAN Components	Ring Topology; Bus Topology; Star Topology; UTP Cable; …	Completed
IEEE 802.3 Standard	Ethernet; Active Equipment; Hub; Switch; Router; …	Completed

After completing some modules, a short test is presented to the student to attest what was learned. Table 4 shows some questions, and associated concepts, detached from that test. The system evaluated the student answers as shown in the table. For each concept associated with the questions, the student answer will change the number of correct or wrong answers, whether he answered correctly or not.

Table 4. Some questions, and associated concepts, detached from the test on our example course.

Question	Associated Concepts	Answer given by the student
Is a machine dependent from sharing the physic connection with other machines in a star topology?	Star Topology	✓
Does intermediate equipment analyze the application content?	Active Equipment	✓
Does a Switch increase the speed of a network?	Switch	✗
Is the isolation of traffic inside a network the principal function of a Hub?	Hub; Switch	✓

Table 5 summarizes the data used to compute the sate of each concept referred previously. For each concept, the system stored the four parameters and applied the defined formulas to calculate the concept's state.

Table 5. Data used to compute the state of referred concepts.

Concept	Parameters	Applied Formula	State
Star Topology	CA = 1	**(1)**	Well Learned
	WA = 0		
	CSCO = 1		
	USCO = 0		
Active Equipment	CA = 1	**(1)**	Well Learned
	WA = 0		
	CSCO = 1		
	USCO = 0		
Switch	CA = 1	**(2)**	Learned
	WA = 1		
	CSCO = 1		
	USCO = 0		
Hub	CA = 1	**(1)**	Well Learned
	WA = 0		
	CSCO = 1		
	USCO = 0		

As we can see in Table 5, the "Switch" concept has one correct answer and one wrong answer, so the percentage of correct answers is no more than 50% and the system evaluated the formula, setting the well learned state as false. Because this concept was associated with one passed SCO, the system then evaluated the formula, setting the learned state as true. Finally, it sets this concept as learned.

Through this sequence example, the system can perceived that a student answered badly to a question related with a concept that he/she should understand, because he completed an SCO associated with this concept. Based on this information, the system can suggest the student to make a revision on the module associated with the referred concept, in order to consolidate his knowledge on the subject.

Based in this tracking process, the system can evaluate the student's progress during his interaction with the system, in order to adapt the course sequence to reflect the specific needs and characteristics of the student.

6. RELATED WORK

This work relates to several different research areas, from which we have made an extensive research work in order to develop our approach. There are two basic ideas that we use as basis of our literature search: the student as a system user, and the approaches that consider the user modeling for web systems. So, we have investigated works in the following areas: authoring systems; user modeling; web adaptive systems for eLearning; adaptive hypermedia; intelligent tutoring systems; and ontologies for eLearning systems.

In the area of authoring systems, project BiTE [7] aims to develop eLearning lessons, and it identifies two main types of eLearning: individual eLearning, focusing the personal aspect of learning; and group eLearning, that deals with collaborative learning. This project also defines several guidelines for developing eLearning systems. Power et. al. [8] describe the MOT system, which is an authoring system for developing eLearning contents in a personalized way. The other advantage of this system is that it uses the SCORM standard. Another important authoring system is REDEEM [9], which categorizes students in several different classifications. The contents can then be personalized according to the student category. Webster [5] analyzes the student reflective abilities in her/his initiative and autonomy in learning. This paper mentions also several cognitive styles and how they relate to learning.

Baldoni [10] describes how SCORM contents can be adapted using metadata, and how it can be used with the semantic web. The Personal Reader [11] is also another important work in the eLearning field, it also uses the semantic web to personalize and enrich eLearning contents. In the same research line, Dolog [12] argues that the semantic web can be important for user modeling, since it enables a more richer knowledge about the user or, in the case of eLearning, the student. Further works on personalization using the semantic web, are presented in [13, 14]. In his work, Kay [15] stresses that the student model should be available to the student, so that s/he can be aware of the learning goals, and how they are being achieved. Kobsa [16] and Benyon [17] present works on user modelling and some user model implementations.

Another important work is the one presented by Triantafillou [18], which uses the cognitive style of a student as the basis for eLearning personalization. Based on these cognitive styles, he develops a student model. Masthoff [19] presents a system capable of generating a study plan depending on the student and her/his progress along the learning sessions. Romero [20] presents an approach for the personalization of web-based educational hypermedia systems, based on genetic algorithms. Other important works on adaptive hypermedia are presented by Brusilovsky [21, 22], which defines a taxonomy for hypermedia adaptation and an architecture for an adaptive hypermedia system. Conlan [23] also presents a good state of the art on adaptive hypermedia, where he summarizes several important aspects of this type of systems.

Our approach reuses several ideas from other research works and integrates them in a coherent model, trying to cover a wide spectrum of personalizing eLearning systems.

7. CONCLUSIONS AND FUTURE WORK

This paper presents an approach to eLearning personalization based on an ontology and a student model. We have presented the detailed student model and how it was implemented. Another important aspect of our work is the use of an ontology to map the student knowledge to course concepts, so that we can better access her/his progress and to adapt contents and navigation structure to a particular student. Future work involves the experimentation of our system in the FORMARE platform with real courses.

8. ACKNOWLEDGMENTS

This work was funded by PT Inovação under the program IDEIA 2005.

9. REFERENCES

1. Souto, M.A.M., et al., *Towards an Adaptive Web Training Environment Based on Cognitive Style of Learning: an Empirical Approach*. 2003.
2. ADL, *The SCORM Overview*, P. Dodds, Editor. 2001, ADL.
3. Moore, M., *Theory of Distance Education*, in *Distance Education*. 1991.
4. Keegan, D., *Foundations of Distance Education*. 3rd ed. 1996, London: Routledge.
5. Webster, R., *Metacognition and the Autonomous Learner: Student Reflections on Cognitive Profiles and Learning Environment Development*, in *ICED 2004*. 2004.
6. Fensel, D., *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*. 2004: Springer.
7. Ravotto, P., *Developing e-Learning Lessons*, in *Bridging the Gap: from the face-to-face to e-learning environment*. 2003: Ipswich, UK.
8. Power, G., et al., *Goal Oriented Personalisation with SCORM*, in *5th IEEE International Conference on Advanced Learning Technologies (ICALT 2005)*. 2005.
9. Ainsworth, S., *REDEEM:ITS Authoring Tools and Human Teaching*, in *ITS2000*. 2000.
10. Baldoni, M., et al., *Reasoning about learning object metadata for adapting SCORM courseware*. 2004.
11. Dolog, P., et al., *The Personal Reader: Personalizing and Enriching Learning Resources using Semantic Web Technologies*, in *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*. 2004.
12. Dolog, P. and W. Nejdl, *Challenges and Benefits of the Semantic Web for User Modelling*, in *12th World Wide Web Conference*. 2003.
13. Henze, N., *Personalization Functionality for the Semantic Web: Identification and Description of Techniques*. 2004.
14. Baldoni, M., C. Baroglio, and N. Henze, *Personalization for the Semantic Web*, in *Summer School Reasoning Web 2005*. 2005.
15. Kay, J., *Learner Control*, in *User Modeling and User-Adapted Interaction*. 2001, Kluwer Academic Publisher.
16. Kobsa, A., *Generic User Modeling Systems*, in *User Modeling and User-Adapted Interaction*. 2001, Kluwer Academic Publishers.
17. Benyon, D. and D. Murray, *Applying user modelling to human-computer interaction design.* 1993.
18. Triantafillou, E., P. A, and Georgiadou, *AES-CS: Adaptive Educational System base on cognitive styles*, in *AH2002 Workshop*. 2002.
19. Masthoff, J., *Automatic Generation of a Navigation Structure for Adaptive Web-Based Instruction*, in *AH'2002 Workshop on Adaptive Systems for Web-based Education*. 2002.
20. Romero, C., et al., *Using Genetic Algorithms for Data Mining in Web-based Educational Hypermedia Systems*, in *AH'2002 Workshop on Adaptive Systems for Web-based Education*. 2002.
21. Brusilovsky, P., *Adaptive Hypermedia*, in *User Modeling and User-Adapted Interaction*. 2001, Kluwer Academic Publishers.
22. Brusilovsky, P., C. Karagiannidis, and D. Sampson, *Layered evaluation of adaptive learning systems.* International Journal of Continuing Engineering Education and Lifelong Learning, 2004.
23. Conlan, O., *State of the Art: Adaptive Hypermedia*. 2003.