# On the Application of the Semantic Web Concepts to Adaptive E-learning

Juan M. Santos, Luis Anido, Martín Llamas, and Judith S. Rodríguez

Departamento de Enxeñería Telemática, Universidade de Vigo
E.T.S.E. Telecomunicacións, Vigo, Spain
{jsgago, lanido, martin, jestevez}@det.uvigo.es

**Abstract.** Adaptive learning means offering customized educational environments for each individual. For this to be possible in virtual environments, like the Internet, actual standardized data models for e-learning must be enriched with explicit "machine-processable" semantic information and appropriate inference rules. This would allow Learning Management Systems (LMS) to automatically process and integrate these models meaningfully in order to offer alternative behaviours depending on each learner and his particular needs. This paper analyses the impact of the upcoming Semantic Web, an extension of the current Web in which information is given an explicit meaning, on the consecution of real high-quality adaptive e-learning environments.

## 1 Introduction

In the last years more and more institutions have adopted the Information and Communication technologies, particularly the Internet, for delivering education and training. During these few years, Internet-based e-learning systems have progressively evolved from the very basic repositories of simple learning documents accessible via Web to the advanced learning environments that include sophisticated tools like pedagogical simulators or complex communication facilities, enabling different educational approaches and methods, like "learning by doing" [1] or "collaborative learning". Adaptive learning seems to be the next step in this evolving process.

Future e-learning systems should provide each apprentice with learning experiences that are unique and tailored to his needs, interests, preferences and learning style in order to maximize the effectiveness of learning. For this personalized learning to be possible a set of complex data structures are needed to express alternative behaviours for the Learning Management System (LMS) depending on each learner and his particular needs. Currently, an active learning technology standardization process is dealing with the lack of interoperability among heterogeneous systems. Nevertheless, expressing complex behaviour (e.g. data needed to offer adaptive learning) is incompatible with standards established by consensus.

The organization of the paper is as follows. Section 2 gives an overview of the adaptive learning requirements and briefly describes some systems that deal with

M.H. Shafazand and A M. Tjoa (Eds.): EurAsia-ICT 2002, LNCS 2510, pp. 536–543, 2002.

these requirements. Section 3 introduces the learning standardization process, illustrating some meaningful proposed models. Some intermediate conclusions are explained in section 4. Section 5 deals with the concepts involved in the Semantic Web. In section 6 it is showed how those concepts can be applied in Learning Managing Systems to provide better personalization of the learning experiences of students. Section 7 concludes and summarizes the paper.

## 2 Adaptive Learning

The personalised attention from teachers to their students is, too often an impossible to reach goal in the classic educational environments, where a teacher has to convey his/her classes to groups, habitually very populated, of students. However, in the modern Internet-based electronic learning environments that homogeneity among students is not so feasible. The access to the flexible "virtual classrooms" can be done from anywhere in any time and by heterogeneous students. Due to this heterogeneity, traditional teaching methods, which have been systematically reflected in the e-learning environments, miss their efficiency and they have difficulties to adapt to these new settings.

Thus, the future of e-learning is in the adaptive e-learning systems. These systems are characterized by offering educational experiences dynamically customized to the real and particular needs of each student at any instant and therefore maximizing the effectiveness of learning. This new generation of e-learning systems will be possible with the introduction of the inferential technologies, which will allow the development of intelligent systems that model the behaviour and profile of the student and personalize the educational session according principally to his learning style and his background knowledge in the subject of study. Other parameters can be considered too to obtain a fully comfortable environment, but they are not exclusive of the learning domain. Hence, it is desirable to get:

– Personalization in content presentation or rendering according to the personal taste of the individual.
– Adaptability in the presentation of contents according to the characteristics of the visualization device. This kind of personalization is mainly emerging because of the proliferation of handheld computers having small displays and reduced computational capacity.
– Accessibility for people with disabilities. For example, making text fonts bigger for short-sighted people.

The e-learning systems that overpower the current market don't present these possibilities of personalization or else they tend to present them in a very limited way. However, several platforms, in an experimental state, have been developed by the research community. Here, some of them are briefly described:

– IDEAL [2]: IDEAL is a prototype of an intelligent agent assisted environment for active learning. In the system, students' learning-related profiles, such as

learning styles and background knowledge, are used for selecting, organizing and presenting the learning materials to individual students and in supporting active learning. The student models are inferred from the performance data using a Bayesian belief network.
- SAC [3]: SAC is a self-paced and adaptive courseware system developed at the Hong Kong Polytechnic University. This system provides dynamic navigational guidance to students taking online courses. Specialized data models not compatible with those presented in the next section are used to perform the adaptation.
- MATS [4]: The Multi-Agent Tutoring System (MATS) is an agent based educational platform that models a "one student - many teachers" learning situation. Each MATS agent represents a tutor, capable of teaching a distinct subject. All MATS tutors are also capable of collaborating with each other for solving learning difficulties that their students may have.

## 3 The E-learning Standardization Process

Recently, several organizations and institutions (e.g. IMS, ADL, AICC, IEEE's LTSC, CEN/ISSS/LT) have been working towards the development of standards and recommendations aimed to solve the interoperability problems currently found in the e-learning domain. The broad adoption of these proposals by the e-learning platform manufacturers will ensure, in the near future, the development of educational resources that can be exploited by a wide variety of compliant platforms. This feature will significantly reduce the time and the costs related to the development of on-line courses. Some of the most outstanding fields of standardization are showed in the next subsections.

### 3.1 Educational Metadata

Metadata is one of the most prolific fields in the e-learning technology standardization process. Generally speaking, metadata can be defined as "data about data". The educational metadata provide descriptions and additional information about learning resources. This information can be used not only for characterizing the resources but also for searching, cataloguing and use improvement.

One of the main contributors in this field has been the IEEE's LTSC with the *Learning Object Metadata* (LOM) specification [5]. LOM defines a total amount of 60 elements grouped into nine meaningful categories that try to reflect all aspects that must be considered in a pedagogical environment.

### 3.2 Learning Resources Organization and Packaging

The need to exchange educational resources between e-learning platforms and authoring tools has caused the development of content packaging formats and procedures. In this way, the definition of a single entity (e.g. a file) that encapsulate the educational content together with its organization and its related metadata will make easier the course transfer among different systems.

The most outstanding recommendation in this field is that proposed by the ADL initiative: the *SCORM Content Aggregation Model (CAM)* [6]. CAM is an extension of the IMS *Content Packaging specification* [7] including some results from the *AICC CMI Guidelines for Interoperability* [8]. A CAM *Package* is composed of two elements: the *Manifest*, an XML file that describes the encapsulated resources and their organization, and the *Resources Collection*, i.e. the physical files that store the resources. The *Organization* element of the *Manifest* defines the static relations among the resources of the aggregation (often a whole course) in a hierarchical tree of *items*. The *Prerequisites* subelement of each *item* defines what other parts of the content aggregation must have been completed before starting the *item*. The LMS is responsible for interpreting these prerequisites and controlling the actual sequencing of the learning resources at run-time.

### 3.3 Student Profiles

As it happens in conventional educational environments, e-learning platforms must handle information about students. The proposals in this field define information models whose aim is to provide not only a structured way to represent this information but a standardized way to interchange it among different systems. The first recommendation made in this field is the one from the IEEE's LTSC, the *Public and Private Information for Learners* (PAPI Learner) [9]. PAPI includes a subset of useful elements organized into five categories: *Personal*, *Preference*, *Performance*, *Portfolio*, *Relations* and *Security*. The IMS consortium has made its own proposal based on the PAPI work. The IMS recommendation, *Learner Information Package* (LIP) [10], identifies eleven categories to describe student-related information: *Identification*, *Securitykey*, *Transcript*, *Goal*, *Qcl*, *Activitity*, *Interest*, *Competency*, *Relationship*, *Affiliation* and *Accessibility*.

## 4 Interim Conclusions

First proposals from the e-learning standardization process are currently in scene and some manufacturers have begun to adopt them. These proposals sound useful in order to solve the interoperability problem existent in the e-learning domain. However, they present some troubles for using them in the next generation of adaptive e-learning systems:

- They are not flexible enough. Particularly, the proposed educational content organizations are too rigid, being difficult the delivery of alternative contents to the student depending of his particular needs.
- They are hardly scalable. Changes on them, for example to include a new element in an educational metadata specification, may imply important modifications at software level in the LMS. It is due to the lack of explicit machine-processable semantic in the recommendations.
- Relations among the different proposals are very poor. For example, it is not easy to match the *Preferences* category from the PAPI specification to

the LOM metadata elements. It makes difficult searching the most appropriate content for a particular leaner. So, such relations should be made more explicit.

## 5 Semantic Web

Currently, the web is a large and decentralized source of information designed primarily for human consumption. The coming Semantic Web [11] is an extension of the current web in which information in a machine-processable form can coexist and complement the existing human-readable information, better enabling computers and people to work in cooperation.

For this emerging Web to function, facilities to put machine understandable data must be developed. Ontologies [12] figure prominently (c.f. Figure 1) as a way of representing the semantics of documents and enabling that semantics to be used by web applications and intelligent agents. An ontology defines the terms used to describe and represent an area of knowledge (like medicine, tool manufacturing, automobile repair, financial management, etc.), including computer-usable definitions of basic concepts in the domain and the relationship among them. They encode both knowledge in a domain and also knowledge that spans domains. In this way, they make that knowledge reusable.

It must be clear that ontologies declare not just hierarchical categories or taxonomies but they may also include other logical rules. Logical rules in ontologies can be used by inference engines to derive new knowledge from current data. For the sake of illustration below we include a very simple example:

A particular ontology about Natural Sciences can state:

```
CARNIVORE "is subclass of" ANIMAL; FOX "is subclass of" CARNIVORE
HERBIVORE "is subclass of" ANIMAL; RABBIT "is subclass of" HERBIVORE
```

A property for the class `CARNIVORE`, defined as a logical rule, can be:

```
CARNIVORE "eat" ANIMAL
```

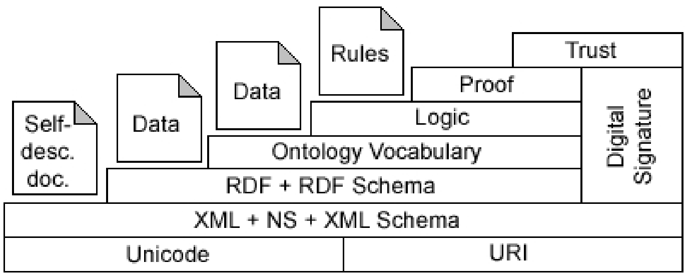


**Fig. 1.** The Semantic Web layers

From these declarations, an inference engine can reason that a fox eat rabbits (if no other rule contradicts this deduction).

In order for ontologies to fulfill their role in the semantic integration of the Web, there will need to be some standardization of Web ontology languages. The W3C is already moving in this direction with languages such as RDF and RDFS. However, in order to achieve the widest possible acceptability, these languages have deliberately been kept very simple and have relatively weak semantics. Much richer ontology specification languages [13] will be needed in order to exploit the full potential of the Semantic Web.

## 6 Semantic Web Concepts in E-learning

The LTSA [14] states that, from a conceptual viewpoint, every Learning Management System is composed by the elements showed in Figure 2. Briefly, the overall operation has the following form [14]: (1) the learning styles, strategies, methods, etc., are negotiated among the learner and other stakeholders and are communicated as learning preferences; (2) the learner is observed and evaluated in the context of multimedia interactions; (3) the evaluation produces assessments and/or learner information; (4) the learner information is stored in the learner history database; (5) the coach reviews the learner's assessment and learner information, such as preferences, past performance history, and, possibly, future learning objectives; (6) the coach searches the learning resources, via query and catalog info, for appropriate learning content; (7) the coach extracts the locators from the available catalog info and passes the locators to the delivery process, e.g., a lesson plan; and (8) the delivery process extracts the learning content from the learning resources, based on locators, and transforms the learning content to an interactive multimedia presentation to the learner.

The Coach is the process that incorporates information from several sources to search and select learning content for appropriate learning experiences. In essence, it is an inference engine, i.e. an automatic control mechanism that applies the axiomatic knowledge present in the knowledge base to determine: the adequate learning contents to be delivered to the learner.

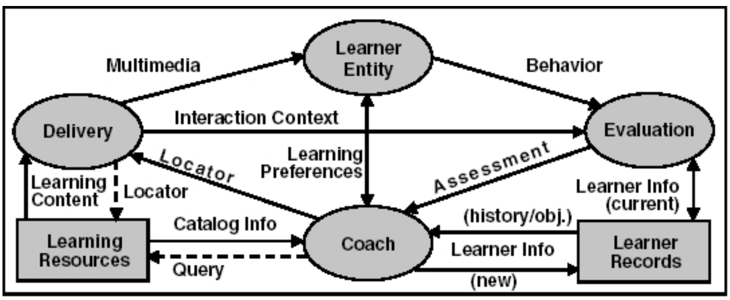


**Fig. 2.** The LTSA architecture components

The Coach work is usually facilitated by the definition of a specific organization for the resources each course is compliant with (c.f. Section 3.2). So, the searching of the next resource to be delivered to the learner in a concrete learning session consists just in finding the next item in hierarchical tree containing the course organization. One of properties of this *item* immediately specifies the resource to deliver. After that, the Coach only has to check that the learner fulfills the prerequisites established for that *item*. This checking is based on the data stored in the learner profile. However, as it was stated in Section 4, this approach is very restrictive because it does not allow the choice among different resources depending on the learner characteristics, knowledge and preferences.

A more open and flexible approach consists in not linking each *item* with a particular resource. On the contrary, each *item* should explicitly define the characteristics of the resource to be delivered in a machine-readable way. In this sense, inference rules constructed with the terms and vocabularies from one or more ontologies become the most useful tool. Such rules will allow the Coach to automatically carry out the search of the suitable resources from the LMS repository and then to select the one that best fits to the learner needs. This approach also permits changing already existing resources and even adding new ones without updating the course structure.

The selection process also needs explicit inference rules to meaningfully carry out the automatic matching between the learner profile data (c.f. Section 3.3) and the resource descriptions (c.f. Section 3.1). The definition of these logic rules in an ontology-based format would avoid the implementation of complex and subjective (i.e. conceived by every implementor) software algorithms in the Coach that would have to be modified on changes in the underlying data models.

Obviously, for this approach to be possible it is necessary to properly annotate the resources in a machine-readable way. The metadata schemes presented in Section 3.1 must include a vocabulary with a huge amount of terms to reduce ambiguity. Such a complex vocabulary would be expensive to construct and difficult to maintain as an only and isolated entity. A most practical approach is to define several small specific ontologies and to establish the possible relations among the identified terms on them through logic rules. Those specialized ontologies developed by different institutions and developers would be populated in one or several repositories accessible to the whole Internet Learning Community. This is precisely the basis of the emerging Semantic Web.

## 7 Conclusions

The notion of the Semantic Web, as thought by Tim Berners-Lee, inventor of the classic Web, is to transform the current World Wide Web so that the information can be understood and used both by computers and humans. The Semantic Web will provide an environment where software agents can perform advanced tasks on behalf of humans. This notion seems very interesting in the e-learning domain. As the Web is in essence a large and decentralized knowledge base, building that knowledge in a "machine-readable" format would permit the development of specialized agents (tutor agents) that asynchronously explore, gather, extract, combine and transform information from several sources in order to present adequate and personalized contents depending on each individual particular learning needs.

Perhaps, this will be possible in a distant future. However, at present, the concepts involved in the Semantic Web can be used to improve the current e-learning systems. In this paper some restrictions of the proposed recommendations from the learning standardization process were identified to deal with adaptive e-learning. Such restrictions would be relaxed by the introduction of distributed ontologies and appropriate inference rules allowing the meaningful integration of different models.

**Acknowledgments.** We want to thank "Xunta de Galicia" and "Ministerio de Ciencia y Tecnología" for their partial support to this work under grants "Arquitecturas distribuidas para Teleservicios" (PGIDT00TIC32203PR) and "CORBALearn: Interfaz de Dominio guiada por Estándares para Aprendizaje Electrónico" (TIC2001-3767).

## References

1. Anido, L., Llamas, M., Fernández, M.J.: Internet-based Learning by Doing. IEEE Transactions on Education, vol. 44, n. 2 (2001).
2. Shang, Y., Hongchi, S., Su-Shing, C.: An Intelligent Distributed Environment for Active Learning. Proc. of 10th Int. WWW Conference, Hong Kong, May 1–5 (2001).
3. Chan, A., Chan, S., Cao, J.: SAC: A Self-paced and Adaptive Courseware System. Proc. of IEEE Int. Conf. on Advanced Learning Technologies, August (2001).
4. Solomos, K., Avouris, N.: Learning From Multiple Collaborative Intelligent Tutors: An Agent Based Approach. Journal of Interactive Learning Research, vol. 10, n. 3.4. (1999).
5. Hodging, W.: Draft Standard for Learning Object Metadata V. 6.4. IEEE LTSC Technical Report, March (2002).
6. Dods, P.: SCORM Content Aggregation Model. ADL Tech. Report, October (2001).
7. Anderson, T., McKell, M.: IMS Content Packaging Information Model V. 1.1.2. IMS Technical Report, August (2001).
8. Hyde, J.: AICC/CMI Guidelines for Interoperability V. 3.5. AICC Technical Report, April (2001).
9. Collet, M., Farance, F.:Draft Standard for Learning Technology - Public and Private Information (PAPI) for Learners (PAPI Learner) V. 7. IEEE's LTSC Technical Report, November (2000).
10. Smythe, C., Tansey, F., Robson, R.: IMS Learner Information Packaging Information Model Specification. IMS Technical Report, March (2001).
11. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. In: Scientific American, May Issue (2001).
12. W3C Web Ontology Working Group at http://www.w3.org/2001/sw/WebOnt
13. Bechhofer, S.: Ontology Language Standardisation Efforts. OntoWeb Rep (2002).
14. Farance, F., Tonkel, J.: Learning Technology Systems Architecture (LTSA). IEEE LTSC Technical Report, November (2001).