
Juan M. Santos, Luis Anido and Martín Llamas
University of Vigo, E.T.S.E. de Telecomunicación
E-36310, Vigo (Pontevedra), Spain  Juan.Santos@det.uvigo.es

Abstract - The Semantic Web 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 co-operation. In this way, some processes that need to be carried out manually today will be performed automatically. E-learning is one of the domains that may benefit from this technology enhancement. We present some of the outcomes from the first author’s PhD, which contributes to the e-learning domain through the proposal of a Reference Service Architecture for web-based brokerage using Semantic Web technologies. In particular, this paper focuses on the underlying ontology that is required for that, introducing several sub-ontologies about, for instance, courses and learning objects, on-line service providers, content providers, learners, etc.

Index Terms – E-learning Brokerage, Ontologies, Open Architectures, Standardization.

INTRODUCTION

Today’s technology enhanced learning landscape is characterized by a high and growing number of heterogeneous educational service providers. For a user with a particular educational need, a typical scenario involves the user visiting one or several online educational centers, browsing their offers, collecting information about the courses (study programs, requirements, needed tools, prices, etc.), selecting the most appropriate course for his/her needs and preferences and, finally, registering it. This manual browsing is too time-consuming and, typically, a user will visit just a very few online centers before making a decision. The existence of specialized e-learning brokers or intermediaries, which gather and integrate the existing educational offers, can improve this situation.

We are working on the development of an architecture for an innovative brokerage system in the e-learning domain that, bringing together the last standards and recommendations defined in the Learning Technologies (LT) standardization process and making use of the new techniques related to the emerging Semantic Web, provides high level services to people looking for appropriate online courses.

The proposed architecture extends the previous works of the authors in the field [1][2] with semantic and inference practices. Among the most important features of our proposal, it can be outlined the next ones:

• The use of the latest results from the LT standardization process, like metadata models for describing learning resources, schemas for storing learner information or data models for describing accessibility issues.
• The use of contextual information (like enrolment dates, description of the institutions that provide the educational services, description of the online platforms and tools used to deliver the courses, etc.) to parameterize the queries.
• The improvement of the traditional search mechanisms with the incorporation of techniques from the Semantic Web area.
• The conformance with common protocols, practices and interoperability mechanisms.

This paper focuses on the description of the development of the supporting ontology, which is the core base of the broker architecture. The organization of the paper is as follows: Section 2 gives an overview of the concepts involved in the Semantic Web and how these concepts can be applied in the e-learning brokerage field. The conceptual framework of the semantic brokerage architecture is outlined in Section 3. In Section 4 ELEARNING-ONT, an integrative ontology for the e-learning domain is briefly described. Finally, Section 5 concludes and summarizes the paper.

USE OF THE SEMANTIC WEB TECHNIQUES IN THE E-LEARNING BROKERAGE

Internet has experienced a great evolution since its birth. Some of the most important changes are the accessibility increase to the general public, the growth of the contents quantity that is available to Internet users or the number of on-line transactions that are performed. Internet has become part of the daily life of a vast number of people over the world. This increase of popularity makes Internet a powerful communication tool; however the growth of the amount of available information makes it difficult for users to find the resources they are looking for.
Due to that problem, some solutions have been implemented that try to make it easier browsing Internet. These solutions are search engines and some examples could be Google or Yahoo!. It is clear that the Web would not have been the huge success it was, were it no for search engines. However, there are serious problems related to their use (identified in [3]):

- It may happen that a search engine does not get any answer for a particular request (this is no frequent, but it is possible, in the most popular search engines) or results from the query do not include relevant pages (a more frequent situation).
- It often happens that a lot of mildly or lowly relevant results are returned from a particular request. Even if the main relevant results are retrieved, they could be almost useless if many others mildly relevant or irrelevant documents were also retrieved.
- Very commonly the initial keywords queried do not get the desired results because the relevant documents use different terminology from the original query. This is unsatisfactory because semantically similar queries should return similar results. Results are highly sensitive to vocabulary.
- When the needed information is spread over several web resources, it is necessary to do several queries to gather all the relevant documents, and then the partial information must be manually extracted and aggregated.

The main obstacle to providing better support to Web users is that, at present, the meaning of Web content is not machine-accessible. Although keyword-based techniques have had an outright success, results from these engines can be improved by conveniently describing the resources in the search space in a machine-readable way.

The coming Semantic Web [4] 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 the Semantic Web to function, facilities to put machine-understandable data on the web must be developed. Ontologies [5] figure prominently in the emerging Semantic Web as a way of representing the semantics of documents and enabling that semantics to be used by web applications and intelligent software 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-readable definitions of basic concepts in the domain and the relationship among them. They encode knowledge in a domain and also knowledge that spans domains. In this way, they make that knowledge reusable.

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 (Resource Description Framework) and RDFS (Resource Description Framework Schema). 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 are needed in order support the design, sharing and integration of the complex ontologies. Although during the last years various languages for the definition of ontologies have been proposed [6], on February 2004 the W3C has standardized OWL [7] as the language for the semantic description of Web resources.

For the construction of a semantic LT Broker, we need the definition of a specific ontology that includes all the relevant terms required to describe all the involved entities (courses, providers, clients, e-learning platforms, etc.). ELEARNING-ONT, an OWL integrative ontology will be described in Section 3. Data collected from Educational Services Providers and supplied by Clients will be conveniently transformed by the Broker into OWL statements that make use of the terms identified in ELEARNING-ONT. This transformation process, as well as other deductive tasks (e.g. triggering of notification messages), is supported by inference rules, it is, axioms that specify an action if certain conditions are met. Currently, several rule languages coexist, but, for our prototypes, we use Notation 3 –or N3– [8] because of its simplicity: N3 is a simplified ontology language with support for rules and queries.

Information stored in the Broker, as OWL statements, can be queried through a semantic search engine. Semantic Web query languages, like RDQL [9], can be used to formulate complex queries in order to retrieve exactly the information in which we are interested. Subsequently a RDQL query is shown that would permit to locate the “courses” of “medium difficulty” on “literature” that are written in “Castilian”.

SELECT ?a, WHERE {?a, <rdfs:type>, learn-ont:Course}, {?a, <lom-edu:difficulty>, lom-edu:MediumDifficulty}), {?a, <dc:subject>, "LITERATURE"}, {?a, <lom-edu:language>, learn-ont-idiom:Castilian})

It must be noted that courses declared to be written in “Spanish” will be also returned, because ELEARNING-ONT defines that “learn-ont-idiom:Spanish” and “learn-ont-idiom:Castilian” are equivalent.

THE FRAMEWORK

In the LT brokerage context, Brokers collect the descriptions of the courses from the academic institutions, as well as the referring contextual information in which they are delivered (data about the institution, delivering tools, dates for enrolment/delivery, etc.). These institutions or Educational Services Providers (ESPs) should be registered in a Broker and provide, among other data, the profile of the institution and the mechanism for accessing the catalogues of the products they offer. The customers of a Broker can access to the different services offered by the intermediation system

Session S3H
making use of different devices, such as personal or portable computers, pocket computers or advanced mobile telephones.

Figure 1 shows the functional elements of a scalable and adaptable Semantic E-learning Brokerage architecture. It makes use of particular ontologies (described in Section 3) and inference rules that can be refined without structural changes in the infrastructure as new statements are identified. Next, the most important elements in the architecture are briefly described (for a deeper description see [10]):

- **Knowledge Base**: This is the basic and core element of the intermediation system. Here, it is available all the information collected and inferred by the Broker, both from the ESP and from the different types of Clients. It is a repository where Ontologies, Inference Rules, Educational Resources and Course Descriptions, Service Provider Profiles, User Profiles and E-learning Platform Descriptions are stored.

- **Search Engine**: It is the software component that provides an API with methods for querying the Knowledge Base. Although there are many ontology query languages, currently RDQL [9] is the most used until a recommended language will be issued by the W3C.

- **Inference Engine**: This component is responsible for inferring new facts (new OWL statements) from a set of previous OWL facts taking into account additional information defined by a particular ontology and in a set of inference rules (or axioms).

- **Data Collector**: It is the component that gathers information from the registered ESPs automatically.

- **Services**: Different services are offered by the described infrastructure. Some of them are Anonymous Searches, Personalized Searches, Notification Service, Course Annotation, Relevance Estimation Service, Taxonomy Management and Supporting Services.

- **Access Interfaces**: Different interfaces are provided to the clients in order to support different access devices (PCs, PALMs, Pocket PCs, WAP devices, etc.). Administrators of academic institutions are provided with particular access interfaces to register and modify their data. Likewise, in Figure 1 it is shown an access entry point for software agents. This Agent Interface consists of a set of Web Services [11] that conforms to a standardized access model [12].

**THE ONTOLOGY**

Several organizations and institutions (e.g. IMS, ADL, AICC, IEEE’s LTSC, CEN/ISSS/LT, ARIADNE, etc.) have been working towards the development of standards and recommendations aimed at solving the interoperability problems currently found in the e-learning domain [13]. The result of this effort is a basic set of information models that allow the representation of several entities involved in the e-learning area. Several of these standards, like the metadata models for describing learning resources, formats for defining competencies, schemas for representing learner information, data models for describing accessibility issues, are the basis for the definition of ELEARNING-ONT, a set of interconnected OWL ontologies that facilitate the automatic management of the implicit semantic present in the instances of the standardized data models.

ELEARNING-ONT, which is briefly depicted in this Section, includes the definition of the concepts, and their interrelations, necessary to develop brokerage services in the e-learning domain. Due to the great quantity of identified terms, the ontology is organized in a range of namespaces (or sub-ontologies). There exists a basic namespace, where fundamental concepts such as “Educational Resource”, “Course” or “Educational Services Provider” are defined. A series of sub-ontologies include the properties, with their corresponding vocabularies, that can be used to describe in detail the instances of the most basic classes.

**Methodology**

In order to identify the most suitable terms to be included in a draft proposal of a domain OWL Ontology for educational brokerage, we defined a systematic methodology. This methodology is based on the guidelines proposed by Noy and McGuinness in [14] and the recommendations described in the Unified Software Development Process [15].

The first stage of the development process involves the capture and documentation of the most basic functional requirements from clients' viewpoint. Starting from a set of core requirements, we successively redefine the “Search for Course in Broker” use case (c.f. Figure 2) in order to capture new and different query possibilities. For each stage we apply the steps proposed by Noy and McGuinness:
Identification of the aim and the scope of the ontology.
Consider to reuse existing vocabularies (in our case, we make use of the elements defined on the data models identified by the LT standardization process).
Enumerating the most important terms in the ontology.
Defining the classes and their hierarchy.
Defining the properties of the classes.
Defining the features of the properties.
Creating instances.

In this way, the development of the ontology is an iterative process, centred on the architecture and driven by use cases, where each stage refines the previous one. As the use cases mature and are refined and specified in more detail, more of the ontology terms are discovered. This, in turn, can lead to new use cases. Therefore, both the ontology and the use cases mature together.

Users Ontology
This sub-ontology includes the properties and classes directly related to the characterization of the users of the brokerage system. The terms identified in this namespace have been mainly extracted from the data models Learning Information Package [16] and Accessibility [17], both developed by the IMS Consortium. The first of these standardized models identifies the necessary elements to describe the characteristics of a (potential) student, whereas the second one extends the previous model with elements that allow us to specify certain user preferences. The existence of the user’s ontology in ELEARNING-ONT makes it possible to the intermediation system the accomplishment of searches adapted to the user needs and preferences in order to obtain more relevant results for the client.

Courses and Educational Resources Ontology
Metadata is one of the most prolific fields in the LT standardisation process. Almost all the institutions and organizations involved in this process have made their own proposals in this field. Currently, the Learning Object Metadata [18] model, developed jointly by several of the institutions involved in this process, is already an official standard of the IEEE. This standard, and in particular its RDF binding, developed by Nilsson et al. [19], has been used as the basis for the sub-ontology of ELEARNING-ONT that includes the needed classes and properties to characterize academic courses.

This ontology is composed of 10 namespaces that group classes and properties related to a particular feature of the courses: lom-base (which includes general classes used in the other namespaces), lom-general (with properties that allow specifying the aggregation level and the type of structure of the educational resources), lom-lifecycle (with classes and properties for the description of the resource life cycle), lom-metametadata (that contains classes and properties to describe the metadata scheme used), lom-technical (which references the technical requirements for the execution of the educational resource), lom-educational (this is the most important namespace. It defines classes and properties which describe pedagogical aspects of the resources), lom-rights (referring basically to the costs and legal restrictions of the educational resource), lom-relation (it contains a only property, isBasisFor), lom-annotation (that will be used add comment about courses) and finally lom-classification (whose main target is to show the classification system in which the resource can be categorized).

Educational Services Providers Ontology
This sub-ontology gathers some terms that allow making descriptions about educational services providers. These are entities or organizations that deliver courses on-line throughout a particular e-learning platform. Due to the lack of standardized conceptual models in the e-learning domain related to this topic, we have taken from the e-commerce domain common use schemes that allow describing enterprises. Particularly, our sub-ontology is based on the Enterprise Ontology [20], developed by the Artificial Intelligence Applications Institute from the University of Edinburgh.

Educational Platforms Ontology
On-line courses are offered to students throughout e-learning platforms. We assume that an e-learning platform is a Web application that includes Internet tools and services within an enclosed space specifically configured and organized to provide learning in a convenient and satisfactory way.
Many educational platform surveys have been used to elaborate the sub-ontology that allows the characterization of these applications and the terms considered to be more convenient have been taken from them. The experience of the authors related to the construction of e-learning platforms has been essential in this field. Mostly, the terms in this sub-ontology allow defining the available tools in a platform.

Other Ontologies and Taxonomies

Besides the mentioned sub-ontologies, some other vocabularies and taxonomies have been used. Among them, we can mention a subset of the Universal Decimal Classification [21] scheme, to use it as a vocabulary of several of the properties defined in ELEARNIG-ONT. The DAML-Time [22] ontology has also been imported to represent temporal concepts (for example, course calendars). Several other data models are currently under study, like ontologies that let us to describe user’s devices.

SUMMARY AND FUTURE WORKS

The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. In this paper, we have outlined the basic issues of an intermediation system in the e-learning domain that makes use of the Semantic Web techniques in order to improve the searching and location processes. This Broker can perform queries in its knowledge base taking into account the user profile, i.e. the needs of the user and its preferences regarding time availability, difficulty level, obtained degree, etc.

This paper have briefly introduced the underlying ontology that is required for that, introducing several sub-ontologies about, for instance, courses and learning objects, on-line service providers, content providers, learners, etc. This ontology, named ELEARNING-ONT (c.f. Figure 3), provides the needed semantics to let computers automatically deal with adaptive intermediation in the e-learning domain.

Some work is taking place and will be developed in the future: Finalization of the prototypes of the architecture.
presented, final specification and validation of the support ontologies and the countersign of the proposal in some forums.

Acknowledgment

We want to thank “Ministerio de Educación y Ciencia” for its partial support to this work under grant “MetaLearn: methodologies, architectures and languages for E-learning adaptive services” (TIN2004-08367-C02-01). We also want to thank the European Commission, Directorate-General for Education and Culture (DG EAC) for its support through the Learning Interoperability Frameworks for Europe (LIFE) project.

References