KNOWLEDGE-BASED E-LEARNING ON THE SEMANTIC WEB

Stefan Trausan-Matu, Valentin Cristea, Octavian Udrea

University Politehnica of Bucharest
Computer Science and Engineering Department

Abstract: The SINTEC environment for learning through the web is presented. Intelligent, knowledge-based tutoring facilities are provided, together to more classical learning management and colaborative learning facilities. The usage of ontologies and metadata, that use standard annotations languages based on XML and RDF enables the extension of the facilities of SINTEC for the future semantic web.

Keywords: e-learning, semantic web, ontologies, knowledge-based processing, intelligent tutoring systems.

1. INTRODUCTION

The paper presents how a learning environment, ("SINTEC", developed at National Centre for Information Technology of the Bucharest "Politehnica" University, under the national program "INFOSOC") may be developed according to the semantic web ideas and how he can, as a consequence, offer, knowledge-based intelligent e-learning services.

The semantic web is the next generation of the WWW, in which knowledge, organized in ontologies, will be used for giving to information in web pages "well-defined meaning, better enabling computers and people to work in cooperation" (Berners-Lee et al., 2001). The semantic web enables intelligent e-learning systems, that dynamically build and monitor knowledge (ontology)-based learner models that typically contain lists of known or unknown concepts. Such models can be further used to adapt the instruction strategies (sequences of learning objects) to learner characteristics. In addition, they track students' activities and interactions with the learning material, analyses their answers and texts they write, identify needs or interests and evaluate their psychological profile and learning style. Socio-emotional intelligence issues should also be considered for tailoring the learning process.

The SINTEC environment includes a collection of tools and repositories that integrates collaborative techniques on the web with knowledge-based methods, agents, and multiple purpose XML-based annotation (metadata, exchange of reusable components, knowledge representation) that empowers personalization.

Knowledge management in the environment uses general ontologies (reusable on the semantic web) combined with ontologies specific to the domain of interest of the trainees. It is able to integrate newly extracted knowledge by text and data mining methods. Graphical knowledge visualization and editing is also provided. Specific ontologies can be developed considering also the existing databases. Intelligent, personalized retrieval of latest information is provided.

A group of tools permit the automatic generation of systematic collections of web pages (Trausan-Matu et al., 2002). These may be also used in connection

with the intelligent tutoring system or for knowledge navigation on the knowledge server

The next section introduces the basic concepts of the semantic web. Ontologies, one of the main ingredients of the semantic web, are presented in the third section. The following section discusses the problematics of metadata and content creation, in relation to e-learning on the semantic web. Section 5 analyses the specific aspects of personalized learning on the web. In addition to the references to SINTEC, present in almost all sections, in the sixth section are presented some general facts about the environment that constitute the subject of this paper.

2. THE SEMANTIC WEB

The semantic web is the new generation of the World Wide Web (Berners-Lee et al., 2001). In addition to the facilities offered by the web, that was created for the inter-human communication, the semantic web is designed also for the usage of autonomous computer programs. Such programs search information on the web, extract needed data, build new documents and manage the available resources. What is very important is that they may communicate without human intervention, in a network of web services.

The "semantic" attribute is emphasizing that the new generation of tools must enable the explicit representation of concepts from the web documents, what meaning, what knowledge is inside the documents. This facilitates resource search, composition and restructuring. The first generation of the web did not provided much in this idea. In HTML, for example, there are only rudimentary semantic annotations (the "meta" tags). Metadata, that means data about data, and the possibility of defining new, semanticaly rich annotations (that was impossible in HTML) is, therefore, one of the most important issues for the semantic web.

One of the main ideas in introducing XML ("eXtensible Markup Language", see http://www.w3.org/xml), is exactly the possibility of having and extensible, emanticaly rich annotation languages. For example, for expressing the semantic content of web documents, an XML-based language was defined: RDF ("Resource Description Framework", see http://www.w3.org/rdf), in which assertions about the meaning of web resources may be stated. XML and RDF are the substrate for complex declarative knowledge representation in ontologies, enabling the usage of powerful artificial intelligence techniques.

3. ONTOLOGIES

Ontologies are one of the most important ingredient of the semantic web. The word "ontology" was first used in philosophy to denote the theory about what is considered to exist. Any system in philosophy starts from an ontology, that means from the identification of the concepts and relations considered as fundamental. An ontology may not be explicit but is always implicit in thinking (Sowa, 1999)

In artificial intelligence, the same word is now often used for declarative knowledge bases in a given domain: "An ontology is a specification of a conceptualization....That is, an ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents" (Gruber).

For example, a part of the algorithms domain ontology may contain the following fragment of concept taxonomy:

Algorithm

  Graph algorithm

  . . .

  Divide&Impera algorithm

    Quick-sort algorithm

    Merge-sort algorithm

    . . .

  Greedy algorithm

    Dijkstra algorithm

    . . .

Each concept has attributes. For example, Dijkstra algorithm may have the following attributes:

- data structures,
- applying constraints,
- complexity.

Each concept may be related with other concepts. For example, the Dijkstra algorithm is related to Graph algorithm.

Viewing knowledge bases as ontologies determines important advantages for developers of knowledge-based systems. First of all, an ontology is developed as a coherent framework for the reality and therefore it facilitates knowledge acquisition and machine learning. A new concept is easy to add in such a framework by finding one or some more general concepts and defining some differences between the new concept and the more general ones.

Specialized editors and other interfaces are provided in SINTEC for the management of

relevant ontologies for the learning domains. Standard ontology and knowledge representation are used (e.g. OIL, see http://www.ontoknowledge.org/oil/).

4. CONTENT CREATION AND METADATA

Content creation and reuse is one of the main problems for e-learning systems. Professor find difficult to develop learning modules according to e-learning standards. Particulary problems are the structuring of texts in conceptual units and defining metadata. The semantic web, through ontologies and metadata offers strong handles to help content authors in these tasks. Related to the semantic web, another important idea is to have programs that could automaticaly find and reuse learning resources from the web.

In SINTEC, content creation is facilitated by semantics-oriented graphical editors that enable the authoring of learning components annotated according to metadata standards IMS, IEEE-LTSC, ARIADNE, and Dublin Core. There is also the possibility of reusing learning components from the web and integrate them with learning components that are stored locally in information repositories. In SINTEC, content creation and management is directly related to the semantic web through ontologies.

The SINTEC platform is currently applying these principles, therefore allowing exchange of the following types of information with other similar applications:

- Exchange of user profile and background information, including estimated preparation level and full training history is accomplished through the implementation of the IMS Learner Profile Information (LIP) specification
- Exchange of several types of leaming content (e.g. lecture notes, practical exercises, course support materials) is accomplished through the use of IMS Metadata and Content Packaging specifications
- Exchange of test information, including questions, tests, grading and evaluation information, as well as full result history is accomplished through the use of IMS Question & Test Interoperability (QTI) specification.

The knowledge-based services are also used for the intelligent retrieval of relevant documents on the web and for text mining. For these purposes statistical natural language techniques and the lexical ontology WordNet are used. The information retrieval agent is able to travel in Internet and access multiple information sources for seeking relevant data. In contrast to traditional web search engines, the information agent is capable of a semantic interpretation of the retrieved information, of filtering this information according to user's preferences and criteria, and of a heuristic classification of data based on the user's profile.

The environment is in the process of including also a multi-agent system that will provide another dimension of the collection of personalized services. For example, the search of relevant documents on the web is performed by agents (Trausan-Matu et al., 2002). An automatic tutor agent will be used to guide the student through the on-line course material and to give different hints to questions that are to be answered by the student.

5. KNOWLEDGE-BASED LEARNING PERSONALIZATION ON THE WEB

The use of the semantic web for personalized training in a given domain means, in the approach presented here, the extraction and adaptation of knowledge from the web, for a given learner, in a given context. Knowledge bases (ontologies), documents (texts), and learning objects in standard XML-based languages are the sources for content creation.

For providing knowledge-based, intelligent tutoring, adaptive learning tools for the development of student models are needed. These tools must track student's activity and interactions with the learning material, must analyse his answers and texts he writes, must identify needs or interests and evaluate his psychological profile and learning style. One important component of the student model is (what is named as "student model" in intelligent tutoring systems): what knowledge he has, what knowledge he does not have and what knowledge he has wrongly. This facts are derived from answers at questions, from the analysis of essays written by students, from students interactions. The facts may be further used for dynamic web pages and test generation and lesson planning.

The web is a huge, permanently changing hypermedia on Internet, browsable with very simple, direct manipulation interfaces. Its explosive growth in only several years is the best proof of its usefulness. Two of the causes of this phenomenon are, probably, the ease of "publishing", of communicating something through text and/or images on the web. From the other direction of the communication process, it is very easy for everybody to explore the network of web pages. As a consequence, we notice the extremely dynamic character of information nowadays, the availability of the web today having definitely changed the information scenario. The time between the

appearance of new information in some domain and the use of this information by people has extremely shortened comparatively with some years ago. Therefore, information may become obsolete very quickly or be replaced by some other information. A good tutoring system should consider this scenario, and consequently be able to update its information continuously.

The process of extracting and using the most relevant information from the web involves three phases: information acquisition performed by searching the web, knowledge identification by semantic editing and the usage of this knowledge (Trausan-Matu et al., 2002).

The domain ontology plays important roles in each of the three mentioned phases. The keywords used by a web spidering agent for the search of relevant documents are obtained from the domain ontology. The process controlled by this agent searches on the web by means of activating a number of search engines. During this phase, data mining techniques may be applied in order to better select automatically the fit between the requested information and the retrieved one.

6. SINTEC

The architecture of the SINTEC system is ilustrated in figure 1. From the knowledge-based perspective, it comprises three main groups of modules:

- Content creation and management,
- Knowledge server
- Intelligent tutoring

Each of these three groups of modules has connections with the resources on the semantic web (ontologies and learning modules anotated with metadata).

The SINTEC environment is under development at the National Center for Information Technology (University Politehnica of Bucharest). The design aimed at obtaining platform independent components that permits a rapid deployment on different premises. The service has been tested and deployed on two platforms using the following configurations:

- an IBM Netfinity machine running Linux, Websphere Application Server and IBM DB2 UDB
- a Sun Enterprise 10000 machine running Solaris, iPlanet Application Server, Oracle 8i

The application is still under development and thanks to its modularity it can be extended to provide more advanced features. The virtual classrooms environment, the assessment engine, the user security service, the student-tutor communications, and the content authoring tools are currently undergoing their test stages.

The training environment:

- provides a flexible and easy to use environment for both students and tutors
- uses adaptive content based on user preferences and preparation level, both for course and test preparation and analysis
- adapts easily to a specific domain by incorporating an adequate specific ontology
- provides interoperability with other applications conforming to a similar set of standards contain both presentation and content authoring services Flexible, standardized, adapted to enterprise needs and to trainees profiles (including emotional intelligence)

7. CONCLUSIONS

The usage of the semantic web is very well suited for intelligent, knowledge-based e-learning. Knowledge representation and processing is facilitated by the existance of reusable ontologies. Content creation and annotation is also supported by the semantic web. A prototype version of a system (SINTEC) developed according these ideas is in the testing phase.

REFERENCES

Berners-Lee, T., Hendler, J., and **Lassila, O.,** (2001), The Semantic Web, Scientific American, May

Trausan-Matu, St., Maraschi, D. and **Cerri, St.** (2002), *Ontology-Centered Personalized Presentation of Knowledge Extracted From the Web*, in S.Cerri, G.Gouarderes (eds.), Intelligent Tutoring Systems 2002, Springer, Lecture Notes in Computer Science number 2363, pp 259-269.

Gruber, T., *What is an Ontology*, http://www-ksl.stanford.edu/kst/what-is-an-ontology.html

Sowa, J. (1999), *Knowledge Representation: Logical, Philosophical and Computational Foundations*, Brooke Cole Publishing Co., Pacific Grove, CA,.

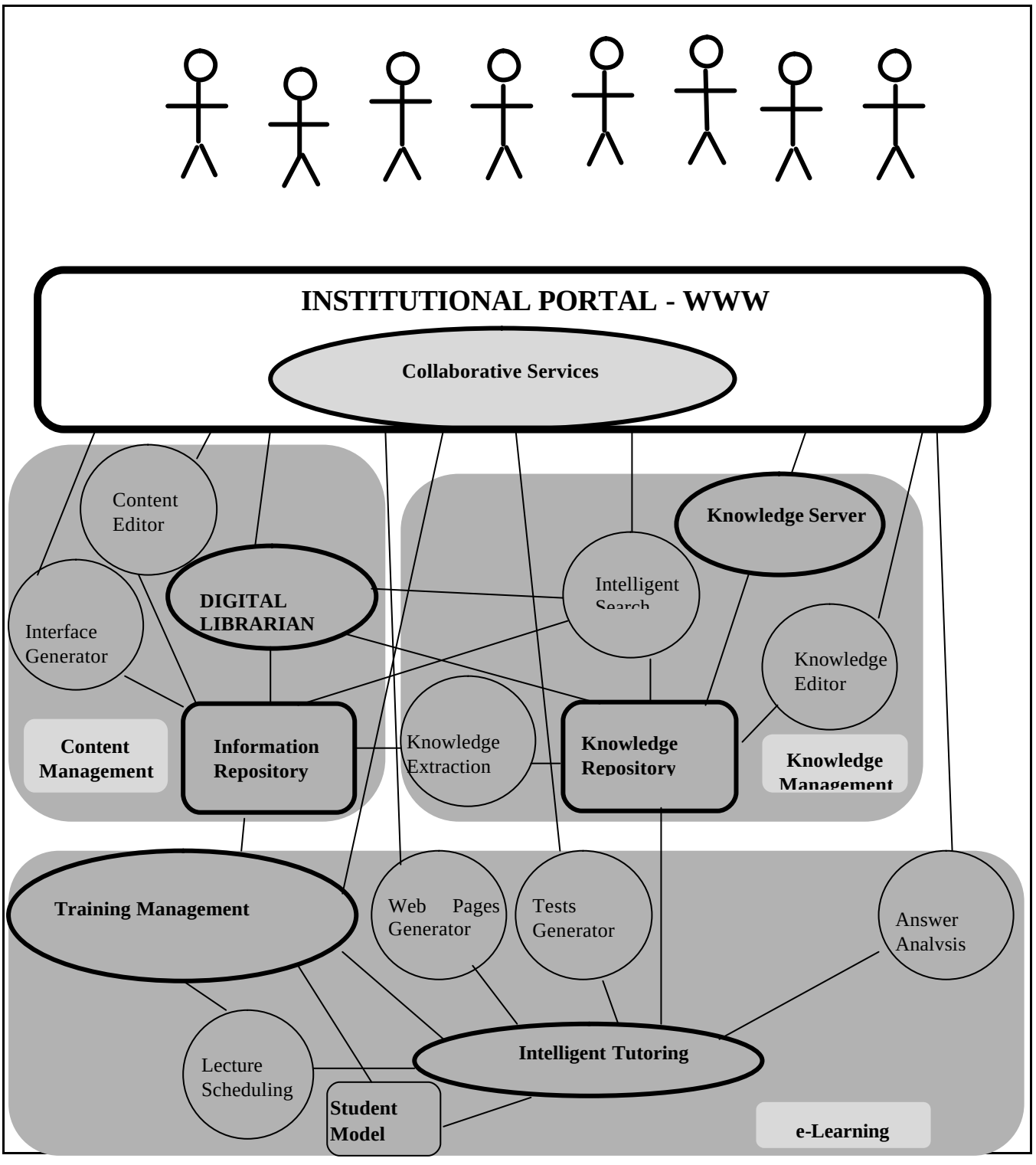

Fig.1 SINTEC Architecture