Chapter XIV
Personalized Information Retrieval in a Semantic-Based Learning Environment

Antonella Carbonaro
University of Bologna, Italy

Rodolfo Ferrini
University of Bologna, Italy

ABSTRACT

Active learning is the ability of learners to carry out learning activities in such a way that they will be able to effectively and efficiently construct knowledge from information sources. Personalized and customiz- able access on digital materials collected from the Web according to one’s own personal requirements and interests is an example of active learning. Moreover, it is also necessary to provide techniques to locate suitable materials. In this chapter, we introduce a personalized learning environment providing intelligent support to achieve the expectations of active learning. The system exploits collaborative and semantic approaches to extract concepts from documents, and maintaining user and resources profiles based on domain ontologies. In such a way, the retrieval phase takes advantage of the common knowledge base used to extract useful knowledge and produces personalized views of the learning system.

INTRODUCTION

Most of the modern applications of computing technology and information systems are concerned with information-rich environments, the modern, open, large-scale environments with autonomous heterogeneous information resources (Huhns & Singh, 1998; Cooley, Mobasher, & Srivastava, 1997). The effective and efficient management of the large amounts and varieties of information they include is the key to the above applications.
The Web inherits most of the typical characteristic of an information-rich environment: information resources can be added or removed in a loosely structured manner, and it lacks global control of the content accuracy of those resources. Furthermore, it includes heterogeneous components with mutual complex interdependencies; it includes not just text and relational data, but varieties of multimedia, forms, and executable code. As a result, old methods for manipulating information sources are no longer efficient or even appropriate. Mechanisms are needed in order to allow efficient querying and retrieving on a great variety of information sources which support structured as well as unstructured information.

In order to foster the development of Web-based information access and management, it is relevant to be able to obtain a user-based view of available information. The exponential increase of the size and the formats of remotely accessible data allows us to find suitable solutions to the problem. Often, information access tools are not able to provide the right answers for a user query, but rather, they provide large supersets thereof (e.g., in Web search engines). The search for documents uses queries containing words or describing concepts that are desired in the returned documents. Most content retrieval methodologies use some type of similarity score to match a query describing the content, and then they present the user with a ranked list of suggestions (Belkin & Croft, 1992). Designing applications for supporting the user in accessing and retrieving Web information sources is one of the current challenges for the artificial intelligence community.

In a distributed learning environment, there is likely to be a large number of educational resources (Web pages, lectures, journal papers, learning objects, etc.) stored in many distributed and differing repositories on the Internet. Without any guidance, students will probably have great difficulty finding the reading material that is relevant for a particular learning task. The metadata descriptions concerning a learning object (LO) representation provide information about properties of the learning objects. However, the sole metadata does not provide qualitative information about different objects nor provide information for customized views. This problem is becoming particularly important in Web-based education where the variety of learners taking the same course is much greater. In contrast, the courses produced using adaptive hypermedia or intelligent tutoring system technologies are able to dynamically select the most relevant learning material from their knowledge bases for each individual student. Nevertheless, generally these systems cannot directly benefit from existing repositories of learning material (Brusilovsky & Nijhavan, 2002).

In educational settings learning objects can be of different kinds, from being files having static content (like HTML, PDF, or PowerPoint presentation format) or in sophisticated interactive format (like HTML pages loaded with JavaScript or Java applet, etc.). Audio files, video clips, or Flash animations could also constitute learning objects. An LO comprises a chunk of content material, which can be re-used or shared in different learning situations. Such a re-use of content from one system to another makes LO standardized so that it can be adopted across different computer platforms and learning systems. The IEEE Standard for Learning Object Metadata (LOM) is the first accredited standard for learning object technology.

Presently there are countless LOs available for commercial and academic use. Because of time and capability constraints, however, it is almost impossible both for a learner and a teacher to go through all available LOs to find the most suitable one. In particular, learning object metadata tags may facilitate rapid updating, searching, and management of content by filtering and selecting only the relevant content for a given purpose (Carbonaro, 2004). Searchers can use a standard set of retrieval techniques to maximize their chances of finding the resources via a search engine (Recker,