Applications

Managing concepts inside e-learning systems

Antonella Carbonaro and Rodolfo Ferrini

Department of Computer Science – University of Bologna

carbonar@csr.unibo.it; ferrini@csr.unibo.it

Abstract

In this paper we introduce an update of our work on developing a distance learning environment based on a collaborative bookmark management system approach. We introduce EasyInfo, an hybrid recommender architecture extension, which processes resources to extract concepts (not just words) from the documents using semantic capabilities. Then, the classification, recommendation and sharing phases take advantage of the word senses to classify, retrieve and suggest documents with high semantic relevance with respect to the student and resource models.

1. Introduction

The Semantic Web will add meaning, or semantics, to Web contents in order to make it easier finding and using information for both humans and machines. Adding formal semantics to the Web will aid in everything from Web searching to resource discovery, till the automation of all sorts of tasks (Koivunen and Miller, 2002). The paper will outline about how the Semantic Web can be used as a fitting technology inside sophisticated distance learning scenarios.

In Bighini and Carbonaro, 2004 we introduced the InLinx (Intelligent Links) system, a Web application that provides an on-line bookmarking service. InLinx is the result of the integration of three filtering components, corresponding to the following functionalities:

1. *bookmark classification* (content-based filtering): the system suggests the more suitable category which the user can save the bookmark in, based on content of the document; the user can accept the suggestion or change the classification, selecting another category that he considers best for the given item;
2. *bookmark sharing* (collaborative filtering): the system checks for newly classified bookmarks and recommends them to other users with similar interests. Recipient users can either accept or reject the recommendation when they receive the notification;
3. *paper recommendation* (content-based recommendation): the system periodically checks if a new issue of some on line journal has been released; then, it recommends the plausible appealing documents, according to the user profiles.

Over the years we have introduced several extensions of the original architecture such as personalized category organization and mobile services (Andronico, Carbonaro, Colazzo and Molinari, 2004). Most recently, we have introduced concepts for classification, recommendation and document sharing to provide a better personalized semantic-based resource management. Generally, recommender systems use keywords to represent both the users and the resources. Another way to handle such data is using hierarchical concept categories. This issue will enable users and the system to search, handle or read only those concepts of interest in a more general manner, providing a semantic possibility. For example, synonymy and hyponymy can reveal hidden similarities, potentially leading to better classify and recommend. We called the extended architecture EasyInfo.

The advantages of a concept-based document and user representation are: (i) ambiguous terms inside a resource are disambiguated, allowing their correct interpretation and consequently a better precision in the user model construction (e.g., if a student is interested in computer science resources, a document containing the word «bank» in the financial context will not be relevant); (ii) synonymous words belonging to the same synset can contribute to the user model definition (for example, both «mouse» and «display» bring evidences for computer science documents, improving the coverage of the document retrieval); (iii) finally, classification, recommendation and sharing phases take advantage of the word senses in order to classify, retrieve and suggest documents with high semantic relevance with respect to the user and resource models.

The Semantic Web appears as a promising technology for implementing distance learning environments. One of its primary characteristics, shared understanding, is based on ontologies as its key element. Ontologies enable the organization of learning materials around small pieces of semantically annotated (enriched) Learning Objects (LOs) (Nejdl, 2001). Items can be easily organized into customized learning courses and delivered on demand to the user, according to her/his profile and business needs.

Filtering and recommending relevant LOs can be useful to address issues like trying to determine the type or the quality of the information suggested from a personalized learning environment. In this context, standard keyword search is of very limited effectiveness. For example, it cannot filter the type of information (tutorial, applet or demo, review questions, etc.), the level of the information (aimed at secondary school students, graduate students, etc.), the prerequisites for understanding the information, or the quality of the information.

Sometimes, the traditional model implemented inside distance learning systems demonstrated some limitations and inconsistencies, assigning fixed roles to subjects participating into educational processes.

These roles are normally included in a vision of the training process that we could define «transfer model»: the teacher owns the knowledge and this is transferred to students via a sequence of lectures. The student learns from references or books while guided by the teacher's lectures. Moreover, if we consider a technical course, where more practical skills must be acquired and demonstrated, very often students have to develop a project. This could be an individual work, but more frequently it is a joint effort among students of a group. It would be more fruitful to let students work together also using cooperative tools that allow them to interact among themselves and with teachers/tutors, but this kind of collaboration «freedom» (with all the administrative problems behind the scenes) is very often extraneous to a typical e-learning system. This latter aspect is particularly relevant in teaching technological disciplines, and specifically, computer science topics.

The architecture of InLinx and of its specialized components presented in the following pages is based on a student-centered approach. We present tools for personalized distribution of educational resources, as important technology to adapt and personalize e-learning system on the basis of the effective user needs and modifiable user behaviour and interests, addressing to the above cited collaboration and active role requirements.

Generally, recommendations are generated by using two main techniques: content-based information filtering, and collaborative filtering (Belkin and Croft, 1992). If, by one hand, a content-based approach allows to define and maintain an accurate user profile, that is particularly valuable when a user encounters new content, on the other hand it has the limitation of concerning only the significant features describing the content of an item. Differently, in a collaborative approach, resources are recommended according to the rating of other users of the system with similar interests. As there is no analysis of the item content, collaborative filtering systems can deal with any kind of item, not just limited to textual content. This way, users can receive items with content that is different from the one received in the past. Since a collaborative system works well if several users evaluate each one of them, new items cannot be recommended until some users have taken the time to evaluate them and new users cannot receive recommendation until the system has acquired some information about the new user in order to make personalized predictions. These limitations often referred to as the sparsity and start-up problems (Melville, Mooney and Nagarajan, 2002). By adopting a hybrid approach, a recommender system is able to effectively filter relevant resources from a wide heterogeneous environment like the Web, taking advantage of common interests of the users and also maintaining the benefits provided by content analysis.

The paper is organized as follows. Section 2 provides a summary of related works. Section 3 introduces how to extract and to use senses from the documents, proposing word sense disambiguation and similarity measure processes. Following, the paper introduces the learning object recommendation process, obtained considering student and learning material profiles and adopting filtering criteria based on the value of selected metadata fields. Some final considerations and comments about future developments conclude the paper.

2. Related works

Several related research projects intend to filter and recommend interesting Web pages to user, but few of them tackle with the problem of collaborative bookmark management.

The most similar approach to our recommendation module are RAAP (Research Assistant Agent Project; Delgado, Ishii and Ura, 1998) and CoWing (Collaborative Web IndexING system; Kanawati and Malek, 2001), even if, none of them introduce the paper recommendation component.

Usually, this idea has been widely and successfully developed for specific domains, like movie or film recommendations, and it's been rarely used for recommending LOs on e-learning environments. Our system uses a hybrid approach and suitable representations of both available information sources and user's interests in order to match user informative needs and available information as accurate as possible.

Current research in the Semantic Web area is very dynamic. The scientific literature and industrial services offer various tools and applications in the Semantic Web scenario that have been developed and are currently used by different communities.

The eXtensible Markup Language (XML)[1] and the Resource Description Framework (RDF)[2] languages are used for adding structure and meaning, along with relationships between Web information sources. Web Ontology Language (OWL)[3] is providing a standard for ontology preparation. Protégé[4] and Ontology Editor (OntoEdit) (Sure, Erdmann, Staab, Studer and Wenke, 2002) are examples of ontology editors; Ontobroker (Fensel, Decker, Erdmann, Studer, 1998) is an ontology server; TextToOnto (Maedche and Staab, 2000) is a tool suite supporting the semi-automatic construction of ontologies by NLP and text mining procedures; Melita (Ciravegna, Dingli, Petrelli and Wilks, 2002) allows the annotation of textual data interactively by users and makes use of a separate training phase to learn annotation rules that are used to make suggestions to the user for subsequent texts. Moreover, some tools are able to cluster or map the inter-relationships among major websites or concepts in a graphic form as a network of nodes of different sizes; see Clusty[5], Kartoo[6] and Mooter[7] as examples. KAON (Karlsrude Ontology)[8] provides a common framework for Semantic Web. SNOBASE[9] is an ontology management environment developed by IBM Alphaworks that provides a mechanism to query ontologies and a programming interface to interact with ontologies, written in RDF Schema and OWL.

The research on e-learning and Web-based educational systems traditionally combines research interests and efforts from various fields, in order to tailor the growing amount of information, to the needs, goals, and tasks of the specific individual users. Semantic Web technologies may achieve improved adaptation and flexibility for users and new methods and types of courseware compliant with the Semantic Web vision. Adaptive Web-based Educational Systems form the basis of the emerging Educational Semantic Web. Among the promising experiences in Educational Semantic Web we would like to mention Denaux et al. (2005) proposing adaptive task recommendations and resource browsing on the Semantic Web, Muna et al. (2005) reporting on the work to develop a framework

1 http://www.w3schools.com/xml/default.asp
2 http://www.w3schools.com/rdf/default.asp
3 http://www.w3.org/2004/OWL
4 http://protege.stanford.edu/plugins/owl/
5 http://www.clusty.com
6 http://www.kartoo.com
7 http://www.mooter.com
8 http://kaon.semanticweb.org/
9 http://alphaWorks.ibm.com/tech/snobase

for Semantic Web mining and exploration, Woukeu et al. (2003) an ontological hypertext framework for building generic web-based educational portals, Porto et al. (2004) presenting ROSA, a data model and a query language to access LOs, based on their semantic properties, Henze et al. (2004) showing how rules can be enabled to reason over distributed information resources in order to dynamically derive hypertext relations.

The following paragraph describes both implemented word sense disambiguation process and the ontology used to retrieve the exact concept definition and to adopt some techniques for semantic similarity evaluation among words.

3. Semantic capabilities

We have improved our previous version of InLinx by introducing the EasyInfo module exploiting semantic abilities. The sense extraction from documents is based on the following tools.

WordNet (Fellbaum, 1998) is an online lexical reference system, in which English nouns, verbs, adjectives and adverbs are organized into synonym sets. Each synset represents one sense, that is one underlying lexical concept. Different relations link the synonym sets, such as IS-A for verbs and nouns, IS-PART-OF for nouns, etc. Verbs and nouns senses are organized in hierarchies forming a «forest» of trees. For each keyword in WordNet, we can have a set of senses and, in the case of nouns and verbs, a generalization path from each sense to the root sense of the hierarchy. WordNet could be used as a useful resource with respect to the semantic tagging process and has so far been used in various applications including Information Retrieval, Word Sense Disambiguation, Text and Document Classification and many others.

GATE (Cunningham, Maynard, Bontcheva, Tablan, 2002) provides a number of useful and easily customizable components, grouped to form the ANNIE (A Nearly-New Information Extraction) component. These components eliminate the need for users to keep re-implementing frequently needed algorithms and provide a good starting point for new applications. These components implement various tasks from tokenization to semantic tagging and co-reference, with an emphasis on efficiency, robustness, and low- overhead portability, rather than full parsing and deep semantic analysis.

The general architecture of the developed semantic module consists of two levels. The first one is a Java level that manages the user interaction, the documents to analyze and GATE functions. The second one is a C level that enquiries WordNet and stores document representation. More detailed, the Java level is able to:

- interact with the users to obtain a new document to process;
- retrieve and analyse the document using specified parameter setting;
- extract document information using a multithread architecture;

- transmit obtained information to C level, using Java Native Interface (JNI) to integrate Java code and native code. We implemented an ad hoc function to interface WordNet libraries and JNI.

We want to find a subset of words related to each document that the system manage, which helps to discriminate between concepts. In such a way, two documents or two users characterized using different keywords may result similar considering underling concept and not the exact terms. Web documents are the collection text written in natural language. To extract important information from documents, we use the following feature extraction pre-process. Firstly, we label occurrences of each word in the document as a part of speech (POS) in grammar. This POS tagger discriminates the POS in grammar of each word in a sentence. After labelling all the words, we select those ones labelled as noun and verbs as our candidates. We then use the stemmer to reduce variants of the same root word to a common concept and filter the stop words.

A vocabulary problem exists when a term is present in several concepts; determining the correct concept for an ambiguous word is difficult, as is deciding the concept of a document containing several ambiguous terms. To handle the word sense disambiguation problem we use similarity measures based on WordNet. Budanitsky and Hirst (Budanitsky Hirst, 2001) give an overview of five measures based on both semantic relatedness and semantic distance considerations, and evaluate their performance using a word association task.

We considered two different similarity measures; the first one is proposed by Resnik (Resnik, 1995) while the second is proposed by Leacock-Chodorow (Leacock and Chodorow, 1998).

The Resnik similarity measure is based on the information content of the least common subsumer (LCS) of concepts A and B. Information content is a measure of the specificity of a concept, and the LCS of concepts A and B is the most specific concept that is an ancestor of both A and B. The Leacock and Chodorow similarity measure is based on path lengths between a pair of concepts. It finds the shortest path between two concepts, and scales that value by the maximum path length found in the is–a hierarchy in which they occur.

We propose a combined approach based on the two described measures considering both a weighted factor of the hierarchy height and a sense offsets factor.

Generally, similarity measures applies to the hyponymy relations (the IS-A or HAS-PART relation in WordNet), that is to the syntactic categories of noun and verb. After the nouns and verbs sense disambiguation, our documents are represented using a matrix storage format containing a row for each WordNet subject code and a column for each WordNet object code. The position i, j of the matrix maintains a list containing the WN verb code that relates the subject i and the object j and the number of times the triple is present in the document. For example:

SUBJi - LIST{VERBi,j#occurrences} - OBJj
After this process documents are represented as vectors of triples: (concept i, list of WN verb codes and their occurrences, concept j).

4. E-learning recommandation

In the context of an e-learning system, additional readings in an area cannot be recommended purely through a content analysis in order to fit learners' interests, but also pick up those pedagogically suitable documents for them. Recker, Walker and Lawless (Recker, Walker, Lawless, 2003) present a web-based system where teachers and learners can submit reviews about the quality of web resources, without considering the pedagogical features of these educational resource in making recommendation. Tang and McCalla (Tang, McCalla, 2005) propose two pedagogical features in recommendation: learner interest and background knowledge. They also study two pedagogy-oriented recommendation techniques, content based and hybrid recommendations, arguing that the second one is more efficient to make «just-in time» recommendations. The following paragraphs describe how we consider the resource content to propose a fitted technique in a hybrid recommendation framework.

The automatic recommendation of relevant learning objects is obtained considering student and learning material profiles and adopting filtering criteria based on the value of selected metadata fields. Our experiments are based on SCORM compliant LOs. For example, we use the student's knowledge of domain concept to avoid recommendation of highly technical papers to a beginner student or popular-magazine articles to a senior graduate student. For each student, the system evaluates and updates his skill and technical expertise levels. The preprocessing component developed to analyze the information maintained in LOs is able to produce a vector representation that can be used by the collaborative recommendation system.

In SCORM, the organization and learning resources must be included with the course and placed in an XML file with the name imsmanifest.xml. The structure required for this file is detailed in the SCORM content aggregation specification.[10] We analyze the imsmanifest.xml file in order to extract .htm and .html files and examine the content, to obtain the loading of some didactical source and its classification. We consider the following metadata to provide the corresponding technical level:

[10] http://www.adlnet.org

- difficulty: represents the complexity of the learning material, ranging from «very easy» to «very difficult»;
- interactivity level: represents the interactive format, ranging from «very low» (only static content) to «very high»;
- intended end user role: represents the user type (for example student or teacher);
- context: represents the instructional level necessary to take up a LO.

The difficulty level is explicitly loaded into our database (in most cases, LMSs use this value). Difficulty and the other above cited values are combined to characterize technical level of learning material ranging from 0 to 1 and representing how demanding is the LO. If some of these fields are not present in the manifest file (they are not required), we consider their average value.

Our system also considers the user's skills to express cleverness as regards different categories. This value ranges from 0 to 1 and it initially depends on the context chosen from the user during his/her registration (primary education, university level, and so on). During the creation of a new category (for example, when a lesson is saved) we consider the user's skill value equal to the resource technical level, presuming that if a user saves a learning material then he could be able to make use of it. The user's skill level is updated when a new resource is saved, taking into account its technical level and the user's skills in that category. Starting value for user's skills parameter, its update frequency, the increment or decrement value and the difference between technical level and user's skills necessary to obtain a recommendation outcome from the following experimental tests. They are easily adaptable, though.

We use artificial learners to get a flavour of how the system works. We have created a SCORM compliant learning material using the abstract of several papers in .html version from scientific journals published on the web. We have linked an *imsmanifest* SCORM file to each paper. Then, we have simulated ten users with different initial profiles (based on the field of interest and the skill level) and saved, in four turns, ten learning resources for each user, obtaining 400 LOs. The main advantage of the described approach is the semantic accuracy growth. To give a quantitative estimation of the improvement induced by a concept based approach, we are executing a comparative experiment between word-based user and resource models on one side and concept-based user and resource models on the other one.

5. Considerations

The paper addresses a key limitations with existing courseware on the Internet. Humans want immediate access to relevant and accurate information. There has been some progress in combining learning with information retrieval, however, these advances are rarely implemented in e-learning courseware. With this objective in mind, we described how to perform automatic recommendation of relevant learning objects considering student and learning material profiles, adopting filtering criteria based on the value of selected metadata fields and capturing not only structural but also semantics information. The proposed method uses hierarchical concept categories, enabling users and system to search, handle or read only those concept of interest to them in a more general manner, providing semantic possibility. Our experiments to test the system's functionality are based on SCORM compliant LOs.

Summarizing, the key elements of the described system could be highlighted as follows. The system provides immediate portability and visibility from different user locations, enabling the access to personal bookmark repository just by using a web browser. The system assists students in finding relevant reading material providing personalized learning object recommendation. The system directly benefits from existing repositories of learning material providing access to open huge amount of digital information. The system reflects continuous ongoing changes of the practices of its member, as required by a cooperative framework. Exploiting a word sense based document representation, the system proposes resources and student models based on word senses rather that on simply words.

BIBLIOGRAPHY

Andronico, A., A. Carbonaro, L. Colazzo, A. Molinari, (2004), Personalisation services for learning management systems in mobile settings, International Journal Continuing Engineering Education and Lifelong Learning.

Belkin, N. J., W. B. Croft, (1992), Information Filtering and Information Retrieval: Two Sides of the Same Coin, Commun. ACM 35, 12, 29-38.

Bighini, C. & A. Carbonaro, (2004), InLinx: Intelligent Agents for Personalized Classification, Sharing and Recommendation, Int. Journal of Computational Intelligence. Int. Computational Intelligence Society. 2 (1).

Budanitsky A., G. Hirst, (2001), Semantic distance in wordnet: An experimental, application-oriented evaluation of five measures, In Workshop on WordNet and Other Lexical Resources. Second meeting of the North American Chapter of the Association for Computational Linguistics, Pittsburgh.

Ciravegna, F., A. Dingli, D. Petrelli, Y. Wilks, (2002), User-System Cooperation in Document Annotation based on Information Extraction. Proceedings of EKAW'02.

Cunningham, H. D. Maynard, K. Bontcheva, V. Tablan, (2002), GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In Proceedings 40th Anniversary Meeting of the Association for Computational Linguistics (ACL 2002). Budapest.

Delgado, J., N. Ishii, T. Ura, (1998), Content-based Collaborative Information Filtering: Actively Learning to Classify and Recommend Documents, M. Klush, G. Weiss (Eds), LNAI 1435, Springer-Verlag, 206-215.

Denaux, R., V. Dimitrova, L. Aroyo, (2005), Integrating open user modeling and learning content management for the semantic web. In: Proceedings of UM2005.

Fellbaum, C., ed (1998), WordNet: An Electronic Lexical Database. MIT Press, Cambridge, Mass.

Fensel, D., S. Decker, M. Erdmann, R. Studer: (1998), Ontobroker: The Very High Idea. In Proceedings of the 11th International Flairs Conference (FLAIRS-98), Sanibal Island, Florida, USA, 131-135.

Henze, N., P. Dolog, & W. Nejdl, (2004), Reasoning and ontologies for personalized e-learning in the semantic web, Educational Technology and Society, 82-97.

Kanawati, R., M. Malek, (2001), CoWing A Collaborative Bookmark Management System, M. Klush, F. Zambonelli (Eds), LNAI 2182, Springer-Verlag, 38-43.

Koivunen M., E. Miller, (2002), W3C Semantic Web activity. In E. Hyvonen, editor, Semantic Web Kick-Off in Finland, pages 27–44, Helsinki. HIIT Publications.

Leacock, C., M. Chodorow, (1998), Combining local context and WordNet similarity for word sense identification. In Fellbaum 1998, pp. 265–283.

Maedche, A., S. Staab, (2000), Semi-automatic Engineering of Ontologies from Text. In: Proceedings of the 12th International Conference on Software Engineering and Knowledge Engineering.

Melville, P., R. Mooney, R. Nagarajan, (2002), Content-boosted collaborative filtering for improved recommendations. Proceedings of the 18th National Conference on Artificial Intelligence (AAAI-2002). Canada.

Miller G. E.,Walter A. Charles, (1991), Contextual correlates of semantic similarity. Language and Cognitive Processes, 6:1–28.

Muna, S. Hatem, Haider, A. Ramadan and Daniel, C. Neagu, (2005), e-Learning Based on Context Oriented Semantic Web, Journal of Computer Science 1 (4): 499-503.

Nejdl W., (2001), Learning Repositories – technologies and Context. Proc. of ED-MEDIA 2001. World Conference on Educational Multimedia, Hypermedia & Telecomuunications, June 25-30, Finland.

Porto, F., AMC. Moura, FJC da Silva, (2004), ROSA: a Repository of Objects with Semantic Access for e-Learning Proceedings of the International Database Engineering and Applications Symposium (IDEAS'04).

Recker, M., Walker, A., & Lawless, K. (2003), What do you recommend? Implementation and analyses of collaborative filtering of Web resources for education. Instructional Science, 31, 229-316.

Resnik, P., (1995), Disambiguating Noun Groupings with Respect to WordNet Senses. Sun Microsystems Laboratories, Chelmsford, MA 01824-4195 USA.

Salton G., (1989), Automatic Text Processing: the Transformation, Analysis and Retrieval of Information by Computer, Addison-Wesley, Reading, Mass.

Tang, T., & McCalla, G. (2005), Smart Recommendation for an Evolving E-Learning System: Architecture and Experiment. International Journal on E-Learning. 4 (1), pp. 105-129.

Sure, Y., M. Erdmann, J. Angele, S. Staab, R. Studer, and D. Wenke, (2002), OntoEdit: Collaborative Ontology Development for the SemanticWeb I. Horrocks and J. Hendler (Eds.): ISWC 2002, LNCS 2342, pp. 221–235.

Woukeu A., G. Wills , G. Conole, L. Carr, S. Kampa, W. Hall, (2003), Ontological Hypermedia in Education: A framework for building web-based educational portals. In Proceedings of ED-MEDIA 2003 – 12th World Conference on Educational Multimedia, Hypermedia & Telecommunications. Honolulu, Hawaii, USA, June 23-28, 2003.