Collaborative and Semantic Information Retrieval for Technology-Enhanced Learning

Antonella Carbonaro

Department of Computer Science, University of Bologna,
Mura Anteo Zamboni, 7, Bologna, Italy
antonella.carbonaro@unibo.it

Abstract. The paper presents an ontological approach for enabling personalized searching framework facilitating the user access to desired contents. Through the ontologies the system will express key entities and relationships describing resources in a formal machine-processable representation. An ontology-based knowledge representation could be used for content analysis and concept recognition, for reasoning processes and for enabling user-friendly and intelligent content retrieval.

Keywords: Ontology, Semantic Web Applications, collaboration in information searching, personalized searching framework.

1 Introduction

Technological advances in information and communication systems have challenged educational institutions to adopt the opportunities of distributed knowledge acquisition and delivery. Among the most recent trends, the availability of wireless communication standards and of mobile devices gives rise for a new landscape of learning as a networked, situated, contextual and life-long activities. In this scenario, new perspectives on learning and teaching processes must be developed and supported, relating learning models, learning methods, didactics, team organization and situational behavior models.

In a distributed learning environment, we usually have a large number of educational resources (web pages, lectures, journal papers, learning objects, social networks, …) stored in many distributed and different repositories on the Internet. Without guidance, students will probably have great difficulties in finding the reading material relevant for a particular learning task. This problem is becoming particularly important in Web-based education where the variety of learners taking the same course is much greater. Vice versa, 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 can't directly benefit from existing repositories of learning material.

The Web is increasingly becoming important than ever, moving toward a social place and producing new applications with surprising regularity: there has been a shift from just existing on the Web to participating on the Web. Community applications and online social networks have become very popular recently, both in personal/social and professional/organizational domains [1]. Most of these collaborative applications provide common features such as content creation and sharing, content-based tools for discussions, user-to-user connections and networks of users sharing common

interest, reflecting today's Web 2.0 rich Internet application-development methodologies.

The Semantic Web offers a generic infrastructure for interchange, integration and creative reuse of structured data, which can help to cross some of the boundaries that Web 2.0 is facing. Currently, Web 2.0 offers poor query possibilities apart from searching by keywords or tags. There has been a great deal of interest in the development of semantic-based systems to facilitate knowledge representation and extraction and content integration [2], [3]. Semantic-based approach to retrieving relevant material can be useful to address issues like trying to determine the type or the quality of the information suggested from a personalized environment. In this context, standard keyword search has a very limited effectiveness. For example, it cannot filter for the type of information, the level of information or the quality of information.

By exploiting each other's achievements the Semantic Web and Web 2.0 together have a better opportunity to realize the full potential of the web [4].

Potentially, one of the biggest application areas of social networks might be personalized searching framework (e.g., [5],[6]). Whereas today's search engines provide largely anonymous information, new framework might highlight or recommend web pages created by recognized or familiar individuals. The integration of search engines and social networks can lead to more effective information seeking [7]. In fact, the system we want to propose can find application in any context in which the group collaboration is a requisite, and we believe that a Web-based learning system is an ideal application domain.

Additionally, we can consider semantic information representation as an important step towards a wide efficient manipulation and retrieval of information [8], [9], [10]. In the digital library community a flat list of attribute/value pairs is often assumed to be available. In the Semantic Web community, annotations are often assumed to be an instance of an ontology. Through the ontologies the system will express key entities and relationships describing resources in a formal machine-processable representation. An ontology-based knowledge representation could be used for content analysis and object recognition, for reasoning processes and for enabling user-friendly and intelligent multimedia content search and retrieval.

In this work we explore the possibilities of synchronous, semantic-based collaboration for search tasks. We describe a search system wherein searchers collaborate intentionally with each other in small, focused search groups. Developed framework (SWS2 – Semantic Web Search 2.0 - project) goes beyond implementation of ad hoc user interface. It also identifies information that one group member searches and uses it in realtime to improve the effectiveness of all group members while allowing semantic coverage of the involved domain. The semantic approach is exploited introducing an ontology space covering domain knowledge and resource models based on word sense representation.

There are many scenarios in which small groups of users collaborate on Web search tasks to find information, such as school students or colleagues jointly writing a report or a research, or arranging joint travel. Although most search tools are designed for individual use, some collaborative search tools have recently been developed to support such collaborative search task [11]. These tools tend to offer two classes of support: i) awareness features (e.g., sharing and browsing of group members' query histories, and/or comments on results and on web pages rating), ii) division of labor features (e.g., to manually split result lists among group members, and/or algorithmic techniques for modifying group members' search results based on others' actions) [12]. Collaborative search tools are relatively novel and thus not widely available.

2 Personalized Searching Framework

One of the areas in which information retrieval is likely to see great interest in the future is synchronous collaborative search. This concerns the common scenario where two or more people working together on some shared task, initiate a search activity to satisfy some shared information need. Conventionally, this need is satisfied by independent and uncoordinated searching on one or more search engines, leading to inefficiency, redundancy and repetition as searchers separately encounter, access and possibly re-examine the same documents. Information searching can be more effective as a collaboration than as a solitary activity taking advantage of breadth of experience to improve the quality of results obtained by the users [13]. Community-based recommendation systems [14], [15] or user interfaces that allow multiple people to compose queries [12] or examine search results [16] represent various forms of collaboration in search.

Traditional approaches to personalization include both content-based and user-based techniques. If, on one hand, a content-based approach allows to define and maintain an accurate user profile (for example, the user may provides the system with a list of keywords reflecting him/her initial interests and the profiles could be stored in form of weighted keyword vectors and updated on the basis of explicit relevance feedback), which is particularly valuable whenever 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 user-based approach, resources are processed according to the rating of other users of the system with similar interests. Since there is no analysis of the item content, these information management techniques can deal with any kind of item, being not just limited to textual content. In such a way, users can receive items with content that is different from that one received in the past. On the other hand, since a user-based technique works well if several users evaluate each one of them, new items cannot be handled until some users have taken the time to evaluate them and new users cannot receive references until the system has acquired some information about the new user in order to make personalized predictions. These limitations often refer to as the sparsity and start-up problems. By adopting a hybrid approach, a personalization 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. A hybrid approach maintains another drawback: the difficulty to capture semantic knowledge of the application domain, i.e. concepts, relationships among different concepts, inherent properties associated with the concepts, axioms or other rules, etc [17].

In this context, standard keyword search is of very limited effectiveness. For example, it does not allow users and the system to search, handle or read concepts of interest, and it doesn't consider synonymy and hyponymy that could reveal hidden similarities potentially leading to better retrieval. The advantages of a concept-based document and user representations can be summarized as follows: (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 user is interested in computer science resources, a document containing the word 'bank' as it is meant in the financial context could not be relevant); (ii) synonymous words belonging to the same meaning can contribute to the resource model definition (for example, both 'mouse' and 'display' brings evidences for computer science documents, improving the coverage of the document retrieval); (iii) synonymous words belonging to the same meaning can contribute to the user model matching,

which is required in recommendation process (for example, if two users have the same interests, but these are expressed using different terms, they will considered overlapping); (iv) 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.

For example, the system could support Computer Science last-year students during their activities in courseware like Bio Computing, Internet Programming or Machine Learning. In fact, for these kinds of courses it is necessary an active involvement of the student in the acquisition of the didactical material that should integrate the lecture notes specified and released by the teacher. Basically, the level of integration depends both on the student's prior knowledge in that particular subject and on the comprehension level he wants to acquire. Furthermore, for the mentioned courses, it is necessary to continuously update the acquired knowledge by integrating recent information available from any remote digital library.

2.1 Use case analysis

A first level of system analysis can be achieved through its functional requirements. Such functional requirements are described by the interaction between users and the systems itself. Therefore, users may be interested in semantic-based search or collaborative semantic-based search.
We define an interaction between users as a collaborative search session managed by the system using specialized components: in particular, the system should cover both user manager and sessions between users manager roles.

2.2 System modules

In the following we list the components able to handle user data:

i) User Interface Controller: it coordinates the information flow between interface control and other system components and allows to perform data presentation for the GUI visualization.
ii) Semantic searcher: it implements semantic-based searches extracting concepts related to introduced keywords using a thesaurus and searching in the underling ontology corresponding documents.
iii) Interest coupler: it performs intersection between user interest matching relevant terms extracted from semantic searcher.
iv) User Manager: it deals with user. For example through the User Manager, it is possible to register new users or to search for their information. Moreover, it is able to associate mail boxes to user to enhance communication.
v) Session Manager: it manages collaborative search sessions allowing user insertion and search terms shared between users. It allows to maintain consistency between session views and creates message boxes for the specific session whose content is available to all the participants.

2.3 Data analysis

i) OWL

The ontology developed to test implemented framework maintains relation between courses, lessons, teachers and course material. Ontology is a representation

model in a given domain that can be used for the purposes of information integration, retrieval and exchange. The ontology usage is widely spread in not only the artificial intelligent and knowledge representation communities, but most of information technology areas. In particular, ontology has become common in the Semantic Web community in order to share, reuse and process domain information between humane and machine. Most importantly, it enables formal analysis of domain knowledge, for example, context reasoning becomes possible by explicitly defining context ontology.

There are several possible approaches in developing a concept hierarchy. For example, a top-down development process starts with the definition of the most general concepts in the domain and subsequent specialization of the concepts, while a bottom-up development process starts with the definition of the most specific classes, the leaves of the hierarchy, with subsequent grouping of these classes into more general concepts. The hybrid development consists in a combination of the top-down and bottom-up processes. Due to our personal view of the domain we took the combination approach. Once we have defined the classes and the class hierarchy we described the internal structure of concepts defining the properties of classes. Over the evolving ontology we perform diagnostics to determine the conformance to common ontology-modeling practices and to check for logical correctness of the ontology.

ii) User Data

It maintains data of the users handled by the system.

iii) Session Data

It maintains data corresponding to collaborative search sessions.

2.4 Developed system interaction

The developed system proposes three different interaction between the users.

i) Search interaction

This interaction starts when a user performs a search proposing one or more keywords. The Semantic searcher module returns a list containing relevant documents and recommends terms for the possible following searches. Therefore, the User Interface Controller is able to find similar user with similar interest in performed searches using Interest Coupler module.

ii) Collaborative search session interaction

A user can decide to contact another user, proposed by the system similar to his interests, to start collaborative search session. The request produces an Invitation message in the message box of the target user. Concurrently, a listening permanent loop allows to User Interface Controller to advise target user. In the case of positive response, the User Interface Controller creates a new collaborative search session and a Session Join request is sent.

iii) Interaction during a collaborative search session

The user could modify the list of search terms adding or removing some keyword. The request, managed by the User Interface Controller, is forwarded to Session Manager that updates search terms, replacing term list and requiring GUI updates. The same interaction can be used to implement a session chat, allowing more collaboration value to the system.

2.5 System GUI

The search home page is showed in Figure 1. Box A allows to the user to insert his nickname to use during SIG sessions dynamically showed in box B.

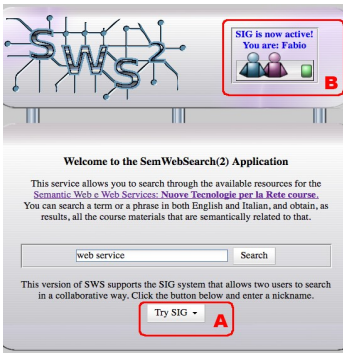

Fig. 1. SWS2 home page

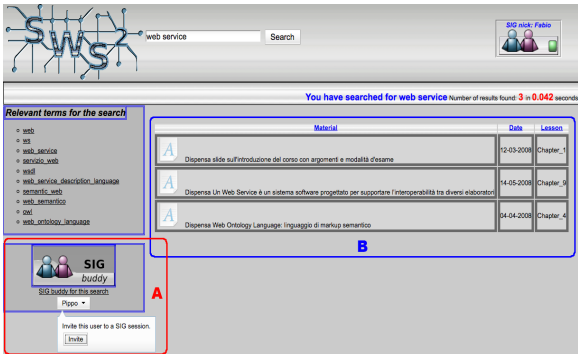

Fig. 2. Collaborative search session

If a user participate to collaborative search sessions, the system proposes in his search result page a new box containing similar users (Figure 2, box A). This button also allows to send Invitation message to target user; a background function verify the presence of new Invitation messages and, eventually, notify them to the user.

Figure 3 box A shows communication facilities offered to system users, while Figure 3 box B shows terms actually used to search session. Using components showed in box B1 the user may add search terms, while using the component showed in box B2 the user may remove session search terms. The button showed in box B3 is twofold: on one hand, it allows users to accept the lists of terms created by the system. On the other, through such button, it is possible to perform the described semantic searches.

3. Considerations

Golovchinsky et al. [7] distinguish among the various forms of computer-supported collaboration for information seeking, classifying such systems along four dimensions: intent, depth of mediation, concurrency, and location.

The intent could be explicit or implicit. In our framework two or more people set out to find information on a topic based on a declared understanding of the information need, which might evolve over time. So, our framework implements explicit information seeking scenarios.

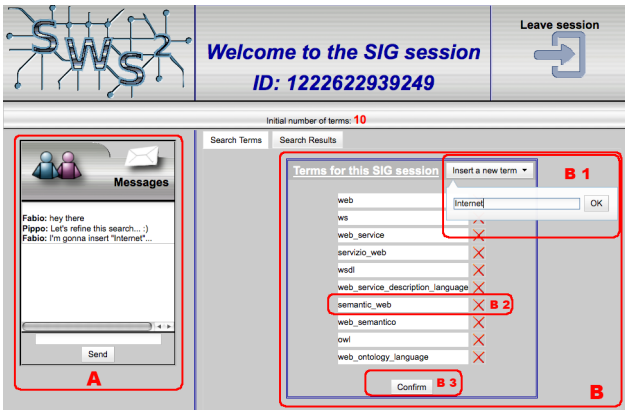

Fig. 3. Collaborative search terms specification

The depth of mediation is the level at which collaboration occurs in the system. Our system implements algorithm mediation at the search engine level explicitly consider ongoing collaboration and coordinate users activities during the search session.

People can collaborate synchronously or asynchronously. In our system the collaboration is synchronous involving the ability of people to influence each other in real time.

Finally, collaboration may be co-located (same place at the same time) or, as in our framework, distributed, increasing opportunities for collaboration but decreasing the fidelity of possible communications.

An important step in the searching process is the examination of the results retrieved. In order to test developed framework we have collected over 50 different documents concerning actual domain. We have extracted several concepts used during the annotation phase and performed tests to verify searching functionalities. It is currently difficult to replicate or make objective comparisons in personalized retrieval researches, so to evaluate search results we have considered the order used by the framework to present retrieved results. During this step, the searcher browses through the results to make judgments about their relevance and to extract information from those found to be relevant. Because information is costly (in terms of time) to download, displays of result lists should be optimized to make the process of browsing more effective. We have also evaluated the effect that the proposed framework has on collaboration and exploration effectiveness. Using implemented tools, searchers found relevant documents more efficiently and effectively than when working individually and they found relevant documents that otherwise went undiscovered.

The work described in this paper represents some initial steps in exploring semantic-based search retrieval collaboration within a focused team of searchers. It could be considered as one possible instance of a more general concept. While the initial results are encouraging, much remains to be explored. For example, most of the current research on sensemaking has been at the individual level, with little understanding of how sensemaking occurs in collaborative search tools.

References

1. Kolbitsch J., Maurer H.: The Transformation of the Web: How Emerging Communities Shape the Information we Consume, in Journal of Universal Computer Science, vol. 12, no. 2, 187-213 (2006)
2. Henze N., Dolog P., Nejdl W.: Reasoning and Ontologies for Personalized E-Learning in the Semantic Web, Educational Technology & Society, 7 (4), 82-97 (2004)
3. Bighini C., Carbonaro A.: InLinx: Intelligent Agents for Personalized Classification, Sharing and Recommendation, International Journal of Computational Intelligence. International Computational Intelligence Society. 2 (1), (2004)
4. Bojars U., Breslin J.G., Finn A., Decker S.: Using the Semantic Web for linking and reusing data across Web 2.0 communities, Web Semantics: Science, Services and Agents on the World Wide Web 6, pp. 21–28 (2008)
5. Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., and Back, M. Algorithmic Mediation for Collaborative Exploratory Search. To appear in Proceedings of SIGIR
6. Freyne J., Smyth B.: Collaborative Search: Deployment Experiences, in The 24th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence. Cambridge, UK, pp. 121-134 (2004)
7. Golovchinsky G., Qvarfordt P., Pickens J.: Collaborative Information Seeking, Computer, vol. 42, no. 3, pp. 47-51, (2009)
8. Calic J., Campbell N., Dasiopoulou S. and Kompatsiaris Y.: A Survey on Multimodal Video Representation for Semantic Retrieval, in the Third International Conference on Computer as a tool, IEEE (2005)
9. Carbonaro A., “Defining Personalized Learning Views of Relevant Learning Objects in a Collaborative Bookmark Management System”, In Z. Ma (Ed.), Web-based Intelligent ELearning Systems: Technologies and Applications (pp. 139-155). Hershey, PA: Information Science Publishing, (2006)
10. Bloehdorn S., Petridis K., Simou N., Tzouvaras V., AvrithisY., Handschuh S., Kompatsiaris Y., Staab S., Strintzis M. G.: Knowledge Representation for Semantic Multimedia Content Analysis and Reasoning, in Proceedings of the European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology, (2004)
11. Paul S. A., Morris M. R.: CoSense: enhancing sensemaking for collaborative web search by, in CHI '09: Proceedings of the 27th international conference on Human factors in computing systems, pp. 1771-1780 (2009)
12. Morris M. R. and Horvitz E.: Searchtogether, an interface for collaborative web search, In Proceedings of UIST, pp. 3–12 (2007)
13. Baeza-Yates R. and Pino J. A.: A first step to formally evaluate collaborative work, in GROUP ’97: Proc. ACM SIGGROUP Conference on Supporting Group Work, pp. 56–60 (1997)
14. Carbonaro A., “Defining personalized learning views of relevant learning objects in a collaborative bookmark management system” Web-Based Intelligent e-Learning Systems: Technologies and Applications, Idea Group Inc. (2005)
15. Smyth B., Balfe E., Boydell O., Bradley K., Briggs P., Coyle M., and Freyne J.: A live-user evaluation of collaborative web search, in Proceedings of IJCAI, pp. 1419–1424, Ediburgh, Scotland (2005)
16. Smeaton F., Lee H., Foley C., McGivney S., and Gurrin C.: Collaborative video searching on a table, in Multimedia Content Analysis, Management, and Retrieval, San Jose, CA, (2006)
17. Carbonaro A., Ferrini R.: Considering semantic abilities to improve a Web-Based Distance Learning System, ACM International Workshop on Combining Intelligent and Adaptive Hypermedia Methods/Techniques in Web-based Education Systems, (2005)