Peer Reviewed Papers

Semantic Management Systems for the Material Support of E-learning Platforms

Vincenzo Di Lecce[1], Marco Calabrese[1], Domenico Soldo[2], Antonella Giove[3]

[1]DIASS – Politecnico di Bari, [2]myHermes s.r.l., [3]II Facoltà di Ingegneria - Politecnico di Bari

Keywords: semantic, WordNet, chatbot

Abstract

The paper describes a semantic-based system for e-learning platforms, designed to automate some highly time-consuming activities for the account of the cognitive process actors. To this end the system helps the instructor organize the support material for Learning Objects and guides the student in learning the structure of specific knowledge domain. Concerning the first of these aspects, the teacher is required only to indicate the resources in digital format (e-books, documents, Web pages) strictly connected to the lesson topic delivered in a structured way. At the student level, an interaction with a chatbot was analyzed in order to make the research and exploration of content more interactive. The virtual assistant guides the student in understanding structured contexts at different levels of detail (book, chapter, paragraph, sentence and phrase) and their mutual logical organization. The knowledge base was developed on the basis of the Semantic Web paradigm and uses WordNet as reference ontology. A prototype system is currently under test and evaluation at the AeFLab Laboratory of the Politecnico di Bari.

1 Introduction

The Web is now the world's largest collection of documents with tens of thousands of new entries each day and an outlook of services and technology solutions in continuous evolution.

In recent years, the progressive affirmation of tools able to provide everyone with the ability to publish and share content in collaboration with other online users (the so-called phenomenon of Web 2.0) has involved a cultural change in adopting new strategies of communication, knowledge and learning. In this context, the new Web generation continues to have a major impact on online training. For instance, teachers begin to explore and appreciate the potential of blogs, media-sharing services and other social software, which, although not designed specifically for e-learning, can be used to encourage students to try new exciting learning opportunities.

Even if the learning process is facilitated by the multiplicity of advanced instruments, however, its basic feature remains above all the ability to negotiate new meanings. That's why the rapid growth of content distribution thanks to the Internet has made the role of semantic information management predominant while emphasizing its strategic value even in e-learning systems.

In an e-learning environment, educational content should focus on small modules called Learning Objects (LOs) with associated semantics (or meta-data) in order to easily allow the retrieval of content by meaning and contexts. Thus, these modules should be linked together in a sort of "conceptual Web" for facilitating the construction of a customized learning path. Nevertheless, in daily practice, the effort to make semantic annotation of metadata required by the teacher is rarely accomplished. At the same time an important concern has to be addressed and regards the problems of information retrieval experienced by unskilled student due to their inability to overcome the linguistic barrier that stands between an effective and targeted content exploration and the knowledge of that specific domain terminology which, paradoxically, should really help them to find what they may be searching for.

In this scenario, this paper describes a system for semantic-based e-learning platforms designed with the dual target of facilitating the teacher in organizing the support material referred to the delivered LOs and guiding the learner, especially the unskilled ones in learning the structure of specific knowledge domain.

2 Ontology based systems and chatbot in literature

DGiven a domain of knowledge referred to learning–oriented goal, there are various possible representations of learning paths: this implies a reflection

about the presentation suitability of the material delivered in relation to the students' profile and their learning objectives (Baker, 2000).

For instance, the FAQs (which have been under study for several years (Whitehead, 1995; Sneiders, 1999), force the learners to carefully read the list of options for finding the right answer in line with their request. This process becomes so slow and imprecise, if the user's knowledge about the topic to be searched is even lower. While the experienced user is prone to analyze the keywords of the FAQ, or domain specific terms near to the most suitable couple of question/answer, for the unskilled users the process is different i.e. it is probable that the they behave as "random surfers", since they are driven by a general understanding free of expressions characterizing the conceptual domain.

In this view, modern IT tools should play a complex role both on the production side and content organization and also on support and guide for the students throughout the learning process. In both cases, knowledge representation has a key role in the light of recent studies about ontology especially with reference to the Semantic Web application.

2.1 Semantic Web and Ontologies

The statement "Semantic Web " by (Berners-Lee, 2001) especially in the academic field has aroused increasing interest around the development of ontologies during the last decade (Gruber, 1993), they represent, in fact, an interesting perspective to look at just for supporting the learning process (Adorni *et al.*, 2010).

Under the concept of "ontology" we conceive a form of knowledge representation that consists of two necessary elements to structure the so-called "semantic network": the concepts and relations between them (Gruber, 1995). Depending on the structure of this network, whether it is a tree (as in taxonomy) or a graph (as a concept map), the complexity and quality of represented knowledge can vary. Di Lecce and Calabrese (Calabrese & Di Lecce, 2008) discuss how a good compromise between expressiveness and computational complexity is represented by the electronic dictionary "WordNet" (Fellbaum, 1998).

WordNet is a lexical-semantic dictionary, created by a team of psycholinguistics at Princeton University. Some of its main uses (Navigli, 2009) concern the development of knowledge base in systems oriented to NLP (Natural Language Processing) and WSD (Word Sense Disambiguation). WordNet is based on the idea of synset i.e. a semantic element that can described through lists of words having a precise meaning in a given context. Unlike traditional dictionaries and thesauri, WordNet is characterized by a browsing feature by lexical-semantic relations that represent recursive mappings in the set of synsets. These struc-

tures within the semantic network of WordNet allows for applying semantic disambiguation algorithms (Navigli & Velardi, 2005) that, given a free text as input, are able disambiguate the context i.e. they associate it to the list of synsets, which are more suitable for representing its semantics. The text can be an ordinary office document or one with specific reference to the Semantic Web and also an HTML page (Di Lecce *et al.*, 2009). In this regard, WordNet is the ideal technological tool to support knowledge representation and processes of context disambiguation.

2.2 Chatbot and human-machine interfaces

The knowledge base due to the limitations imposed by the semantic gap, usually allows only for a very faint representation of reality if compared with the perception of a human observer. This limit is reduced, however, when the reference context is sufficiently narrow and driven (task-oriented).

An innovative tool for human-machine communication, particularly in Web environments, is represented by the chatbot, which are programs capable of sustaining a conversation with a human natural language (Kerly *et al.*, 2008) on restricted domains. Although the actual effectiveness of such systems runs into significant practical and theoretical difficulties, the chosen direction, also in e-learning framework seems to be just that (Kerly *et al.*, 2007).

In literature typical examples of chatbots are ELIZA (late '60) and ALICE that was Loebner Prize winner (after 2000 year) and it is based on AIML markup language (AIML 1.0.1 2005). The AIML allows for defining quite simple stimulus-response pattern and can be obtained as a result of an automatic process of knowledge extrapolation starting from pre-processed forms of information such as FAQ and glossary (De Gasperis, 2010); more sophisticated Chabot architectures following the probability-based approach have been presented recently (Bentivoglio *et al.*, 2010).

3 Proposed System

The proposed system is configured as a tool for the semantic organization and use of knowledge extracted from digital format material. Hereinafter the detailed features of the system and the salient aspects of the process are indicated as follows: knowledge semantic indexing and management of system-user interface by means of chatbot.

3.1 System features

As with any e-learning platform, the system provides for interfacing two categories of user: the teacher and the learner. The teacher is responsible for

loading the material within the platform; students are required to look up the contents, trying to identify, in relation to their learning paths, the ones that best match the issues of interest.

The main element of the system lies in the increasing efficiency and effectiveness of those individual activities performed by the actors of the cognitive process that is the shortening of time for the content semantic annotation (by teachers) and use of that (by learners). This objective, which will be after specified, is achieved through a process of automatic semantic indexing based on the use of WordNet.

To better define the performance of the system from the user side, a UML diagram, depicting the cases of use, is proposed. Figure 1 shows the two actors interacting respectively with the cases of uses for the submission of content and the query of them. The case of use about the indexing extends that of content submission thus expanding the repository of unstructured or partially structured knowledge to be indexed. In turn, the indexed contents extend the knowledge base accessible by the student.

Fig. 1 - Use case UML Diagrams of the proposed system the two actors interfacing with the systems are the teacher and the student.

3.2 System functioning

The proposed system consists of two separate processes that share the same knowledge base (Figure 1). The semantic indexing process involves the acquisition of knowledge from each digital document indicated by the teacher, data extraction and structure, mining of semantic tags from data and association of these tags to the document structure. The document is then indexed and stored in the knowledge base. The process of guided browsing of the content allows the learner to query the knowledge base by using a chatbot that employs a semantic disambiguation engine for processing ambiguous queries. Hereinafter a more detailed description of both processes is provided.

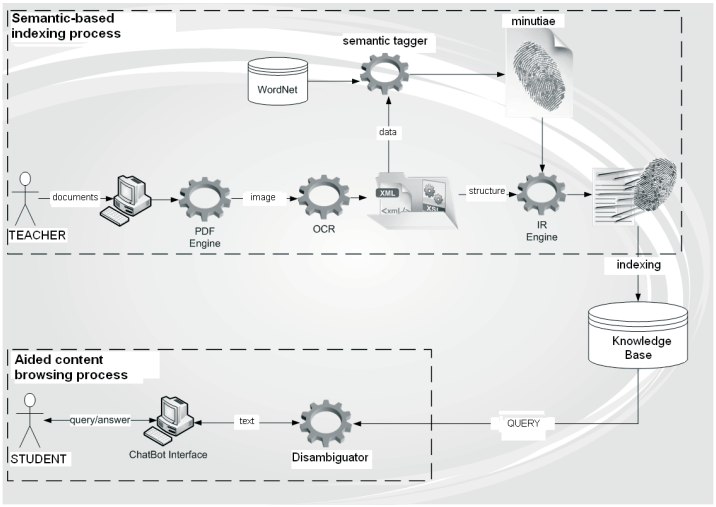

Fig. 2 - Processes characterizing the proposed systems

3.2.1 Knowledge semantic indexing

The contents specified by the teacher are digital resources organized according to their internal structure. In general, regardless of the information source size, it is possible to identify a tree structure, which is common to all resources. There are many elements to be considered around this structure (... book, chapter, paragraph, sentence, phrase...). The differences in "syntax" structuring of knowledge amid various formats (HTML, XML, e-books, etc...) depend mainly on the nature of links among different sections of the resource. For example, within an HTML page, the reference between an index entry and its content is produced by a link, while in an open document, e.g. PDF the link is provided by the section identification numbers and / or page.

Making use of common software for extracting text and syntactic structure from documents belonging to the different types under consideration (OCR), the system obtains the language data base where it can operate. The second semantic analysis techniques developed previously by the authors (Di Lecce, 2009) are applied just on these data. This phase leads to the assignment of those tags called "minutiae" to the linguistic data previously obtained. The minutiae are lexical-semantic relations between the terms of the linguistic base. The minutiae thus obtained lead to a semantic network that represents a specialized sub-graph of WordNet in relation to the domain of knowledge.

At this point, considering the tagged terms, the minutiae are assigned to the different elements of structure (book, chapter, paragraph, sentence, and phrase)

that characterize the analyzed text. In this way, through typical Information Retrieval techniques, a projection operation of the WordNet ontology structure is performed. Following this procedure the semantic paths typifying the context of "belonging" of the analyzed resources are obtained.

At the end of this process it is possible to obtain a graph structure similar to ontology of context for each resource. The lexical-semantic structure is, because of the process that characterizes its definition, a subpart of the WordNet ontology enriched with references to the indexed documents. Each contextual lexical-semantic entity is thus, for a resource its meta-description, because it describes the context. This allows for performing a semantic indexing of resources indicated by the teacher and a clearer consultation for the student (for more details on this concern see the next subsection).

In the current version of the system, the texts are treated in English and the instrument of knowledge representation used, as already mentioned, and is WordNet. However, the proposed approach is inherently multilingual, the development by the scientific community of the lexical-semantic dictionaries following WordNet model, into other languages, including Italian, is in fact under advanced engineering phase (for instance the MultiWordNet project http://multiwordnet.fbk.eu/english/home.php).

3.2.2 Chatbot supporting the content browsing

The access to the contents by the student is characterized as a human-machine dialogue process. Anyone, who is interested to know or deepen a topic, rarely knows in advance what it will be possible to find (otherwise with the extent of the imprecision varying from case to case) and at the same time this one does not know the terms of the context; therefore for an in-depth examination of a topic, a repeated query process is required so that the system can find the contents of interest and the mutual relations between them. In the proposed system, the mediator of this human-machine dialogue process is realized by a chatbot developed in a prototype form through a CGI Matlab® engineered specifically for this purpose.

The chatbot, by means of the semantic knowledge described above, receives the user queries formulated in natural language and tries to disambiguate the content in order to return text elements being semantically closer to the meaning of the request.

The sequence of words received from the chatbot is compared with those entries available in the lexical-semantic vocabulary (words not included in the dictionary are ignored.) Commonly, a word can have more than one meaning; in this case a well-known disambiguation technique is applied (Navigli & Velardi, 2005) to find the most probable interpretations. In general the more

ambiguous is the query; the more the semantic matching is operated at a high level of lexical-semantic taxonomy i.e. the query refers to a form of a very abstract and imprecise knowledge representation. In this case the system suggests the user the appropriate level of abstraction (book, chapter, paragraph, and sentence) from which it is possible to proceed for an in-depth analysis about the topic of interest.

Figure 3 shows an example of chatbot response to the query of the word "architecture" of course in English language. The domain knowledge base is structured on a manual of "Linux" available online and previously indicated by the teacher. The manual consists of over 6000 pages organized into 16 first-level structures (Chapters), each of which constitutes an average of about 70 second-level sections, and so on up to reach the granularity of a single sentence.

Fig. 3 - Prototypal GUI chatbot. The user performs the requests to access content through a textbox (left).The system returns those results semantically related to the request according to the detail level that is appropriate to disambiguate the response (right). The students are thus led to the search for a more suitable granularity level of ambiguity of their query thus establishing a dialogue process with the system under the supervision of the chatbot.

4 Conclusions

This paper has introduced a system for the automatic structuring and guided use of that support material for the learning modules in e-learning platforms. The innovative aspect of the proposal is the system ability to provide contents according to a lexical-semantic structure that indexes the linguistic contents following the different levels in which the text is organized (for example book, chapter, paragraph, paragraph, sentence). This feature has a profound impact on both the timing of semantic content annotation and searching of them. In fact, students, without a thorough knowledge of the learning topic, inevitably

produce ambiguous queries to the system. Consequently, the system responds by returning the text portions according to that level of granularity, which is more suitable for disambiguating the request thereby ensuring not only the delivery of content but also evidence of how the knowledge is structured.

A prototype of the system with Matlab® CGI technology is currently under further test phases at the AeFLab Laboratory of the "Politecnico di Bari".

REFERENCES

Adorni G., Battigelli S., Brondo D., Captano N., Coccoli M., Mirando S., Orciuoli F., Stanganelli L., Sugliano A.M., Vivanet G. (2010), *Approcci basati su ontologie per l'apprendimento per tutti, in qualunque momento e in ogni luogo: studio dei casi CADDIE e IWT*, Journal of e-Learning and Knowledge Society, 2(6), pp. 53-68.

AIML 1.0.1 (2005), http://www.alicebot.org/TR/2005/WD-aiml, ALICE Artificial Intelligence Foundation.

Bentivoglio C.A., Bonura D., Cannella V., Carletti S., Pipitone A., Pirrone R., Rossi P.G., Russo G. (2010), *Agenti intelligenti supporto dell'interazione con l'utente all'interno di processi di apprendimento*, Journal of e-Learning and Knowledge Society, 2(6), pp. 27-36.

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

Baker M. (2000), *The roles of models in Artificial Intelligence and Education research: a prospective view*, Int. Journal of Artificial Intelligence in Education, 11:122-143.

Di Lecce V., Calabrese M. (2008), *Taxonomies and Ontologies in Web Semantic Applications: the New Emerging Semantic Lexicon-Based Model*. IEEE Int. Conf. on Intelligent Agents, Web Technologies and Internet Commerce, pp. 277-283.

Di Lecce V., Calabrese M., Soldo D (2009), *A Semantic Lexicon-based Approach for Sense Disambiguation and Its WWW Application*. Int. Conf. on Intelligent Computing. LNCS-LNAI. Vol. 5755/2009 pp. 468-477, Springer-Verlag Berlin Heidelberg 2009

Fellbaum C. (1998), *WordNet: An electronic lexical database*, MIT Press, Cambridge.

De Gasperis G. (2010), *Costruzione di Basi di Conoscenza AIML per chatter bot a partire da FAQ e Glossario*, Journal of e-Learning and Knowledge Society, 2(6), pp. 79-87.

Gruber T.R. (1993), *A translation approach to portable ontologies*. Knowledge Acquisition, 5(2):199-220.

Gruber T.R. (1995), *Toward principles for the design of ontologies used for knowledge sharing*, Int. Journal of Human and Computer Studies, Vol. 43, pp.907–928.

Navigli R. (2009), *Word Sense Disambiguation: a Survey*. ACM Computing Surveys,

41(2), ACM Press, pp. 1-69.

Navigli R., Velardi P. (2005), *Structural Semantic Interconnections: a Knowledge-Based Approach to Word Sense Disambiguation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(7), pp. 1063-1074.

Kerlya A., Hallb P., Bull S. (2007), *Knowledge-Based Systems*, Vol. 20, Issue 2, pp. 177-185.

Kerly A., Ellis R., Bull S. (2008), *Conversational Agents in E-Learning*, in T. Allen, R. Ellis & M. Petridis (eds), Applications and Innovations in Intelligent Systems XVI – Proc. of the 28th SGAI Int. Conf. on Innovative Techniques and Applications of Artificial Intelligence, Springer, London, 2008, 169-182.

Sneiders E. (1999), *Automated FAQ Answering: Continued Experience with Shallow Language Understanding. Question Answering Systems*. AAAI Fall Symposium. Technical Report FS-99-02, North Falmouth, Massachusetts, USA, AAAI Press, pp.97-107.

Whitehead S.D. (1995), *Auto-FAQ: an experiment in cyberspace leveraging*, Computer Networks and ISDN Systems, 28(1-2), pp. 137-146.