Applying Ontologies to Educational Resources Retrieval driven by Cultural Aspects

Regina Motz[1], Jacqueline Guzmán[1], Claudia Deco[2] and Cristina Bender[2]

[1]*Instituto de Computación, Facultad de Ingeniería,*
Universidad de la República, Montevideo, Uruguay
rmotz@fing.edu.uy

[2] *Facultad de Ciencias Exactas, Ingeniería y Agrimensura,*
Universidad Nacional de Rosario, Argentina
{deco, bender}@fceia.unr.edu.ar

Abstract

This work proposes the inclusion of cultural aspects management into an educational resource retrieval system in order to guide the semi-automatic semantic retrieval of electronic-materials. To this end, we define an ontology of cultural aspects that will basically provide a sound specification of the diverse cultural aspects and a semantic query refiner. Then, we explain why the ontology of cultural aspects is a key piece in this system.

Keywords: Web-Based Educational System, Ontology, Cultural Aspects.

1. Introduction

As the web has become one of the biggest repositories of knowledge, easily accessible for everyone, Information Retrieval has stopped to be an exclusive field for specialists, becoming something accessible to any person. In parallel, the use of electronic Educational Resources is increasing since e-learning became popular. Nowadays, students and professors are faced with the necessity of finding electronic educational resources that are more qualified according to their needs and characteristics, known as Cultural Aspects. This is usually a big task because of the following basic reasons: a) the great amount of existing electronic educational resources in the web; b) the difficulty to automatically manage different cultural aspects; and c) the difficulty for the user to correctly specify his/her search.

This work describes how these problems are resolved in the project EduCa [1] by using Ontologies. Ontologies are used for the semantic improvement of the educational resources search process. Moreover, the user´s background, objectives, learning styles and cultural environment are specified by an ontology and used in the clustering process of the educational resources.

The following section presents a brief overview of ontologies and information retrieval. Section 3 presents the considered cultural aspects. The system´s architecture is described in Section 4. Section 5 illustrates our approach with an example. Finally, some conclusions and future work are presented in Section 6.

2. Ontologies and Information Retrieval

Ontologies are generally used to specify and communicate domain knowledge in a generic way. While in a formal sense, "ontology" means study of concepts, one can use the word "ontology" as a concept repository about a particular area of interest. Tom Gruber [2] has defined “an ontology is a formal, explicit specification of a shared conceptualization”. Ontologies are very useful for structuring and defining the meaning of the metadata terms that are currently collected inside a domain community. They are a popular research topic in knowledge engineering, natural language processing, databases, intelligent information integration and multi-agent systems. Ontologies are also applied in the World Wide Web community where they provide the ability for making the semantics of metadata machine understandable.

Information Retrieval (IR) deals with the representation, storage, organization of, and

access to information items. The representation and organization of the information items should provide the user with easy access to the information in which he is interested [3]. Given a collection of documents and a query, the objective of a search strategy is to retrieve all the relevant documents to a user query while retrieving as few non-relevant documents as possible. A search strategy is a logical expression composed by different concepts combined with logical connectors: conjunction, disjunction and negation.

Unfortunately, characterization of a user's information need is not a simple problem. It is not simple due to the semantic complexity of vocabulary. IR faces with several problems. On one hand, authors and users frequently use different words or expressions when they refer to one concept. For example, in mathematics, "matrix" can also be expressed as "array". If in a document appears "array" instead of "matrix" this document would not be retrieved. This problem can be solved making use of synonyms. On the other hand, some words can have different meanings. For example, the word "matrix" can refer to a rectangular array of elements set out by rows and columns or to a container into which liquid is poured to create a given shape when it hardens. This is solved disambiguating the sense of the word.

Some statistics [4] indicate that the great majority of users do not know search techniques, and they have difficulty of clearly expressing their information needs, and therefore, they do not obtain the wanted results. Although users do not have to know IR techniques, they would improve their search results using an interface that implements these techniques using ontologies. Section 4 shows how WordNet [5] is used to this end in the EduCa project. Moreover, the educational resource retrieval in the EduCa project is driven by user cultural aspects, so that, an ontology that describes user's cultural aspects is also used to enhanced the search strategy.

3. Cultural Aspects

Cultural aspects are preferences and ways of behavior determined by the person's culture. In this project, the cultural aspects are just the features that distinguish between the preferences of students or professors from different regions. We decided to work over a group of characteristics identified in research about the person's preferences in the learning activities. Table 1 shows the cultural aspects selected up to this moment. Some aspects described in this section can be consider personal characteristics, cultural aspects, or cultural aspect of a kind of auditory or group of people like: professor of Engineer, student of Fine Art, or High School Students, etc.

Cultural Aspects
1. Degree of Impatience
2. Degree of Colorful
3. Attitude (Active, Passive, Reactive)
4. Treatment (formal, colloquial)
5. Language a. Idiomatic expressions b. Common Usage Verb Tenses
6. Learning Styles a. Type of styles (Verbal Holistic, Visual Holistic, Verbal Analytic, Visual Analytic) b. Degree of Generality (High, Medium, Low) c. Real case, or Not
7. Activity a. Type of Activity b. Type of Participation (Individual, Group, Collective) c. Resources used (documents, Diagrams, Exercises, etc.)

Table 1: A Cultural Aspects Taxonomy.

Degree of Impatience. The degree of impatience affects the user's preferences for a educational resource. For example: if the resource has extensive materials, if the materials require a special hardware ability in order to answer with a good performance and quality, or if the students have to navigate through various windows before finding the information, etc. the students' degree of impatience might be affected.

Attitude. If the student is a reactive person, the educational resource should offer dynamic activities, which change the work scenario to cause the reaction of the student. If the student is active, the resource should have a high level of interaction. And if the student is a passive person, he will prefer activities with a low level of interaction, for example reading material.

Language: The best way to communicate with a person is using her/his mother language. Moreover, it is better to use the idiomatic expressions and

common usage verb tenses of her/his culture. However, when using the mother-language is not possible, the professional profile of a user will be determinate that she/he feels good using other languages.

Treatment: In some cultures the interpersonal communication is quite *formal*, so students belonging to these cultures could find a more colloquial way of communication aggressive. On the contrary, cultures where relationships are established in a more *colloquial* way, could find a formal way of communication cold or impersonal. Consequently, being able to identify if the content of a course, the material or an activity are expressed in a formal or colloquial way, is another element to take into account when you want to find the best course, materials or activities. The aim of this research is not to transform contents, but to measure aspects such as conjugation and language like elements that will affect the quality of the final result, and that will be considered when choosing between two learning objects (courses, topics, exercises, etc.).

Learning Styles. The learning style determines the form to organize and represent the information to the student for his or her better comprehension and fast absorption. We use the following styles: Holistic Visual, Holistic Verbal, Analytic Visual, and Analytic Verbal. The Holistic style is associated with the parallel process of the information. In this kind of process the person adopt a global boarding of the apprenticeship. The holistic student preferred to see real applications or examples as soon as possible, because they go through between the theory and the real word since the beginning of the learning process. In this style you can find people that preferred the information presented with declarative text, they called *Holistic Verbal*, and the people that preferred the information presented with graphics, image, etc., they called *Holistic Visual*. The Analytic style is associated with a lineal or sequential process of the information. In this kind of process the person adopt a focal boarding, they exam the topics, one per time, in sequential order. This kind of student see real examples only if it is necessary to understand. In this style you can find people that preferred information in plane text, organized in small paragraph with one idea, this kind of Analytic style is called Verbal. People that absorbed quickly information presented through visual resources like image, diagrams, etc., have a learning style called Analytic Visual.

Activities. Activities can be classified according to the level of reaction and interaction that it will be required. Some activities require that the student implement a solution *(Make a Solution*). Other activities require less degree of reaction like selection a solution (*Selected Solution*). And finally we can find passive activities like hear, read or see some material (*See*, *Hear*, *Read*). Moreover, may be activities that require interaction between students; these aspects affect the degree of student preferences. We can find *individual* activities which don't require interaction, the *collective* activities required interaction between students but the end solution is individual, and finally the *group* activities that not only required a high degree of interaction between student but the student should be agree in one end solution. We call this aspect *Type of Participation.* Activities can be implemented using different tools, which can also affect the student's affinity with the proposed course. Some of these tools can be, for example, the use of: forum, chat, and additional software like a simulation, among others. This aspect is associated with the degree of interaction that the student is required into the system. Finally other important aspect is the material or *resources* used to present an activity and required to do this. The activities will be presented through text documents, diagrams, figures, etc., and will be required make exercises, read complementary material, etc. In general, these materials are called resources (text document, figure, diagrams, exercises, etc.).

All these cultural aspects are specified in our MultiCultural Aspects Ontology following the standard Learning Object Metadata (LOM) [6] and using OWL [owl7]. A complete description of this ontology can be found in [8] or the implementation in http://www.fing.edu.uy/inco/grupos/csi/esp/Proyectos/Educa.

4. Architecture for Educational Resources Retrieval

The key issue is that cultural aspects modeled by ontology are used to generate the user profile and to refine the semantic resources searching strategy. The user profile is obtained directly from the user by a set of queries driven by the ontology. A module called *Educational Resource Finder* performs the educational resources search process. This module has the *Terms Semantic Refner* module that generates the search strategy and the *Cultural Aspects Refiner* module that

incorporates to the previous strategy cultural aspects conditions of the user, in order to retrieve only those courses that match with his/her cultural characteristics, as depicted in Figure 1. In the scope of this paper we assume that exists a learning object (LO) repository with the educational resources enhanced with metadata that describes their cultural aspects values. This task is performed inside the EduCa project by an ongoing Master´s thesis work [9].

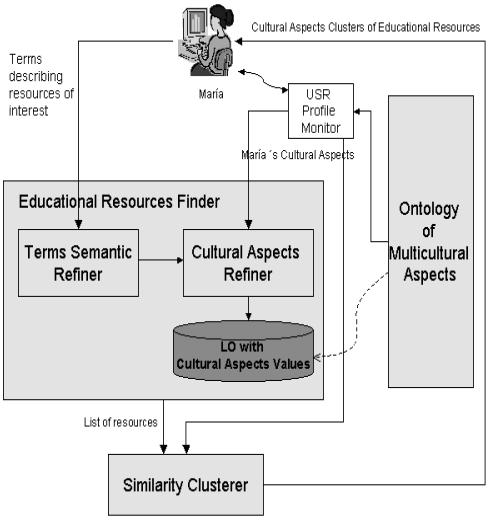

Figure 1: Architecture for Educational Resources Retrieval.

The *Terms Semantic Refiner* guides the user for sense disambiguation of the concepts submitted by him/her. This module allows the user to select concepts hierarchically related in order to reduce the amount of resources to retrieve, and it expands semantically concepts in order to increase the amount of resources to be retrieved. The architecture of the *Terms Semantic Refiner* module is presented in Figure 2. Modules painted in gray co lour indicate that a user's participation is needed.

When a user formulates a query, he/she gives to the refiner a set of terms $\{C_i\}$ with i = 1, … , n, where each term represent a concept. The result given by the refiner is a search strategy associated to these concepts. The module *Orthographic correction* verifies that terms are correctly spelled. A corrected term C'_i is given for each concept C_i. If C_i is orthographically correct, C'_i coincides with C_i. If C_i is orthographically incorrect, then it could be replaced by the correct form C'_i, with the user's previous acceptance.

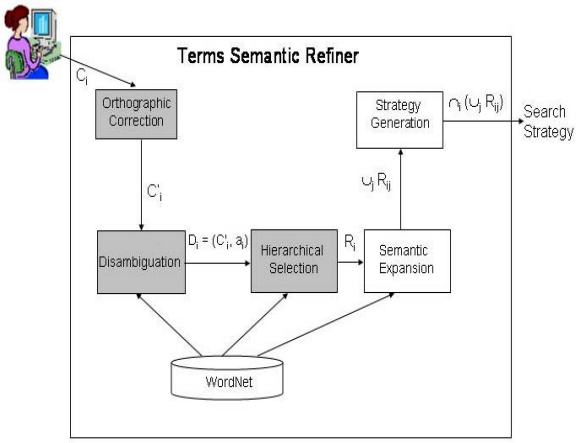

Fig. 2: Terms Semantic Refiner Architecture

The *Disambiguation* module takes for each concept C'_i, different senses associated to this concept and show them to the user. The user selects the sense that corresponds to his/her search interest. Each sense of a concept has an associate conceptual hierarchy. The result of this module is a disambiguated concept in the way $D_i = (C'_i, a_i)$, where C'_i is the initial concept, orthographically corrected, and a_i is the sense chosen by the user.

Hierarchical selection: It shows, for each given concept i, concepts hierarchically related with this concept. If there exist related concepts for some C_i, the user could move into this associated conceptual hierarchy. This would allow him/her to locate a nearer concept to his/her information need. This nearer concept could be a broader or a narrower term which could be in a higher or a down level, or moving across the hierarchy tree, it could be possibly in another branch of this tree. This hierarchical selection increase precision, and it is interactive because the user can choose these related concepts provided by the refiner. So, for each $D_i = (C'_i, a_i)$ in the hierarchical selection we obtain a R_i, which could be:

- D_i if the user has decided not to change the hierarchical level;

- J_i if the user has decided to replace the given concept with another J_i hierarchically related;

- $\bigcup_{j=1}^{s} J_{ij}$ if the user has decided to replace the given concept with a set of related concepts.

In general, the third possibility is presented when

the initial term is a general term, and the user is interested only in several hyponyms, that is to say, several narrower terms. In this navigation across a conceptual hierarchy, the user could decide to select a narrower concept, which could be ambiguous. In order not to require again user's participation, disambiguation here is automatically performed using the initial sense of the concept selected by the user.

The objective of the *Semantic Expansion* module is to recover resources that are also relevant, even though they do not correspond exactly with the terms given by the user. It consists on incorporating terms that are conceptually equivalent to the search terms. For example, a search involving the term "matrix", it could be expanded adding it synonymous "array". This expansion is automatic and it allows increase the amount of courses to retrieve. The result of this module is a group of r terms $\{R_{i1},..., R_{ik}, \ldots, R_{ir}\}$ semantically related with each concept R_i, with i = 1,…n.

The *Strategy Generation* module produces the search strategy associated to the user's interest. This strategy consists on the disjunction of the expansions of each concept and the conjunction of these expansions. If a certain concept is wanted to be discard, the concept is expanded in order to discard this concept and all its synonyms and related terms. Then the result is the AND NOT of these ORs. For a search that involves concepts: C_1 and.... and (not C_h) and ... and C_n the following strategy is obtained:

$$(R_{11} \text{ OR } R_{12} \text{ OR... OR } R_{1\,r1}) \text{ AND ….. AND } (\text{NOT } (R_{h1} \text{ OR } R_{h2} \text{ OR... OR } R_{h\,rh})) \text{ AND ..... AND } (R_{n1} \text{ OR } R_{n2} \text{ OR... OR } R_{n\,rn})$$

where $(R_{i1} \text{ OR } R_{i2} \text{ OR... OR } R_{i\,ri})$ is the expansion of the concept C_i.

A more detailed description of the *Terms Semantic Refiner* module can be found in [10].

Each of these strategies generates a cluster of documents according to the amount of the cultural aspects conditions satisfied by courses. These clusters are showed to the user by the *Similarity Cluster* module in a descending order of the amount of conditions considered.

5. An example

Let us suppose that Maria, an uruguayan physics student, wants to find courses about "dynamics" and she decides to ask, in English, for the more general concept "mechanics". The Semantic Refiner takes "mechanics" and verifies that it is orthographically correct. If the user had written "mecanics", the Orthographic Corrector would have suggested her the word "mechanics", which is orthographically correct. Then the Disambiguation module shows different senses of that word. In this case, it has two senses. Mechanics is the branch of physics concerned with the motion of bodies in a frame of reference; and mechanics are the technical aspects of doing something, e.g. mechanisms of communication. In this case, María chooses the first sense. After this, the Hierarchy Selection module expands this concept with its hyponyms using the ontology WordNet. María moves in the hierarchy and selects the term "dynamics", because she is interested in the branch of mechanics concerned with the forces that cause motions of bodies. The Semantic Expansion module takes this phrase, expands it, and automatically incorporates the term "kinetics", in order to incorporate a synonym. The Strategy Generator takes this set of terms and automatically builds the following search strategy:

dynamics OR kinetics

If a search involves several concepts, the Semantic Refiner does the process described above with each concept and then the Strategy Generator module combines them. As a result, the search strategy associated with this search consists on the disjunction of each one of the expansions and then the conjunction of the resulting sets of expansions.

The resulting search strategy is used by the Cultural Aspects Refiner module, which incorporates to this strategy, cultural aspects conditions of the user, in order to retrieve only those courses that match with his/her cultural characteristics. The generic search Strategy is:

Subject = (dynamics OR kinetics) AND Language = Spanish AND FigurePreference = High AND ExercicePreference = Low

In this example, there are three cultural aspects conditions:

C1	Language = "Spanish"
C2	FigurePreference = "High"

C3 ExercisePreference = “Low”

The strategies that generate each cluster are:

E1 = Cs AND C1 AND C2 AND C3

Which generates a cluster with courses that satisfies the subject condition and the three given cultural aspects conditions.

E2 = ((Cs AND C1 AND C2) OR (Cs AND C1 AND C3) OR (Cs AND C2 AND C3)) – E1

Which generates a cluster with courses that satisfies the subject condition and two cultural aspects conditions.

E3 = (Cs AND C1) OR (Cs AND C2) OR (Cs AND C3) - E2

Which generates a cluster with courses that satisfies the subject condition and one cultural aspects condition.

6. Conclusions and Future Work

There are a number of E-Learning Systems based on ontologies , ones of the most relevant are Edutella Project [11], SeLeNe Project [12] and the work of Aroyo [13] These projects include ontology-driven subject domain, repository of learning resource and adaptation. A domain ontology provides a broadly agree vocabulary for domain knowledge representation and how they are interrelated. Course/learning tasks are typically described in terms of subject domain concepts and some instructional relationships (such as “pre-requisites”, ”uses”, etc.) between the involved concepts. Their emphasis is on the reusability of the educational material and they offer personalized access to the educational material according to the learning model. However, the idea of the proposed Cultural Aspects Model described in this paper is not to prescribe any learning style or properties, but rather provide a formal specification where different cultural aspects can be described and used to characterize the cultural aspects of learners and web-base materials and activities. Therefore, by means of the cultural aspects definitions the system would be able to be retrieving the educational resources with relevant cultural aspects associated to them, according to each user’s profile. The Semantic Refiner acts as the specialist in information sciences, and prepares an appropriate strategy, and it solves most of the problems related with search contingencies, such as disambiguation of ambiguous terms, correct use of disjunction and conjunction, correct use of parenthesis, inclusion of synonyms and words with different ways of spelling, use of specific terms, correct use of negation and typing errors.

Future work is the development of a prototype of the architecture described in this paper. This prototype must include the generation of the repository of learning objects with cultural aspects values and the development of the user’s profile monitor.

7. References

[1] Red de Educación con Calidad Cultural (EduCa) www.fing.edu.uy/inco/grupos/csi/esp/Proyectos/Educa

[2] GRUBER T. R. A translation approach to portable ontology specifications. Knowledge Acquisition, 5:199-220 1993.

[3] Baeza-Yates, R., and Ribeiro-Neto, B. (eds.), *Modern Information Retrieval*. ACM Press, New York, 1999.

[4] M. Kobayashi, and K. Takeda, “Information Retrieval on the Web”, *IBM Research*, Tokyo Research Laboratory, IBM, Japan, 2000.

[5] Voorhees, E., Using Wordnet for Text Retrieval, in Fellbaum C. “WordNet, an electronic Lexical Database”, Mit Press. 1998.

[6] The Implememtation of some LOM elements: http://kmr.nada.kth.se/el/ims/schemas/lom-educational and http://mkr.nada.kth.se/el/ims/schemas/lom-technical

[7] OWL Web Ontology Language Guide http://www.w3.org/TR/owl-guide/

[8] J. Guzmán and R. Motz. Towards an Adaptive Cultural Web-based Educational System. *Submitted to publication.*

[9] Mariusa Warpechowski.Metadata retrieval for Electronic Object: the case of Learning Objects. Masther´s Thesis. Universidade Federal de Rio Grande do Sul, Brasil, *to appear*.

[9]PROTÉGÉ: Ontology editor: http://protege.stanford.edu/download/download.html

[10] C. Deco, “Propuesta de un refinador semántico para la recuperación de información desde la web”, *Tesis de Maestría*. Instituto de Computación. Facultad de Ingeniería. Universidad

de la República, Montevideo, Uruguay. Diciembre 2004.

[11] Edutella Project: Wolfgang Nejdl, Boris Wolf, Changtao Qu, Stefan Decker, Michael Sintek, Ambjorn Naeve, Mikael Nilsson, Matthias Palmer, and Tore Risch. Edutella: A P2P networking infrastructure based on RDF. In Proceedings of WWW-2002. ACM Press, 2002. Proyect http://edutella.jxta.org/

[12] SeLeNe Project: Kevin Keenoy, Alexandra Poulovassilis1, Vassilis Christophides, Philippe Rigaux, George Papamarkos, Aimilia Magkanaraki, Miltos Stratakis, Nicolas Spyratos, and Peter Wood. Personalisation Services for Self e-Learning Networks. Disponible en http://www.dcs.bbk.ac.uk/selene/reports/

[13] Aroyo,L. and Dcheva, D. *Authoring Support in Concept-based Web Information Systems for Educational Applications.* International Journal of Continuous Engineering Education and Life-long Learning. 14(3),2004.