

Learning Objects Recommendation System: Issues and Approaches for Retrieving, Indexing and Recomend Learning Objects



Ricardo Azambuja Silveira^a, Rafaela Lunardi Comarella^a,
Ronaldo Lima Rocha Campos^b, Jonas Vian^c, and Fernando
de la Prieta^d

^aUniversity of Santa Catarina, Department of Informatics and Statistics, Campus Reitor João David Ferreira Lima s/n, 88040-900, Florianópolis, Brazil. email: {ricardo.silveira, jerusa.marchi}@ufsc.br

^bInstituto Federal Catarinense, Campus Sombrio s/n, 88960-000, Sombrio, Brazil. email: guilherme.klein@ifc-sombrio.edu.br

^cUniversity of Santa Catarina, PPGCC, Campus Reitor João David Ferreira Lima s/n, 88040-900, Florianópolis, Brazil. email: thiago.gelaim@posgrad.ufsc.br

^dDepartment of Computer Science and Automation Control, University of Salamanca, Plaza de la Merced s/n, 37008 Salamanca, Spain. e-mail: fer@usal.br

KEYWORD

ABSTRACT

Learning Objects; Repositories; Search; Recommendation Systems; Multiagent Systems; Ontology This paper discusses some important issues regarding the the management of Learning objects covering searching over repositories and different approaches of recommendation systems and presents a multiagent system based application model for indexing, retrieving and recommending learning objects stored in different and heterogeneous repositories. The objects within these repositories are described by filled fields using different metadata (data about data) standards. The searching mechanism covers several different learning object repositories and the same object can be described in these repositories by the use of different types of fields. Aiming to improve accuracy and coverage in terms of recovering a learning object and improve the relevance of the results we propose an information retrieval model based on a multiagent system approach and an ontological model to describe the covered knowledge domain.

1. Introduction

The efforts and the necessary investments held by educators and the educational institutions to produce qualified learning content are considerable. This is especially important in the production of content to be used in Distance Education. Because of the lack of the teacher physically near the student, all the information and knowledge available to the student must be complete and should be explicitly covered to fill the subject matter and help



the student to achieve the expected learning objectives. Owing to the large cost of producing learning content, there is a strong advantage in improving the possibility of reusing all the produced learning material. Therefore, the development of mechanisms to facilitate the reuse of learning content has attracted the interest of several research groups, organizations and educational institutions around the world (?; ?).

Nowadays the most well-known approach used to improve the reuse and the organization of learning content is that of Learning Objects (LO) (?). A learning object is any digital resource or artifacts designed to support the learning process, promoting reuse and adaptability (?). Learning objects may be since small parts or packages of knowledge, such as images, sounds, videos until more complex devices, such as complete courses or software application for virtual laboratories (?).

There is a large volume of educational content on the web, which could not be directly accessible through conventional search engines. This information or knowledge is said to belong to the so-called hidden, deep, or invisible web, as opposed to the contents of the website which are referred to as visible. Usually, LOs belong to this class of knowledge because they are stored in widespread digital repositories called Learning Objects Repositories (LORs). These repositories usually belong to academic institutions and universities (?).

According to these authors (?), There are several types of LORs' architectures designed to achieve interoperability between LOs they keep stored. The main standards used in these architectures are: IMS DRI (?), CORDRA (?) or IMS LO Discovery and Exchange (?). These standards generally provide some kind of search interface that allow to access the available LOs in the repository through queries which can be made to the database of the repository that hosts them. Any interaction for retrieving LOs can be done manually or be automated through different software systems.

Repositories of learning objects are therefore mechanisms designed to store and organize learning objects so that they can be located and retrieved. The learning object repository provides the interface for submitting, reviewing and recovering any learning object hosted in its knowledge base (?; ?). The recovery of a LO is usually done through an appropriate search engine, and the objects retrieved from a repository is delivered and can be used directly by students, teachers and learning designers for authoring courses or used directly by any Learning Management Systems (LMS).

The IEEE (?) defines learning objects as any material, digital or not, that can be used, reused and referenced in e-learning, but, according to McGreal (?), certain features of a learning object should be highlighted: accessibility, interoperability, adaptability, reusability, durability, retrieval, assessment and interchangeability. Learning objects can be based on text, animations, presentations, images, software, but it is essential that it should be described by a set of metadata (data about data), according to some formal specification (?). From a practical point of view, a learning object is some learning content material formally described by its metadata, stored in some repository, which can be combined with other learning objects to create larger objects such as lessons and courses (?).

During recent years, many research groups of reputable organizations have proposed formal models as standards and specifications for learning objects. Vian and Silveira list some of these organizations (?):

- Ariadne Foundation (?)
- Aviation Industry Computer-Based Training Committee (?)
- Dublin Core Metadata Initiative (?)



- Institute of Electrical and Electronics Engineers - Learning Technology Standards Committee (?)
- Instructional Management Systems Global Learning Consortium (?)
- International Organization for Standardization (?)
- Advanced Distributed Learning (?) (ADL 2010)
- Obaa (?)
- ISO-IEC (?)

According to (?), the metadata set used to describe the educational content within the repository should be managed and, whatever the internal mechanism used by de LOR to the management of these LOs, it requires a communication protocol for metadata searching. These protocols enabled in the repositories contain search functionality and/or harvesting. There us at least five different protocols available to achieve interoperability between repositories:

- Z39.50 (?),
- Open Archives Initiative Protocol for Metadata Harvesting (?),
- the simple query interface (SQI) (?),
- the search/retrieval via URL (?),
- simple publishing interface (SPI) (?).

And last but not least, there are several different technologies used to storage and management of the knowledge base of these repositories. Some repositories are based on database management technologies, others use Lightweight Directory Access Protocol (LDAP), or eXtensible Markup Language (XML) technologies for ad hoc implementation but, recently specific specialized frameworks such as DSpace (DSPACE 2010) and FEDORA ((Flexible Extensible Digital Object Repository Architecture) (FEDORA 2010) have been used.

Some important projects about learning object repositories were undertaken an must be mentioned in order to provide a comprehensive view of the complexity and intricacy of diferent type of the architectures of these repositories: CESTA - Collection of Entities Support the use of Technology in Learning (?), Lume - digital repository of the Federal University of Rio Grande do Sul (?), the International Bank for Educational Objects (?) RIVED (Network Interactive Virtual Education) (?) and MERLOT (Multimedia Educational Resource for Learning and Online Teaching) (?). Currently there is a tendency to organize the repositories in hierarchical federations such as: the Federation Educa Brazil (?) and Repository Federation of Learning Objects Colombia - FROAC (?).

This scenario explains the heterogeneity of the available technologies used to produce and store learning objects. This heterogeneity makes painful and laborious the interoperability among the different tools used to



produce, store and retrieve learning objects. Moreover, it is clear that the lack of effective and specialized LO search tools does not allow wide reuse of the learning objects produced.

Useful search tools that effectively recover an expected LO in a large sets of repositories should deal with issues such as treating with different types of metadata, different architectures and repository technologies, resolving semantic ambiguities and automate as much as possible the indication of objects based on anthologies. Furthermore, it is expected that such tools are not limited solely to the specific search using keywords provided by the user, but also act as recommendation systems

Most of existing tools used to retrieve information about learning objects are usually solely based on syntax search. This type of search is not the most effective way to get learning objects as it usually is when retrieving documents on the web. Research in this area has highlighted many different approaches to solve the problem of search and recommendation of learning objects in federated set of repositories. These approaches can be based on robust search mechanisms fired from users demand combined with recommendation systems based on user profile, similarity of others users, pair recommendation and ontological similar knowledge.

This paper discusses some important issues regarding the management of Learning objects searching mechanisms and different approaches of recommendation systems and presents a multiagent system based application model for indexing, retrieving and recommending learning objects stored in different and heterogeneous repositories which provides to the user a ranking of learning objects based on their profile and semantic treatment of the keywords provided by the user

This paper is organised as follows. Section 2 presents an overview of the theoretical background and related works found in the literature that are slightly related with our proposal. Section 3 presents the proposed model. In Section 4 we discuss some evaluation scenarios of application and presents some final remarks.

2. Background

As we pointed out in previous works (?; ?), the theoretical background of the researches about learning objects recommendation system covers most of the different issues and approaches for retrieving, indexing and recommend learning objects including the basic concepts related to learning objects as well as formal specification of metadata for the LOs.

Furthermore, multi-agent systems technology is often used to allow the combination of different search techniques to use diverse protocols for access to heterogeneous databases. In addition, retrieval techniques, usually used for Semantic Web, could help to solve the problem of retrieving objects in distributed and heterogeneous environments using semantic aspects of learning object. The obtained results from research involving semantic search point out to using representation of domain ontology in order to contextualize the domain knowledge in search problems over specific concepts and also using indexing and weighting information techniques to improve the accuracy, coverage and the performance of search tools.

As mentioned above, Learning Objects are educational resources that can be used in the learning process supported by technology (?). Learning Objects can be based on text, animations, presentations, images, or software application, and have to be described by a set of metadata, according to some formal specifications (?). The learning object is the learning content formally described by its metadata, stored in repositories, which can



be combined with other learning objects to create larger objects such as lessons and courses (?).

Digital content repositories or Learning Repositories are software systems specially developed to store and organize digital resources, and thus providing search and content retrieving mechanisms (?; ?). Digital repositories have to deal with a series of requirements such as storage, cataloging, dissemination, harvesting, search and retrieval of Learning Objects by using user interface for submission or cataloging content, according to one or more of the aforementioned metadata standards to describe it as well as using mechanism for dissemination and harvesting interfaces, communication protocols, and search and retrieval interfaces.

It is not currently possible to distinguish, at the current state of the art, just one particular accepted and adopted model to be used to guide all of these policies in repositories implementation. There is thus, a heterogeneous scenario which has been addressed in several ways and, according to (Gil et al., 2011), to create highly sustainable repositories they must deal with various types of storage systems, different kind of access to objects and methods for consulting. According to Fabre, heterogeneity is more technically related to highlighting differences in both the implementation of the repositories (software) as well as differences in hardware that supports them (Fabre et al., 2003).

Thus the concept of heterogeneity for search engines can be set upon two major pillars: access to the resources, and metadata standards used to describe objects.

In this article we address these two areas in the model of agents, while mainly focusing on techniques that deal with different metadata standards.

Such approach is justified by the difficulty in representing the object semantic in its recovery and integration with other systems, (?) expose that the adoption of only one metadata standard for the integration of various systems is complicated, the demands are different from one system to another and for that reason a mapping technique called Crosswalks is used.

To ensure interoperability, the repositories are adopting one or more disseminating protocols. These protocols came out of the efforts to find efficient forms to replicate not only the structure, but the object of the repositories and end up defining a standard form to obtain access to these objects. In the same way these repositories are different according to the technology and the protocols used: Lightweight Directory Access Protocol (LDAP), Open Archives Initiative protocol for Metadata Harvesting (OAI-PMH; Open Archives Initiative Object Reuse and Exchange (OAI-ORE).

What can be noticed is that even though the access mechanisms to the resources (objects and metadata) deal with a range of technologies, the emergence and adoption of standards to make the systems inter-operable has helped to standardize such accesses. Regarding the description of objects in the current scenario, there is also a range of patterns that have emerged about demands and different ideas and to attend several purposes. Thus the concept of metadata is related to information structures that describe on various aspects, its own resources and, such concept is commonly referred as being the data of the data, or the information about the information (?).

One of the main reasons to use metadata is to simplify information recovery in a relevant way. Moreover, it helps organizing, facilitates interoperation and resource integration, digital identification, filing and preserving. Among the main existing Standards, the most relevant ones and, therefore most adopted are: Dublin Core, LOM and OBAA.

Dublin Core metadata set emerged in 1995 in a workshop sponsored by OCLC and NCSA. Its continuing development and specifications are managed by Dublin Core Metadata Initiative (DCMI) [9]. The initial objective



was the creation of some fields to describe web resources by their original authors. With the growth of electronic resources and the lack of descriptors to catalog them, the standard defined some elements and simple rules to enable such cataloging.

Created by IEEE Learning Technology Standards Committee (LTSC), the Learning Objects Metadata (LOM) is an IEEE standard meant for the reuse and description of educational resources (?). LOM defines a minimum set of attributes to manage, locate (language) and validate educational objects.

OBAA (?) is a Brazilian technical and functional requirement specification standard for the production, edition and distribution of interactive digital content, allowing them to be used in Web platforms, mobile devices and digital television. This standard was developed by the Federal University of Rio Grande do Sul (UFRGS) in partnership with Vale dos Sinos University (UNISINOS) in response to a call from the Education, Communication, Science and Technology Ministries and used based on the LOM standard. OBAA is an extension of LOM. There were included some elements in the Technical and Educational categories and added another two; Accessibility and Segment Information Table to meet the Brazilian needs with relation to these segments (?).

Regarding theoretical background of software agents we consider a multiagent system with a network of problem solvers who work in conjunction to solve problems which go beyond their individual capability (?). These problem solvers are essentially autonomous, distributed and heterogeneous in nature. We also adopt the FIPA reference model of agents (?) and use JADE (Java Agent Development Framework) (?) to construct it.

Besides the theoretical background related to multiagent systems and learning objects we consider retrieval techniques used for finding documents. These techniques aim to find the location of information that satisfies specific needs, from digital media (?) and techniques for knowledge representation to ontologies consisting of entities that can be classes, concepts, instances, individuals, relationships, properties, data types and values.

3. Proposed Model

The objective of this model is to promote better recovery of learning objects, through support of software agents in different repositories with heterogeneous access and metadata. The proposed multiagent system model was developed to be able to index, classify and retrieve learning objects in different repositories besides ensuring good coverage and recovery for areas of knowledge and prioritize relevant results.

According to Figure 1, the main components are: the multi-agent system, the system of web search, the set of repositories of learning objects, and the foundations of knowledge and profiles. The web search system is responsible for the interface of the multiagent system with the user, receiving user's profile information, and the search terms. Repositories can be completely different, both in content and in technology that was developed and for each type of repository there is an Agent Indexer adapted to handle a heterogeneous environment. Knowledge bases are accessed by the Agent Profile and are responsible for expanding the terms to search and filter the objects returned, since the base profile is responsible for storing user information and is consulted by the Agent Profile. Figure 2 indicates the information flow system.

Figure 2 indicates the information flow as follows: When the system user accesses the web search system, he is identified and thereafter he may send a search request, through terms (keywords). The web search system receives this request and passes the data request to the multiagent system using a remote procedure call, in a



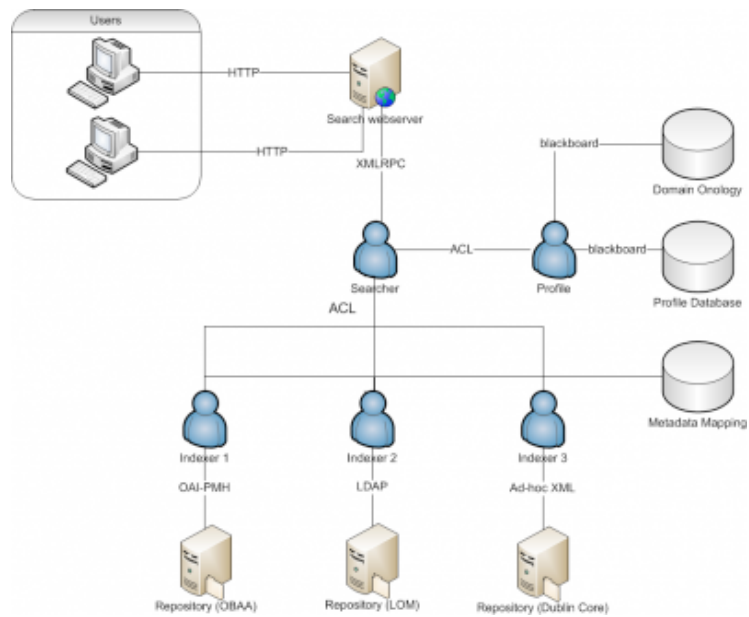


Figure 1: Multiagent system architecture

vector containing the search terms and the user ID (if existent). The Search Agent receives the request and sends a message containing the terms and user ID to the Profile Agent, prompting which databases to search and the search terms treated semantically. With this information, the Profile Agent sends a message to the Profile Database requesting if the user profile exists and user's information. Thereafter, he asks the user about the preferred databases and chooses the specific knowledge base, subjecting it to the search terms to be expanded and detailed. Otherwise, a set of default database is selected and he tries to infer the knowledge area of the terms, if possible, the terms are subjected to a knowledge base for expanding and detailing and following the flow pattern, if not, the terms do not suffer modifications. After, to define the databases and search terms, the Agent Profile informs the selected Searcher Agent about the databases which triggers the Indexers Agents by sending messages containing the search terms. Such agents will carry out the search in the repositories indexed by them, based on the fields abstract, title and keywords. Once the search is performed, the agents send messages to Searcher Agent, containing the identification of the objects recovered by them. The Search Agent receives the list of objects and sends it to the Profile Agent. The Profile Agent checks the user information to sort the list of objects. Then the objects are sorted based on the user profile and the sorted list is sent for viewing. If the user profile is not found, the objects are subjected to a similarity function, to avoid possible objects out of context and duplicates.

3.1 Description of Agents

The conceptual model was designed to index any type of repository, using a specialized index agent. The implementation of the model was performed using the JADE framework. The communication between agents is established by exchanging messages in ACL (Agent Communication Language). Specifically, the communication between the agent Profile and profile database is performed by blackboard and similarly the communication of agent Profile with the knowledge base is also established by blackboard. We have implemented an agent who is specialized in communication protocol and an OAI-PMH in the LDAP protocol. Both are capable of indexing any repositories that disseminate their data according to these protocols and adopt the metadata standards: LOM, Dublin Core or OBAA. The prototype is used to perform indexing in two repositories, UNA-SUS (UFSC) and CESTA 2 (UFRGS). This choice was made because both implement the OAI-PHM protocol and use the LOM metadata standard. The CESTA 2 repository was chosen because it is a multidisciplinary repository with a large set of objects. The UNA-SUS UFSC repository was chosen as well because of its specificity in public health, so you can use all the features of the knowledge bases to expand the search terms since the repository was indexed using the terms of DECS (Brazilian version of MESH). Searcher Agent The Searcher Agent acts as a coordinator for the entire process. It interacts with the web system, receiving search requests, and returning the results; ordering the databases to be searched and the terms extended to Agent Profile; requesting objects that meet the search parameters to the Indexer Agents; returning the selected objects to Profile Agent and sorts them by relevance. The Searcher Agent receives requests from the users, who enter keywords to search and request information about the user agent profile. This information, when available, is stored and used to classify objects. After receiving the search terms the agent uses a system of domain ontologies to expand the terms and build a network of knowledge and then the Searcher Agent requests the Indexer Agents for objects and their description (metadata). With the information from the Indexer Agents, along with the terms and expanded user information, the Searcher Agent ranks the results and sends them to the web search system. For this purpose the



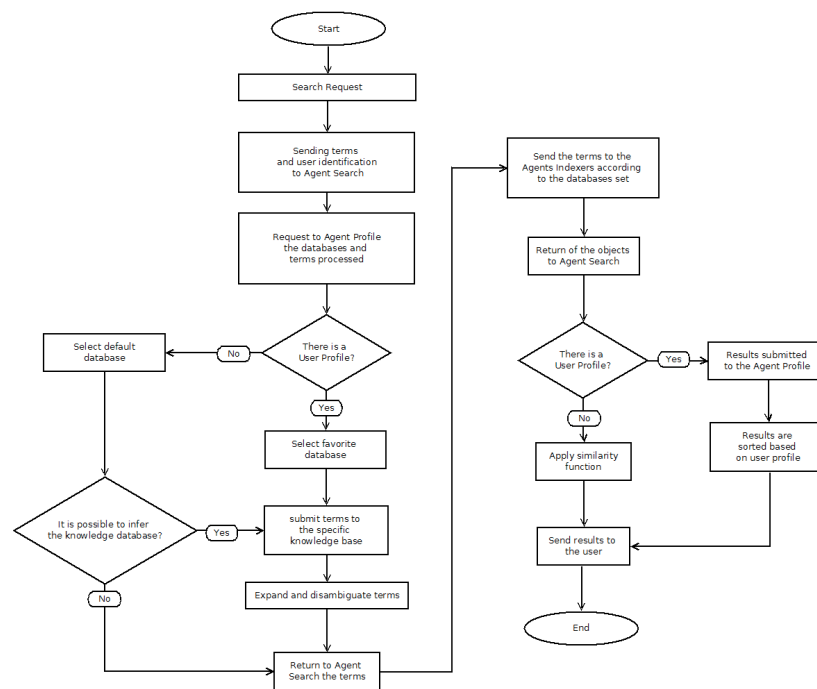


Figure 2: Multiagent system architecture



agent communicates with the web search system using a remote procedure call protocol in XML (XMLRPC). The prototype interface was developed in PHP language, with the possibility of user identification. The system is requested using protocol by calling the Search method with a vector containing text information to identify the user and the terms to be searched. As an answer to the call, the web system also receives a text vector, with information of the objects returned by the multiagent system.

Profile Agent The Profile Agent, in the multiagent system, is responsible for enriching the search, producing specialized and customized search terms according to the user. To perform this task, the agent interacts with two bases: the user profile and a dictionary of terms and knowledge. The Profile Agent is able to retrieve the user's profile, select the appropriate databases according to the user, expand and detail the terms to search, sort and filter the returned objects by relevance and assist in the construction and refinement of profile user. The Agent Profile begins operating when it receives a request from the Searcher Agent, containing user ID information and the search terms. With the user's identification, the Profile Agent requests for database profiles, if the user's profile exists. Thus the technique that defines user ID must be known by both the agent and the database, and both must be able to communicate in a standard way, using ad-hoc language or blackboard. Being able to identify and obtain information from the user's profile, the agent sends a message to the profile database requesting which databases the user tends to use in the search and which knowledge base that it uses to treat the terms to search. To choose the knowledge base, the Agent Profile first checks if there is a user profile and if so, it indicates a base, otherwise it tries to infer which base. This inference is carried out by subjecting the terms of a similarity function with the terms contained in the database; the selected base is the one that has the highest degree of similarity. After choosing the knowledge base, the agent communicates with it by ad-hoc language or blackboard. To this base, the agent will submit the terms and receive the expanded terms for the search. As an example, the agent provides a communication with the base specializing in medical terms, submits the terms CVA (Cerebrovascular Accident) and Blood Circulation, the base receives the terms, synonyms and expands the communicating agent, in this case the term is synonymous CVA, Cerebral Vascular Accident, Cerebral Stroke, Brain Vascular Accident, etc. and that Blood Circulation is related to Cardiovascular System. In the second phase, the Agent Profile is tasked with sorting out objects, resulting in the repositories search for relevance. If there is no user's profile selected, the agent will use a similarity function of terms with objects and thus ordains them. If the user's profile exists the agent will also apply a similarity function, albeit using metadata values most used and stored in the user's profile as a filter, giving those objects that fit those values with greater relevance. As an example, chosen the year 2005, the user's profile contains the date field set to dates greater than 2005, when sorting objects, with objects whose metadata Date meet this definition they receive a higher weight and therefore are ordered as having greater relevance. Once the results are classified the agent prepares the objects to be sent as an indexed list with relevant information about each object, its metadata fields content and its location. The list is submitted to the Searcher Agent, which passes on to the web search system responding to its request. According to this information the system assembles a web visualization of the learning objects information to the user, highlights the most relevant objects and builds a profile update function according to each object shown. This function retrieves the user profile information as it accesses some items shown in the web interface, analyzes the profile fields that refer to metadata, and establishes standards for the calculation of the frequency of the field values, thus updating the profile.

Ontology The expansion of the search terms, responsible for ensuring greater coverage, is performed by the Profile Agent using the available knowledge bases expressed by specific domain ontology. This



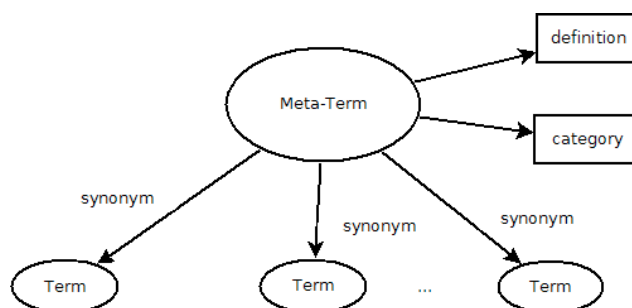


Figure 3: Multiagent system architecture

ontology is a dictionary of terms that expands beyond the terms, through their synonyms, defines and categorizes them in order to qualify by avoiding ambiguities and terms outside their area of expertise. As an example, in the medical field CVA can mean Cerebral Vascular Accident (more common term) or a reference to a chemotherapy protocol. Viruses can be computer virus or biological virus or the medical field, or even the social area. This ontology follows a standard model, to be consulted by the Agent Profile, as shown in Figure 33.

According to Figure 3, the terms are related to a meta-term relationship for a synonym. This meta-term defines a concept or idea at a high level, containing its definition, organizational category and for terms that represent that concept. Since the relationship of meta-term with terms is determined by synonyms, locating a term, you can get your meta-term and following all its synonyms. Thus accomplishing the expansion of the terms. Making use of the same knowledge base, the agent uses the profile definition and synonyms of the terms to calculate the relevance of the objects. Profile Database The database profile is used by the agent profile for user information. It can be implemented with different techniques, since the technique accepts messages from the Profile Agent. Each profile contains a series of specific information that define the characteristics of the objects to be recommended, they are: the most used database (preferred) default values for metadata fields, relations with other profiles and knowledge databases. Initially the system has a set of specialized profiles, focusing on the specifics of each knowledge area addressed by the system. This profile contains information about the most frequently used specialized search databases referring to the area in question and information about which knowledge base is used to treat the terms. Every user who signs up for the system indicates the area that the users belongs to and thereby inherits the specific information in this profile, given the characteristics form relationships with other profiles. As an example, a new user accesses and fits the system, selects the medical field and thus his/her profile is related to the specific profile of the medical field, previously registered in the system. This new profile will inherit the information about which databases are from the medical field and knowledge base to expand the search terms. As the system indexes different databases, the user can choose to search on a subset, removing those search databases that are less relevant, or adding those databases that can bring additional results, those building a list of more useful and complete databases. This interaction is done directly by the web search system, using advanced search fields. The web system is then responsible for updating the base

profile. One of the important information fields contained in the user profile is the identification of the knowledge base. This knowledge base was previously built through the model specified in the search. This database is referenced by a specialized profile and is inherited by the user when he/she creates his/her profile. Technically, it is a reference, and may be implemented in several ways accessible by the Profile Agent via messaging. In the prototype the following fields are stored and refined: Date (range), Educational Background, Knowledge Area, Type, Target, Interaction and didactic strategy. These fields were defined based on interviews with two experts, who considered them the most relevant for identifying and filtering of learning objects. Indexer Agent The Indexer Agent was modeled to index different object repositories with heterogeneous access mode and several types of objects description. Thus, for each repository indexed in the system, there is a specific Indexer Agent. To perform this task the agent should be able to access the object repositories; obtain the information about the objects it contains, a description of these objects (metadata) and if available, collaborative information such as the number of hits and qualifications and build an index of the collected information. Access to the repository, objects and their descriptions is performed by a dissemination protocol. The tendency of the repository is to use patterns of dissemination and collection protocols such as the OAI-PMH. However there are cases such as repositories of BVS (Biblioteca Virtual em Saúde, medical repository), in which the mechanism is designed in a specific way (ad-hoc). There are repositories where access is via a data structure in LDAP, or search protocols such as SPARQL. Each Indexer Agent should then be specialized in a repository and its dissemination protocol mechanism, ensuring access to the objects it contains. Related to the harvesting protocol's capabilities are accessing the repository and obtaining the objects and their descriptions, because each protocol type has a different kind of technical metadata description, for example, the metadata collected by the OAI-PMH are represented in a XML structure. By defining the protocol used by the agent, the object descriptions structures and their metadata, should be also be defined, allowing the agent to understand and retrieve them. After obtaining the information about the objects and their known structure, the agent builds an index with the information collected. It is through this index that the Indexer Agent returns the objects that were requested by the Searcher Agent. This index must be standardized so that the various objects, from the various repositories, are described and available in a standard way. Using a mapping base, a list of relevant metadata is created to search and map the known standards for this list, creating a standard group of metadata. Thus, not all metadata contained in the objects are mapped, only the relevant ones, making the content simpler and robust. By adopting a standard, such as LOM, and map the objects to this standard, the entire system operates on a single standard, simplifying the implementation stage of the model and its dissemination. However, if there are some objects that are rarely described or which are represented in a simpler pattern (such as Dublin Core) some fields are not being described. Beyond this, if some objects contain fields that are not in default to be mapped, or its description is incomplete, the search results are impacted. Using the original description format of the object, instead mapping it, simplifies the agent's work, and ensures complete representation of the object. However brings additional complexity to the other agents in the system, they should be able to understand all metadata standards known to the system. In the prototype the constructed agents are specialized in the OAI-PMH protocol and are able to recognize Dublin Core, LOM and OBAA patterns, adopting the mapping 1-1 to the OBAA, namely the Dublin Core and LOM patterns are transformed into OBAA. This transformation is simple, because the standard OBAA is an extension of LOM, in other words, all fields present in LOM, are within OBAA. Concerning the Dublin Core standard, this mapping is made by the semantic agreement between fields, for example, the title metadata in the Dublin Core is



represented by dc.title and OBAA is represented by obaa.general.title. Performance Repositories can contain a large number of objects. Analyzing all of them to select the most relevant to a job search can be very costly. In systems where the response time is not a factor limiting the problem can become irrelevant, but in systems with time related response while large indexes can hinder the response time. To deal if this issue, we have two alternatives; using search algorithms on large indexes or using an auction system for Indexer Agents. To reduce the search time of the agent in the index objects can adopt any specific algorithm to search large indexes, such as a graph inverted indexes. Divide the index into sub-index ordered by areas or subareas of knowledge, or store the results of searches conducted more for a faster query. By using an auction system ensures the response time because the Searcher Agent opens an auction of a number of selected objects to the terms and finish in a certain time. Thus, each Agent Indexer query table as auction and is consulting its index, responds with the objects with greater relevance. At the end of the Auction Agent Searcher collects the results.

4. Final Remarks

The proposed model allows searching heterogeneous repositories using semantic features and user's information, increasing coverage and relevance of their results. The option of using an open protocol based on XML for communication between the multiagent and the web search application incorporated features distributed throughout the system, making it more accessible since any application using the protocol is able to communicate with the system. This feature allows the use of the tool in a more transparent way in other systems, like the LMS (Learning Management System). Thus a module for LMS Moodle is being developed, which will allow search and content implementation through its own platform. At the current stage of development the system is already functional, having implemented the following agents: Searcher; Index (LDAP and OAI-PMH protocol) for CESTA 2 and UNASUS/UFSC; and agent Profile is also integrated with the JENA framework being able to expand and establish relationships between the terms and synonyms referred to it. The Human-machine interface is through a web application that communicates with the Agent Searcher via XML protocol (XMLRPC library). It is already possible to search the repositories CESTA 2 and UNASUS/UFSC. For now there are two domain ontologies created, information security and medical subjects. To validate the model we submitted a list of terms to search engine repositories CESTA 2, UNASUS/UFSC and the multiagent system. Of the 22 collected and analyzed results of the first 10, in 5 cases the SMA results are relevant, 4 are the same as the other systems and 1 is irrelevant. Of the remaining results 9 are considered good (they meet the search terms) but not as good as the first 10. The model for classifying objects based on user's information, statistics and collaborative evaluations of objects, despite being in the initial phase, has already showing a good potential. And because of this potential, is being improved the ontology for the domain of medicine, with the aim of preparing a survey with specialists. The CESTA 2 repository recently changed your harvest protocol to OAI-PMH they was LDAP at the moment of our tests.

