International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 5 (2013) pp. 243-251

Personalization on E-Content Retrieval Based on Semantic Web Services

A.B. Gil[1], S. Rodríguez[1], F. de la Prieta[1] and De Paz J.F.[1]

[1] Department of Computer Science, University of Salamanca,
Plaza de la Merced, Salamanca 37008, Spain
{abg, srg,fer,fcofds} @usal.es

Abstract*: In the current educational context there has been a significant increase in learning object repositories (LOR), which are found in large databases available on the hidden web. All these information is described in any metadata labeling standard (LOM, Dublin Core, etc). It is necessary to work and develop solutions that provide efficiency in searching for heterogeneous content and finding distributed context. Distributed information retrieval, or federated search, attempts to respond to the problem of information retrieval in the hidden Web. Multi-agent systems are known for their ability to adapt quickly and effectively to changes in their environment. This study presents a model for the development of digital content retrieval based on the paradigm of virtual organizations of agents using a Service Oriented Architecture. The model allows the development of an open and flexible architecture that supports the services necessary to dynamically search for distributed digital content. A major challenge in searching and retrieving digital content is also to efficiently find the most suitable content for the users. This model proposes a new approach to filtering the educational content retrieved based on Case-Based Reasoning (CBR). It is based on the model AIREH (Architecture for Intelligent Recovery of Educational content in Heterogeneous Environments), a multi-agent architecture that can search and integrate heterogeneous educational content through a recovery model that uses a federated search. The model and the technologies presented in this research exemplify the potential for developing personalized recovery systems for digital content based on the paradigm of virtual organizations of agents. The advantages of the proposed architecture, as outlined in this article, are its flexibility, customization, integrative solution and efficiency.

***Keywords*:** Learning Object, Repositories, Federated Search, Web Services, CBR, Recommendation, Multi-Agent System**.**

I. Introduction

There is a large volume of educational content on the Web that is not directly accessible through conventional search engines. This information is said to belong to the so-called hidden, deep, or invisible Web, as opposed to the contents found in the more accessible surface web. The solutions developed by conventional search engines are very efficient for retrieving the visible Web contents. The simplified method based on centralized recovery model works well when the information sources have left their content exposed to web crawlers. However, this does not apply in the deep Web, where information can only be accessed via search mechanisms adapted to specific sources. This paper presents our research within this context as related specifically to educational content repositories.

The current environment presents an increasing variety of distributed repositories of educational content. Repositories are often highly heterogeneous, with different storage systems, access to objects with their own methods of consultation, etc. The problem of heterogeneity in database systems is an open issue in educational repositories. A large number of initiatives have been brought forth to standardize the processes and technologies that comprise the issue of heterogeneity, such as Content Object Repository Discovery and Registration/Resolution Architecture (CORDRA) [1], Digital Repositories Interoperability (DRI) [2] or Learning Object Discovery & Exchange (LODE) [3].

This paper is focused on Learning Object Repositories (LOR). Many of these repositories do not have a system that allows a higher level of abstraction between the internal and end users of the stored data. Others form networks using architectures to facilitate interoperability. Most are based on various standards that assign an abstraction layer that connects their internal characteristics with the exterior characteristics, allowing for greater automation and computerization for containing LO. Others, such as the MERLOT repository, implement architectures that enable interoperability of their contents by providing offline consultation mechanisms through federated searches using a web customer service Simple Query Interface (SQI) [4] (via WSDL specification [5]) or through Restful Web services [6,7]. Using the applications that make use of these web services, it would be possible to access the tagged information for learning objects. This information could be displayed in any of the metadata standards that exist, mainly Learning Object Metadata (LOM) [8] or o Dublin Core [9]. Consulting metadata repositories is the main way to obtain the information needed to locate learning objects, evaluate their usefulness, and retrieve them.

The objective of this paper is to present certain singularities in the effort to adapt semantic web technologies while recovering information in the field of online education. Mainly to show the importance of content recommendation systems based on available semantic information in the search for an ordered management of educational resources for online education systems. This study presents the AIREH tool (*Architecture for Intelligent Recovery of Educational content in Heterogeneous Environments*) [10], which makes it possible to search and recover educational resources encapsulated in the form of a LO. Similarly, a system can use a CBR (Case-Based Reasoning) system to recommend which educational resources might be of particular interest to the user, based on information from previous uses and searches. This system is based on Multi-Agent Systems (MAS) using virtual organizations (VO).

II. Gaps in Current Educational Repositories

The emergence of what can already be considered as the Learning Object (LO) paradigm has brought with it a number of advantages regarding the reuse of learning content. While LOs offer facilities related to content specification, and the search and recovery of educational resources, the process of innovation has also produced different challenges that have not yet been resolved. The problems impeding commitment to the interoperability of educational content can be grouped into two general areas. The first is related to the problems associated with the monolithic structure of learning object repositories such as lack of reliability or availability, high access times in some cases, erroneous results, poor results, etc. In summary, LORs do not allow comprehensive user management to solve the LO recovery task with the flexibility and power necessary to ensure easy interoperability of dispersed and heterogeneous sources.

The second important problem is related to the absence of automatic mechanisms that control the technical quality, semantics and syntax of Learning Objects, ensuring their correct specification in any of the metadata schemas that describe them. For example, the IEEE LOM standard (IEEE Draft 1484.12.1, 2002), specifies the conceptual schema that defines the structures of the data for instances of LO metadata. The basic schema of LOM [8] is composed of 9 categories (General, Life Cycle, Meta-Metadata, Technical, Educational, Rights, Relation, Annotation and Classification) and 47 elements. Although these 9 categories can describe the resources very well, LOM is able to embed other metadata standards using XML namespaces, like Dublin Core, etc. But the syntactical definition alone is insufficient, since there is no obligation for the attributes to be specified to ensure that any LO has a minimum quality that can be used within a particular educational context, as shown in Figure 1.

A series of problems in the repositories requires solutions that are adapted to the heterogeneity of each of these repositories, that are isolated, and that ensure real and effective interoperability of educational content globally. A solution will enable a centralized global search and the effective reuse of resources by the end user in a personalized way to access the contents. This requires raising the level of abstraction and looking at the classification of systems storing and searching for LOs. While in theory this can be seen as an advantage because it increases the number of results, in practice it has two drawbacks. The first relates to the response time, which increases considerably, and the second involves the repeated occurrence of LOs in the results.

Furthermore, this can be considered an additional challenge in the efficient management of services and elements involved in this type of platform. This is due to the inclusion of efficient management techniques using labeled tags that facilitate the storage, search, retrieval, etc. within the educational content.

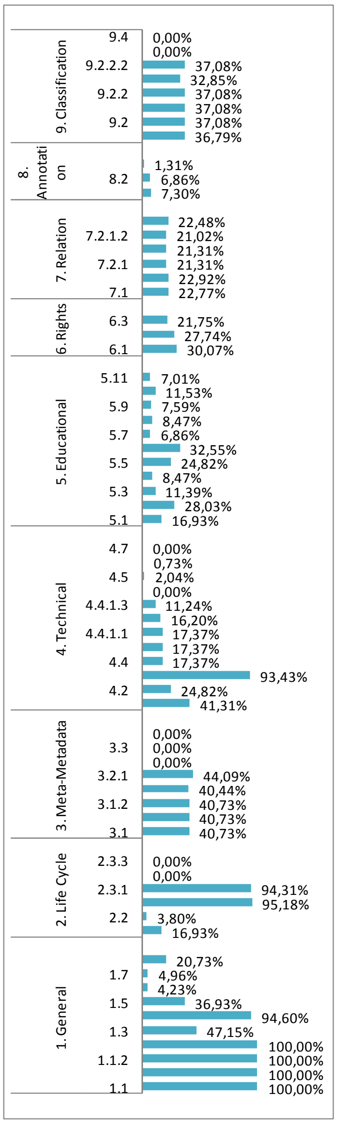

Figure 1. Percentage of LOM tag elements in use

A comprehensive solution for the problem of educational content retrieval goes beyond any simple recovery. What is needed is a filtering mechanism that includes semantic aspects of the objects retrieved and that can be evaluated by generating the most suitable results according to the user.

III. Related Work

With so many LOR, a major challenge is to find the most suitable LOs for the users as efficiently as possible. This objective has attracted much research in the field of the selection and recommendation of LO.

Researchers and developers of e-learning have begun to apply information retrieval techniques with technologies for recommendation, especially collaborative filtering [11], or web mining [12], for recommending educational content. A recent review of these applications can be seen in [13]. The features that handle these information filtering techniques in this context are the attribute information of education items (content-based approach) and the user context (collaborative approach).

One of the first works in this context was developed by Altered Vista: a system in which instructional techniques are evaluated based on collaborative filtering recommendation with close neighbors [14, 15]. These works explore how to collect user reviews of learning resources and propagate them through word-of-mouth recommendations.

RACOFI (Rule-Applying Collaborative Filtering) proposes a collaborative filtering by rules, with an architecture for the custom selection of educational content [15]. The author's recommendation is to combine both approaches to reduce recommendation by integrating a collaborative filtering algorithm that works with user ratings of a set of rules of inference, which creates an association between the content and rate of recommendation.

McCalla [16] has proposed an improvement to collaborative filtering called the ecological approach to designing e-learning systems. Key aspects of this proposal take into account the gradual accumulation of information, and focus on end users.

Manouselis *et al.* [17] have conducted a case study with data collected from the CELEBRATE portal users to determine an appropriate collaborative filtering algorithm.

Some solutions take a hybrid approach. [18- 21] make use of algorithms based on reviews from other users according to interests which are extracted through nearest neighbor algorithms. These correlation-based algorithms are used to calculate an index score on the usefulness of learning objects through the analysis of comments from students with similar profiles. These algorithms improve preference-based selection algorithms by incorporating aspects of student preferences. The preference pattern of each student is recorded in a history of preferences that is generated and updated according to comments from the student's preference. If a selected learning object has been given a positive score, its preference score increases for all the features of the learning object. The combination of the scores for a learning object is determined by the two algorithms that decide the position of the learning object according to the outcome of the recommendation. However, all of the selected learning objects are treated equally without any distinction between them, which would allow more precise assessment criteria of the user, affecting the very pattern of preference of the user.

The use of algorithms based on biological models, such as ACO (Ant Colony Optimization) is the basis of [22], which proposes a set of attributes based on a colony of ants (attributes- based ant colony system, AACS) to help students find their way through an adaptive model of learning objects more efficiently. This mechanism is based on the use of learning activities and educational elements to predict the optimal trajectories associated with the ACO algorithm, and recommend the sequence of learning objects. This work is interesting, but bases its recommendation on a path of learning through a different set of learning objects. The ultimate goal is to attain certain knowledge. The recommendation is the sequence produced by the optimal route between the different LOs.

The works by [23, 24, 25] suggest the need for selecting learning objects by taking into account the educational content described by their metadata, which falls in line with this thesis. They propose a mechanism called Contextualized Attention Metadata (CAM) to capture information about the actions along the life cycle of learning objects, including their creation, labeling, supply, selection, use and maintenance. These studies proposed four metrics to LOM and CAM for classifying and recommending the learning objects retrieved: Link Analysis Ranking, Similarity Recommendation, Personalized Ranking and Contextual Recommendation. These metrics classify learning objects according to criteria such as popularity ranking, the similarity of objects based on the number of downloads, and more. How these rankings contribute to the selection of learning objects and how they combine with each other are still open questions and a highly interesting field of study.

Based on semantic aspects that consider contextual information from the student's cognitive activities and the LO content structure, Qiyan et al. [26] propose a framework for recommending learning objects to suit the student's cognitive activities through an approach based on ontologies. The same approach follows the work of Ruiz-Iniesta [27] with a framework that simplifies the development of recommendations for LO.

There are other approaches, mechanisms or criteria for the categorization of educational content that require direct human intervention in their assessment, but list some criteria for assessing the quality of the content. Among these is the assessment contained in the MERLOT repository or LORI tool. The MERLOT repository (Multimedia Educational Resource for Learning and Online Teaching) offers the best current example of widespread application in the evaluation of educational content for Web-based education [28]. Content ratings are obtained through comments and ratings on a five point scale by users and reviewers appointed by MERLOT. This evaluation is based on three general properties: quality of content, potential effectiveness as a tool for teaching and learning, and ease of use. The classification is based on the quality of search results using a weighted average of these three classifications. The peer review process in MERLOT is carried out by two experts working asynchronously who return the descriptions of the contents recovered in a list sorted by the ranking, which has been established in turn by evaluating the quality of the content in descending order of assessment, where the contents are not evaluated at the end.

LORI (Learning Object Review Instrument) is a tool known to assess the quality of education resources on-line. It is simply an assessment protocol for learning objects in nine areas on a bridges point scale that can be implemented on-line by using rubrics, rating scales and comment fields. As an assessment tool, it is available at its website, which can be used to assess an individual or a panel of experts from a range of LOs, based on the advice of [28, 29]

There are a growing number of papers proposing systems to recommend learning resources, as evidenced by the lack of operational solutions and confirmed by recent work [30]. The evaluated proposals all concluded that the incorporation of mechanisms to assess attributes related to the educational content, as well as aspects of user context and their interaction with the content, create effective recommendation mechanisms.

However, a closer look at the revised proposals underscores the lack of applications on real systems and educational content. Most of the jobs listed in this section are based on simulations or have been applied to a local case study or a particular repository, with a priori control, for small groups of parameters that are usually local. The solution proposes that not all display results are from real context. Some of the recommended educational content objects are not learning objects, as defined in the present study, and the great majority do not therefore address aspects of semantic tagging of resources in their approach.

The architecture proposed in this paper provides multiple perspectives to assess the recovery of educational content from a real, open and scalable environment, and will also will be a support mechanism to implement the recommendation or ranking for the recovered LOs.

IV. Architecture Overview Aspects

The situation in the present context of education urgently requires a new type of application that can search for educational content in a distributed environment across different formats, servers and networks. This paper proposes AIREH (Architecture for Intelligent Recovery of Educational content in Heterogeneous Environments), as an intermediary architecture that introduces several needed components designed to simplify the problem:

- A translation feature which transforms a particular query language into one that is valid in existing repositories.
- A federation feature that sends queries to multiple repositories and reflects their responses.
- An aggregate feature that can unify metadata from different repositories, thus allowing the user the best possible choice.

In this environment, the architecture provides the optimal use of intelligent agents, which can now apply their characteristics (autonomy, status, reactivity, rationality, intelligence, coordination, mobility and learning) to a stable system, and can also react intelligently to the needs of the environment along several features. The idea of modeling the architecture as a virtual organization stems from the notion that an organization can adapt its actions to any change in order to achieve its goals and interact with heterogeneous components.

Given the heterogeneous and changing technological situation that accompanies the proposed educational context, and the need to reuse the data in a real operational context, we were motivated to design a model of an integrated architecture in which an organization of agents can execute search and retrieval actions of educational content based on a federated search model. The innovation of this architecture will be to provide an organization of agents with the self-adaptive capabilities needed to address the current problems, and with the ability to adapt to future changes in highly dynamic environments such as those discussed.

This model will solve the problems of the distribution and integration of different repositories, the abstraction of the internal logic of each repository, and the classification, storage and retrieval of LOs. In addition it will add the capacity of simple scalability, possible situations for use of new protocols, internal logical repositories, and cataloging or heterogeneous applications designed to cover service-related features.

A. Federated Search

The main contribution of a federated search is that the search process is done through search mechanisms in individual information sources. In addition, the search refers to the location of each source and provides a distributed control of information related to the different sources of hidden information. The federated search mechanism is thus a much more complex, rich and comprehensive centralized recovery model.

A federated search used to recover content in distributed heterogeneous systems, such as LO repositories, can be described as the sequence in the resolution of the three following subproblems:

1) Selection of Repositories

During this phase it analyzes the description of resources in the repositories and studies how to represent information that is distributed in them (response times, efficiency, etc.).

2) Selection of Resources

This phase determines the need for information and provides a set of descriptors in order to recover the results and decide which results are most likely to satisfy the query using a recovery algorithm.

3) Merger of Results

This phase builds the integration and combination of results returned by queries on the n-repositories, forming a single list that gives the user a ranked list of results.

The cornerstone of this architecture is the recovery of LO in a real environment using federated searches in different repositories. It is necessary to provide the user with a framework that unifies the search and retrieval of objects, thus facilitating the learning process that filters and properly classifies the learning objects retrieved according to a set of rules. The generation of the rules for the organization of the items recovered is based on educational metadata and will provide useful content to the end user. Mechanisms will provide documentation of the recovered objects, which can be evaluated, and will generate the most suitable position according to the user. The architecture provides multiple perspectives to assess the recovery of educational content.

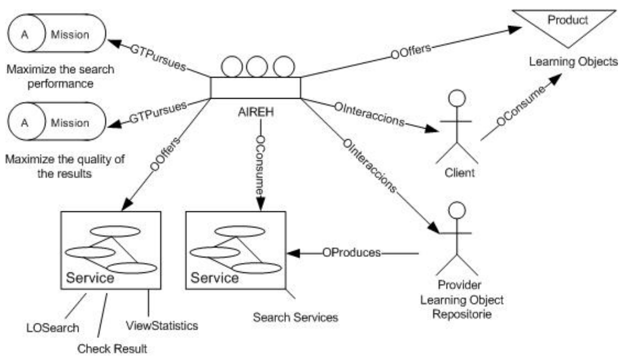

Figure 2. Diagram of the organizational model (by function)

B. Model Overview

The design and development of SMA methodologies need to support designers, and be both robust and reliable. Many traditional approaches detail the structure of the SMA in terms of a role model, which identifies the roles that agents play in the system and the interaction protocols in which they participate.

These methodologies can be classified as agent-oriented since they assume an individualistic perspective by using an agent with clearly defined tasks and skills to help the other agents achieve their individual goals. Closed systems, on the other hand, do not allow the participation of agents with behavior that is selfish or unauthorized.

The proposed AIREH architecture is seen as an intermediary communication point between the Learning Object Repositories (LOR), the LOs that they store, and users who use them. The system provides a federated search system. In addition, once the results from the different repositories have been received, an identification phase and filtering process adapt the results to the user preferences.

Figure 2 details the elements of the organizational model (functional view), showing the results (products and services) offered by the system, the type of environment, and interest groups.

To provide these services the platform requires providers, represented by the LOR, to offer search services that enable information to be harvested. Moreover, the product also offers statistical information on the performance of the repository and the use of LOs, identifying those that are used according to the search patterns.

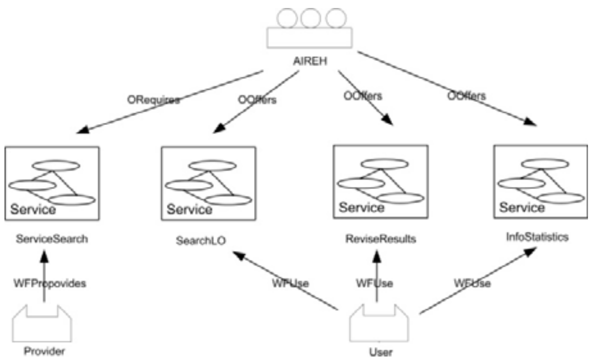

Figure 3. Organizational architecture model diagram

The mission of the organization is to maximize the system performance of queries by reducing time and increasing performance, and to maximize the quality of results.

Figure 3 shows the functional view (external function) model for AIREH associated to the organization, where services are connected to each other with associated roles and relationships (WFProvides/WFUses).

C. Roles Acquisition by Service Facilitator

The dynamism of the system, which is designed as an organization, can be reflected by the registration in different stages: registration of new players, new services, new protocols, service requests, and expulsions from the system.

The roles of management and the services associated to the organizational units of this particular organization will be available through the OVAMAH platform [33]. OMS (Organization Management System) provides the necessary services for the proper functioning of the agent organization. It also provides a range of services to register or unregister structural components, in particular, the roles, norms and existing units in the system, and offers facilities to report on those components.

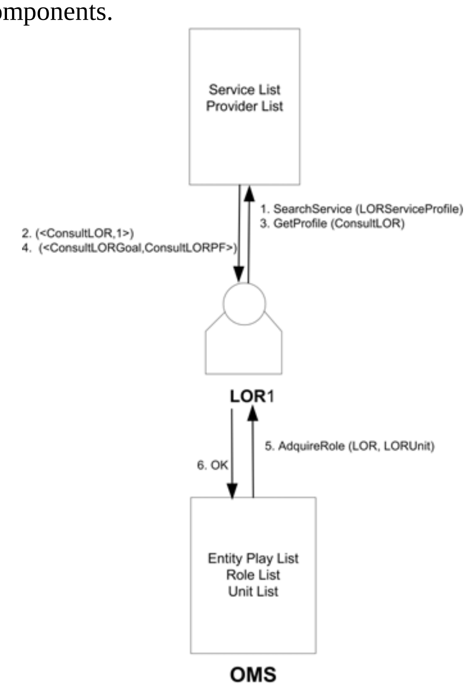

Figure 4. Example acquisition role by an external agent

Figure 4 provides a scenario in which a new LOR is registered in the organization. The ranking value indicates the degree of alignment between the service and the specified service proposed.

The units contain the Acquire Role, Report Unit and Stop Role services, in addition to the dependent domain services, which have already been identified above. For example, if an external LOR wants to contact them, it is first necessary to go through the process of acquiring the corresponding role.

D. AIREH Recovery Performance Evaluation

The search process is integrated into the agents of the organization and connected by services in a way that is totally transparent to the user. This integration addresses problems regarding the distribution and integration of different repositories, the abstraction of the internal logic of each, and the classification, storage and search for LOs. Moreover, the system adds capabilities such as easy scalability scenarios, use of new protocols, logical internal repositories, cataloging, and heterogeneous applications designed to cover services with related features.

The algorithm used to select the effective content for the user takes into account the semantics of LOs and the technical aspects for the search in the LOR. This influences the cataloged results, which are retrieved automatically through several mechanisms involving user assets. In this paper, the processing of the retrieved LO metadata addresses three aspects: completeness, reliability and ranking accuracy. Given a set of n metadata for a single LO recovered for a given repository J, the set of these metadata files is determined by $O_J=\{ O_1 \dots O_i\}$ with i=1,…n.

The relevance is related to profit or the potential use of recovered materials in relation to achieving the goals, interests or problems intrinsic to the user. Based on this approach and in the context of this work, the metadata of the LO recovered were categorized by the criterion of relevance based on the same binary operation: R = {0,1}. For example if the LO cannot be recovered because it lacks the tagged information indicating the source of the resource (the category attribute <technical> <location> LOM), it is described as irrelevant and is credited with null value (0). Otherwise, it qualifies as a relevant value (1). This approach allows the calculation of the accuracy of the search engines for each query in LOR J, by using Equation 1.

$$P_j = \frac{\sum_{1}^{n} R_i \cdot O_j}{n} \quad (1)$$

$$E_j = \frac{\sum_{1}^{n} R_i \cdot O_j}{\sum_{j=1}^{m} \sum_{i=1}^{n} R_{ij} \cdot O_i} \quad (2)$$

$$G_j(t) = \frac{\sum_{1}^{n} \frac{numberLO_n}{time_n}}{n} \quad (3)$$

Given a query Q in a series of Repositories, the full set of metadata recovered will be the union of all the metadata repositories recovered in m. To calculate the Relative Recall, take the denominator of the equation; the sum of the LOs judged relevant to each search for the overall system is determined by the Equation 2.

It is particularly relevant to meet the demands for the LOs that meet the requirements of a user request in real time. The dynamics of the environment in the recovery of resources allow for the user to be provided with a large number of LOs very quickly, so it is necessary to have some measure that allows us to evaluate this feature. At present there is no published system which allows this type of control over the content of what is proposed as a new measure. Equation 3 represents the temporary Gain, $G_J(t)$, for repository J which contains the measure that relates the number of LO retrieved for n queries over time.

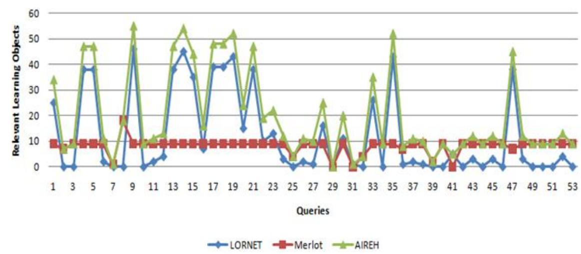

Figure 5. Relevant LO recovered

The architecture was evaluated by performing a battery of tests to validate its efficiency in real environments. The system is robust against failure because it incorporates several methods in different agents in the organization throughout the query time by planning the management of repositories based on the performance of the agents. Once each instantiated LOR agent performs the query in each repository, each LOR is in charge of canceling the query and reporting any problem affecting the established QoS levels, such as query time, performance of the repository, and so on. This data reveals the significant increase in the number of relevant LOs recovered (Figure 5) while the number of LOs to recover in time decreases (Figure 6). The proposed architecture increases the temporary gain in the system by 15% on average over isolated repositories.

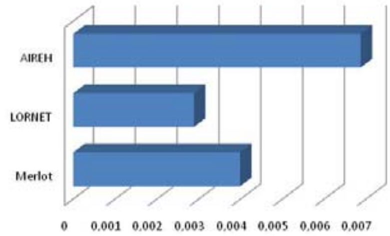

Figure 6. Comparative average temporary gains

V. Recommendation Strategy

A recommendation system is a tool that predicts user likes according to their characteristics, interests or abilities, based on previously obtained information. There are various techniques based on Artificial Intelligence (AI) which are oriented to carrying out these tasks. One of them is Case Base Reasoning (CBR). Recovery techniques and their adaptation to CBR techniques have become effective for the development of recommender systems [31, 32].

The purpose of CBR is to solve new problems by adapting solutions that have been used to solve similar problems in the past [33]. A CBR manages cases (past experiences) to solve new problems. The way cases are managed is known as the CBR cycle, and consists of four sequential steps which are recalled every time a problem needs to be solved: retrieve, reuse, revise and retain.

A CBR depends largely on the structure and content representation and its collection of cases. The developed system is characterized by working with cases defined by the characteristics of the educational context. Each case is divided into the following main components:

- A set of attributes referred to as target, which contains the definition of the problem, that is to say, the query.
- A set of attributes associated to the previous user interactions.

Once the definition of the problem is formed in terms of attributes, the objective of CBR is to generate the ranking of these learning objects in response to user characteristics that are reflected in the characteristics of learning objects available, such as educational level LO, the format or the language of the resource. The CBR system is initiated by a new request made by the user who is searching for LOs. At that moment, the CBR system is executed. The information contained in the new case at the beginning of the execution cycle of the CBR system is defined by the following tuple:

$$c = \{T, u_i, x_i\} \quad (4)$$

Where T refers to the set of attributes defined in the target extracted mainly from the information in the markup language in accordance with standard tagged used (LOM, DC, etc.) i.e. T = {*title, language, keywords, format...*}. The user identifier is u_i and x_i is the value associated with the final solution.

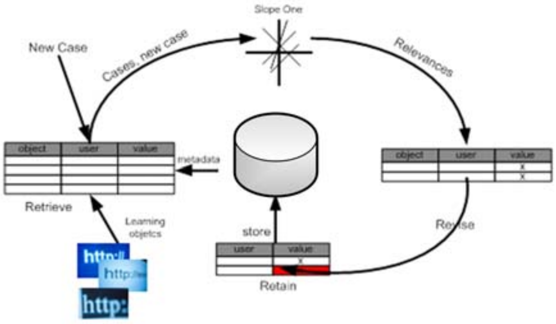

Figure 7. CBR system implemented in AIREH

Using the information defined by equation 4, the reasoning cycle for the CBR system is initiated. Figure 7 illustrates the reasoning cycle, the CBR system is.

During the retrieve phase the metadata for the learning objects are downloaded from different repositories using simultaneous searches through a federated search procedure based on an organization of agents, as explained in previous sections. The Slope One method is applied during the reuse phase in order to predict the degree of relevance of the recovered LO. Finally, during the revise and learning phase, information related to the user's final assessment is stored. The different steps for the reasoning cycle will now be explained in greater detail.

Once the information has been recovered from the repositories and the retained environment, different cases are obtained according to the structure indicated in Equation 4. The information listed in table 1 is obtained from the data found during the retrieve phase. Each cell contains a value v_{ij} that represents the user's evaluation of the learning object.

User	LO_1	LO_2	…	LO_m
u_1	v_{11}	v_{12}	…	v_{1m}
u_2	v_{21}	--	…	v_{2m}
...	...	...	…	...
u_n	v_{n1}	0.27	…	v_{nm}

Table 1. Information retrieved from the cases.

The average is calculated for each pair of individuals, as seen in equation 5. The final averaged values could be combined according to equation 6, with a weighted average relative to the number of predictions that exist for each article.

$$\bar{d}_{ij} = \frac{\sum_{k=1}^{m-1}(v_{ik} - v_{jk})}{m-1} \quad (5) \qquad x_{ik} = \frac{\sum_{j=1}^{n-1} m_j \bar{d}_{ij}}{\sum_{j=1}^{n-1} m_j} \quad (6)$$

Where v_{ik} represents individual i for which the unknown value is being calculated, m is the number of values that exist for both articles i and j (if v_{ik} is unknown, v_{jk} will not be considered in the calculation), and v_{jk} is individual j.

Where v_{ik} represents individual i for whom the unknown variable for k is calculated, m_j is the number of values that exist for category j, v_{jk} is individual j.

During the revise and retain phase, the user rates the objects retrieved during the reuse phase. The values are then stored in the cloud for future retrievals.

A. Evaluation and Results

The recommendation is made by implementing the CBR proposed mechanism and according to the group of recovered cases. A series of queries were made based on a selection of 60 different keywords from the computer science ground extracted from UNESCO codes. To validate the recommendation proposal, we evaluated the results obtained by the AIREH assessment with Merlot [1] and Lornet [2] repositories over a period of 6 months with 40 users. Each user input a key word and then analyzed the predictions made for the previous 15 predictions. The values were assigned to each item on a scale of 1 to 5. The implementation of the algorithms was based on the Apache Mahout library, which provides techniques such as Map Reduce, allowing a high level of efficiency in multiprocessing systems.

Elements	KNN	Slope One	SVD
500.000	43s	39s	38s
5.000.000	6:37s	5:36s	5:52s

Table 2. Comparison of results of the calculation times.

The first step was to compare the execution times for different alternatives to collaborative filtering in order to determine the viability of the different solutions. The execution times were based on simulated data, starting with the first test of 500,000 pieces of data and a second of 5,000,000. Table 2 lists the calculation times to obtain the recommendations.

In order to analyze the efficiency of the CBR system, the predictions were compared with other methods of collaborative filtering. The techniques selected were KNN (K-Nearest Neighbor) and SVD (Single Value Decomposition).

KNN	Slope One	SVD
1.30	0.76	0.78

Table 3. Comparison of efficiency results.

While the times for constructing the recommendations are very similar, the difference is due to the fact that the KNN algorithm needs the same execution time for any prediction made for a different user, while the Slope One and SVD have a prediction time for execution of less than one second, regardless of the user. The results shown in Table 3 indicate the average error values obtained by the methods indicated in each column. The weighted values are based on a scale of 1 to 5.

The results in Table 2 indicate that Slope One provided the best results, although very similar to those obtained by SVD. The reason for not using SVD is that it is necessary to statistically determine the number of elements that reduce the dimensionality, which would involve the analysis of the value with subsequent executions.

We also evaluated user perception regarding the quality of the recommendations made by the proposed mechanism throughout the evolution of the CBR. The evolution of the number of cases in the case base allows for greater knowledge and appreciation of potential LO as shows Figure 8. This improvement is due to the system's ability to learn and adapt to lessons learned. Likewise, the experiences allow a better adaptation to the user profile.

System success is evaluated through user interaction with the recommended LO, as well as the assessment it makes of each. The user perceives an improvement in the time of the

[1] MERLOT, Multimedia Educational Resource for Learning and Online Teaching (http://merlot.org)

[2] LORNET, Learning Object Repository Networks (http://www.lornet.com).

recommended LO.

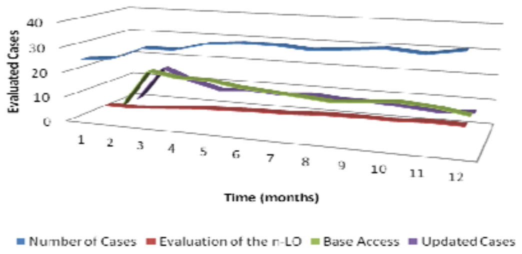

Figure 8. Evaluation of the recommendations of the CBR

Figure 8 also shows that the number of updated cases decreases as the system acquires experiences (the X axis represents the evaluations during the time period and the y axis quantifies the number of cases concerning the aspects evaluated (number of cases, access to the base and updates of cases). By increasing the number and variability of the cases captured, the ability to find cases similar to the query that the user requires increases and may validate the recovered LO criteria that define user tastes and/or needs.

VI. Conclusions and Preliminary Results

An important objective in the development of recovery technology in the deep Web content is to improve the quality, scope and accuracy of existing visible Web engines through the use of structured descriptions of resources, i.e., through semantic rich metadata. This is possible if the metadata of these resources are accessible.

The proposed architecture can search multiple repositories simultaneously, a complex problem that is further exacerbated by the heterogeneity of digital repositories. The AIREH architecture provides multiple perspectives to assess the recovery of educational content from a real, open and scalable environment, and also supports mechanisms that will implement the recommendation or ranking for recovered LOs. The development of a single ordered list of Learning Objects that incorporates a user's relevance criteria in this work is one of the tasks that the AIREH agent model implements with a CBR reasoning model.

This paper has presented a recovery architecture based on educational content partner organizations. The main novelty in the proposed architecture is its dynamic capability. This ability confers adaptive planning to carry out an optimal distribution of the tasks of the organization's member agents, enabling the retrieval of intelligent content and flexibility in highly dynamic environments for which it was created. In summary, the architecture presented in this study can define the actions that an organization of agents must carry out, anticipate the changes that may occur during the execution of a given query, and use adaptive planning within an organization of agents according to context characteristics (users, profiles, features, content, variability of learning object repositories', etc.).

The system is still in a process of development and undergoing more detailed testing, which will allow for more extensive results in the future. With AIREH it is possible for the user to retrieve LO efficiently and simply, since it allows the retrieved elements to be filtered according to each user and their previous actions. Some new aspects to consider in future studies could include the integration of richer semantic aspects for the recovery and cataloging of educational content.

Acknowledgment

This work has been partially supported by the project MICINN TIN 2009-13839-C03-03. We would like also to thank the participants of our user study.

References

[1] ADL CORDRA (Content Object Repository Discovery and Registration/Resolution Architecture) from: http://www.adlnet.gov/Technologies/adlr/ADLRDocuments/ADL%20Registry%20Documentation/ADL%20Registry%20Documentation.aspx

[2] IMS Digital Repositories Specification. V.1 2003. http://www.imsglobal.org/digitalrepositories/

[3] David Massart, Nick Nicholas, Nigel Ward. IMS GLC Learning Object Discovery & Exchange (LODE). Version 1.0, (Base Document) Draft 14, March 2, 2010.

[4] B. Simon, D. Massart, F. Van Assche, S. Ternier, & E. Duval. “A simple query interface specification for learning repositories”, CEN Workshop Agreement (CWA 15454). Available on Web ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/WS-LT/CWA15454-00-2005-Nov.pdf, November 2005.

[5] Web Services Description Language (WSDL) 1.1. W3C Note 15 March 2001. Available on Web http://www.w3.org/TR/2001/NOTE-wsdl-20010315

[6] Roy T. Fielding and Richard N. Taylor. “Principled design of the modern Web architecture”. In Proceedings of the 22nd international conference on Software engineering (ICSE '00). ACM, New York, NY, USA, 2000, pp. 407–416.

[7] Roy T. Fielding. “Architectural styles and the design of network-based software architectures”. PhD Thesis, University of California, Irvine, 2000.

[8] IEEE 1484.12.1-2002, Draft Standard for Learning Object Metadata. The Institute of Electrical and Electronics Engineers, Inc. Retrieved March 23, 2010,

[9] DCMI Specifications 2010. http://dublincore.org/documents/dcmi-terms/

[10] Gil, A.-B.; Rodriguez, S.; de la Prieta, F.; Martin, B. Educational content retrieval based on semantic Web services. *Next Generation Web Services Practices (NWeSP), 2011 7th International Conference on* , pp.135-140, 19-21 Oct. 2011

[11] Bobadilla, J., Serradilla, F., Hernando, A., & MovieLens. (2009). Collaborative filtering adapted to recommender systems of e-learning. Knowledge-Based Systems.

[12] Khribi, M. K., Jemni, M., & Nasraoui, O. (2009). Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval. Educational Technology & Society, 12(4), 30–42.

[13] Manouselis, N., Vuorikari, R. and Van Assche, F. (2010), Collaborative recommendation of e-learning resources: an experimental investigation. Journal of Computer Assisted Learning, 26: 227–242

[14] Recker, M., Walker, A., and Lawless, K. What do you recommend? Implementation and analyses of collaborative information filtering of web resources for education. Instructional Science, 31(4-5), 2003, pp. 299-316.

[15] Lemire, D., Boley, H., McGrath, S., and Ball, M. Collaborative Filtering and Inference Rules for Context-Aware Learning Object Recommendation. Technolodgy and Smart Education, 2(3). 2005. pp. 179-188.

[16] McCalla, G. The Ecological Approach to the Design of E-Learning Environments: Purpose-based Capture and Use of Information about Learners. Journal of Interactive Media in Education, 2004 (7) Special Issue on the Educational Semantic Web. Volume 1, p. 18, 2004.

[17] Manouselis, N., Vuorikari, R., and Van Assche, F. Simulated Analysis of Collaborative Filtering for Learning Object Recommendation. SIRTEL Workshop, EC-TEL 2007.

[18] Aijuan, D., & Baoying, W. (2008). Domain-based recommendation and retrieval of relevant materials in e-learning. In IEEE international workshop on semantic computing and applications 2008 (IWSCA '08) (pp. 103–108)

[19] Ghauth, Khairil, Abdullah, Nor. Learning materials recommendation using good learners' ratings and content-based filtering. Educational Technology Research and Development. 2010. Springer Boston. SN 1042-1629. http://dx.doi.org/10.1007/s11423-010-9155-4

[20] Tsai, K. H., Chiu, T. K., Lee, M. C., and Wang, T. I. A learning Object Recommendation Model based on the Preference and Ontological Approaches. Proceeding of the Sixth International Conference on Advanced Learning Technologies (ICALT'06) 2006.

[21] Wang, T. I., Tsai, K. H., Lee, M. C., and Chiu, T. K. Personalized Learning Objects Recommendation based on the Semantic Aware Discovery and the Learner Preference Pattern. Educational Technology and Society, 10 (3), 2007. pp. 84-105.

[22] Yang Y. An evaluation of statistical approaches to text categorization. Journal of Information Retrieval 1 (1999) 67-88.

[23] Kerkiri, T., Manitsaris, A., & Mavridou, A. (2007). Reputation metadata for recommending personalized e-learning resources. In Proceedings of the second international workshop on semantic media adaptation and personalization (pp. 110–115). Uxbridge.

[24] Ochoa, X. and Duval, E. Use of Contextualized Attention Metadata for Ranking and Recommending Learning Objects. Proceedings of 1st International Workshop on Contextualized Attention Metadata: Collecting, Managing and Exploiting of Rich Usage Information. 2006. pp. 9-16.

[25] Wolpers, M., Najjar, J., and Duval, E. (2007). Tracking Actual Usage: the Attention Metadata Approach. Educational Technology & Society, 10(3), pp106-121.

[26] Qiyan Han; Feng Gao; Hu Wang, "Ontology-based learning object recommendation for cognitive considerations," 8th World Congress on Intelligent Control and Automation (WCICA), 2010, vol., no., pp.2746-2750, 7-9 July 2010

[27] A. Ruiz-Iniesta, G. Jiménez-Díaz, M. Gómez-Albarrán. Personalización en Recomendadores Basados en Contenido y su Aplicación a Repositorios de Objetos de Aprendizaje. IEEE-RITA Vol. 5, pp. 31-38, Núm 1, Feb. 2010.

[28] Vargo, J., Nesbit, J. C., Belfer, K., & Archambault, A. (2003). Learning object evaluation: Computer mediated collaboration and inter-rater reliability. International Journal of Computers and Applications, 25 (3), 198-205.

[29] Nesbit, J. C., Belfer, K., & Vargo, J. (2002). A convergent participation model for evaluation of learning objects. Canadian Journal of Learning and Technology, 28 (3), 105-120.

[30] Manouselis, N., Drachsler, H., Vuorikari, R., Hummel H. and Koper R. Recommender Systems in Technology Enhanced Learning. Recommender Systems Handbook, pages 387-415. Springer, 2011

[31] Miquel Montaner, Beatriz López, and Josep Lluís de la Rosa. 2002. Opinion-Based Filtering through Trust. In Proceedings of the 6th International Workshop on Cooperative Information Agents VI (CIA '02), Matthias Klusch, Sascha Ossowski, and Onn Shehory (Eds.). Springer-Verlag, London, UK, 164-178.

[32] Corchado J. M. and Laza R. (2003). Constructing Deliberative Agents with Case-based Reasoning Technology, International Journal of Intelligent Systems. Vol 18, No. 12, December. pp.: 1227-1241.

[33] J. Kolodner, Case-Based Reasoning, Morgan Kaufmann 1993.

Author Biographies

Ana B. Gil Doctorate in Computer Science from the University of Salamanca. Undergraduate studies in Physics and MSC from the University of Salamanca. She is currently an Assistant Professor in the Department of Computer Science at University of Salamanca. Her current research interests include applied semantic web, software agents, human-computer interaction and community-enabling technology.

Sara Rodríguez Doctorate in Computer Science and Automation from the University of Salamanca. Obtained degree in Technical Engineering and Computer Science Systems in 2004 and Computer Science Engineering in May 2007 also from the University of Salamanca. Assistant Lecturer in the Department of Computer Sciences and Automation. Her research experience focuses on the application of multi-agent systems in different contexts, and on the use of Digital Animation techniques.

Fernando de la Prieta holds a MCS from the University of Salamanca (Spain) where he is currently pursuing her PhD in Computer Science. Undergraduate degree in Computer Science from the University of Salamanca. Research focuses on multi-agent systems, virtual organizations, Cloud Computing and educational technology. He has published articles in prestigious national and international conferences, and has participated in different research projects both at a national and European level.

Juan F. de Paz Doctorate in Computer Science and Automation from the University of Salamanca (2010). Currently Assistant Lecturer in the Department of Computer Science and Automation at the University of Salamanca. Undergraduate degree in Engineering and Computer Science (2005) and Statistics (2007) from the University of Salamanca. His research work focuses on the coordination and development of projects in the field of Bioinformatics and Biomedicine, applying reasoning techniques and Artificial Intelligence algorithms.