
Learning object retrieval in heterogeneous environments

Ana B. Gil*, Sara Rodríguez,
Fernando De la Prieta and
Juan Manuel Corchado

Computer and Automation Department,
University of Salamanca,
Plaza de la Merced s/n, 37008, Salamanca, Spain
E-mail: abg@usal.es
E-mail: srg@usal.es
E-mail: fer@usal.es
E-mail: corchado@usal.es

*Corresponding author

Abstract: This paper presents a solution to the problem of the search and retrieval digital tagged content in heterogeneous learning object repositories through architecture for intelligent retrieval of educational content in heterogeneous environments (AIREH) framework. This architecture unifies the search and retrieval of objects, thus facilitating the personalised learning search process by filtering and properly classifying learning objects retrieved for an approach for semantic-aware learning content retrieval based on abstraction layers between the repositories and the search clients. The use of federated databases techniques by using an organisation of agents allows those agents to work in a coordinated manner to solve a common problem, allowing the agents to adapt to the constantly changing environment (users, content repositories, etc.). Combining a complete agent-based architecture that implements the concept of federated search along with IR technologies may help organising and sorting search results in a meaningful way for educational content.

Keywords: multi-agent systems; distributed computing; e-learning; learning objects; repositories; simple query interface; SQI; learning technology standards; web services.

Reference to this paper should be made as follows: Gil, A.B., Rodríguez, S., De la Prieta, F. and Corchado, J.M. (2013) 'Learning object retrieval in heterogeneous environments', *Int. J. Web Engineering and Technology*, Vol. 8, No. 2, pp.197–213.

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

Sara Rodríguez received her Doctorate in Computer Science and Automation from the University of Salamanca and her degree in Technical Engineering and Computer Science Systems in 2004 and Computer Science Engineering in 2007 also from the University of Salamanca. She is an 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 received his MCS degree from the University of Salamanca (Spain) where he is currently pursuing his PhD in Computer Science. He received his undergraduate degree in Computer Science from the University of Salamanca. His research focuses on multi-agent systems, virtual organisations, cloud computing and educational technology. He has published articles in prestigious national and international conferences, and has participated in different research projects both at national and European levels.

Juan Manuel Corchado is a Spanish Computer Scientist, Professor, Researcher and author. He is a Full Professor with Chair at the University of Salamanca, where he has been twice elected as the Dean of the School of Sciences. In addition to a PhD in Computer Sciences from the University of Salamanca, he received his PhD in Artificial Intelligence from the University of the West of Scotland. He specialises in both artificial intelligence and ambient intelligence, and has authored more than 300 papers in leading international scientific journals.

This paper is a revised and expanded version of a paper entitled ‘Automatic learning object extraction and classification in heterogeneous environment’ presented at the 9th Conference on Practical Applications of Agents and Multi-Agent Systems, Salamanca, Spain, 6–8 April 2011.

1 Introduction

The field of distance education today evolves by the incorporation of standards and technologies in a meaningful way to facilitate reusability and interoperability of the educational resources (Hatala et al., 2004). One of the most widely accepted approaches of this development is based on the learning object (LO) paradigm (Lujara et al., 2007). These LOs incorporate both a self-contained modular resource and a set of educational descriptive metadata. There are different standards to labelled educational metadata like Dublin Core (DCMI Specifications, 2008), MPEG-7 (2010) and mainly learning object metadata (LOM) (see IEEE LOM 1484.12.1-2002, 2002).

Usually, LOs are stored in widespread digital repositories called learning objects repositories (LORs). These repositories mainly belong to academic institutions and universities. There are several configuration architecture repositories of LOs in order to achieve interoperability of its contents. The main standards are IMS DRI (2003), CORDRA (2005) or IMS *LO Discovery and Exchange* (Massart et al., 2010). These standards generally provide some kind of search interface that makes it possible to access its LOs through queries to the database that hosts them. Any interaction for retrieving LOs can be done manually or be automated through different software systems. Interoperability is based on different types of protocols, which control the interaction between systems and enable the efficient exchange of information. The metadata describing educational content, are managed by the repositories. Whatever the mechanism of management of these LOs requires a communication protocol for metadata search in these LOR. These protocols enabled in the repositories contain search

functionality and/or harvesting. In the field of e-learning are mainly five protocols to achieve this interoperability between repositories: Z39.50 (2003), Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH, 2008), the simple query interface (SQI) (see Simon et al., 2005a, 2005b), the search/retrieval via URL (SRU, 2010), and simple publishing interface (SPI) (Ternier et al., 2010).

There is a large volume of educational content on the web, which 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 of the website which are referred to as visible. The solutions developed by conventional search engines are very efficient for retrieving the visible web contents. These engines make use of a centralised retrieval model that copies the information in a single centralised database, indexes and sorts the contents by classified documents in the database for user queries. This method 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, such as the educational content repositories, which are the subject of this work. These repositories implements architectures that enables the interoperability of its contents by providing online consultation mechanisms through federated searches using a web customer service. These consultation mechanisms, mainly SQI through simple object access protocol (SOAP) or restful web services (Fielding and Taylor, 2000), provide the repositories with an abstraction level between the internal and external heterogeneity that includes a specification with a wide range of possibilities with different service-based communication protocols, synchronous and asynchronous queries, etc.

In all, a big effort has been focused on the creation of sustainable repositories that are highly heterogeneous due to their different storing systems, access to objects, query methods, etc. It is necessary to work towards and develop solutions that enable the efficient search of heterogeneous content with regards to the distributed context where they lodge. This paper presents architecture for intelligent retrieval of educational content in heterogeneous environments (AIREH) framework, that unifies the search and retrieval of objects, thus facilitating the learning search process by filtering and properly classifying LOs retrieved according to some rules (Gil et al., 2011). The generation of such rules for the organisation of the items recovered will be based on educational metadata and will provide useful content to end users. The architecture will provide multiple perspectives to assess the retrieval of educational content.

The remainder of this paper is organised as follows: Section 2 describes relevant works related to LO retrieval and selection. Section 3 present the federated search provided for LOs in digital repositories developed. The proposed architecture is introduced in Section 4. Selected results of a comprehensive evaluation of the architecture are presented in Section 5. This paper closes with relevant conclusions and an outlook to future work in Section 6.

2 Related work

The current increase in resources for online education calls for the efficient management of those services and the elements involved. Bates (2005) suggests that one of the

problems that must be overcome in order to standardise the use of LOs is the difficulty in actually finding them.

The greatest efforts within the context of e-learning are nowadays then focused on achieving the interoperability by adding labelling techniques in the educational content creation in a way to facilitate the storage, search, retrieval, etc. Many successful works have been proposed to meet the user's required LOs. Main of them will be presented at this section.

Casquero et al. (2008) proposes BILDU architecture based on a set of horizontally integrated vertical search engines through one interface, which supports a network incorporating the features of social recommendation and mechanisms of association. Qu and Nejdil (2004) propose in Edutella encapsulating other content integration and searching the collections through peer-to-peer mechanisms. Zhou et al. (2010) propose in LESSON the use of a metasearch engine and system P2P sharing, and Santiago and Raabe (2010) also propose a P2P network between institutions, with the possibility of searching and retrieving LOs through an API within the content managers.

Some others such as Kärger et al. (2006), Almaraz (2008) or Santiago and Raabe (2010) works, employ an architecture based on mediators to integrate LO repositories. Other proposals for the retrieval of educational content are based on methods adapted from the area of information retrieval such as Yen et al. (2010) by proposing similarity metric calculations on LO or Mayorga et al. (2010) by including clustering techniques as formal concept analysis (FCA) with pedagogical aims to search the web. Other approaches such as Dietze et al. (2007), Hanafy and Fakhry (2008) or Li et al. (2009), make use of retrieval architectures based web services.

Proposals such as Margain Fuentes et al. (2010) or Soto et al. (2007) offer the solution of an architecture based on educational content retrieval using agent technology.

The successful retrieval of educational content can provide in large part the characteristics, restrictions or conditions on the proper use of learning resources. The expressive power of queries to select LO repositories or queries according to user are more accurate and better present the results that meet user expectations. Various works such as Kärger et al. (2006), Lee et al. (2008) or Segura et al. (2010) show how a little power of expression in the retrieval of LOs in their LORs can provide suggestions for improvement by incorporating many semantic aspects in the expansion of consultative mechanisms. Several tools appears such as Shih et al. (2006) that has developed a metadata authoring tool to support the search and reuse of LOs. Or Lee et al. (2008) that proposes an architecture with an ontological approach for LO retrieval by expanding the query algorithm with semantic aggregation, which is called ontology query expansion algorithm. Lee et al. (2008) define the context ontology and infers user's intentions through short questions that expand the query in the LOR. But the simulation of these proposals and performance tests on html documents on the visible web, do not make use of the metadata characteristics of the real LO, or of the LOR search functionalities from the initial storage. Most of them make new contributions to the mechanism of semantic expansion of queries, but it remains a proposal that does not contrast real LOR or their metadata tagged files. Recent Segura et al. (2011) work defines a query expansion as a dynamic mechanism for searching and filtering the results that were validated in the MERLOT repository to search for LO related to genetics using Gene Ontology. The mechanism used is connected to MERLOT Restful service to recover the contents.

Avancini and Straccia (2005) propose CYCLADES, an environment where users search, access and evaluate digital resources available in the repositories that are accessed

through the open archives initiative (OAI). This system makes it possible to offer recommendations on resources that are stored in different repositories and through an open system. The recommendations offered by CYCLADES were evaluated through a pilot study with 60 users, which focused on the test performance (prediction accuracy) of various collaborative filtering algorithms.

The proposed solutions are many and varied, most do not get results on real environments and the majority does not provide educational content in the particular case of LOs. But they share a common denominator, which is that they promote interoperability and the reuse of resources including the development aspects related to adaptation, or they use their own content or environmental mechanisms to allow for better access including some user aspects. Real e-learning is mainly web-based supported by means of Internet connect learners, teachers, courses and all technologies implies. All the evaluated proposals 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 with educational content. Most of the related works we have revised have been applied to a local case of study, for small and usually local parameters. There are a growing number of papers proposing systems to recommend learning resources, evidenced by the lack of operational solutions as confirmed by Manouselis' (2011) recent work.

It is needed bring the user with an unify framework that search and retrieval LO process easier; that filters and classify the retrieved LO according some rules to give useful content based on the educational metadata provided; to be browsed in order to view multiple perspectives in an open and free environment for learners, teachers, etc. This paper proposes an architecture that facilitates the searching and cataloguing of LO in different and real LOR. Next section present the foundations of the architecture that provides multiple perspectives to assess the retrieval of educational content from a real, open and scalable environment, and will also be a support mechanism to implement the recommendation or ranking for the recovered Los.

3 Federated search system for LO

With the situation that arises from the context of education, it is clear that a new type of application is needed; one that allows the search for educational content in a distributed environment across different formats, servers and networks at the hidden web. Callan (2000) exposes that distributed information retrieval or federated searches attempt to respond to the problem of information retrieval in the hidden web. The main aim of federated search is to develop models and strategies to get the most benefit from these distributed sources. Thus, users perceive the system as a single point of access to information they require, regardless of the number of sources that exist, their location or management mechanism. The process is completely transparent to the user, who does not perceive the complexity involved, and gives a uniform treatment to the information retrieved. The main contribution of 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 centralised retrieval model than the traditional search engines that do not provide real solutions to the problem of the hidden web. Federated search can be applied to the retrieval of LO repositories such as the sequence described in the resolution of three subproblems. The procedure consists of three steps:

- 1 *Selection of repositories*, e.g., to analyse the description of resources in the repositories and to study how to represent information that is distributed in such repositories. It is very important to know the subject, area, and the context that the information stored in repositories belongs to, as well as many other key features of the various sources of hidden information (response times, efficiency, etc.).
- 2 *Selection of resources*, given a need for information and a corresponding set of descriptions, a retrieval algorithm is used to determine the results most likely to satisfy the query.
- 3 *Merger of results* is 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 sort order of the results will produce a ranking algorithm, or priority ranking in the presentation of the contents recovered. This last phase estimates the labelling data and the agreement between each reference user profile and the corresponding recovered LO

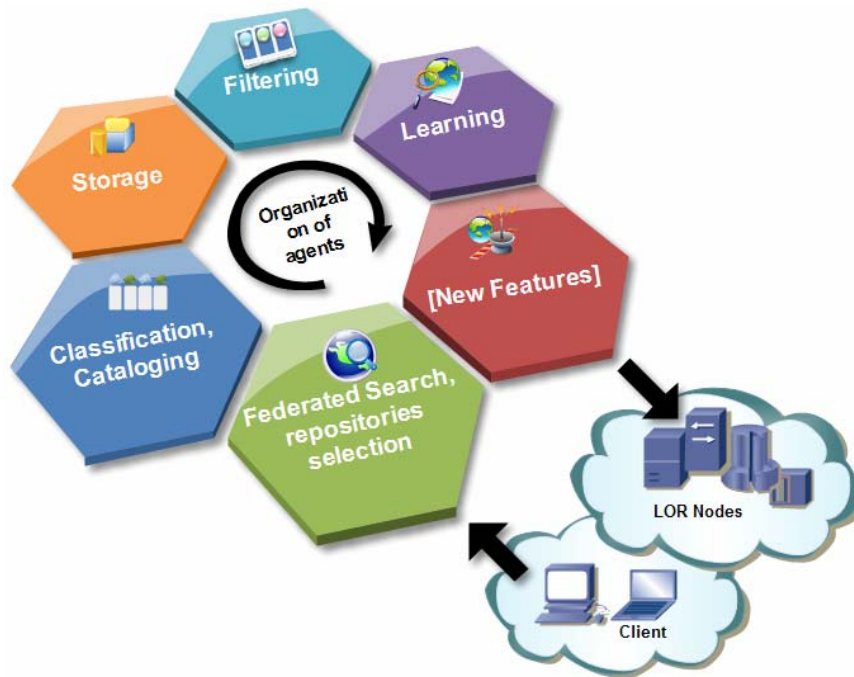
The use of federated search mechanisms for retrieval educational content from heterogeneous sources solves the problem of information retrieval from heterogeneous LOR. Conflicting needs must be met in order to create the federation and thus integrate several components: the repositories must be distributed across physically dispersed locations; access to these repositories is read-only; information systems are operating under different platforms; the types used are distributed repositories (metadata stored in one place and LO in the other) and they are managed by autonomous systems; in some cases access to the repositories requires the establishment of a session, in others direct access, etc.

Multi-agent systems are known for their ability to adapt quickly and effectively to changes in their environment. This work proposes a model for the development of digital content retrieval based on the paradigm of virtual organisations of agents. The model allows the development of an open and flexible architecture that supports the services necessary to conduct a search for distributed digital content dynamically. In this sense, the use of a system architecture based on the use of intelligent agents is ideal, since they can apply their characteristics (autonomy, status, reactivity, rationality, intelligence, coordination, mobility and learning) to develop a stable system with the ability to react intelligently to the needs of the environment. The idea to model the architecture as a virtual organisation, on the assumption that an organisation can adapt its actions to achieve its goals and interact with heterogeneous components, is proposed as a theoretically efficient solution.

From a heterogeneous and changing technological situation that accompanies the proposed educational context, and in order to reuse educational content in a real operational context, there is motive to design a model for an integrated architecture in which an organisation of agents can execute actions for the search and retrieval of content based on a federated search model. The innovation of this architecture will be to provide an organisation of agents with the self-adaptive capabilities required to address the current problems in highly dynamic environments such as those discussed, and to adapt

to future changes. The features of the proposed architecture are presented in Figure 1. This solves the problems of distribution, the integration of different repositories, the abstraction of the internal logic of each repository, and the classification, storage and retrieval of LOs, in a completely transparent way. In addition count with capacities, such as simple scalability due to the use of new protocols or heterogeneous applications designed to cover services related features. The following sections describe how the proposed organisation intends to perform the functions of architecture.

Figure 1 General scheme of distribution of tasks in architecture (see online version for colours)



4 The AIREH architecture

The proposed architecture, called AIREH, is seen as a communication point midway between the LOR, the LO that they store and users who use them. The system provides a federated search system that makes it possible to search multiple geographically dispersed repositories simultaneously. In addition, once the results of the different repositories have been received, an identification phase is developed and filtered in order to adapt the results to the user preferences (see Gil et al., 2012; Prieta et al., 2012).

The organisational model used in the architecture of the platform uses OVAMAH to define the structure and rules (see Rodríguez et al., 2009). Adaptation in this type of model is based on coordination between the participants of the organisation. The agents have access to the infrastructure offered by OVAMAH through a number of services included in what is called organisation manager service (OMS). OMS provides the services necessary for the proper functioning of an organisation; it provides a range of

services to register or unregister structural components, in particular, the roles, rules and existing units in the system. It also offers facilities to report on what those components are.

Figure 2 Overall system architecture of the e-learning retrieval system

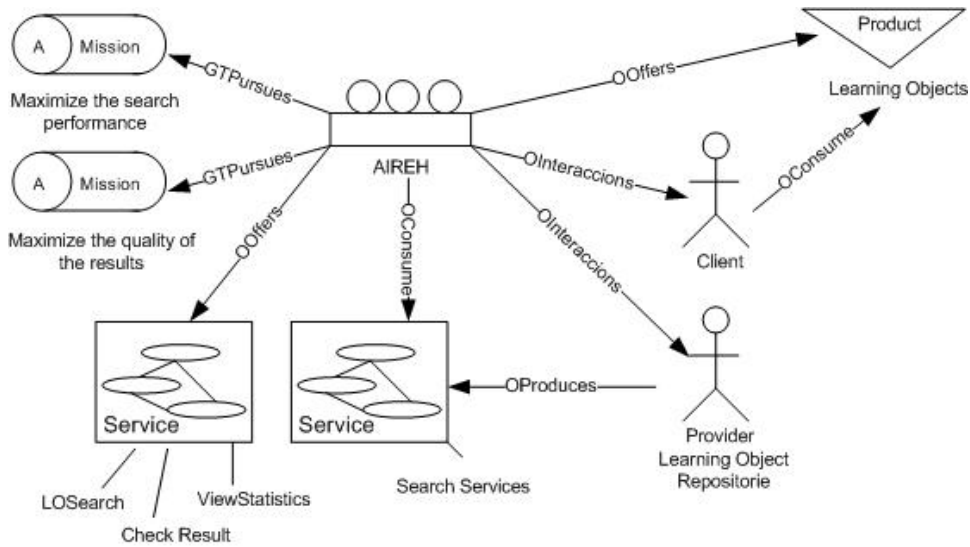


Figure 2 details the elements of the organisational model (functional view), as it shows the results (products and services) offered by the system, the type of working environment, and the interest groups. Also justifies the rationale for the architecture, which aims to facilitate the search for educational resources through the implementation of a federated search system. The functional view of the organisational unit shows AIREH SearchLO, ViewStatistics and ChechResults services. In order to provide these services, the represented providers, in this case the LOR, must offer search services that enable the collection of information. The system offers educational resources as a final product encapsulated in the form of LO. These will be requested by users through a search process. It also provides statistical information such as product performance and use repositories of LO, identifying those that are used according to the search patterns. The mission of the organisation will first maximise the system performance of queries by reducing time and increasing performance, and also, maximise the quality of results.

Figure 3 shows the organisation model, indicating the external agent (A-agents), internal (agents) and the roles they play. LOR roles, user, results and translator, as seen in the figure above, are accessible by external agents that require a procurement role.

The dynamism of the system, designed as an organisation, can be reflected in different stages: registration of new players, new services, new protocols, service requests, expulsions from the system, etc.

As an example, if an external LOR agent wants to contact any of these services, it is first necessary to acquire the corresponding role. Figure 4 shows a scenario where a new LOR is registered in the organisation. LOR1 has been registered as a member of the platform and asks the service facilitator (SF) what services have already defined a similar profile. This request is carried out using the service SearchService (message 1) of the SF,

which corresponds to the profile *LORServiceProfile* search service *LOR1* implemented by the repository. *SF* returns the service identifiers to meet the search requirements with a value (rank value) for each service (message 2). The ranking value indicates the degree of alignment between the service and the specified service proposed. Then *GetProfile* *LOR1* (message 3) obtains detailed information about the service *ConsultLOR*. The outputs of the service are ‘service goal’ and ‘profile’ (message 4). *ConsultLOR* profile specifies that the service should be provided by agents who assume the role in *LORUnit* *LOR*. For this reason, organisation management systems (OMS) asks the *LOR1* service *AcquireRole* to acquire that *LOR* role (post 5). *AcquireRole* service is carried out successfully (message 6), as *LORUnit* is accessible from the virtual organisation, and finally, *LOR1* is registered as *LOR*.

Figure 3 Diagram of the organisational model (internal and external agents)

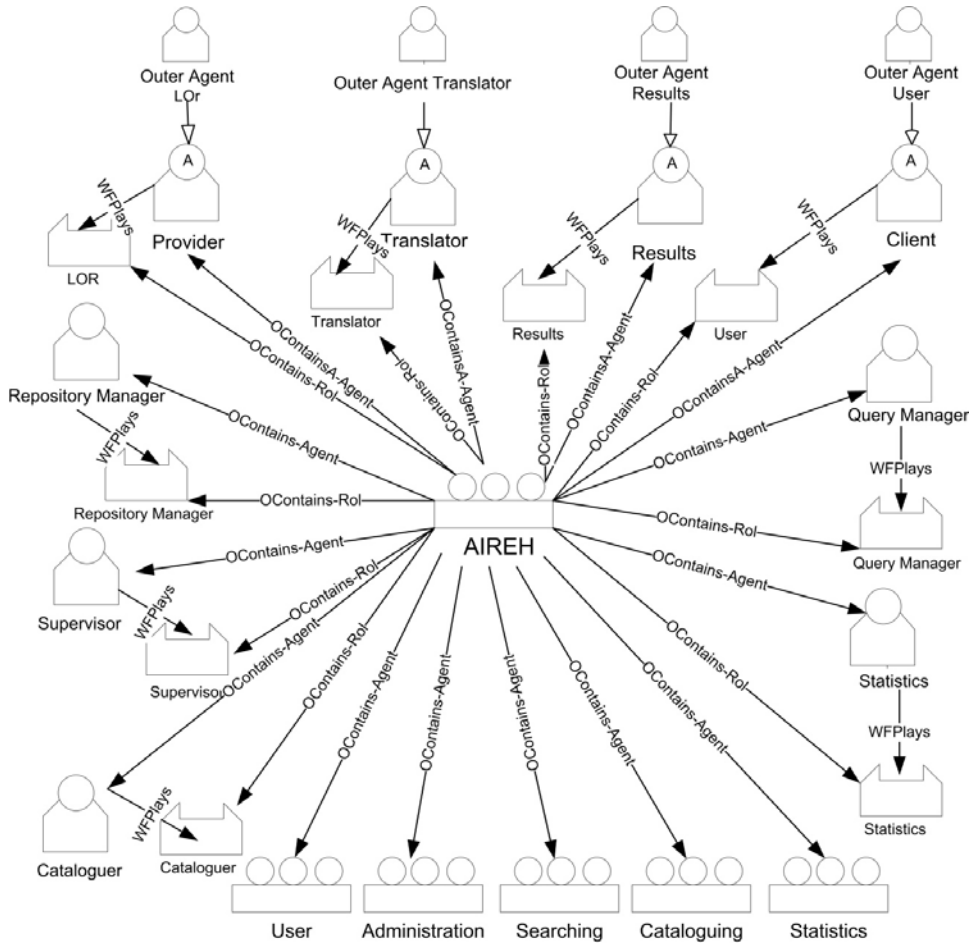
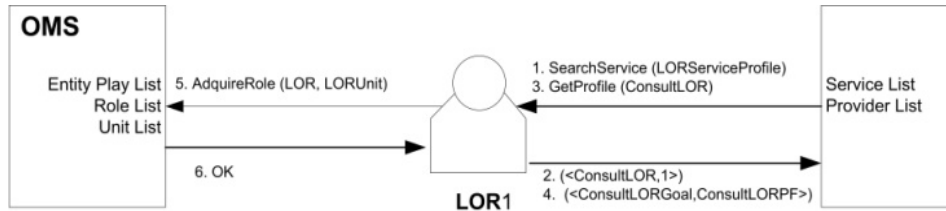


Figure 4 Example acquisition role by an external agent

5 AIREH evaluation

Educational metadata constitute a simple mechanism to describe content by using the tags to which they are associated. The semantic-based knowledge representation is used for content analysis and LO recognition. The retrieved LOs, tagged with any standard, are used for reasoning processes and for enabling user-friendly and intelligent content search and retrieval.

The problem of heterogeneity affects not only to LOR systems but also to the educational content creation. The LO creation involves authors in labelling each one of the content (i.e., LOM). This work is a high cost and non-automatic activity. Consequently, it generates semantic mismatch like lack of important labelled fields, erroneous fields, etc. that makes more difficult the retrieval processes. As a result, a new pruning heuristic was established that can select some of the possible elements and quickly go through our system and filter out the majority of the irrelevant recovered objects. Once this process has taken place, the remaining elements are filtered according to certain parameters related to the quality of the retrieved metadata documents, and the user context information is integrated into the use of the LOs. The process is complex and the association of more aspects helps to improve the quality of the contents recovered. An algorithm for selecting effective content for a particular user should take into consideration the semantics of LO use and the technical aspects for a search in the LOR, and should also classify the results retrieved automatically through mechanisms involving smart features. The processing of metadata documents retrieved in this work addresses three aspects: completeness, reliability, and the accuracy of managing results (ranking accuracy):

- *Completeness.* The description of the LO through its metadata must be as complete as possible to provide results on which to plan automatic reasoning mechanisms that generate an acceptable level of confidence for their recommendation. The attributes in the missing or mislabelled metadata represent a major hole in the precision with which they have been recovered and can be treated. Some fields may be considered irrelevant LOM while others are absolutely essential (title, location, language, etc.) in order to automate the process of searching and filtering relevant LO to the user.
- *Reliability.* A good quality metadata tagging is fundamental for LO retrieval. This functionality automates the application that links all sides of the object (file, metadata and educational content) and equips the LOR hosting them with a higher rate of trust in the system. Any LO can have syntactic or semantic errors. Syntax errors show a lack of conformity to specification. In the case of controlled

vocabularies, it is assumed that one should use any of the terms available and not others. Other problems are incompatibility with the attribute values, misspellings, and so on. Semantic errors can only be detected by inspection heuristics in the context. Some attributes may be ambiguous or inadequate so that the processing of such information is subject to a certain degree of subjectivity and must therefore be limited as much as possible with that metadata.

- *Ranking accuracy.* The set of possible LOs that a user requests can be very large. The classification algorithm for optimal results should show the user retrieved LOs by an ordered list according to a relevance criterion based on the request and the educational context in which the query is performed. The information displayed to the user about the relevant objects located is important to facilitate the user election process. The learning system attributes related to user domain is a key mechanism in the process of managing the implemented LO.

The developed system besides the search of LOs makes a ranking of the objects returned after a first filter process (by eliminating duplicates and non-reusable LOs). This facilitates a homogeneous organisation of LO from heterogeneous sources facilitating the final selection for the users of the application. This final cataloguing is based on two variables. The first one does an estimation of the quality of the objects depending on such different parameters as size, incorporation of key characteristics, metadata measures, completeness, etc. The second one of the variables in the classification of the recovered objects imposes the use of collective aspects observed by final users of the application. This second one allows providing the classification by adding collaborative recommendation. User activities in the LO search and selection are monitored in AIREH configuring both implicit and explicit aspects that form the user profile. User also could vote the LOs retrieval and system makes feedback to the ranking algorithm. This mechanism implement the use of social recommendation attending to the LO context is an open branch we are working more widely to improve.

5.1 The quality assessment

Evaluation metrics from information retrieval field have been adopted. The two most used evaluation measures are precision and recall. Precision, P , for any LOR j is the fraction of the retrieved metadata documents (O) that are relevant to the query. Relative recall, E , for each LOR is the fraction of retrieved relevant documents respect to the total number of relevant documents in the accessed repositories according to a priori knowledge.

$$P_j = \frac{\sum_1^n R_i O_i}{n} \quad (1)$$

$$E_j = \frac{\sum_1^n R_i O_i}{\sum_{j=1}^n \sum_{i=1}^n R_{ij} O_i} \quad (2)$$

If we denote R the set of returned documents from each one of the LORs while T the consulting and retrieval time, we define a new evaluation metric called temporary gain.

$$G_J(t) = \frac{\sum_1^n \frac{R_n}{T_n}}{n} \tag{3}$$

These parameters characterise the ability of the system to retrieve relevant documents and avoid irrelevant ones.

5.2 Evaluation and results

Here, we apply the proposed validation strategy to assess the AIREH performance. The proposed architecture was evaluated by performing a battery of 53 queries to validate their efficiency in real environments. Tests were performed on two repositories considered as the most relevant today; the MERLOT repository and the LORNET repository. The search patterns or topics were chosen at random among topics in English for Science and Technology of UNESCO codes developed by experts. These topics are mainly composed of sets of words (longer queries) due to the specific nature of this terminology, but single words (short queries) were also used to assess different behaviours on the search engines repositories.

Figure 5 Temporary gain values (see online version for colours)

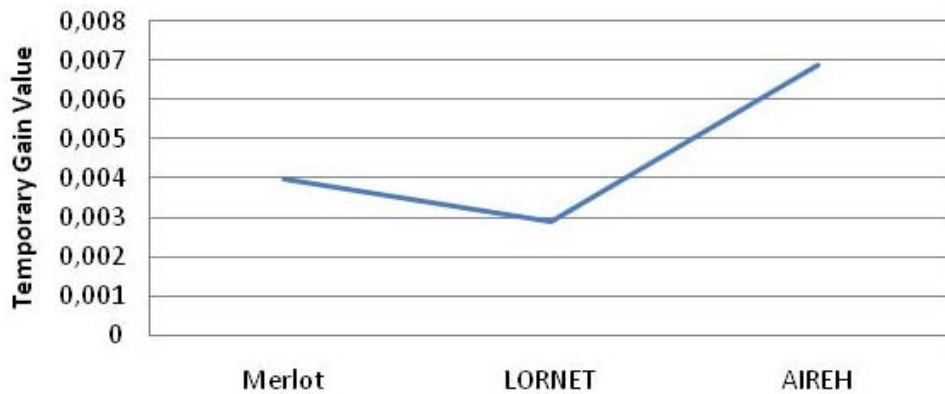


Figure 6 Relative recall comparative (see online version for colours)

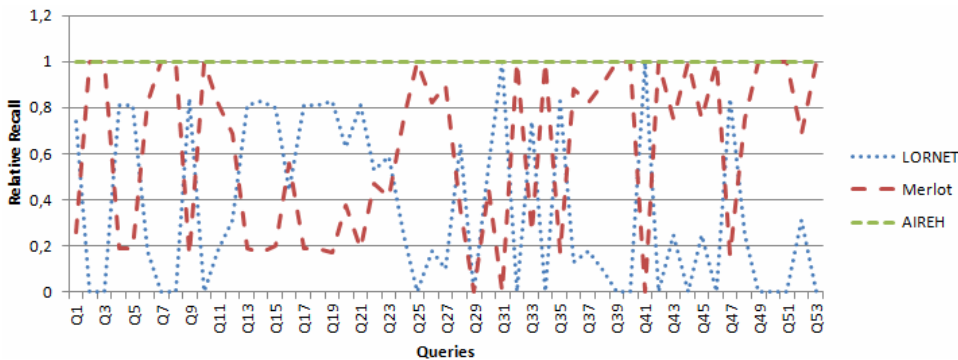


Figure 7 Precision evaluation (see online version for colours)

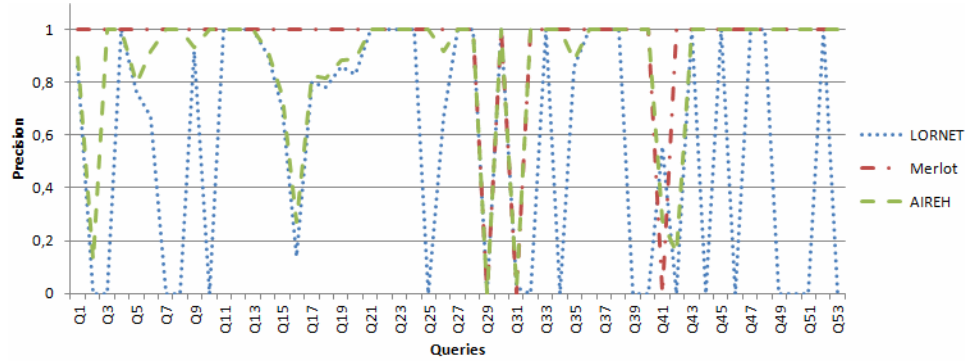
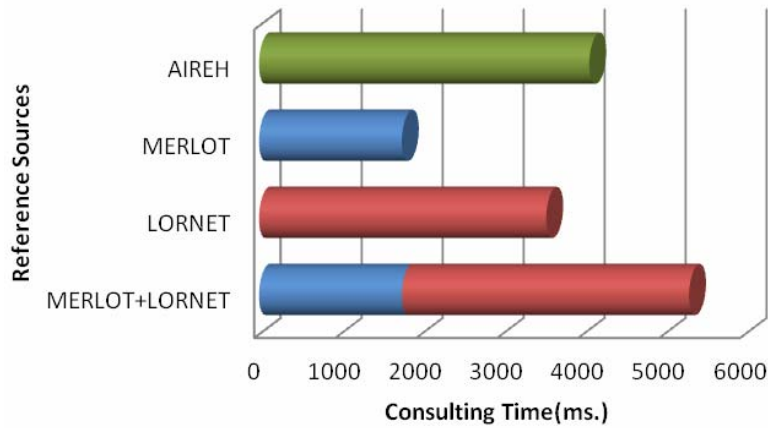


Figure 8 Evaluation of query time for several repositories (see online version for colours)



Data reveals that the proposed architecture significantly increases the number of LO to recover in time, and increases the temporary gain value of the system by about 15% on average as compared to isolated repositories such as Merlot or LORNET. Figure 5 shows the average temporary gains for the series of tested queries.

Related to measures of relevance of the retrieved content Figure 6 shows relative recall increases to its maximum value, 1, because the system filters those LO that have no information about the location of the resource metadata.

The system is robust against failure because it incorporates several methods within different agents in the organisation during the time that the query is being processed. Prior to organising the repositories according to performance, the repository manager agent checks the active repositories for any request made by the query manager. Once the LOR agents that perform the query in each repository have been instantiated, each agent is responsible for cancelling the query and reporting to the statistics agent if there are any situations that affect the established QoS levels, such as query time, performance of the repository, and so on. Although some of these LOR cancel the consultation process, there would still be other agents operating the instantiated LOR, which is always monitored by the repository manager. Because the system filters all the repositories that are not

answered, the number of LO that the user has access to consists of the sum recovered from LO repositories at the time of consultation increasing the number of LO while centralised the recovered at same point.

6 Conclusions

The search and location for educational contents, LOs at this work, turns out to be an active investigation field besides the development of educational systems based on open networks for the reutilisation of educational resources. A new approach for intelligent search of educational content was introduced by proposing an architecture based on the application of virtual organisations of multi-agent systems.

Taking into account the results obtained from the construction of the AIREH system, it is possible to conclude that the agent-based architecture is efficient to solve the problem of federated searches in heterogeneous repositories, due to the adaptation and learning facilities involved. The above mentioned architecture allows the adjustment of the system in a changeable environment depending on the load of the system and of the condition that is observed in the repositories depending on the on fly performance and includes several aspects related to LO context and user preferences. As an example, it has obtained a reduction of the delay time for the final results by optimising the query while increases the temporary gain and relative recall.

This work demonstrate the possibility of modelling an efficient system for managing open systems from a model of an adaptive organisation that provides distributed planning, and is flexible and dynamic. The model and the technologies presented in this research are an example of the potential for developing retrieval systems for digital content based on the paradigm of virtual organisations of agents. The advantages of the proposed architecture are its flexibility, customisation, integrative solution and efficiency.

There are still many different topics to be covered for enhancing the performance of searching and retrieval educational content with personalisation aspects. However, the research presented in this paper has provided suitable strategies for the future research towards enhanced content based retrieval systems.

Acknowledgements

This paper is a revised an expanded version of a paper entitled ‘Automatic learning object extraction and classification in heterogeneous environments’ presented at 9th International Conference on Practical Applications of Agents and Multiagent Systems to be hold in Salamanca, 28–30 March 2012.

This work has been partially supported by the project MICINN TIN 2009-13839-C03-03.

References

- Almaraz, Á. (2008) *Búsquedas Federadas Significativas en repositorios de Objetos de Aprendizaje*, Disertación para MsC en Ciencias de la Computación.
- Avancini, H. and Straccia, U. (2005) 'User recommendation for collaborative and personalised digital archives', *International Journal of Web Based Communities*, January, Vol. 1, No. 2, pp.163–175.
- Bates, T. (2005) 'Questioning the value of re-usable learning objects in education: the need for a business case', *II Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables*, Barcelona.
- Callan, J. (2000) 'Distributed information retrieval', in Croft, W.B. (Ed.): *Advances in Information Retrieval*, Chapter 5, pp.127–150, Kluwer Academic Publishers.
- Casquero, O., Portillo, J., Benito, M. and Romo, J. (2008) 'BILDU: compile, unify, wrap, and share digital', *Learning Resources Interdisciplinary Journal of E-Learning and Learning Objects*, Vol. 4, pp.97–111.
- CORDRA (2005) *ADL CORDRA (Content Object Repository Discovery and Registration/Resolution Architecture)* [online] <http://lab.usgin.org/standards/cordra-content-object-repository-discovery-and-registrationresolution-architecture> (accessed 30 April 2013).
- DCMI Specifications (2008) [online] <http://dublincore.org/specifications/> (accessed 30 April 2013).
- Dietze, S., Gugliotta, A. and Domingue, J. (2007) 'A semantic web service oriented framework for adaptive learning environments', *Lecture Notes in Computer Science*, Vol. 4519, No. 2007, pp.701–715.
- Fielding, R.T. and Taylor, R.N. (2000) 'Principled design of the modern web architecture', in *Proceedings of the 22nd International Conference on Software Engineering (ICSE '00)*, ACM, New York, NY, USA, pp.407–416.
- Gil, A., Prieta, F. and Rodríguez, S. (2011) 'Automatic learning object extraction and classification in heterogeneous environments. Highlights in practical applications of agents and multiagent systems', *Advances in Intelligent and Soft Computing*, Vol. 89, No. 2011, pp.109–116, Springer Berlin Heidelberg, DOI: 10.1007/978-3-642-19917-2_14; ISBN 978-3-642-19916-5.
- Gil, A., Rodríguez, S., De la Prieta, F., de Paz, F. and Martín, B. (2012) 'CBR proposal for personalizing educational content', *Advances in Intelligent and Soft Computing*, Vol. 152, pp.115–123, Springer, Berlin, Heidelberg, DOI: 10.1007/978-3-642-28801-2_14, ISBN: 978-3-642-28800-5.
- Hanafy, M. and Fakhry, M. (2008) 'A proposed architecture for learning object selection and discovery based on WSMO', in the *Proceedings of the 2008 4th International Conference on Next Generation Web Services Practices*, pp.10–14.
- Hatala, M., Richards, G., Eap, T. and Willms, J. (2004) 'The interoperability of learning object repositories and services: standards, implementations and lessons learned', in *Proceedings of the 13th international World Wide Web Conference on Alternate Track Papers & Amp, Posters, WWW Alt. '04*, New York, NY, USA, 19–21 May, ACM, New York, NY, pp.19–27.
- IEEE LOM 1484.12.1-2002 (2002) Draft Standard for Learning Object Metadata, The Institute of Electrical and Electronics Engineers, Inc.
- IMS DRI (2003) *IMS Digital Repositories Specification. V.1* [online] <http://www.imsglobal.org/digitalrepositories/> (accessed 30 April 2013).
- Kärger, P., Ullrich, C. and Melis, E. (2006) 'Integrating learning object repositories using a mediator architecture', *Proceedings of EC-TEL LNCS*, No. 4227, pp.185–197.
- Lee, M-C., Tsai, K.H. and Wang, T.I. (2008) 'A practical ontology query expansion algorithm for semantic-aware learning objects retrieval', *Computers & Education*, May, Vol. 50, No. 4, pp.1240–1257.

- Li, Y., Chen, Z., Huang, R. and Cheng, X. (2009) 'An integrated learning resource management system with web services', *New Trends in Information and Service Science, 2009. NISS '09. International Conference on New Trends in Information and Service Science*, Vol. 1, pp.863–868.
- Lujara, S.K., Kissaka, M.M., Bhalalusesa, E.P. and Trojer, L. (2007) 'Learning objects: a new paradigm for e-learning resource development for secondary schools in Tanzania', in *Proceedings of World Academy of Science, Engineering and Technology*, October, Vol. 24, ISSN 1307-6884.
- Manouselis, N., Drachler, H., Vuorikari, R., Hummel, H. and Koper, R. (2011) 'Recommender systems in technology enhanced learning', *Recommender Systems Handbook*, pp.387–415, Springer.
- Margain Fuentes, M.d.L., Muñoz Arteaga, J., Álvarez Rodríguez, F., Vanderdonk, J. and Orey, M. (2010) 'MIRROS: intermediary model to recovery learning objects', *Computación y Sistemas*, Vol. 13, No. 4, pp.373–384, ISSN 1405-5546.
- Massart, D., Nicholas, N. and Ward, N. (2010) IMS GLC Learning Object Discovery & Exchange (LODE), Version 1.0, (Base Document) Draft 14, March.
- Mayorga, J., Cigarrán, J. and Rodríguez-Artacho, M. (2010) 'Retrieval and clustering of web resources based on pedagogical objectives', *Actas de los Talleres de las Jornadas de Ingeniería del Software y Bases de Datos*, Vol. 4, No. 3, pp.124–130.
- MPEG-7 (2010) [online] <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm> (accessed 30 April 2013).
- Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) (2008) *The Open Archives Initiative Protocol for Metadata Harvesting*, Protocol Version 2.0 of 2002-06-14, Document Version 2008-12-07T20:42:00Z [online] <http://www.openarchives.org/OAI/openarchivesprotocol.html> (accessed 30 April 2013).
- Prieta, F., Paz, F., Gil, A. and Boureux, M. (2012) 'Retrieving learning resources over the cloud', *Advances in Intelligent Systems and Computing*, Vol. 173, pp.47–56, Springer, Berlin, Heidelberg, DOI: 10.1007/978-3-642-30859-8_5, ISBN: 978-3-642-30858-1.
- Qu, C. and Nejd, W. (2004) 'Interacting the Edutella/JXTA Peer-to-peer network with web services', *SAINT 2004*, pp.67–73.
- Rodríguez, S., Pérez-Lancho, B., De Paz, J.F., Bajo, J. and Corchado, J.M. (2009) 'Ovamah: multiagent-based adaptive virtual organizations', *12th International Conference on Information Fusion*, Seattle, Washington, USA, July.
- Santiago, R.d. and Raabe, A.L.A. (2010) 'Architecture for learning objects sharing among learning institutions – LOP2P', *IEEE Transactions on Learning Technologies*, April–June, Vol. 3, No. 2. pp.91–95.
- Segura, N.A., Sanchez, S., Garcia-Barriocanal, E. and Prieto, M. (2011) 'An empirical analysis of ontology-based query expansion for learning resource searches using MERLOT and the Gene ontology', *Knowledge-Based Systems*, February, Vol. 24, No. 1, pp.119–133.
- Segura, N.A., Vidal, C.C. and Prieto, M.M. (2010) 'Query expansion based on domain ontology for learning objects search', *2010 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, 9–11 July, Vol. 1, pp.446–451.
- Shih, T.K., Chang, C-C. and Lin, H.W. (2006) 'Reusability on learning object repository', *Proceedings of ICWL LNCS*, No. 4181, pp.203–214.
- Simon, B., Massart, D., Van Assche, F., Ternier, S. and Duval, E. (2005a) 'A simple query interface specification for learning repositories', *CEN Workshop Agreement (CWA 15454)* [online] <ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/WS-LT/CWA15454-00-2005-Nov.pdf> (accessed 30 April 2013).
- Simon, B., Massart, D., Van Assche, F., Ternier, S., Duval, E., Brantner, S., Olmedilla, D. and Miklos, Z. (2005b) 'A simple query interface for interoperable learning repositories', *1st Workshop on Interoperability of Web-based Educational Systems*, Chiba, Japan.

- Soto, J., García, E. and Sánchez-Alonso, S. (2007) 'Semantic learning object repositories', *International Journal of Continuing Engineering Education and Life Long Learning*, Vol. 17, No. 6, pp.432–446.
- SRU (2010) Search/Retrieve via URL, Draft SRU 2.0, 9 December [online] <http://www.loc.gov/standards/sru/> (accessed 30 April 2013).
- Ternier, S., Massart, D., Totschnig, M., Klerkx, J. and Duval, E. (2010) 'The simple publishing interface (SPI)', *D-Lib Magazine*, September/October, Vol. 16, Nos. 9/10, DOI:10.1045 [online] <http://www.dlib.org/dlib/september10/ternier/09ternier.html> (accessed 12 June 2013).
- Yen, N.Y., Shih, T.K. and Jin, Q. (2010) 'A new paradigm of ranking & searching in learning object repository', in *Proceedings of the Second ACM International Workshop on Multimedia Technologies for Distance Learning (MTDL '10)*, ACM, New York, NY, USA, pp.1–6, DOI: 10.1145/1878052.1878054.
- Z39.50 (2003) *Information Retrieval (Z39.50): Application Service Definition and Protocol Specification* [online] <http://www.loc.gov/z3950/agency/Z39-50-2003.pdf> (accessed 30 April 2013).
- Zhou, S., Xu, M. and Guan, J. (2010) 'LESSON: a system for lecture notes searching and sharing over internet', *Journal of Systems and Software*, October, Vol. 83, No. 10, pp.1851–1863.