

## Article

# Smart System for the Retrieval of Digital Educational Content

Ana B. Gil <sup>1,\*</sup> , Fernando de la Prieta <sup>1</sup> , Sara Rodríguez <sup>1</sup> and Juan M. Corchado <sup>1,2,3,4</sup> 

<sup>1</sup> Bisite Research Group, University of Salamanca, Calle Espejo 2, 37007 Salamanca, Spain; fer@usal.es (F.d.l.P.); srg@usal.es (S.R.); corchado@usal.es (J.M.C.)

<sup>2</sup> Air Institute, IoT Digital Innovation Hub (Spain), Carbajosa de la Sagrada, 37188 Salamanca, Spain

<sup>3</sup> Department of Electronics, Information and Communication, Faculty of Engineering, Osaka Institute of Technology, Osaka 535-8585, Japan

<sup>4</sup> Pusat Komputeran dan Informatik, Universiti Malaysia Kelantan, Karung Berkunci 36, Pengkaan Chepa, Kota Bharu 16100, Kelantan, Malaysia

\* Correspondence: abg@usal.es; Tel.: +34-670-584-983

Received: 15 August 2019; Accepted: 11 October 2019; Published: 17 October 2019



**Abstract:** The education sector is a major generator, consumer, and depositary of educational content. Thanks to technological advances, today's educators and learners have ubiquitous and on-demand access to information. Technology has made it possible for us to communicate and share information effortlessly from anywhere in the world. However, the availability of vast amounts of heterogeneous educational content will not be useful unless we search, retrieve and integrate it, creating interoperable educational environments. The current challenges to integrating educational content arise from its distribution over several repositories. This research proposes AIREH (architecture for intelligent retrieval of educational content from heterogeneous environments), for the retrieval of digital content through agent-based virtual organizations. This flexible architecture facilitates the search for and integration of heterogeneous content through an information retrieval model that involves both case-based reasoning and federated search. Moreover, AIREH is based on an adaptive organization model for distributed planning, thanks to which, it manages open systems flexibly, dynamically, and effectively. The conducted case study gives very promising results and demonstrates the advantages of using agent-based virtual organizations in the retrieval of labeled digital content. The proposed model is flexible, customizable, comprehensive and efficient.

**Keywords:** Multiagent systems; educational content; case-based reasoning; artificial intelligence

## 1. Introduction

Educational institutions are major generators, consumers and depositaries of didactic content. It is important, therefore, to implement content management solutions in the education sector, where they can facilitate the search and dissemination of learning content. Digital content management has evolved visibly in the last decade, introducing new methods for the management and organization of large volumes of digital content. The Internet is used as the generator and the possessor of a large number of data warehouses. This has led to the emergence of phenomena such as globalization and decentralization. Data warehouses classify information according to several criteria and store it in specialized data repositories. Each repository has different access mechanisms, resulting in dozens of communication protocols. It is, therefore, necessary to develop different techniques, tools, and methodologies whose combination could provide a technological solution, facilitating access to the educational content stored in repositories. The solution must adapt to an environment that is composed of the total number of (i) highly dynamic, (ii) heterogeneous, (iii) open contexts.

In educational technology, one of the most widespread approaches to making educational content reusable and interoperable is to fragment it into independent, modular units that can be used in different environments and applications [1]. This modular approach makes it possible to easily access and combine educational content items for a single learning objective, this is termed as a learning object (LO) [2,3]. The ability to access simple didactic resources specifically designed for use in different contexts facilitates the process of creating educational content. This reduces the time of creating and compiling educational resources, enabling the efficient development of comprehensive educational courses. The change in the creation process takes place when it is based on the search and aggregation of these small modular pieces, as simple didactic resources specifically designed to be used in different contexts. Normally, LO are stored in digital repositories, called learning object repositories (LORs). These systems must also have the functionalities of importation, exportation, identification, and retrieval of content [4,5]. The user usually interacts with the information contained in the repository via their web browser.

To support the reuse process (speed, flexibility, and uniformity), the learning object has external descriptors or labels, which enable the identification, organization, search and retrieval of content items. These descriptors are created with standard-based metadata, ensuring the interoperability of systems and the description of content items using an adequate set of attributes. Currently, there are many different standards. The Dublin Core Metadata Element Set [6] and the IEEE Learning Object Metadata (IEEE LOM) [7] are the most popular metadata standards. There are also others such as Metadata for Learning Opportunities (MLO), eX-changing Course Related Information (XCRI) along [Schema.org](https://www.schema.org/) vocabularies and more. There have been many initiatives aimed at finding international [8] and national [9–11] metadata standards. The Learning Resource Metadata Initiative (LRMI), is a noteworthy project which began in 2011 and became known as the Dublin Core Metadata Initiative (DCMI) in October 2014, which oversees the management of the specification [12]. All the efforts to make knowledge management more sustainable are driven by the same conviction. That standard facilitate the management of resources. Compatible standards make knowledge management sustainable because they enable system interoperability and lower system development costs.

Technology has made it possible for us to communicate and share information effortlessly from anywhere in the world. However, the availability of vast amounts of heterogeneous educational content will not be useful unless we search, retrieve, and integrate it, creating interoperable educational environments. Integrating content is challenging because there tends to be a vast amount of distributed educational content with multiple characteristics, and multiple users must be able to access it. In addition, new web applications manage information that is distributed in different geographical areas, this further complicates the integration of data and the incorporation of security measures. In many cases, such applications are unable to fully guarantee the security of the data.

This paper presents the AIREH framework (architecture for intelligent retrieval of educational content in heterogeneous environments), which is capable of searching, retrieving and sorting LOs, according to the case-based reasoning (CBR) [13] system rules established by a user. After retrieving the learning objects, AIREH filters and classifies them to facilitate the search process. The framework is based on an agent-based virtual organization [14] that is capable of operating in a heterogeneous environment.

The rest of the paper is structured as follows: Section 2 reviews the state-of-the-art literature associated with the current problem. Section 3 presents the architecture of the proposed model. Section 4 describes the case study that has been conducted in a real environment to validate the performance of the developed model. Section 5 analyzes the experimental results obtained from the case study. Finally, Section 6 draws conclusions from the results and describes future lines of research.

## 2. Related Work

There are many state-of-the-art developments that are based on the LO paradigm [15–17] and there are many digital content repositories that store such content through the LO format. However, neither traditional web interfaces nor conventional search engines are able to manage all the labeled

LOs directly. Educators are not aware of the existence of different metadata standards and even if they were aware the content is often not labeled correctly. There is a large volume of educational content in the Web, however, standard web engines are only able to retrieve a part of the items, because they can access the visible Web but not the deep Web. In response to a user query, standard search engines make use of a centralized retrieval model that copies the information in a single centralized database and then indexes and sorts the contents by using classified documents in the database. However, this is not the case in the LORs where information can only be searched via the access mechanisms of each repository.

The educational context generates huge volumes of digital resources. The maturity of current technology has led to a significant increase in institutional repositories which form part of the hidden network of large databases. It is, therefore, necessary to find solutions that allow for the efficient search of heterogeneous contents in a distributed context. A single repository would not provide a general knowledge base. It is, therefore, necessary to bring together all the available digital repositories, creating in this way a knowledge network with a distributed repository infrastructure. The topology of distributed repositories may vary, ranging from total decentralization, through possible intermediate situations, to absolute centralization.

Distributed information retrieval, or federated searches [18] attempt to respond to the problem of retrieving information from the hidden Web. The main aim of a federated search is to develop models and strategies to retrieve information from distributed sources. The process is completely transparent to the user, who does not perceive the complexity involved and uniformly processes the retrieved information. The main contribution of a federated search is that the search is performed in multiple LO repositories. In addition, the location of each source is provided, and information extracted from hidden sources is controlled distributedly. Federated search is a much more complex and comprehensive centralized mechanism than traditional search engines, which do not provide real solutions to the problem of the Hidden Web.

Federated searches can be applied in the retrieval of LOs from LORs, resolving the following three problems in the sequence in which they are given: (1) Selection of repositories, (2) Selection and retrieval of resources relevant to the query, and finally and most importantly (3) Integration of the content retrieved from multiple sources, forming a single list which is delivered to the user. Moreover, the results are personalized for each user, adapting to their needs and preferences.

To arrive at this solution, the particularities of each of the involved elements will be described.

### *2.1. Standards for the Learning Object Paradigm*

E-learning standards still have much room for improvement and development. More time is needed before standards become generalized and consolidated, enabling greater interoperability, reusability, manageability, accessibility, durability, scalability and reliability.

In the case of LO, metadata describe the characteristics of educational content items by means of labels in order to guarantee technical interoperability. Labels enhance the usefulness of the LO, making it easier to reuse items by facilitating their storage, retrieval, localization, and exchange. Moreover, labels make it possible to quickly recognize the characteristics of the content items stored in the repositories. Nevertheless, the metadata are not useful if users do not use metadata standards or use them incorrectly. The adequate use of standards is essential to establish standardized norms for these descriptors and to achieve interoperability between content items. It is evident that the LO paradigm has contributed to the emergence of many technological perspectives on the reuse of learning content. Beyond the standard metadata formats of the previously listed LOs, there are a large number of other standards that facilitate the management and communication of platforms and their integration into more complex models. There exist several standards proposing conceptual models for the aggregation of learning resources, such as the Resource Aggregation Model for Learning, Education, and Training (RAMLET) [19]. It aims to facilitate interoperability by providing an ontology that allows in the mapping of existing aggregation formats [20,21]. Many are the challenges that have emerged during

the innovation and development processes, such as including adaptive mechanisms in the use of LOs [22], semantic web services [23], and developing complex services [24]. These challenges are yet to be overcome. For now, the problems that impede the interoperability of educational content can be divided into three main groups:

1. The biggest problem is the LOR's tight structure, which prevents external management from becoming flexible and powerful. These features are essential if we are to ensure interoperable systems and easy access to dispersed and heterogeneous sources. This tight structure also impedes external users from managing resources. This tight structure also impedes comprehensive user management according to user interests in line with other users in the same context. Furthermore, the LOR's search interface does not display accurate and reliable LO information because it cannot retrieve data from the Deep Web.
2. Another problem is that the internal logic of the majority of LORs is not understandable to the applications that access them. Most of them need some intermediate abstraction layer, such as web services or related technologies. The extraction of LOs is, therefore, a complicated and slow process that sometimes requires the user to intervene manually. The systems that include a middleware layer also encounter problems, such as high response times, unavailable LORs, erroneous results, etc.
3. The following group of problems is directly related to the absence of automatic mechanisms that would control the quality of the labels and of the contents, according to technical, semantic and syntactic aspects of the LOs, ensuring the correct specification of these LOs in any of the metadata schemes that describe them. This improvement would provide simple channels for the user to access all possible resources. On many occasions, the same contents may have different metadata, depending on the repository in which the search is performed.

Beyond the problems associated with the LORs, there are some additional problems directly related to the specification of LOs. Standard metadata labeling must accurately define the educational resource, however, current mechanisms do not establish a minimum number of characteristics that an LO must have in order to be used within a particular educational context. This results in two problems. The first one is related to the unique identification of each LO. A learning object is well-defined a priori if it has a catalog ID (repository) and an entry ID to the repository in which it is stored. The second one is associated with the absence of the object located in the LOR, in which case the LO is well described but the content cannot be accessed. Content and labeling are like two sides of the same coin and should go together and function properly. That is why unity should be considered an important aspect of this type of technology.

The repository context experiences two problems. The first problem is the dispersion and heterogeneity of the labeled content items in multiple repositories. The second problem is intrinsic to the content, such as incorrect labelling which impedes the efficient reuse of educational content. To solve these problems, it is necessary to find solutions that will ensure the presence of both, labels and content within LOs. Although the LO paradigm solves the problem to a large degree, some important issues remain unresolved. Also, it is necessary to propose solutions that would adapt to the high number of metadata standards (Dublin Core, IEEE LOM, etc.), query languages—e.g. VSQL (Very Simple Query Language) [25]—, middleware layers—e.g. SQI (Simple Query Interface) [26]—, Open Archives Initiative Protocol for Metadata Harvesting (OAI-MPH) [4], etc., repository architectures [27–29]. These solutions should allow for a centralized global search, enabling the efficient reuse of resources. To this end, it is necessary to achieve a higher level of abstraction and to improve the LO's storage processes. In addition, detailed labelling would further improve the search and discovery of content items.

## 2.2. State of the Art

To standardize the use of LOs it is necessary to facilitate their search and retrieval [30]. Several studies have already pursued the goal of improving the visibility of LOs. BILDU [31] has proposed

an architecture that consisted of a set of horizontally integrated vertical search engines. The BILDU architecture has a single interface, which supports the incorporation of social recommendation mechanisms. Content integration and repository searches through peer-to-peer mechanisms are proposed in Edutella [32]. Edutella's architecture is based on mediators which enable the integration of LORs [33]. Different methods are applied to retrieve educational content [34], including the metric proposed for the calculation of similarity between different LOs.

Content retrieval processes will only be considered successful if they enable the consensual use of educational content and its structures. It is crucial that the research community focus its current efforts on establishing the characteristics, constraints or conditions for the appropriate use of these learning resources. This process must, therefore, involve several disciplines, a study of the semantic aspects of resources, an improvement of the communication standards and the management of large databases.

The Ontology Query Expansion Algorithm [35] adapts an ontological approach to LO retrieval by expanding the query algorithm with semantic aggregation and a dynamic mechanism for searching and filtering LOs whose theme is genetics. This mechanism has been validated in the MERLOT (Multimedia Educational Resource for Learning and Online Teaching) repository where the mechanism searched for LOs using Gene Ontology [36]. The mechanism is connected to the MERLOT Restful service for the retrieval of contents. Many varied solutions have been proposed [37–40], although most of them have not been implemented in real environments, and the majority do not focus specifically on using the LO paradigm for the retrieval of educational content. Search results have been improved by proposals of more complex solutions that implement different technologies and social recommendations. Only two proposals, Fuentes et al. [41] and Carrion et al. [42], make use of agent technology, but neither of the two approaches coordinates the completion of tasks by agents within the system.

Our review of the state of the art confirms that it is necessary to continue working for greater interoperability and efficiency in the reuse of resources. It has also been established that to achieve interoperability and efficiency, it is important to develop features such as adaptability, flexibility, and accessibility. Due to the vast number of LORs, a major challenge with regard to LO recommendation is to find the LOs that best suit the users' search criteria.

E-learning researchers and developers have begun to integrate information retrieval techniques with recommendation technologies [43,44], collaborative filtering [45], web mining [46]. The information extracted by information filtering techniques is associated with the attributes of the educational resource (content-based approach) and the user context (collaborative approach). One of the first works developed in this context was Altered Vista: a system in which teaching techniques are evaluated by a recommendation engine with collaborative filtering [47]. This work researches ways of collecting user reviews of learning resources and of propagating them in the form of word-of-mouth recommendations. RACOFI is a Rule-Appling Collaborative Filtering system whose architecture enables custom educational content selection [48,49], it improved collaborative filtering in its proposal of the ecological approach to designing e-learning systems. The key aspects of this proposal are that it considers the gradual accumulation of information and that it focuses on end users.

Some solutions take a hybrid approach, such as Dong et Wang [50] and Ghauth et Abdullah [51], which make use of algorithms based on the user reviews of learning objects. The most similar users are extracted through nearest neighbor algorithms. These correlation-based algorithms are used to calculate an index score on the usefulness of LOs through the analysis of comments from students with similar profiles. This algorithm helps prevent the cold start problem when no data are available and there are no values for the LO. However, all of the selected LOs are treated equally and indistinguishably, making it possible to assess user preferences more precisely, affecting the very pattern of user preferences. The use of algorithms based on biological models, such as ACO (Ant Colony Optimization) is the basis of Yang et Wu [52], which proposes a set of attributes based on a colony of ants to help students find their way through an adaptive LO model more efficiently. This work is interesting because it bases its recommendation on the generation of different learning paths through different sets of LOs.



The ultimate goal of these solutions is to help the user learn about a specific topic by recommending a sequence of LOs that creates an optimal learning path.

Older studies evidenced the lack of operational solutions [22,39,47], while more recent works [52–57] point to the need for improving the search mechanism for the selection of LOs. There is a growing number of papers that propose systems for the recommendation of learning resources. Some of them take into account the educational content described by metadata and the user's learning context as the basis of recommendations. They propose a schema called Contextualized Attention Metadata (CAM) for the capture of information on the different processes that occurred during the LO life cycle, including their creation, labeling, supply, selection, use, and maintenance. These studies propose metrics for use in the LOM standard and the CAM schema for the recommendation of the retrieved LOs and their classification according to criteria such as popularity ranking, the similarity of items, the number of downloads, and more. A highly interesting field of study is understanding how these rankings contribute to the selection of LOs and how they combine with each other, questions that have not been answered yet.

All the reviewed state-of-the-art proposals concluded that effective recommendation mechanisms can be created by incorporating mechanisms for the assessment of the characteristics of educational content (LOs in our work), user context and the user's interaction with the content. However, taking a closer look at the reviewed LO search proposals, we have noticed that they lack applications in the real environment. In short, there is no standard that would cover all the needs of the process of searching, retrieving and reusing learning objects. Current technology does not give solutions to the automatic or semi-automatic search and retrieval of LOs and, therefore, the reuse of learning resources.

For this reason, this paper proposes a flexible and modular architecture, allowing for the intelligent retrieval of the educational digital content stored in LO repositories. The architecture enables the interoperability of layers. It would also be important to incorporate efficient management techniques that would be capable of dealing with future challenges for the efficient management of services and platforms.

### 3. Model of the AIREH Architecture

#### 3.1. On the Proposed Solution

It is necessary to search for new paradigms and applications which are going to optimize the search for educational content distributed in different formats, servers, and networks. This paper proposes an Architecture for Intelligent Retrieval of Educational content from Heterogeneous environments, called by its acronym AIREH, as a solution. AIREH is an architecture for federated search of educational contents, specifically the so-called LO.

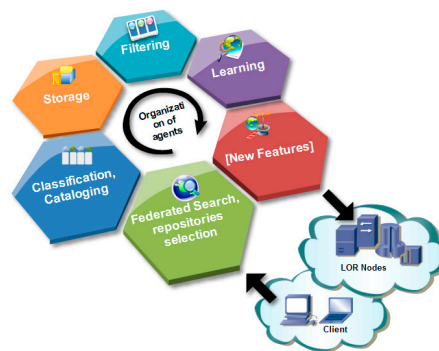
This federated search system consists of a series of elements: the repositories are distributed in physically dispersed places, access to these repositories is read-only, information systems operate under different platforms, repositories are distributed (metadata stored in one site and LO in another) and are managed by autonomous systems, in some cases access to repositories requires the creation of a session, in others the access is direct, etc.

In this sense, the choice of a multi-agent system (MAS) architecture has been ideal for the development of a retrieval system. This is because the characteristics (autonomy, situation, reactivity, rationality, intelligence, organization, mobility and learning) [58] of agent architectures have ensured the development of a stable system, with the ability to react intelligently to the needs of the environment. The architecture has been modeled as a virtual organization, because an organization can adapt its actions to changes in the environment, achieving its objectives and interacting with heterogeneous components. As a result, the authors have considered it to be an effective solution.

Education is a technologically heterogeneous and continually changing sector. The motivation behind this proposal is to enable teachers and students to optimally reuse educational content. To this end, it is necessary to establish effective processes for the search and retrieval of educational content.

The article proposes a model with an agent-based virtual organization architecture, where the tasks associated with the federated search of educational content, are distributed among different agent organizations. The innovation of this proposal lies in the flexibility of the proposed system, a feature that derives from the underlying architecture. Moreover, the ability of agents to self-adapt to highly dynamic environments makes the model adaptable to changes.

The AIREH architecture distributes the different tasks among nodes, as illustrated in Figure 1. As shown, the federated search for LOs involves six separate tasks. The first task (federated search and repository selection) is to initiate and supervise the entire federated search process. In addition, the active repositories that are relevant to the query are detected. The second task (classification, cataloging, etc.) involves the ranking and classification of the retrieved LOs. The next task is to store the LO processing results and statistics. After that, a filtering process is carried out to extract useful information, according to the search criteria established by the user. To provide the user with the most accurate solutions, we have coupled AIREH with a case-based reasoning system that learns from previous search experiences. The CBR system can adapt past solutions to new cases, facilitating the search process and increasing the accuracy of the results delivered to the user. Finally, as a result of completing the tasks described above, new features are applied in the search for an answer to the user's query.



**Figure 1.** Distribution of tasks in the architecture for intelligent retrieval of educational content from heterogeneous environments (AIREH) architecture.

The proposed architecture solves the problems associated with repository distribution, integration, and internal logic. It also provides a solution for the classification, storage, and retrieval of LOs, in a completely transparent way. In addition, AIREH is a scalable architecture, which means that it is going to be compatible with new protocols, the architecture also includes internal components, such as repositories, catalogs or heterogeneous applications designed to cover service-related features.

The cornerstone of this work is the retrieval of LOs in a real environment using federated search in different LORs. To cover the current research gap, this proposal provides the user with a framework that unifies the search and retrieval of objects and establishes a set of rules to facilitate the filtering and classification of the retrieved LOs. The rules that are established for the organization of the retrieved items are based on educational metadata and provide useful and relevant content to end users. In addition to the search feature, AIREH incorporates mechanisms that document the retrieved LOs, making it possible to list the LOs in order of relevance to the user's search. The architecture proposed in this work provides multiple perspectives to assess the retrieval of educational content in a real, open and scalable environment and is a support mechanism for the implementation of recommendations and the ranking of the retrieved LOs according to relevance.

### 3.2. Description of the Model

MAS methodologies provide researchers with robust and reliable architectures in which they can deploy their solutions. In recent years, researchers have proposed new procedures and methodologies for the design of open MAS. Open MAS permit heterogeneous agents to interact with different





The next step to understanding Figure 2, is to describe the role played by each agent and its relationship with other agents. Therefore, agents in Figure 2 have the following roles:

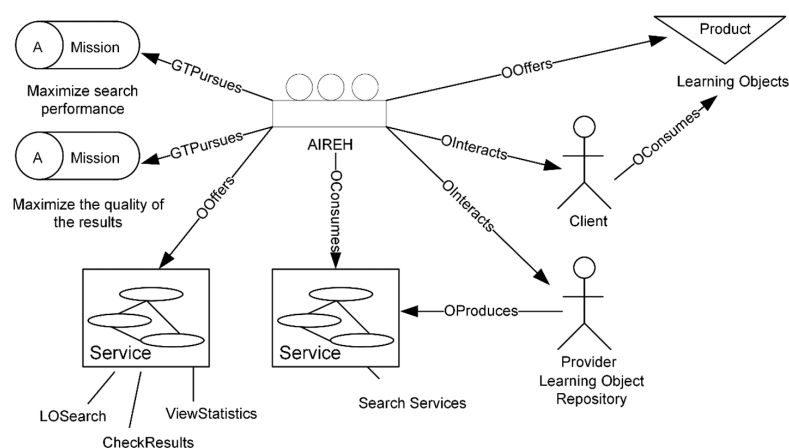
- **User agent.** Defines the user or client in the system. This agent is responsible for launching the federated search process by sending a search request to the Query Manager agent. This agent evaluates the results of the query, in terms of the relevance of the content of the LOs and the order in which they are given. Also, the user agent has access to statistical information about both LORs and LOs.
- **Query Manager agent (QMA).** This agent is in charge of supervising the entire federated search process. As mentioned previously, a federated search is a simultaneous search for LOs in multiple repositories. The agent receives a natural language query from the user agent and is responsible for completing the query by using propositional logic. It initiates the federated search by querying the Repository Manager agent with respect to the received query. When the Repository Manager agent indicates the end of the federated search, the QMA requests the Cataloguer agent to apply cataloguing techniques and collaborative filtering to the results. Then, the QMA sends the results about the federated query, ordering the items according to the preferences to the user agent for evaluation.
- **Repository Manager agent (RMA).** The role of this agent involves the control of the queries sent to different LORs, e.g., it has control over the federated search process, quality aspects. This agent receives formalized queries from the QMA. It checks the repositories that are active and requests all statistic data from the Statistics agent, these data are included in a list of high-performing repositories. It instantiates the LOR agents needed to perform the federated search, one for each of the repositories. Once the LOR agents have been instantiated, the RMA sends a query in formalized language to each LOR agent. The QMA notifies the RMA when the federated search is completed. It is responsible for monitoring the proper functioning of LOR agents and prevents yield loss if the repository stops working properly during the query.
- **Learning Object Repository agent.** Depending on the assigned LOR, there are different types of agents that take on this role. Each single LOR agent is in charge of conducting the query to a single repository. Each agent type implements a different middleware layer; however, this is not a problem as the overall system includes different agents for all possible middleware layers. In a case where a query is sent to a single repository with multiple interfaces, the interface with best performance is chosen. In a federated search there will be many LOR agents. This agent is responsible for requesting individual LOR agents, different instances of this agent will work simultaneously. The agent performs the LOR query, carrying out all the processes defined in the specification of the middleware layer of a given repository. The LOR agent is responsible for sending the LO results received in response to the query, it also sends the Statistics agent the statistical data related to the query.
- **Translator agent.** This agent is responsible for transforming the user query into the formalized language of the repository to which the query is sent. The agent receives the query in propositional logic from its LOR agent and converts it for the LOR, acting as an intermediary between the LOR and the architecture.
- **Results agent.** This agent receives all the LOs retrieved from each of the LOR agents during a federated search. It automatically extracts the information from the metadata schema, eliminating the items whose data are not valid. Although in the theoretical proposal there seems to be a single results agent, in the deployed model the same role is taken on by different agents. Each implements a different standardized metadata scheme. They are responsible for the correct reception of the federated search results by extracting useful information from their assigned LOR. This agent, therefore, extracts metadata and data structures from the LORs. Before storing the extracted LOs, it performs filtering that eliminates the defects that would otherwise impede the use of the LO by the users. It executes an algorithm that evaluates the degree of overlap between retrieved LOs,

avoiding duplicate LO. It collects relevant statistical data such as memory and response accuracy from the LOR agent and sends them to the statistical agent.

- **Cataloguer agent.** In coordination with the RMA, this agent is responsible for preparing the ranking of the LOs obtained as a result of the federated searches in the different LORs. After carrying out a pre-filtration, where it eliminates the incomplete LOs, it stores the rest. This agent implements CBR which uses information from previous searches in order to classify the elements that best suit the user's needs. This CBR incorporates each user's profile information as well as the type of educational content they are looking for (content-based filtering). Subsequently, it makes use of the user's votes as well as the suitability of the previous results classification (collaborative filtering). In this way, this agent orders the retrieved LOs according to the user's preferences, considering their profile and level of education. In order to carry out this process, the Cataloguer agent requests the LO voting histories, previous LO rankings and user feedback from the Statistics agent. With all this information it generates the ranking of the LOs that best adapt to the user who made the query. Like other agents in the organization, it also sends the ranking information to the Statistics agent, who stores it for future recommendations.
- **Statistics agent.** This agent is responsible for collecting and providing statistical data to other agents in the organization. It provides statistical data to the RMA which creates a list of high-performing repositories, making it possible to optimize the efficiency of the system. It helps the Results agent improve the quality of the results and it sends statistical data to the cataloguer agent which ranks LOs in each federated search. The Statistics agent receives query statistics from the LOR agents. Moreover, it receives the LO evaluation and results in relevance feedback from the User agent.
- **Supervisor.** This agent maintains overall control of AIREH. It analyzes the structure and syntax of all the messages that enter and leave the system, supervising the correct functioning of the agents within the architecture.

Following the guidelines outlined in the methodological guide of GORMAS, one of the first tasks is to instantiate the functional vision (mission) of the organizational model, which includes the products and services offered by the system, the type of environment, goals, global aims (mission and justification), and the interest groups that are covered and the information they consume.

The primary mission of the organization is to maximize the performance of the query system by reducing time and increasing performance, as well as the quality of results. Figure 3 details the elements in the organizational model (functional view) by showing the results (products and services), the type of work environment, and the stakeholders of the system. The architecture's logic facilitates the search for educational resources through the implementation of the federated search system. This functional view of the organizational unit shows some of the services offered by AIREH such as LOSearch ViewStatistics or CheckResults.



**Figure 3.** Organizational model diagram of the AIREH architecture (functional view).

Educational resources, LOs, are requested by the user through a simple interface that facilitates the search process in distributed repositories. The system offers educational resources as an end product in the form of an ordered and customized list of LOs relevant to the user. The system includes statistical information, such as system performance or the use of different repositories, identifying those used by the user according to their search patterns.

The GORMAS model also provides the functionality as an open system, meaning that both services should publicize the policies of acquisition and release roles, and determine what functionality must be implemented by internal staff and what is to be supplied by external agents as shown in Figure 4. Thus, the dynamism of the organization is modeled as an open system. LOR, User, Results, and translator agents are accessed by outsiders that require a procurement role. The remaining agents (Query Manager, Repository Manager, catalogue, and Statistics) are insiders and are externally unreachable.

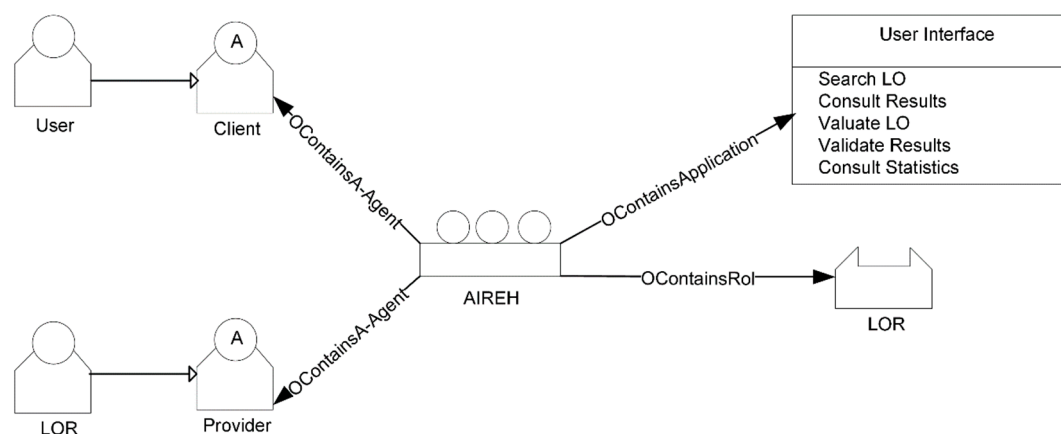


Figure 4. Organizational architecture model diagram (structural view).

## 4. Results

This paper proposed AIREH, an agent-based virtual organization architecture for the intelligent retrieval of educational content. To test the performance of the designed model, a case study has been conducted in which the model carried out a series of federated searches in LORs and then merged the retrieved results. In theory, this model is perfectly fit for its purpose, because it is a dynamic and heterogeneous model of MAS. However, it was necessary to obtain empirical evidence before making such kind of an assumption. The main objective of the conducted case study was to demonstrate the usefulness of the model in practice, assessing its performance under a set of metrics. Specifically, the completeness and accuracy metrics described in the next subsection have been used in the assessment. In addition, the metrics for distributed information retrieval were implemented in the model to ensure the optimal reuse of LOs, they enabled the system to retrieve LOs according to a series of relevance criteria, such as the utility criterion.

### 4.1. Experimental Results

As demonstrated previously, educational metadata describes learning content by using the labels to which they are associated. Metadata can only be used effectively by a user who has knowledge about the different standards and knows how to use them. The labeling process, which is generally manual, results in poorly described documents as their labels often miss out on the key features of the content. This makes it difficult, or impossible in some cases, to directly use that metadata in the process of searching and retrieving content. The extension of standard specification levels produces LO attributes that are either irrelevant or ambiguous and difficult to apply.

A detailed study of the education sector led to the emergence of a new pruning heuristic that allows AIREH to select the most relevant metadata from the set of metadata of a given LO and quickly filter the majority of irrelevant content items by means of the Results and the Cataloguer agents.

Once this process has taken place, the remaining LOs are filtered according to a series of quality parameters of the retrieved metadata documents. After that an algorithm for selecting user-adapted content must take into account the semantics of the LO and the technical aspects of searching in the LOR, cataloguing the retrieved results automatically through intelligent mechanisms. In this work, the metadata documents were processed, and their quality level was established considering three different aspects: completeness, reliability, and accuracy of results management:

- **Completeness.** The metadata of the LO must give a detailed description of its contents. Compliance with this condition will lead to a more rigorous search process, better reasoning mechanisms and a more accurate recommendation system.
- **Reliability.** Good quality metadata labelling is essential for the optimized retrieval of content items. The metadata contains LO access information (file, metadata, and educational content) and gives the LOR that stores them a higher degree of trust in the system.
- **Accuracy of Results.** An answer to a user request may include a large set of LOs. The classification algorithm filters the LOs in this set and sorts them according to their relevance to the user's request and the educational context of the query. Ordering the LOs according to their relevance is important because it facilitates the user's choice. Attributes associated with the user domain are important because they enable the CBR recommender system to make personalized content-based recommendations.

#### Evaluating the Performance of the Content Retrieval Architecture

The case study has been conducted to verify the adequacy of the proposed retrieval architecture plays and its role in the retrieval of educational content. The case study consisted of a federated search system which involved 2 LOR agents: an agent that queries user agent type, and a translator. Both LOR agents implemented the SQI client code repository access interface. Tests have been performed on two repositories, the MERLOT [65] repository (statefull) and the Learning Object Repository Network LORNET [66] (stateless), each repository has different session management. The Query Manager agent implements VSQL.

The case study has been conducted in a real environment. It evaluated AIREH's efficiency in retrieving content items from two LORs. To evaluate the quality of AIREH's information retrieval, a series of evaluation metrics have been used. The determination of LO retrieval performance in the different LORs in this paper is based on the calculation of Relative recall and precision were calculated to determine the system's performance when retrieving LOs from different LORs. Moreover, we have proposed a new metric called gain, which calculates the time it takes to retrieve educational content from LOR. The gain formula is detailed below.

Formally, given a set of  $n$  LO metadata retrieved for a given LOR  $j$ , the set of such metadata files will be determined by  $O_j = \{O_1 \dots, O_i\}$ , with  $i = 1, \dots, n$ . Precision,  $P$ , is related to the utility or potential use of the recovered materials in relation to achieving objectives, interest or problems intrinsic to the user. According to the general approach to content retrieval, the metric has to be adapted to the context of this work. Thus, binary management is performed on the metadata of the recovered LO, categorizing them according to relevance criteria:  $R = \{0,1\}$ . The selection criteria for binary values are:

- If the LO cannot be recovered because it lacks a label that would indicate the source of the resource (mainly the <location> attribute of the <technical> category of LOM), it is qualified as irrelevant and is attributed a 0.
- In any other case, it is qualified as relevant with value 1.

When determining relevance, this binary criterion makes it possible to elaborate our own formula to measure the precision of retrieval systems for LO through the different queries in each repository. The precision measurement in the retrieval of LO of a given LOR  $j$  is then defined by Equation (1).

Once the query  $Q$  has been generated in a series of LORs, the set of recovered LO metadata will consist of all the metadata retrieved from repositories  $O = \bigcup_1^m O_p$ .

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

In Equation (2), Relative Recall,  $E$ , is calculated for a given repository  $j$ , takes as the denominator the sum of the LOs judged relevant in each search, comparing their relevance with that of the relevant LOs retrieved from the total number of repositories.

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

Metadata retrieval from LORs is variable in time, this must be taken into account when estimating the performance of the system. To assess such a dynamic aspect, an analysis has been carried out in the context of educational content retrieval from real LORs. This case study has underscored the great variability in the number of resources stored in different repositories, giving an index of variation of contents over time, the labelling of the resources, the updating of contents, internal retrieval procedures, etc. All these dynamic aspects must be considered in the formulation of the solution. Especially because the user's demand for LO should be met within a reasonably short period of time. This functional requirement makes it necessary to have some metric or formula that would allow us to evaluate the performance of the system in terms of the time it takes to deliver the results. No metric has been found in the bibliography to establish this type of control and a new metric is proposed for this purpose. The term proposed was called Temporary Gain of a repository  $J$ ,  $G_J(t)$ , and it is defined in the Equation (3) as average time value for  $n$  queries, evaluating the number of LOs as results,  $R_n$ , for the  $n$  query that a repository  $J$  returns while  $T_n$  is the query and retrieval time.

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

At this point, it is timely to compare the temporal gain metric at repository  $J$  with the average value obtained from the set of LORs. We call this new function the Average Temporal Gain,  $\overline{G}(t)$ , and calculate it considering the  $m$  studied repositories. Moreover, the traditional formula is used to calculate the mean in Equation (4).

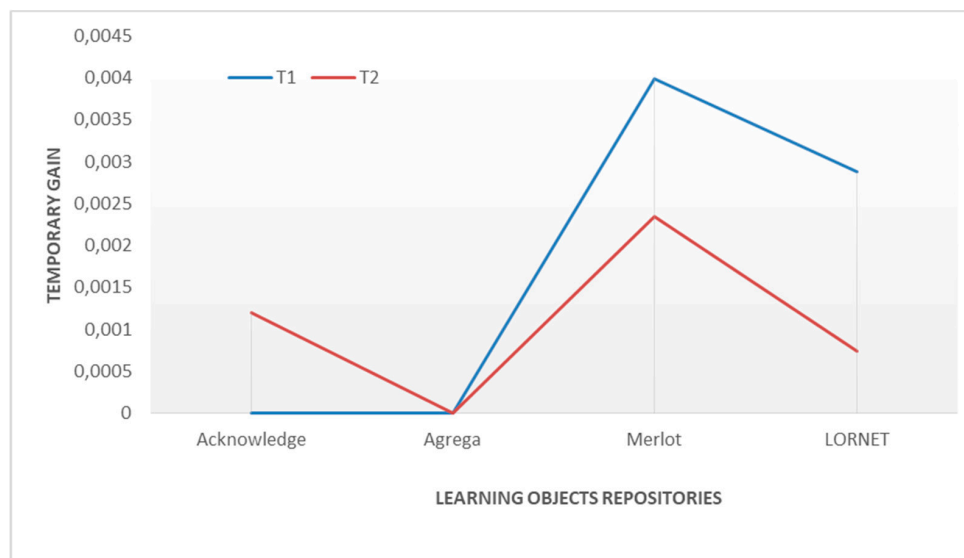
$$\overline{G}(t) = \frac{\sum_{j=1}^m G_J(t)}{m} \quad (4)$$

The goal of retrieval is to obtain as many relevant LOs as possible for the user performing the search. This goal is to achieve the highest average gain at any given time. Another important issue is the estimation of the suitability of the LOs that are delivered to the user. To estimate these issues a set of 53 searches in the LORNET and MERLOT repositories was performed. Experts in information retrieval suggest the creation of queries and agree that this process should be done with the help of experts in the field [67]. That is why a set was formed containing the search performed for the topics related to the Science and Technology of UNESCO codes. Also, according to openDOAR [68], the most widely used language for searching digital content is English (77% of queries). These topics are composed of long and short queries with the aim of identifying the different behaviors of the queried repositories' search engines.

This research has proposed two new metrics, temporary gain and the average gain for a set of LORs, are the metrics that AIREH calculates in each query. Thus, these two metrics can be considered a new metric mechanism in the intelligent retrieval of educational content that allows optimizing in the search and cataloguing of learning objects. In two moments of time,  $t_1$  and  $t_2$ , distant in several months and maintaining the same battery of queries are launched into a set of four repositories performing these 60 case study queries, in four repositories. Figure 5, shows values for temporary gain at testing

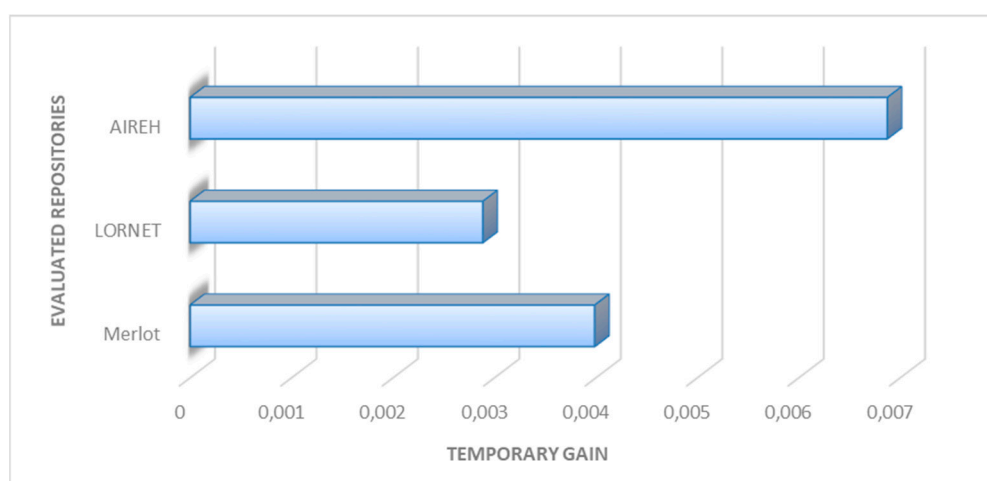


phase. This metric makes it possible to quantify the variability of the number of resources extracted as well as the time used.



**Figure 5.** Temporary gain values for the same 60 queries at two different moments in time.

The objective of the retrieval is clearly to obtain the greatest possible number of LOs relevant to the search criteria introduced by the user. This objective is to achieve the highest average gain at any given time. The average gain will increase if the accesses are eliminated to the repositories that had minor or null temporary gains over a period of time. That is why AIREH, the proposed architecture, is capable of extracting from different repositories, and by including the metrics associated with the temporary gain presented gets to improve the process significantly by increasing the number of LOs retrieved within a time period. Figure 6 shows how AIREH increases the temporary gain value of the system by 15% compared with MERLOT and LORNET repositories.



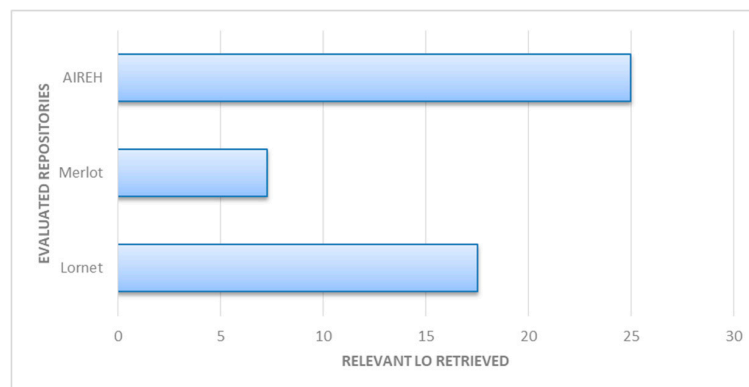
**Figure 6.** Comparison of temporary gain in AIREH, Learning Object Repository Network (LORNET), and Multimedia Educational Resource for Learning and Online Teaching (MERLOT), during the test phase of the proposed architecture.

This significant improvement is achieved thanks to the agent-based virtual organizations which manage effectively all the tasks involved in simultaneous query processing in several LORs. Once the QMA is instantiated with a request, repository verification tasks are performed by the agent called

RMA, which organizes LORs according to their performance which is calculated by the proposed gain metrics.

Another important issue estimating the suitability of the LOs retrieved for the user. The mechanism for organizing the LO extracted from a set of repositories should be considered an essential element of the LO filtering process, enabling the recommendation and customization of the results delivered to the user.

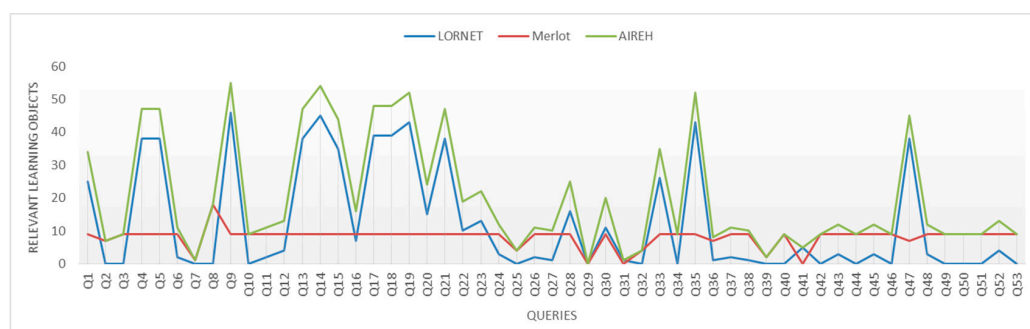
The system is robust against failure during query processing because CBR is incorporated into the organization. Prior to organizing the repositories according to performance, the RMA checks the active repositories for any request made by the QMA. Then, the LOR agent is in charge of extracting the LOs from a selected LOR, it manages the whole process and reports any Quality of Service (QoS) issues to the Statistics agent, such as query time, repository performance, and so on. Although some of the LOR agents could cancel the query to performance levels in the process, there would still be other agents in parallel operating the instantiated LOR, which is always monitored by the Repository Manager. In short, this mechanism provides the user with access to the LOs retrieved from different repositories, improving their user experience with distributed LORs, as demonstrated in Figure 7, where the average in number of retrieved LO is increased in the test phase, with the same 60 queries explained.



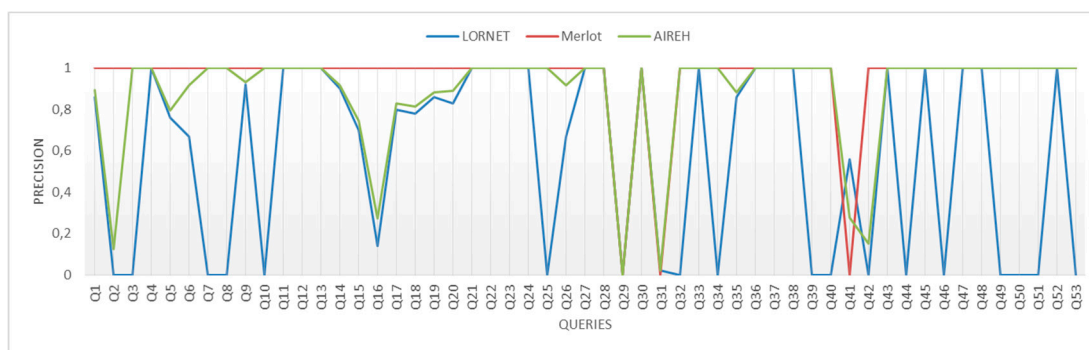
**Figure 7.** Average metadata documents retrieved during the testing phase.

The retrieval process results in multiple LO which respond to the user query. The retrieval of content by the proposed architecture system depends on the number of LOs retrieved by separate LOR. Therefore, in the event that there is no LO to respond to the user's query, the AIREH system cannot resolve the lack of content. However, the system ensures that any content related to the user's request will be retrieved.

Results of the relevance of the retrieved content, Figure 8 shows that AIREH is just under half the MERLOT repository, due to the low relevance of content in LORNET. However, relative recall, Figure 9, increases to its maximum value, 1, because the designed task inside the architecture to filter those LOs that have no information about the location of the resource in the LOM metadata.

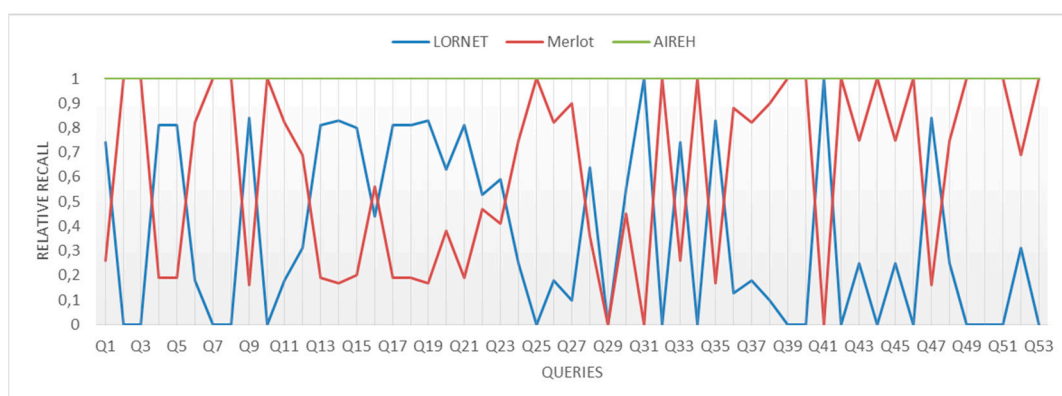


**Figure 8.** Comparison of relevant metadata documents from learning objects (LOs).

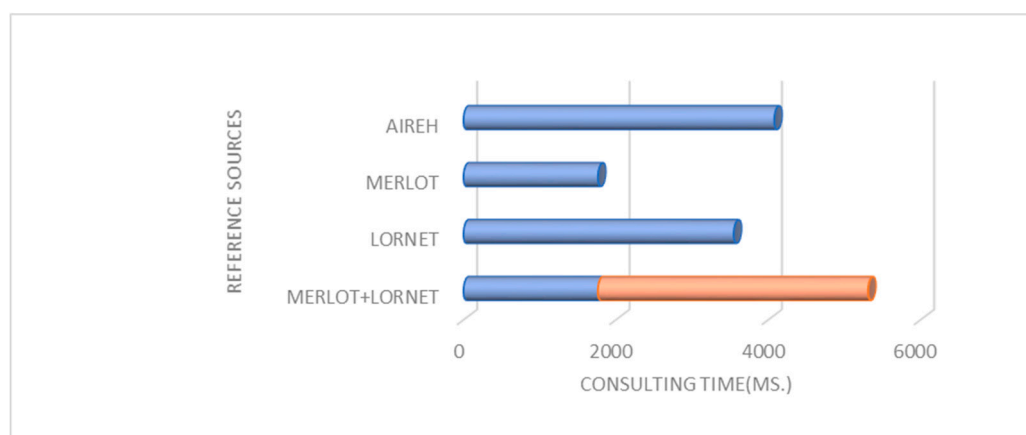


**Figure 9.** Comparative precision results for the architecture proposed.

As shown in Figures 9 and 10 the assessment in the number of outcomes have improved considerably by using a solution based on a federated search. This increase in performance has been achieved without increasing query time shown in Figure 11. This optimization in consulting time is due to the parallelization performed during the retrieval, which minimizes the time involved by 15% on average compared to the repository that required the greatest amount of time. It is important to underscore this fact since the response that AIREH provides to the user query is a set of LOs for the different LORs, having filtered out those that are not reusable and ordering the remaining repositories according to the personalized aspects previously described.



**Figure 10.** Relative recall comparative.



**Figure 11.** Average query time.

The federated search mechanism has been addressed as the resolution of the issue of content retrieval by solving three phases: (1) the selection of repositories, (2) the retrieval of content, and

finally (3) the merging of results. In addition to the centralized search for LOs in different repositories, the system performs a classification of objects retrieved through an initial filter that removes duplicate LOs without impeding access to the source of content. This step addresses the first phase of the proposed federated search by selecting the most appropriate LORs according to statistical parameters and performance during the search and retrieval session of content and is based on user research through key terms. An initial screening of recovered LO is immediately performed for the first organization that provides a homogeneous set of retrieved metadata documents. This pretreatment enhances the recovered achievement of the content for the second phase, the retrieval in federated search. A second stage of filtering incorporates aspects of quality of the retrieved objects according to two criteria: (1) the assessment of the quality of metadata retrieved and (2) the assessment of LO estimated by users through collaborative techniques. The assessment of the quality of the retrieved metadata incorporates parameters such as size (a larger improvement in the quality of information content is estimated), main features completeness, education, etc. The second incorporates the collective assessment on the use of LO through assessments from other users in the application, implemented by social aspects in the recommendation. The sum of these criteria on the ranking of retrieved objects in the system provides a hybrid recommendation that begins with a refined content-based recommendation and collaborative features. This stage ends with the third phase of the federated search and includes recommendations on merging the content, which improves the quality of the retrieved content for the user that generated the query, as explained in next section.

#### 4.2. Recommendation Strategy

The mechanism of organizing the learning objects extracted from the repository set should be included as one more aspect in the retrieval and filtering of LO relevant to the user. The field of educational content imposes new perspectives on the recommendation process in terms of customization aspects. This point delves into recommendation systems as a solution for establishing a classification in the retrieved educational content that shows the user the retrieved set in a personalized way.

The development of a single ordered list of LOs that incorporates the user's relevance criteria in this work is one of the tasks that the cataloguer agent implements with a CBR reasoning model in AIREH [69,70]. This agent is built as CBR-BDI agent, as a kind of agent that works with beliefs, desires, and intentions from the conceptual point of view and with cases from the implementation point of view. Each case based on CBR cases is divided into two parts. The first describes a set of attributes of the so-called objective element as the definition of the problem in CBR terminology. The second of the elements consists of the set of attributes that describes the user's interest in the given topic, the solution through a set of goals in CBR description. For this initial problem is defined based on the elements of the context, as shown in Figure 12.

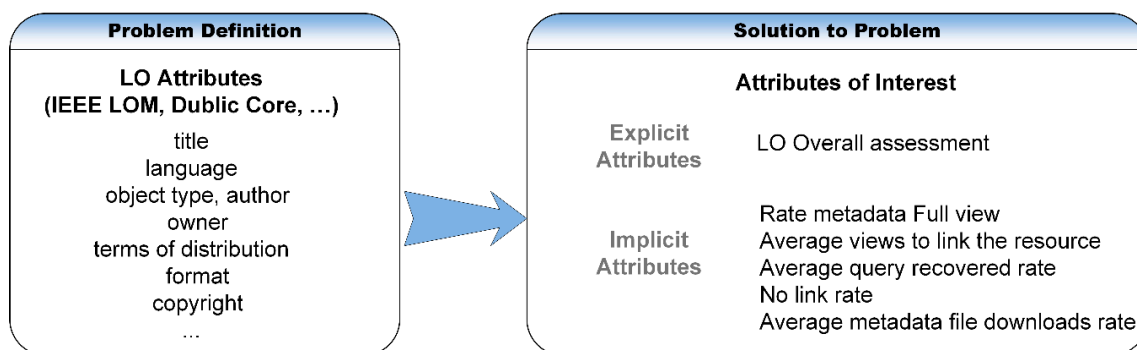


Figure 12. Example of representation of cases in the LO domain.

Given a set of educational contents  $P = \{p_1, p_2, \dots, p_i\}$ , each of them,  $p_i$ , is characterized by a series of objective attributes from the total set  $T$  of possible attributes. So, each unit of educational

content is represented by a vector  $p_i = \{p_{i_1}, p_{i_2}, \dots, p_{i_n}\}$ . In the specific case of LOs, the total set of attributes defined,  $T = \{title, language, keywords, format, \dots\}$ , is extracted fundamentally from the information described in its labelling in accordance with the educational standard (LOM, LOM-Es, DC, etc.). At the same time, other attributes are introduced in order to reflect aspects of the environment in which it has been recovered, such as the repository to which it belongs, date, etc.

Table 1 shows a description of the problem formalized by the USER, QUERY, and PREF attributes. The QUERY field is included to estimate another objective property but removing it from the set of objective parameters favoring a modular design of the ranking algorithm that allows flexibility in implementation. Through these attributes, the system can obtain the description of the problem to be solved by including CBR-BDI agents. These are agents that work with beliefs, desires, and intentions (BDI) from the conceptual point of view and with cases from the implementation point of view. A case consists of a description of the initial problem with attributes such as those described in Figure 12, the proposed solutions and the efficiency for the final solution.

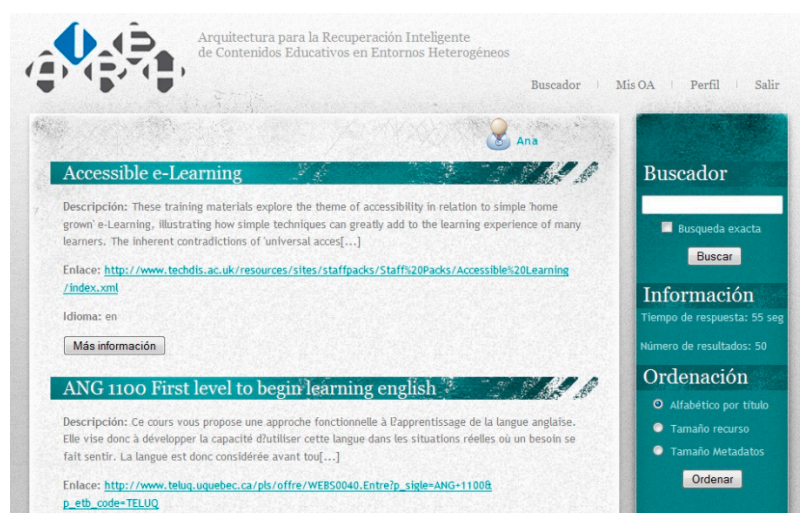
**Table 1.** Case Attributes

Case Field	Element Type
USER	User Profile
QUERY	Initial User Query
PREF	User Preferences
STIME	Time Stamp

Once the definition of the problem has been established based on attributes, the objective of the CBR is to generate the ranking of these learning objects according to the particular characteristics of the user that are reflected in the characteristics of the available LOs, e.g. the educational level of the LO or the language of the resource.

The CBR system retrieves past cases which resemble the present user query, and which have been resolved with a high degree of satisfaction. The cases stored in the case base also contain information about the users' interaction with the system.

The system tracks user interest through the submitted list to learn relevant information, as well as explicit and implicit information, in order to preserve the new case. On the user interface (UI) of presentation of the LOs to the user, Figure 13, the user can interact visualizing the complete tagged information of the object, downloading it, accessing the source of the educational resource, by voting it, etc.

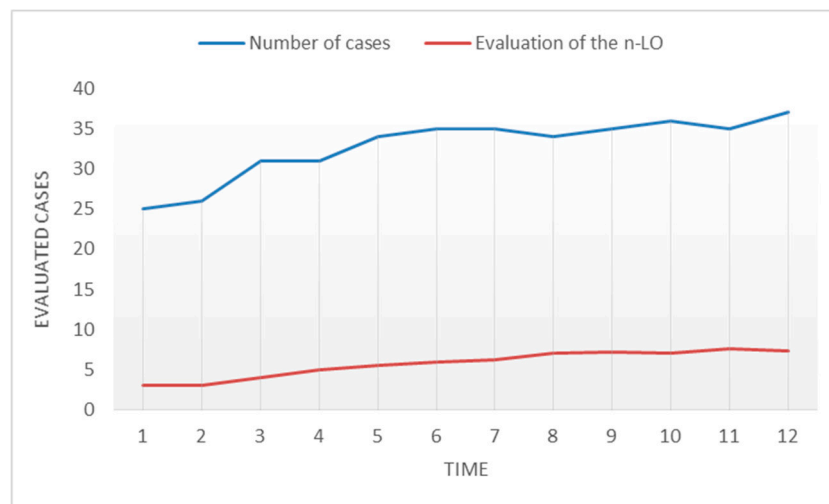


**Figure 13.** Sample results presentation in the User Interface (UI) AIREH tool.



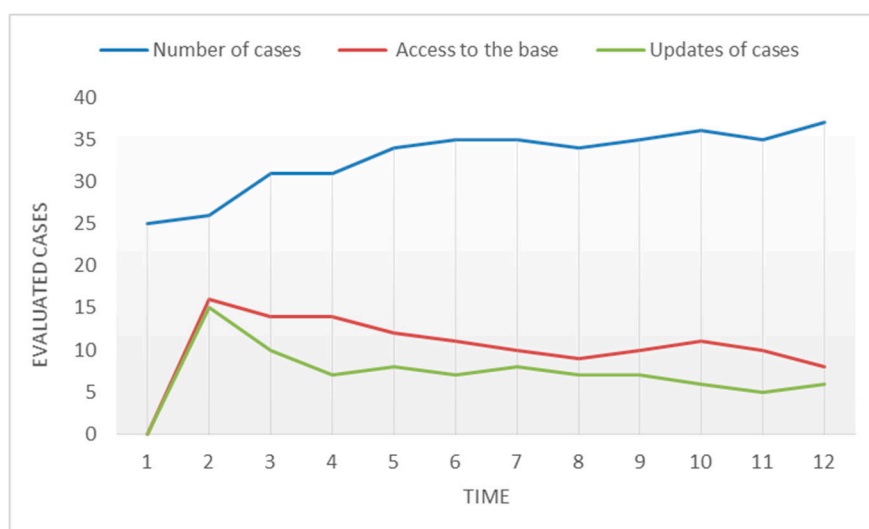
Recommendations are made on the basis of a group of recovered cases. To validate the application, we compared the evaluations of AIREH made by 10 different users over a period of 6 months.

As the number of successfully resolved cases increases in the case base, the system becomes more effective, has more LOs and has more knowledge about its users. Figure 14 shows how the users' positive ratings have increased in comparison to the first n-LO recommendations ( $n = 5$ ). The success of the system is evaluated through user interaction with the recommended LO, as well as the assessment they make of each interaction. This improvement is due to the system's ability to learn and adapt old solutions to new cases. Likewise, past cases enable the system to adapt more accurately to the user profile.



**Figure 14.** User evaluation of the case-based reasoning (CBR) recommendations.

In Figure 15, the number of updated cases decreases as the system acquires experience. The X-axis represents the number evaluations within a time period and the Y-axis quantifies the cases evaluated by the system over time, as the system gains more information (number of cases, access to the case base and case updates). This is logical, as the number of cases stored in the case base increases, it becomes easier for the CBR to find cases in the case base that are similar to the query of the user increases. With time, the CBR may even begin to validate the recovered LO criteria according to user needs and/or tastes.



**Figure 15.** Evolution of the CBR.

The number of cases in the base is limited to 50 cases. This limitation avoids the problem of the unlimited growth of the case base. In situations in which it is necessary to remove a case, it is done with the use of the STIME attribute, which is used as a timestamp. The case with the furthest time attribute compared to runtime is eliminated and replaced by a new one.

## 5. Discussion

AIREH has been designed to manage the content stored in LO repositories. Its flexible architecture makes it possible to develop different functionalities for content management. The architecture provides the user with repository access and management rights. To demonstrate that AIREH is an innovative and highly advanced system, it is necessary to make a comparative study of the proposed model and other state-of-the-art tools. In our study, AIREH and other tools were to solve the problem of recovering sparse educational content. The most important state-of-the-art search tools were: The Global Learning Objects Brokered Exchange (GLOBE) which is used to search through a web client in the repositories that are part of the alliance, and the ARIADNE search tool [71], which is linked to the GLOBE project [27]. Another relevant tool is PALOMAWeb [72], which has emerged from the referenced research work [73]. This tool is viewed as relevant because it is used by several universities in Quebec and in other parts of Canada. PALOMAWeb connects Canadian organizations with GLOBE, enabling the management of LO through its metadata.

Table 2 compares the functionalities of the chosen tools and of AIREH. From the table, it can be observed that AIREH has many more functionalities than the other two, despite this, all three tools are destined for the retrieval of educational content. Only the external aspects have been taken into account because there is no information available on the methodologies and technologies used by these tools.

**Table 2.** Comparative of the characteristics of the LO federated search engines.

Features/Tools	Paloma	Globe	AIREH
Access to distributed repositories	Yes	Yes (just what makes up the Alliance) Currently only ARIADNE (in beta)	Yes
Inclusion of different query languages	NO	NO	Yes
Personalization	NO	NO	Yes
Incorporation of social aspects	NO	NO	Yes
Metadata access by the user	Yes	NO	Yes
Tracing historical storage	NO	NO	Yes
Search with Language criteria	NO	Not Implemented	Yes
Different standards of tagged languages	No, only LOM	Yes	Yes
Web Client	Yes	Yes	Yes

The architecture proposed in this work incorporates important features such as personalization and metadata management, these functionalities improve user experience and the processing of the recovered LOs. The proposal also allows for flexible access to the repositories because the users can set up their preferences and easily incorporate query languages as long as the surveyed repositories meet the required specifications. Flexibility and adaptability to the environment, along with the ability to personalize results, are some of the characteristics that make AIREH an effective solution to the problem of content delivery [74]. Until now a keyword search in the metadata content has been used for the retrieval of the LOs contained in LORs. The use of multilabel classifiers guarantees the construction of powerful tools with a greater ability to search large volumes of data.

An important objective of Deep Web content retrieval technology is to improve the quality, scope and accuracy of Visible Web engines through the use of structured resource descriptions, i.e., through

semantically rich metadata. Moreover, due to restrictive copyrights, the actual digital content is freely available only in exceptional cases. A practical approach is, therefore, to retrieve the metadata from a series of repositories of geographically distributed and centralized LOs. After processing them and enabling users to perform searches from a single user interface, management becomes centralized and allows for the implementation of substantial improvements in the ranking process of recovered LO.

When building a federated search application, it is important to note a large number of available repositories, the need for advanced algorithms, compute-intensive information retrieval, as well as the growing number of users who need custom results for the queries they make to the repositories. All these factors can cause scalability and performance problems when designing this type of application.

Systems for the retrieval of educational technology are a relatively new concept. Thus, the presented proposal is just a starting point and there are many different possibilities for its further development. In this regard, the main line of future work will be the inclusion of new filtering mechanisms in the virtual organization and their subsequent validation through empirical evidence, as has been done in this research with the federated search system. Finally, in the near future, we are going to work on the design and further development of algorithms for improved search and retrieval of LOs, increasing the degree of relevance of the LOs to the search criteria of the user.

Issues such as the internationalization of the search for educational content should be taken into account, in order to solve the problems of multilingual retrieval and the merging of results based on criteria that include elements related to language, and other factors. The areas in which this work could be developed further include:

- **Testing and Validation.** Much more extensive testing is needed in order to assess the proposed architecture in terms of application and design, calculation of response time, quality of LOs, etc. The results could lead to the development of more refined models and robust systems.
- **Resolution of new practical problems.** To more thoroughly check the validity of the proposed model, it must be applied to new, practical problems. In this way, it would be possible to check if it can properly resolve different types of problems, or if the model is limited to the specific problems that have been studied in this work.
- **Integration of semantic aspects in retrieving and cataloguing content.** Even when educational resources are labeled according to a metadata standard, they are mainly descriptive and do not provide semantic information. The search results would benefit greatly from the inclusion of semantic search processes and from LO processing based on knowledge models such as domain ontologies. This would increase the functionality of the proposal because not only would quantitative aspects be taken into account but also the semantic features of the query. The search results would be filtered according to the semantic meaning of the content in the learning objects.

## 6. Conclusions

This research has presented an architecture for the retrieval of educational content from partner organizations, called AIREH (Architecture for Intelligent Retrieval of Educational content in Heterogeneous Environments). To optimize the retrieval of digital content, this work proposes a retrieval model whose core is an agent-based virtual organization architecture. Multi-agent systems are known for their ability to adapt quickly and effectively to changes in their environment. The model allows for the development of an open and flexible architecture that supports dynamic-search-related services for distributed digital content. AIREH is a multi-agent architecture that can search and integrate heterogeneous educational content through a retrieval model that uses federated search. The main novelty of the proposed architecture is that it is based on a model with dynamic and adaptive planning capabilities, making it possible to distribute tasks optimally among the different agents of the organizations, thus enables intelligent content retrieval and flexibility in a highly dynamic environment. of modeling an efficient system for managing open systems. The model and the technologies presented in this research demonstrate the advantages of using agent-based virtual organizations in the retrieval of labeled digital content. The proposed model is flexible, customizable, comprehensive and efficient.

The model of agent-based virtual organizations includes an organization with self-adaptive capabilities for highly dynamic environments at runtime. In conclusion, AIREH proposes a flexible architecture that allows for the development of a solution whose functionalities make it possible to manage the content stored in LO repositories. Nevertheless, this architecture can be developed further and improved through the implementation of other standards.

The objective of this research work has been to propose a tool for the retrieval of educational content, nevertheless, the proposed solution is applicable to any other field given the popularity of labelling mechanisms for large data repositories. The semantics of the context vary in every field of application, however this only changes the labelling of resources and the information that is used to generate a case base. An example of a direct application would be Biobanks or Biorepositories because they are digital collections of metadata that give access to information on biology resources. A representative example of such Biobanks is the one supported by the European Institute of Bioinformatics with web-based solutions for genomic data.

**Author Contributions:** Conceptualization, A.B.G. and F.d.I.P.; methodology, A.B.G., F.d.I.P., and S.R.; software, A.B.G. and F.d.I.P.; validation, A.B.G., F.d.I.P., and S.R.; formal analysis, A.B.G., F.d.I.P., and S.R.; investigation, A.B.G., F.d.I.P., and S.R.; writing—original draft preparation, A.B.G. and F.d.I.P.; writing—review and editing, A.B.G. and F.d.I.P.; visualization, A.B.G. and F.d.I.P.; supervision, S.R. and J.M.C.; funding acquisition, J.M.C. and S.R.

**Funding:** This work was supported by the Spanish government and European FEDER funds, project InEDGEMobility: Movilidad inteligente y sostenible soportada por Sistemas Multi-agentes y Edge Computing (RTI2018-095390-B-C32).

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Begosso, L.C.; Begosso, L.R.; Ribeiro, A.; dos Santos, R.M.; Begosso, R.H. The Use of Learning Objects for Teaching Computer Programming. In Proceedings of the 2015 IEEE Frontiers in Education Conference (FIE), El Paso, TX, USA, 21–24 October 2015; pp. 1–6.
2. Wiley, D.A. Connecting Learning Objects to Instructional Design Theory: A Definition, a Metaphor, and a taxonomy. *Instr. Use Learn. Objects* **2000**, *2830*, 1–35.
3. Silveira, I.F.; Omar, N.; Mustaro, P. *Architecture of Learning objects Repositories*; Learning Objects: Standards, Metadata, Repositories and LCMS; Informing Science Institute: Santa Rosa, CA, USA, 2007; Volume 1, pp. 131–156.
4. Lagoze, C. The Making of the Open Archives Initiative Protocol for Metadata Harvesting. *Library Hi Tech* **2003**, *21*, 118–128. [CrossRef]
5. IEEE. *IEEE Recommended Practice for Learning Technology—Open Archives Initiative Object Reuse and Exchange Abstract Model (OAI-ORE)—Mapping to the Conceptual Model for Resource Aggregation*; IEEE Std 1484.13.6-2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–32. [CrossRef]
6. Weibel, S.L.; Koch, T. The Dublin Core Metadata Initiative. *D Lib Mag* **2000**, *6*, 1082–9873. [CrossRef]
7. IEEE. *IEEE Standard for Learning Object Metadata*; IEEE Std 1484.12.1-2002; IEEE: Piscataway, NJ, USA, 2002; pp. 1–40. [CrossRef]
8. Open Education Consortium. 2019. Available online: <http://www.oeconsortium.org/> (accessed on 14 October 2019).
9. Spanish Ministry of Education. *Application Profile of the LOM Metadata Scheme for the Spanish Educational Community (LOM-ES)*. 2010. Available online: <http://educalab.es/documents/10180/40863/1LOM-ES.pdf/67a11fe2-edc0-487f-b6d5-6a87dc258668> (accessed on 14 October 2019).
10. Vicari, R.M.; Ribeiro, A.; da Silva, J.M.C.; Santos, E.R.; Primo, T.; Bez, M. *Brazilian Proposal for Agent-Based Learning Objects Metadata Standard—OBAA*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 300–311.
11. Cechinel, C. *Scientific Collaboration between Countries in Lacio from a Social Network Analysis Perspective*; Conferencias LACLO: Valdivia, Chile, 2013; Volume 4.
12. Dublin Core Metadata Initiative. *LRMI Metadata*. 2015. Available online: [https://www.dublincore.org/specifications/lrmi/lrmi\\_terms/](https://www.dublincore.org/specifications/lrmi/lrmi_terms/) (accessed on 14 October 2019).

13. Kolodner, J. *Case-Based Reasoning*; Morgan Kaufmann: Burlington, MA, USA, 2014.
14. Ramchurn, S.D.; Huynh, D.; Jennings, N.R. Trust in Multi-Agent Systems. *Knowl. Eng. Rev.* **2004**, *19*, 1–25. [CrossRef]
15. Knapp, M.; Risha, Z.; Gatewood, R.; Van Der Volgen, J.; Brown, R.; Kizilboga, R. *Learning to Love the LOR: Implementing an Internal Learning Object Repository at a Large National Organization*; Medical Reference Services Quarterly: Philadelphia, PA, USA, 2019; Volume 38, pp. 143–155. [CrossRef]
16. Franzoni, V.; Tasso, S.; Pallottelli, S.; Perri, D. *Sharing Linkable Learning Objects with the Use of Metadata and a Taxonomy Assistant for Categorization*; Springer: Cham, UK, 2019; pp. 336–348.
17. Sucunuta, M.; Riofrio, G.; Tovar, E. Information Retrieval Model for Open Educational Resources. In Proceedings of the 2019 IEEE Global Engineering Education Conference (EDUCON), Porto, Portugal, 8–11 April 2019; pp. 1255–1261.
18. Callan, J. Distributed Information Retrieval. In *Advances in Information Retrieval: Recent Research from the Center for Intelligent Information Retrieval*; Croft, W.B., Ed.; Springer US: Boston, MA, USA, 2000; pp. 127–150.
19. Verbert, K.; Hoebelheinrich, N.J.; Blinco, K.; Lewis, S.; Kraan, W. RAMLET: A conceptual model for resource aggregation for learning, education, and training. *D Lib Mag.* **2016**, *22*, 1–11. [CrossRef]
20. Verbert, K.; Duval, E. Towards a global architecture for learning objects: A comparative analysis of learning object content models. In *Proceedings of the 2004 EdMedia+ Innovate Learning*; LearnTechLib: Waynesville, NC, USA, 2004; pp. 202–208.
21. Frantiska, J., Jr. *Creating Reusable Learning Objects*; Springer International Publishing: New York, NY, USA, 2016. Available online: <https://www.springer.com/gp/book/9783319328881> (accessed on 14 October 2019).
22. Guevara, C.; Aguilar, J.; González-Eras, A. The model of adaptive learning objects for virtual environments instanced by the competences. *Adv. Sci. Technol. Eng. Syst. J.* **2017**, *2*, 345–355. [CrossRef]
23. Gil, A.; de la Prieta, F.; Rodríguez, S.; Martín, B. Educational content retrieval based on semantic Web services. In Proceedings of the 2011 7th International Conference on Next Generation Web Services Practices, Salamanca, Spain, 19–21 October 2011; pp. 135–140.
24. Melara Abarca, R.; Perez-Martinez, C.; Gelbukh, A.; López Morteo, G.; Martinez Reyes, M.; Pérez López, M. Wikification of Learning Objects Using Metadata as an Alternative Context for Disambiguation. *Computación y Sistemas* **2014**, *18*, 755–765.
25. Simon, B. *A Simple Query Interface Specification for Learning Repositories*; CEN Workshop Agreement: Chiba, Japan, 2005.
26. Simon, B.; Massart, D.; Van Assche, F.; Ternier, S.; Duval, E.; Brantner, S.; Miklós, Z. A Simple Query Interface for Interoperable Learning Repositories. In Proceedings of the WWW\* 05 Workshop on Interoperability of Web-Based Educational Systems, Chiba, Japan, 10–14 May 2015; pp. 11–18.
27. De Santiago, R.; Raabe, A. Architecture for Learning Objects Sharing among Learning Institutions—LOP2P. *IEEE Trans. Learn. Technol.* **2010**, *3*, 91–95. [CrossRef]
28. Deora, B.S.; Sarangdevot, S.S. E-Learning Standards and their Necessity. *Int. J. Adv. Res. Comput. Sci.* **2011**, *2*, 500–504.
29. IEEE Standards Association. *ECMAScript Application Programming Interface for Content to Runtime Services Communication*; IEEE Standards Association: Piscataway, NJ, USA, 2019. Available online: [https://standards.ieee.org/project/1484\\_11\\_2.html](https://standards.ieee.org/project/1484_11_2.html) (accessed on 14 October 2019).
30. Bates, T. Questioning the value of re-usable learning objects in education: The need for a business case. In Proceedings of the II Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables, Barcelona, Spain, 19–21 October 2005.
31. Casquero, O.; Portillo, J.; Benito, M.; Romo, J. BILDU: Compile, unify, wrap, and share digital learning resources. *Interdiscip. J. E Learn. Learn. Objects* **2008**, *4*, 97–111. [CrossRef]
32. Qu, C.; Nejd, W. Interacting the Edutella/JXTA peer-to-peer network with web services. In Proceedings of the 2004 International Symposium on Applications and the Internet, Tokyo, Japan, 26–30 January 2004; pp. 67–73.
33. Kärger, P.; Ullrich, C.; Melis, E. *Integrating Learning Object Repositories Using a Mediator Architecture*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 185–197.
34. Yen, N.Y.; Shih, T.K.; Jin, Q. A new paradigm of ranking & searching in learning object repository. In Proceedings of the Second ACM International Workshop on Multimedia Technologies for Distance Learning, Firenze, Italy, 25–29 October 2010; pp. 1–6.



35. Lee, M.C.; Tsai, K.H.; Wang, T.I. A practical ontology query expansion algorithm for semantic-aware learning objects retrieval. *Comput. Educ.* **2008**, *50*, 1240–1257. [CrossRef]
36. Alejandra Segura, N.; Salvador, S.; García-Barriocanal, E.; Prieto, M. An empirical analysis of ontology-based query expansion for learning resource searches using MERLOT and the Gene ontology. *Knowl. Based Syst.* **2011**, *24*, 119–133. [CrossRef]
37. Shih, T.K.; Chang, C.C.; Lin, H.W. *Reusability on Learning Object Repository*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 203–214.
38. Barak, M.; Ziv, S. Wandering: A Web-based platform for the creation of location-based interactive learning objects. *Comput. Educ.* **2013**, *62*, 159–170. [CrossRef]
39. Fernández Diego, M.; Gordo Monzó, M.L.; Boza García, A.; Cuenca, L.; Ruiz Font, L.; Alemany Díaz, M.D.M.; Alarcón Valero, F. Metadata, repository and methodology in learning objects. In Proceedings of the EDULEARN15: 7th International Conference on Education and New Learning Technologies, Barcelona, Spain, 6–8 July 2015; pp. 4755–4761.
40. González Ruiz, L.M.; Hermida Carbonell, J.M.; Montoyo, A. Towards a New Proposal to Evaluate the Learning Objects Quality in Learning Strategies for Education (QEEs). June 2012. Available online: <http://hdl.handle.net/10045/35627> (accessed on 14 October 2019).
41. Fuentes, M.D.L.M.; Arteaga, J.M.; Rodríguez, F.Á.; Vanderdonkt, J.; Orey, M. MIRROS: Intermediary model to recovery learning objects. *Computación y Sistemas* **2010**, *13*, 373–384.
42. Carrion, J.S.; Gordo, E.G.; Sanchez-Alonso, S. Semantic learning object repositories. *Int. J. Contin. Eng. Educ. Life Long Learn.* **2007**, *17*, 432. [CrossRef]
43. Manouselis, N.; Vuorikari, R.; Van Assche, F. Collaborative recommendation of e-learning resources: An experimental investigation. *J. Comput. Assist. Learn.* **2010**, *26*, 227–242. [CrossRef]
44. Martins, E. Learning objects and training complex machines. *Work* **2012**, *41*, 184–187.
45. Bobadilla, J.; Serradilla, F.; Hernando, A. Collaborative filtering adapted to recommender systems of e-learning. *Knowl. Based Syst.* **2009**, *22*, 261–265. [CrossRef]
46. Khribi, M.K.; Jemni, M.; Nasraoui, O. Automatic personalization in e-learning based on recommendation systems: An overview. In *Intelligent and Adaptive Learning Systems: Technology Enhanced Support for Learners and Teachers*; IGI Global: Hershey, PA, USA, 2012; pp. 19–33.
47. Recker, M.M.; Walker, A.; Lawless, K. What do you recommend? Implementation and analyses of collaborative information filtering of web resources for education. *Instr. Sci.* **2003**, *31*, 299–316. [CrossRef]
48. Lemire, D. Collaborative filtering and inference rules for context-aware learning object recommendation. *Interact. Technol. Smart Educ.* **2005**, *2*, 179–188. [CrossRef]
49. McCalla, G. The ecological approach to the design of e-learning environments: Purpose-based capture and use of information about learners. *J. Interact. Media Educ.* **2004**, Art. 3. [CrossRef]
50. Dong, A.; Wang, B. Domain-Based Recommendation and Retrieval of Relevant Materials in E-learning. In Proceedings of the 2008 IEEE International Workshop on Semantic Computing and Applications, Santa Clara, CA, USA, 10–11 July 2008; pp. 103–108.
51. Ghauth, K.I.; Abdullah, N.A. Learning materials recommendation using good learners' ratings and content-based filtering. *Educ. Technol. Res. Dev.* **2010**, *58*, 711–727. [CrossRef]
52. Yang, Y.J.; Wu, C. An attribute-based ant colony system for adaptive learning object recommendation. *Expert Syst. Appl.* **2009**, *36*, 3034–3047. [CrossRef]
53. Wan, S.; Niu, Z. An e-learning recommendation approach based on the self-organization of learning resource. *Knowl. Based Syst.* **2018**, *160*, 71–87. [CrossRef]
54. Imran, H.; Belghis-Zadeh, M.; Chang, T.W.; Graf, S. PLORS: A personalized learning object recommender system. *Vietnam J. Comput. Sci.* **2016**, *3*, 3–13. [CrossRef]
55. Kerkiri, T.; Manitsaris, A.; Mavridou, A. Reputation Metadata for Recommending Personalized e-Learning Resources. In Proceedings of the Second International Workshop on Semantic Media Adaptation and Personalization (SMAP 2007), London, UK, 17–18 December 2007; pp. 110–115.
56. Wolpers, M.; Najjar, J.; Verbert, K.; Duval, E. Tracking actual usage: The attention metadata approach. *Educ. Technol. Soc.* **2007**, *10*, 106–121.
57. Klašnja-Milićević, A.; Ivanović, M.; Nanopoulos, A. Recommender systems in e-learning environments: A survey of the state-of-the-art and possible extensions. *Artif. Intell. Rev.* **2015**, *44*, 571–604. [CrossRef]
58. Wooldridge, M. Intelligent agents. *Multiagent Syst.* **1999**, *35*, 51.

59. Zambonelli, F.; Jennings, N.R.; Wooldridge, M. Developing multiagent systems: The Gaia methodology. *ACM Trans. Softw. Eng. Methodol.* **2003**, *12*, 317–370. [[CrossRef](#)]
60. Dignum, M.V. *A Model for Organizational Interaction: Based on Agents, Founded in Logic*; SIKS: Amsterdam, The Netherlands, 2004.
61. Dignum, V. The role of organization in agent systems. In *Handbook of Research on Multi-Agent Systems: Semantics and Dynamics of Organizational Models*; IGI Global: Hershey, PA, USA, 2009; pp. 1–16.
62. Argente, E.; Botti, V.; Julian, V. *Organizational-Oriented Methodological Guidelines for Designing Virtual Organizations*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 154–162.
63. Giret, A.; Julián, V.; Rebollo, M.; Argente, E.; Carrascosa, C.; Botti, V. *An Open Architecture for Service-Oriented Virtual Organizations*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 118–132.
64. Argente, E.; Botti, V.; Carrascosa, C.; Giret, A.; Julian, V.; Rebollo, M. An abstract architecture for virtual organizations: The THOMAS approach. *Knowl. Inf. Syst.* **2011**, *29*, 379–403. [[CrossRef](#)]
65. Schell, G.P.; Burns, M. Merlot: A repository of e-learning objects for higher education. *E Serv.* **2002**, *1*, 53–64.
66. Saliah-Hassane, H.; Kourri, A.; la Teja, I.D. Building a Repository for Online Laboratory Learning Scenarios. In *Proceedings of the 36th Annual Conference Frontiers in Education*, San Jose, CA, USA, 27–31 October 2006; pp. 19–22.
67. Hawking, D.; Craswell, N.; Bailey, P.; Griffiths, K. Measuring Search Engine Quality. *Inf. Retr.* **2001**, *4*, 33–59. [[CrossRef](#)]
68. Norris, M.; Oppenheim, C.; Rowland, F. Finding open access articles using Google, Google Scholar, OAster and OpenDOAR. *Online Inf. Rev.* **2008**, *32*, 709–715. [[CrossRef](#)]
69. Gil, A.; Rodríguez, S.; de la Prieta, F.; Martín, B.; Moreno, M. *Intelligent Recovery Architecture for Personalized Educational Content*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 85–93.
70. Gil, A.; Rodríguez, S.; De la Prieta, F.; De Paz, J.F.; Martín, B. *CBR Proposal for Personalizing Educational Content*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 115–123.
71. Duval, E.; Forte, E.; Cardinaels, K.; Verhoeven, B.; Van Durm, R.; Hendrikx, K.; Haenni, F. The Ariadne knowledge pool system. *Commun. ACM* **2001**, *44*, 72–78. [[CrossRef](#)]
72. Lundgren-Cayrol, K.; Ruelland, D.; Habel, G.; Magnan, F. Modeling for Tools and Environments Specification. In *Visual Knowledge Modeling for Semantic Web Technologies: Models and Ontologies*; IGI Global: Hershey, PA, USA, 2010; pp. 414–438.
73. Paquette, G.; Léonard, M.; Lundgren-Cayrol, K.; Mihaila, S.; Gareau, D. Learning design based on graphical knowledge-modelling. *J. Educ. Technol. Soc.* **2006**, *9*, 97–112.
74. Gil, A.B.; Rodríguez, S.; De la Prieta, F.; Corchado, J.M. Learning object retrieval in heterogeneous environments. *Int. J. Web Eng. Technol.* **2013**, *8*, 197–213. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).