Search and selection of learning objects in repositories: A review

Germán A. Osorio-Zuluaga Néstor Darío Duque-Méndez
Universidad Nacional de Colombia - Sede Manizales, Departamento de Informática y Computación

***Abstract*—This article aims to present a review of the academic literature on the search for learning objects in repositories of the last ten years. To carry it out, first, it defined a methodology for the search and selection of the papers to be included in this study. Then, the bibliographic databases Web of Science, Scopus, and Google Scholar for the implementation of the methodology were used. Next, based on the selected documents, three strategies for the search of learning objects were characterized, such as classic by keywords, full-text search, and hybrids. In this sense, hybrid solutions that use multiple integrated strategies have been gaining importance. Finally, it was verified that the methodology used to determine the papers of the review was adequate and allowed an approach to the state of the art in this topic.**

I. Introduction

The vertiginous advance of the Web has promoted the development of distant education [1], [2]. In this sense, a large number of organizations offer and share educational resources in different formats to learners locally and globally. These resources are called learning objects (LOs) and are usually stored in repositories. Additionally, LOs can be tagged through metadata to facilitate their search and recovery [2]. Accordingly, the metadata provides structured descriptions to the LOs [1]. In the same way, several repositories can be organized as federations, which offer a unified approach to representing these repositories through a hierarchical system that centralizes educational resources in a single portal, increasing their visibility and facilitating the uniform administration of applications to discover and access the contents of the LOs available in a group of repositories [3].

In addition, in order to reuse and share LOs between repositories, standardized protocols have been defined to catalog them [1], [4]. Among the most important ones are the IEEE-LOM, Dublin Core [4], Can Core and OBAA. Each one specifies the syntax and semantics of the attributes needed to describe a LO [3]. For example, the IEEE-LOM standard consists of nine categories and about seventy descriptive elements [5].

On the other hand, within the architecture of the repositories, a component that is responsible for the search and recovery of LOs is necessary. In this sense, to obtain information on the Internet there are several search engines of general purpose, in which keywords are entered and they return, as a result, web pages containing the keywords entered. The paradigms used by these search engines are not the most suitable for the recovery of LOs [6] because they have two main drawbacks: the first is that this keyword approach requires the advance indexing of the content of the learning objects, not only the textual ones but also the multimedia objects; the second is that learning objects are not semantically related to the subject of learning [1]. These situations produce many unrelated results for the learner [7].

Furthermore, the search and retrieval of learning objects in repositories are mainly based on metadata [3]. In other web contexts, full-text searches are used. In the first one, the search is done from the metadata, in which the users pre-select and search the individual topics of a source of information, such as author, title, and subject; the search engine finds matches between the terms of the query with the terms of structured metadata and generates the results. In the second, the full-text search, the system finds matches between the terms of the query with the terms in the individual documents of a repository and classifies the results algorithmically [8].

In the same vein, the use of hybrid methods has been gaining importance, in which several methods are integrated in order to achieve better search results, [1], [9], [10], [11].

Finally, this article proposes a review of the literature about the search and selection of learning objects in repositories in the last ten years. For this purpose, the methodology is presented in the following section. In Section III the literature on the subject is reviewed and characterized. Finally, in Section IV, the discussion is presented; and in Section V conclusions of this work are presented.

II. Papers selection methodology

According to the methodologies proposed by [12] and [13], the following steps were followed in order to identify the academic literature regarding the search and selection of learning objects in repositories:

- Identifying main concepts.
- Listing related terms.
- Determination of the search equation.
- Establishing inclusion and exclusion criteria.
- Selection of papers.

It is important to note that this process is not strictly sequential, it is also iterative [14]. As such, in the development of the process, it was necessary to go back to previous stages in several times.

To start, the bibliographic databases used for this review were Web of Science (WoS) and Scopus which index peer-reviewed publications.

The process conducted is explained below.

A. Identifying main concepts

The summary phrase [15] for this review was "Search and selection of learning objects in repositories". For this reason, the concepts search engine, information retrieval, and learning object were initially taken, which encompass the tasks, subjects or instruments that refer to it. In this sense, the following search equation was applied to the “core collection” of WoS and no results were obtained. The same was done with Scopus and only 5 results were obtained:

- *TITLE (("search engine" OR "information retrieval") AND "learning object")*

Then, the next stage of the methodology was continued.

B. Listing related terms

As proposed in [12], synonyms of the main concepts were searched and used. For example, for learning object, terms such as knowledge object, educative resources, educative materials, and teaching materials were used. Also, it was included the term that encompasses the broadest category of electronic educational technology: e-learning.

In addition, it was necessary to incorporate terms that would mean information retrieval or search engine. In this sense, through an exploratory task in papers, terms such as extract, select, discover, identify, rank, classify, and deliver were found.

C. Determination of the search equation

Based on the terms selected in the Subsections II-A and II-B, and the use of logical operators to establish relationships that occur between them, the search equation was established [13]. Additionally, the wildcard symbol * was used so that variations in the words are included in the search results [12]. For example, educat* retrieves papers that contain terms such as educational and educative in the title. Therefore, the following search equation was proposed:

- *TITLE ((search* OR retriev* OR extract* OR select* OR discov* OR identif* OR rank* OR classif* OR deliver*) AND ("learning object*" OR "knowledge object*" OR elearning OR e-learning OR "educat* resourc*" OR "educat* mater*" OR "teach* mater*"))*

With this last equation of search 182 papers from WoS and 789 from Scopus were obtained; some of them common to the two bibliographic databases.

D. Establishing inclusion and exclusion criteria

First, as mentioned above, the bibliographic databases used for this review were Web of Science (WoS) and Scopus and all types of documents were included: articles, reviews, proceeding papers, letters, book chapters and editorial materials.

Next, with respect to the lapse of time and based on the defined search equation, the papers of the last five years were initially taken. With this refinement, the number of papers obtained in both bibliographic databases was reduced to less than a half. In addition, in a first exploration and despite the fact that the search equation was very accurate, not all the papers were relevant to the investigation. Therefore, it was decided to perform the bibliographic review for the last ten years.

Finally, according to the exclusion criteria, 602 Scopus papers and 151 WoS papers were obtained.

E. Selection of papers

For the review, 20 relevant documents were considered. In this respect, selection criteria were defined in the following order:

1) For all retrieved and relevant papers, eighty percent of the most cited papers per year (number of citations divided by the number of years of publication), ordered from highest to lowest. Then, ordered from highest to lowest, the highest H index of the authors of each paper. Finally, ordered from highest to lowest, the year of publication.
2) Sixteen percent of the papers of the last two years (most recent papers), not including conference proceedings, in ascending order according to the quartile of the journal (from Q1 to Q4). Then, sorted by the impact factor / CiteScore of the journal. Finally, from highest to lowest, the highest H index of the authors of each paper.
3) Four percent of conference proceedings of the last two years, ordered from highest to lowest, the highest H index of the authors of each paper.
4) Other pertinent articles that fulfilled one of the previous characteristics, although they did not comply with the search equation, which were found as bibliographic references in previously reviewed articles.

With regard to criterion number 1, in the paper by Gaona-García et al. [16], the measure of the number of citations per paper is also used. This work differs from theirs, in that the average number of citations per year is used instead of the total number of citations per paper.

Additionally, since the criteria to determine the number of citations per paper used by WoS and Scopus differ, it was necessary to use a bibliographic tool in which all the papers obtained in these two bibliographic databases were present and that also had a unique criterion to define the number of citations per paper. In this sense, Google Scholar complied with these two aspects and was the only one considered. Likewise, with this same tool, the H index of the authors of the papers was obtained. When this information was not found in Google Scholar, the H index was valued at zero.

Furthermore, with respect to Google Scholar, three problems were detected in this bibliographical tool during the study period, which generated a bias in the study. The first one has to do with the fact that all kinds of documents are included in the number of citations of each paper, including self-citations and documents not reviewed by peers. Second, the process by which papers are assigned to a specific author profile is performed by an unsupervised algorithm, starting with the surname and initial letters of the author’s name; it is the case that some authors, even of a different gender, with surnames and initial letters similar or equal, are erroneously assigned. This situation is palliated a bit when the same author updates the information contained in his profile of researcher, but until the date of the study, this was not a very generalized task among the academic community, as can be deduced from the inconsistencies found. The last one is that the researcher profile in Google Scholar is associated with the researcher's email account and if the researcher has several accounts, for example, the personal and institutional accounts, the tool assigns either one to the other, or both; and in many cases, the information differs e.g. the H index, despite being the same researcher. However, despite these drawbacks, it was decided to use it in the review since this bias affects the entire population of papers considered in the review.

III. Review: search and selection of LOs

Based on the review of the papers, three strategies were characterized for the search and selection of OAs in repositories.

- Classic search with keywords in the metadata.
- Search based on content.
- Search based on hybrid methods.

For each category, papers are presented in chronological order from the oldest to the most recent. Table I lists the papers reviewed in this work.

A. Classic search with keywords in the metadata

Ochoa and Duval in [17] aim to improve the present status of the search for learning objects. To do this, they review the literature and make a theoretical proposal to search for learning objects based on a classification by relevance. In the review they describe the different relevance metrics and identify three current approaches showing, mainly, their disadvantages: first, the ranking based on human review presents the difficulties of being a very expensive manual process and static in time, which does not adapt to the different user requirements. The second, the ranking based on text similarity, has the advantage that it is easily calculated but due to the low amount of text in the metadata of the learning objects they lead to low performance in the results. The third, the ranking based on user profile, in which the topics of the user's profile are compared with the classification of the learning object; its outstanding disadvantage is that it does not integrate well with the user workflow. Hence, the authors suggest three characteristics that a new generation of search mechanisms for learning objects should meet, such as: taking into account the information generated by the human user; calculating its value automatically; and does not requiring the conscious intervention of the user.

In addition to the review of the literature, Ochoa and Duval [17] propose LearnRank, a metric based on different relevance metrics. For its implementation, the RankNet algorithm was selected, which uses a neural network to learn the optimal ranking based on the original metrics. Furthermore, in order to evaluate the proposal, an experiment was carried out in which ten users participated and ten lessons were created on topics in the computer area. The tests performed showed a significant increase in the performance of the ranking compared with the reference rank.

In another article by [2], the authors describe ProLearn Query Language (PLQL): a query language for learning objects in repositories. In its definition, an exact search has been combined through the metadata of the learning objects using XML-based query engines and an approximate search through keyword-based searches using information retrieval engines such as Lucene. In this sense, the exact search is executed first and then pruned with the approximate search. In addition, PLQL in combination with a simple query interface (SQI) and an application profile of LOM learning object metadata from a European repositories federation, called Learning Resource Exchange (LRE), performs federated searches in all those repositories.

Biletskiy et al., in [6], propose a personalized search approach for educational Web resources making use of the student profile and the descriptions of the learning objects, based on the IEEE LOM and IMS LIP standards. To carry it out, they propose to develop ontologies of the student and the LO. In addition, the proposed solution allows the student's feedback on the suitability of the educational resources recovered in the personalized search. Finally, for the validation of the proposed approach, a prototype of the learning object search and retrieval system was implemented which provides the student with the twenty best-ranked learning objects that, according to the authors, demonstrate the validity of the proposal.

Yen et al. [18] propose a process flow to help users retrieve learning objects in federated repositories (under the SCORM and CORDRA specifications) that the authors call "Guided Search" based on ranking metrics and a more efficient search algorithm. To carry it out, three steps were proposed. First, the use of weighting metrics of learning objects based on time series and a social assessment mechanism inspired by Web 2.0 and social networks. Then, the ranking metrics were used to retrieve the learning objects in a specific order according to the users' query. Finally, the tool called «Search Guider» helps users to recover relevant learning objects according to their needs.

Based on the process flow proposed by [18], on a developed system called «MINE Registry» that stores and shares around 21,000 learning objects, an experiment is conducted in three stages. In the first stage, the overall performance of the MINE Registry is evaluated, in which high values of the precision measure were not obtained, but on the recall measures were. In the second one, the performance is compared with three other known methods (grid, ontological, and inference network approaches) and close results were obtained. Finally, the guided search is evaluated in which shorter times for users were achieved.

In the work of Hsu [19], a Multi-layered Semantic LOM Framework (MSLF) is proposed to integrate Semantic Web technologies in LOM, in order to overcome the weakness of LOM with respect to its lack of semantic metadata. This framework was used to implement an intelligent prototype LOM, to find relevant learning objects in a repository, called LOFinder; which consists of four main components, namely: the LOM base, the knowledge base, the search agent, and the inference agent. In addition, LOFinder integrates three learning

Table I
LIST OF PAPERS ON SEARCH AND SELECTION OF LEARNING OBJECTS IN REPOSITORIES

Paper	Category	Cit.	Cit./year	Experim./Evaluation	Methods, Techniques, Tools
Ochoa 2008 [17]	Metadata	103	10.3	10 users, 10 lessons	Learning to Rank (Ranknet)
Ternier 2008 [2]	Metadata	62	6.2	European repositories federation	XML-based query engine, keyword-based search, Apache Lucene
Biletskiy 2009 [6]	Metadata	52	5.8	50 LOM documents, an artificial learner profile and another one real	Ontologies
Yen 2010 [18]	Metadata	64	8.0	21,000 LOs, 3 topics, 40 users	PageRank, time series, social evaluation mechanism, cosine similarity
Hsu 2012 [19]	Metadata	30	5.0	Java-based prototype, 125 LOs, 217 relations between LOs, 9 rules	Ontology-based reasoning, rule-based inference
Yigit 2014 [4]	Metadata	16	4.0	The SDUNESA software, 12 instructors, 7 criteria	Multi-criteria decision making (MCDM) method, AJAX, XML and SOA Web Services, Analytical Hierarchy Process method (AHP)
Sabitha 2016 [20]	Metadata	10	5.0	4 learning styles, 1,026 LOs, 35 metadata attributes	Fuzzy C-Means clustering
Barbagallo 2017 [21]	Metadata	1	1.0	13 health professionals, validated in the osteoporosis domain	Ontologies, NLP tool ALCHEMY, MOODLE LMS, MySQL
Koutsomitropoulos 2017 [22]	Metadata	0	0.0	Integration with the eClass LMS	Ontologies, Automatic Query Expansion (AQE)
Hassan 2011 [7]	Full-Text	14	2.0	Dataset in computer science domain, 14 topics	Classifiers: Naive Bayes, Support Vector Machines, and support vector regression
Zeng 2017 [23]	Full-Text	3	3.0	7,510 diabetic questions, 144 diabetic patient educational materials	Topic modeling using Latent Dirichlet Allocation; semantic group-based model; TF-IDF; Vector Space Model; Cytoscape; Apache Tika
Rahman 2017 [24]	Full-Text	0	0.0	36 first-year undergraduate students	Fuzzy classification, C4.5 algorithm, Google search
Gasparetti 2018 [25]	Full-Text	0	0.0	Dataset of three collections of different domains	C4.5 decision tree, multilayer perceptron neural network, Naive Bayes classifier, Tagme annotation tool, Wikipedia like weak ontology, Feature Information Gain to feature selection
Cernea 2008 [26]	Hybrid	29	2.9		Latent Semantic Indexing, collaborative tagging, folksonomy, Pearson Correlation Coefficients
Lee 2008 [1]	Hybrid	95	9.5	2 datasets, 6 topics, Java Learning Object Ontology	Ontologies, Automatic Query Expansion, ambiguity elimination function
Zhuhadar 2008 [27]	Hybrid	41	4.1	A training set, 10 concepts, 28 subconcepts and 2,812 documents	Ontology, clustering, cosine similarity, TF-IDF, semantic web
Khribi 2009 [9]	Hybrid	337	37.4	Data file obtained from 11,542 sessions	Cosine similarity, Apache Nutch search engine, association rules, collaborative filtering; content based filtering
Shih and Seng 2009 [28]	Hybrid	32	3.6	Web-based prototype, 20 elementary school students	Apache Lucene, expert system shell DRAMA, Automatic Query Expansion, ontologies, inverted file indexing, Ontology Building Algorithm, cosine similarity
Zhuhadar 2010 [10]	Hybrid	12	1.5	The HyperManyMedia search engine	Hand-made ontology, VSM, cosine similarity, K-way clustering, Bisecting K-Means, recommendation as Rule-based
Atkinson 2014 [11]	Hybrid	12	3.0	Web-based prototype, 3,300 LOs, 500 web documents, 16 teachers	Machine learning methods: Named-Entity Recognition, Naive Bayes classifier; and Natural Language Processing (NLP). WordNet (AQE), LSA, Formal Concept Analysis, VSM, cosine similarity

object recovery approaches such as LOM metadata, ontology-based reasoning, and rule-based inference (LOM metadata, ontology-based reasoning, and rule-based inference).

In the work of Yigit et al. [4], SDUNESA repository software is presented to store, share and select learning objects. It was implemented with Web 2.0 technologies such as XML, AJAX, and SOA Web Services. Furthermore, the analytical hierarchy process (AHP) was used for the selection of the learning objects, which is part of the multi-criteria decision-making (MCDM) methods. In this sense, for the definition of the AHP criteria for the selection of learning objects, twelve instructors who work in the computer engineering department of a university were interviewed. In addition, seven qualitative and quantitative criteria were defined (with their respective sub-criteria) such as the type of learning resource, format, difficulty, level of interactivity, semantic density, the expected role of the end user, and structure. In this study, no evaluation and comparison tests are done with other methods.

Sabitha et al. [20] propose a data mining approach by clustering, according to the attributes of the metadata and the learning styles of the students, for the particularized delivery of learning and knowledge objects to learners. To achieve this, the learning objects are mapped into four dimensions of learning styles (participation, processing, presentation, and organization) and then grouped by fuzzy c-mean clustering. The authors conclude that this approach achieves a more personalized and authentic learning experience.

In the paper of [21], the ELSE system (E-Learning for the Semantic ECM) is presented, which integrates semantic search methodologies and e-learning technologies, which allows the creation of customized courses according to the student's requirements and preferences. It is based on a reference ontology that contains the concepts and relationships of the particular domain, selected by a panel of experts in that domain. In addition, students can choose a combination of the inductive vs. deductive and sequential vs. global approaches for the course; and also specify their training needs starting from the ontology. In this sense, the semantic similarity method SemSim [29] is used. It also allows integration with the MOODLE platform. Finally, it was validated in the osteoporosis domain and in general, the judgment was positive both in terms of usability and personalization.

In the work by Koutsomitropoulos et al. [22], it is mentioned that instructors and students are faced with two major problems when using digital educational resources: the first, is the difficulty to discover and retrieve material complementary to the courses; and in the same line, the second, the excessive manual work in the queries and the selection of educational material, based on the results. Therefore, they propose a framework and a service to face these problems, making use of the expansion of queries based on keywords that describe a particular course. This service is composed of three main components: the development of a thesaurus under ontologies; a subsystem of management of learning objects; and finally, a semantic middleware to evaluate the semantic relevance between the keywords and the thesaurus, to perform the query expansion and conduct the federated search in remote repositories. In this sense, the authors suggest that the expansion of the query contributes to an improved recovery. Finally, they state that a prototype of the proposed system is in operation for university's online courses.

B. Full-text search

In the paper of [7], the possibility of automatically identifying educational resources was evaluated, for which experiments were carried out with a dataset constructed on fourteen topics of computer science and manually annotated. In these experiments three classifiers were used: Naive Bayes, support vector machines and support vector regression. In this way, the authors conclude that the educational value of a learning object can be automatically assigned with high precision.

In the work of [23], a study is presented in which three information retrieval algorithms are compared for the recommendation of educational materials about diabetes for questions asked by patients. In this sense, the authors point out the importance of satisfying the information needs of patients, in order to facilitate self-management and care of their disease. To carry it out, they assessed the algorithms of vector space modeling (as baseline), Latent Dirichlet Allocation, and semantic group-based model on educational materials from the Mayo Clinic database and questions obtained from patients in the TuDiabetes web forum. In addition, for its evaluation, the precision metric of the top-ranked documents was used. According to this, it was determined that: the vocabulary of the language used in the educational materials is different from that one used in the forum of questions; the topic modeling-based model had better performance and has the potential to accurately recommend educational materials; and, finally, this one can mitigate the difference of vocabularies between the educational materials and the questions.

Rahman and Abdullah [24], in their article, propose a framework for the use of learners, within an institutional instructional environment, and that adds to Google Search an ability to filter their results based on their academic background, the behavior of learning when they use the search engine, and the behavior patterns of other students. To achieve this, the framework makes use of dynamic student profiles and a grouping algorithm. In addition, this proposal seeks to overcome the difficulty of generic search engines in the sense that they do not take into account the differences in the learning profiles of users and that, according to a reported study, only 8% of the results were of educational resources according to the learner's query. On the other hand, the proposed framework consists of two modules: one, the dynamic profile of students created from their academic record; the other, the re-rankig of the results of Google Search based on the profile of the learners. According to that, students are classified as beginner, intermediate or master through the C4.5 algorithm of the decision tree. Furthermore, the framework was tested through a prototype in 36 first-year undergraduate students at the University of Malaya, showing that the application was able to customize Google Search results according to the particular needs of the students.

Gasparetti et at. [25] propose an approach for the identification of prerequisites of textual learning objects, through machine learning. To carry it out, the learning objects are tokenized, their terms labeled and the semantic relationships between the terms are extracted. This last task is carried out with the use of Wikipedia, which is considered a weak ontology, using the Tagme annotation tool. Next, the recognition of requirements is performed using automatic learning classification algorithms such as C4.5 decision tree, multilayer perceptron neural network, and Naive Bayes. Finally, to evaluate the accuracy of the approach, experiments are conducted on real online courses in different domains.

C. Search based on hybrid methods

Cernea et al. [26], in their paper, propose an architecture called SOAF (for its first letters in Spanish of "semantics of learning objects based on folksonomies") for the semantic indexing of LOs in repositories. In this sense, the metadata used in indexing, through Latent Semantic Indexing, are obtained through three sources: the automatic semantic indexing based on the low-level characteristics of the learning objects; the descriptors supplied by the authors; and the collaborative annotations of the learners. In addition, with respect to tags and before being incorporated them into metadata, they are processed through a collaborative filter based on user profiles and linked to an ontology of a specific field.

Lee et al. [1] propose an approach based on ontologies for the semantic-aware retrieval of learning objects, which has two characteristics: the first, an automatic expansion algorithm based on ontologies; and the second, an ambiguity elimination function to adjust the unsuitable query terms. With the proposed model, the authors point out that two drawbacks of traditional information retrieval technology based on keywords of the Salton vector space model are overcome. The first drawback is the need to index the content in advance, which may fail in the case of learning objects given the broad context in which they are immersed, which may contain multimedia elements difficult to include in the index. The other one has to do with the presence of learning objects, not semantically related, that have common keywords. In addition, for the recovery of learning objects in the repository, indexed in a standard way, they are searched in the same way as in the ontology assistant system.

To corroborate the model proposed by Lee et al. [1] experiments were conducted on the automatic expansion algorithm and the semantic-aware learning object retrieval, compared to the performance of traditional keyword-based recovery techniques. The performance was measured based on three metrics used by the information retrieval community: precision, recall, and F-measure. In all the results, the proposed approach overcame the traditional one, based on keywords.

In the paper by Zhuhadar and Nasraoui [27], the authors present an e-Learning platform for the personalized recovery of learning objects that make use of the standards of the Semantic Web to represent the contents and profiles of users as ontologies and re-ranking the search results based on how the terms are assigned to these ontologies. To achieve this, they propose a three-layer architecture: semantics, algorithms, and personalized interface. In this sense, in the first layer, a semantic profile of the student is elaborated based on his search history; queries are constructed with keywords related to the student profile, the courses and the most significant terms of each concept; and a taxonomy of all the documents of the repository is extracted, clustering them in categories finer than those given by the colleges. Then, for a given query of a student, documents similar to the terms of the query are retrieved; these results are re-ranked according to the semantic profile of the student and the most similar clusters of concepts. Finally, the authors state that their results show the effectiveness of the re-ranking of search results based on the semantic profile of the student.

In the work by Khribi et al. [9], a personalized online recommender system for students is proposed, which does not require explicit feedback. The recommended learning objects are calculated based on the student's recent browsing history, as well as the content of the learning object and the exploitation of similarities and differences between the preferences of the students. For the recommendation of the learning objects, a hybrid system is used, where the results of two recommendation approaches are combined and integrated: one, a cascade collaborative filtering, increased by recommendation by content of the learning object; and another, a recommendation approach based on weighted content and collaborative filtering based on the learner's profile. Both techniques are executed separately and the results are integrated into a single recommendation set. Equally important, to improve the recommendation based on the content, made through the Apache Nutch search engine, the content of the LOM educational metadata that provides additional information of the learning objects are added, automatically, to the native index generated by the search engine (inverted index). The results are sorted through the cosine similarity (TF-IDF vector).

In [28], the Shih and Tseng paper, a ubiquitous learning information retrieval system context-based is proposed, founded on instructional strategies and goals. The proposed system is composed of four components: the user interface for the input of the query and the detection of the context, mainly related to the student's location; the expansion of the query by the inference of rules; the recovery of content; and the construction of ontology and the generation of rules. In this context, for the search and classification of teaching contents, a combined technique of similarity based on keywords (full text through the cosine function) and on metadata (based on the number of matching attributes) was used. In addition, to evaluate the system, a prototype was developed that was tested in a primary school and three experiments were conducted with 20 of its students. Finally, the authors present three results: the first, the proposed system can improve the recovery performance based on the context (in a ubiquitous learning scenario); the second, according to a survey conducted, shows that it is feasible for teachers to help the system generate a simple ontology based on a predefined course scheme; and the third, the results show that expanded queries work better than the original.

In the work of [10] the implementation of a hybrid recommender system of learning objects based on two types of recommendation is presented: the first, content-based recommendations through ontology; the second, rules-based recommendations founded on the interests of the student. The results of both recommenders are combined according to different weights. Additionally, for the content-based recommender design, a hand-made ontology from a coarse-grained taxonomy was constructed on 7,424 documents (in English and Spanish) from the repository of Western Kentucky University. Furthermore, the ontology of each learner is extracted, based on their preferences, from the general ontology of the repository and is a subset of it. Finally, based on the results, the authors highlight the effectiveness of personalization in the search engine of the recommender system.

Atkinson et. al [11] propose an approach in which they make use of several paradigms for the semantically guided extraction, indexing, and search of educational metadata found on the Web, in order to identify educational resources and thus help teachers in this task. In this sense, the proposed model incorporates automatic learning techniques, formal analysis of concepts and natural language processing algorithms. To validate the model, a Web-based prototype was implemented and 500 documents were extracted with which three types of experiments were carried out: one, parameters setting; another, classification accuracy; and the third, quality of the extracted metadata. For the latter experiment, 16 secondary school teachers participated. The authors mention that promising results were obtained and pointed out that the semantically guided metadata extraction can improve access and use of educational resources present on the web.

IV. Discussion

According to the results obtained with the methodology used, searches based on metadata represent most of the papers reviewed, followed by hybrid search and, to a lesser extent, full-text search (see Figure 1). However, if the comparison is made by citations, the hybrid search of papers represents a very high percentage of the total (see Figure 2) and the full-text search of papers takes a very low value. This seems to indicate that, although research is still being conducted on metadata searches, the academic community is giving a greater preponderance to research on hybrid searches.

Another aspect that the review shows is the use of the automatic expansion of queries in 20% of the papers reviewed. In this way, it is possible to increase the recall while maintaining precision [22]. Likewise, the use of the student profile is proposed in 30% of the works to achieve more personalized results.

On the other hand, in 40% of the papers, the use of ontologies is considered in order to incorporate the semantic search under specific domains. Thus, it is intended to overcome the weakness of the current LO standards with respect to the lack of semantic metadata [19]. Also, for more general domains, the authors start experimenting with weak ontologies such as Wikipedia [25].

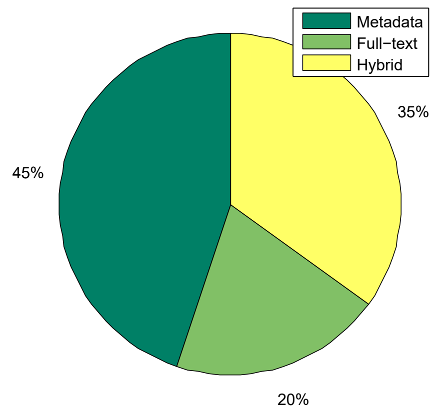

Figure 1. Papers by search category

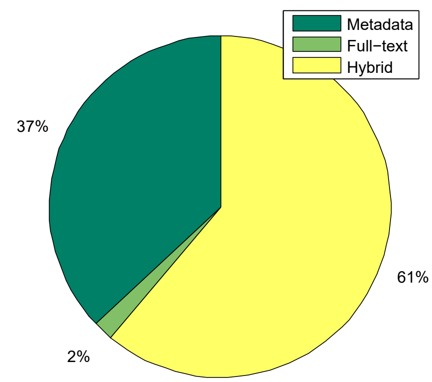

Figure 2. Citations by search category

Furthermore, in hybrid searches, methods and techniques of machine learning, neural networks, fuzzy logic, natural language processing, and data mining have been incorporated.

V. Conclusions

A review of the literature on searches of learning objects in repositories was conducted. For this, a methodology was implemented to select the papers to be reviewed. In this line, it incorporates the use of Google Scholar as a tool to unify the way of counting the number of citations per paper. In addition, it gives the possibility of knowing the H index of the authors, in most cases. However, although it presents some biases, they apply to the entire population. Therefore, it was used. Finally, it was verified that the methodology used to determine the papers of the review was adequate and allowed an approach to the state of the art in this topic.

On the other hand, it was found that although the search for LOs through metadata continues to be preponderant, the hybrid search has attracted the interest of the academic community, as shown by a large number of citations. Similarly, it was verified that in order to improve the search results, the use of ontologies is proposed although it remains, for the most part, in specific domains. Additionally, in several works, the use of the user profile is proposed to achieve more personalized results. Last but not least, also in several papers, it is proposed to integrate the automatic expansion of queries with searches to improve recall.

References

[1] M. C. Lee, K. H. Tsai, and T. I. Wang, "A practical ontology query expansion algorithm for semantic-aware learning objects retrieval," *COMPUTERS & EDUCATION*, vol. 50, no. 4, pp. 1240–1257, may 2008.

[2] S. Ternier, D. Massart, A. Campi, S. Guinea, S. Ceri, and E. Duval, "Interoperability for Searching Learning Object Repositories. The ProLearn Query Language," *D-Lib Magazine*, vol. 14, no. 1/2, 2008.

[3] N. D. Duque Méndez, D. A. Ovalle, and J. Moreno Cadavid, *Objetos de Aprendizaje, Repositorios y Federaciones... Conocimiento para Todos*. Manizales: Universidad Nacional de Colombia, 2014.

[4] T. Yigit, A. H. Isik, and M. Ince, "Web-based learning object selection software using analytical hierarchy process," *IET Software*, vol. 8, no. 4, pp. 174–183, 2014. [Online]. Available: http://digital-library.theiet.org/content/journals/10.1049/iet-sen.2013.0116

[5] IEEE, "Standard for Learning Object Metadata," Institute of Electrical and Electronics Engineers, New York, Tech. Rep., 2002. [Online]. Available: http://ltsc.ieee.org/wg12/files/LOM{_}1484{_}12{_}1{_}v1{_}Final{_}Draft.pdf

[6] Y. Biletskiy, H. Baghi, I. Keleberda, and M. Fleming, "An adjustable personalization of search and delivery of learning objects to learners," *Expert Systems with Applications*, vol. 36, no. 5, pp. 9113–9120, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2008.12.038

[7] S. Hassan and R. Mihalcea, "Learning to identify educational materials," *ACM Transactions on Speech and Language Processing*, vol. 8, no. 2, pp. 1–18, 2011.

[8] J. Beall, "The Weaknesses of Full-Text Searching," *The Journal of Academic Librarianship*, vol. 34, no. 5, pp. 438–444, sep 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0099133308001067

[9] M. K. Khribi, M. Jemni, and O. Nasraoui, "Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval," *Educational Technology & Society*, vol. 12, no. 4, pp. 30–42, 2009.

[10] L. Zhuhadar and O. Nasraoui, "A hybrid recommender system guided by semantic user profiles for search in the e-learning domain," *Journal of Emerging Technologies in Web Intelligence*, vol. 2, no. 4, pp. 272–281, 2010.

[11] J. Atkinson, A. Gonzalez, M. Munoz, and H. Astudillo, "Web metadata extraction and semantic indexing for learning objects extraction," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 41, no. 2, pp. 649–664, 2014.

[12] The_Ohio_State_University, *Choosing & Using Sources: A Guide to Academic Research*. Columbus: The Ohio State University, 2016.

[13] T. Ferreras-Fernández, H. Martín-Rodero, F. J. García-Peñalvo, and J. A. Merlo-Vega, "The systematic review of literature in LIS: An approach," *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality - TEEM '16*, pp. 291–296, 2016. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3012430.3012531

[14] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and G. PRISMA, "Preferred Reporting Items for Systematic Reviews and Meta-Analyses :," *Annulas of Internal Medicine*, vol. 151, no. 4, pp. 264–269, 2009.

[15] L. J. Martínez Rodríguez, *Cómo buscar y usar información científica: Guía para estudiantes universitarios 2013*. Santander, España: Universidad de Cantabria, 2013.

[16] P. A. Gaona-García, D. Martin-Moncunill, and C. E. Montenegro-Marin, "Trends and challenges of visual search interfaces in digital libraries and repositories," *The Electronic Library*, vol. 35, no. 1, pp. 69–98, 2017. [Online]. Available: http://www.emeraldinsight.com/doi/10.1108/EL-03-2015-0046

[17] X. Ochoa and E. Duval, "Relevance Ranking Metrics for Learning Objects," *IEEE Transactions on Learning Technologies*, vol. 1, no. 1, pp. 34–48, 2008. [Online]. Available: https://lirias.kuleuven.be/bitstream/123456789/261082/1/tltMetrics.pdf

[18] N. Y. Yen, T. K. Shih, L. R. Chao, and Q. Jin, "Ranking Metrics and Search Guidance for Learning Object Repository," *IEEE Transactions on Learning Technologies*, vol. 3, no. 3, pp. 250–264, 2010.

[19] I.-C. Hsu, "Intelligent Discovery for Learning Objects Using Semantic Web Technologies," *Educational Technology & Society*, vol. 15, no. 1, pp. 298–312, 2012.

[20] A. S. Sabitha, D. Mehrotra, and A. Bansal, "Delivery of learning knowledge objects using fuzzy clustering," *Education and Information Technologies*, vol. 21, no. 5, pp. 1329–1349, 2016. [Online]. Available: http://dx.doi.org/10.1007/s10639-015-9385-5

[21] A. Barbagallo and A. Formica, "ELSE: an ontology-based system integrating semantic search and e-learning technologies," *Interactive Learning Environments*, vol. 25, no. 5, pp. 650–666, 2017.

[22] D. Koutsomitropoulos, G. Solomou, and K. Kalou, "Federated semantic search using terminological thesauri for learning object discovery," *Journal of Enterprise Information Management*, vol. 30, no. 5, 2017.

[23] Y. Zeng, X. Liu, Y. Wang, F. Shen, S. Liu, M. Rastegar-Mojarad, L. Wang, and H. Liu, "Recommending Education Materials for Diabetic Questions Using Information Retrieval Approaches," *Journal of medical Internet research*, vol. 19, no. 10, p. e342, 2017.

[24] M. M. Rahman, N. A. Abdullah, and F. Aurangozeb, "A Framework for Designing a Personalised Web-based Search Assistant Tool for eLearning," in *International Conference on Information and Communication Technology (ICoICT)*. Malacca, Malaysia: IEEE Xplore, 2017.

[25] F. Gasparetti, C. De Medio, C. Limongelli, F. Sciarrone, and M. Temperini, "Prerequisites between learning objects: Automatic extraction based on a machine learning approach," *Telematics and Informatics*, vol. 35, no. 3, pp. 595–610, 2018. [Online]. Available: https://doi.org/10.1016/j.tele.2017.05.007

[26] D. A. Cernea, E. Del Moral, and J. E. Labra Gayo, "SOAF: Semantic Indexing System Based on Collaborative Tagging," *Interdisciplinary Journal of E-Learning and Learning Objects*, vol. 4, no. 1, pp. 137–150, 2008. [Online]. Available: http://www.ijello.org/Volume4/IJELLOv4p137-149Cernea.pdf

[27] L. Zhuhadar and O. Nasraoui, "Semantic information retrieval for personalized e-learning," in *IEEE International Conference on Tools with Artificial Intelligence, ICTAI*, vol. 1, 2008, pp. 364–368.

[28] W.-C. Shih and S.-S. Tseng, "A Knowledge-based Approach to Retrieving Teaching Materials for Context-aware Learning," *EDUCATIONAL TECHNOLOGY & SOCIETY*, vol. 12, no. 1, pp. 82–106, 2009.

[29] A. Formica, M. Missikoff, E. Pourabbas, and F. Taglino, "Semantic search for matching user requests with profiled enterprises," *Computers in Industry*, vol. 64, no. 3, pp. 191–202, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.compind.2012.09.007