

Article

A Semantic Model for Selective Knowledge Discovery over OAI-PMH Structured Resources

Arianna Becerril-García *  and Eduardo Aguado-López

Dissemination of Science Research Group, School of Political and Social Sciences,
Autonomous University of the State of Mexico, 50100 Toluca de Lerdo, Mexico; eaguadol@uaemex.mx

* Correspondence: abecerrilg@uaemex.mx; Tel.: +52-722-555-7146

Received: 9 May 2018; Accepted: 7 June 2018; Published: 12 June 2018



Abstract: This work presents OntoOAI, a semantic model for the selective discovery of knowledge about resources structured with the OAI-PMH protocol, to verify the feasibility and account for limitations in the application of technologies of the Semantic Web to data sets for selective knowledge discovery, understood as the process of finding resources that were not explicitly requested by a user but are potentially useful based on their context. OntoOAI is tested with a combination of three sources of information: Redalyc.org, the portal of the Network of Journals of Latin America and the Caribbean, Spain, and Portugal; the institutional repository of Roskilde University (called RUDAR); and DBPedia. Its application allows the verification that it is feasible to use semantic technologies to achieve selective knowledge discovery and gives a sample of the limitations of the use of OAI-PMH data for this purpose.

Keywords: ontologies; Semantic Web; OAI-PMH; knowledge discovery; awareness to context

1. Introduction

The World Wide Web (the Web), a gigantic community of documents, has evolved rapidly since Tim Berners-Lee introduced its concept to the world. Finding relevant material for learning, teaching, or research in the flow of existing resources and publications is becoming a major challenge for both students and scientists. In addition, sharing of resources, resource metadata, and data across the Web is a central principle in scientific and educational contexts. Scientific collaboration has long been striving for wider reuse and sharing of knowledge and data [1].

This mass of information that constitutes the Web sometimes feels “a mile wide but an inch deep”. How is a more integrated, consistent, and deeper Web experience built? [2]. This question is where semantics is situated, the process of communicating information with sufficient meaning. It is thus possible to build intelligent applications that provide better understanding by identifying content in greater depth.

Moreover, the variety of information resources on the Web that are useful for a student, academic, teacher, or scientist is very broad, encompassing books, journal articles, reports, conference proceedings, theses, pre-prints, and data files, among others, all available through specialized portals, repositories, and databases, both commercial and non-commercial.

Since 2001, a movement has been consolidating that promotes free and unrestricted access to scientific content, especially if it has been publicly funded. This movement is called Open Access and was formalised through three statements: Budapest [3], Berlin [4], and Bethesda [5]. Repositories, portals, and journals that adhere to this movement adopt, as good practice, an interoperability protocol to exchange information and thus have communication rules and standards for structuring data.

The registration of Open Access mandates ROARMAP shows a growing curve in the global implementation of policies at the level of institutions and countries that encourage the adoption of

Open Access. Similarly, OpenDOAR, the Directory of Open Access Repositories, shows a constant rise in the creation of repositories.

The adoption of policies helps the population of repositories, consolidating these platforms as main disseminators of the knowledge generated by an institution or a country.

Open Access platforms have adopted the OAI-PMH (Open Archives Initiative—Protocol for Metadata Harvesting) protocol as the de facto standard of communication and interoperability. It accounts for this, the Registry of Open Access Repositories—ROAR—with thousands of active data providers displaying millions of articles on the Web. This shows the consolidation of the repositories and the large amount of academic and scientific material available online under the philosophy of free access.

It is thus, that it is relevant to dedicate efforts and apply new techniques such as Semantic Web technologies—which have been positively proven in terms of the benefits they bring to other sectors—in order to add value to the discovery of knowledge contained in these platforms; in such a way that students, professors, or researchers obtain greater benefits when working with them.

The structuring of data, with which these information resources are presented on the Web, provides another semantic level in understanding their nature. Thus, software applications can use them for the benefit and improvement of information retrieval engines. Research discovery is facilitated through the navigation of content, and structured data lend themselves to a variety of applications that can use these data [6].

On the other hand, knowledge discovery (KD) is the most desirable end product of computing. Finding new phenomena or improving knowledge about them is second in value only after preserving the world and the environment [7]. Frawley, Piatetsky-Shapiro, and Matheus [8] define knowledge discovery as the non-trivial extraction of previously unknown and potentially useful implicit information from data. The challenge of extracting knowledge from data demands statistical research, databases, pattern recognition, machine learning, data visualization, optimization, and high-performance computing to deliver advanced business intelligence and web discovery solutions [9].

The OAI-PMH emerged with the goal of interoperability between platforms. In addition to being exchanged between systems, resources structured under this specification are used in search engines that allow users to query. However, these data have untapped potential uses in applications that enable semantic knowledge discovery.

The Dublin Core Metadata Initiative (DCMI) supports the development of interoperability standards at different levels, among which it is a set of metadata for simple and generic descriptions popularized as being part of the OAI-PMH protocol specifications.

The scope of Dublin Core metadata specification is currently characterized in four levels of interoperability

- Level 1: Shared term definitions
- Level 2: Formal semantic interoperability
- Level 3: Description set syntactic interoperability
- Level 4: Description set profile interoperability

The OAI-PMH protocol implements the interoperability level 1 with a metadata set of 15 elements. The protocol also requires serialization in XML, therefore, it is necessary to adapt these outputs to the syntactic coding of RDF/XML or some other RDF serialization format, as a determining factor for the description of resources such as generic scheme for the exchange of metadata between heterogeneous and distributed systems, content description standards (ontologies, thematic maps, thesauri, among others), and a variety of protocols and standards for the exchange of information on the Semantic Web.

The overall objective of the approach presented here is to propose a semantic model for the selective discovery of knowledge about resources structured with the OAI-PMH to verify the feasibility and give evidence of the limitations in the application of Semantic Web technologies to these data sets

in order to discover knowledge, a process understood as identifying resources that were not explicitly requested by a user but are potentially useful.

The OntoOAI model arises from the idea of allowing knowledge discovery through inference processes on OAI-PMH structured resources, considering a user profile to discover potentially useful information for the user. This article shows the complete model and application case. The first findings of this model in its early stages were published in Becerril Garcia, Lozano Espinosa, Molina Espinosa [10]; and Becerril Garcia, Lozano Espinosa, Molina Espinosa [11].

The results consist of a semantic model for selective knowledge discovery, a semantic application that enables the processing of data structured with OAI-PMH, the application of ontologies in the description and verification of the knowledge obtained from OAI-PMH resources and inference testing mechanisms on the resultant dataset.

2. Materials and Methods

The application of Semantic Web technologies to data structured with OAI-PMH for selective knowledge discovery for a user has led to the development of a model named OntoOAI.

The following sequence of steps (Figure 1) represents the methodology to develop the OntoOAI model:

1. Harvesting of resource metadata that are available through repositories that implement OAI-PMH.
2. The conversion of the results obtained in XML to RDF.
3. The enrichment of statements (obtained in Dublin Core) with additional information from authors (FOAF).
4. Validation of the resulting knowledge.
5. Triple storage.
6. Information retrieval based on inference, which will provide service to the end user. One may note that the required information input from a user interface matches the query and the contextual user information.

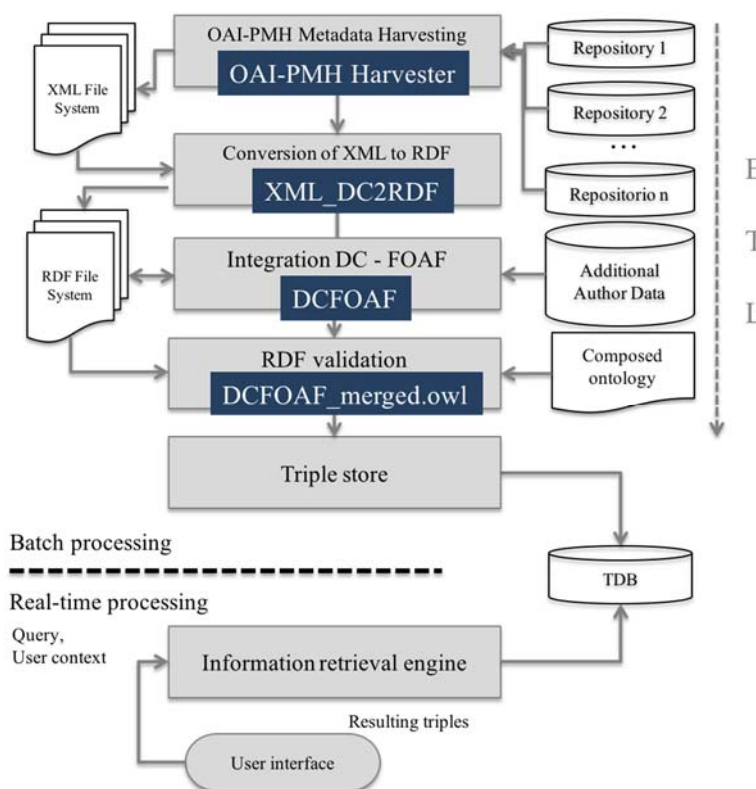


Figure 1. OntoOAI Methodology.

The user context is an input to the information retrieval process for performing selective discovery. The developed software components are:

- OAI-PMH harvester
- XML_DC2RDF
- DCFOAF
- DCFOAF_merged.owl

This approach is part of the logic of the extract, transform, and load (ETL) process, given that the information is collected and organised to later be transformed into RDF, enriched and validated using ontologies, and loaded as triples to a data file. This methodology is designed to keep adding from 1 to n repositories, provided they are OAI-PMH compliant.

2.1. Sources of Information

By using one directory of repositories that implement OAI-PMH, like OpenDOAR, ROAR, or OpenArchives for instance, one can locate Web sites that meet this specification and are therefore candidates to provide information for OntoOAI. Additionally, it is important to analyse the “service providers” that, in turn, present the metadata via the protocol, which would significantly simplify harvesting, given that a service provider already harvests a set of repositories. After selecting the sites that satisfy the protocol within these directories, it is important to locate the base URL of each, which will be input into the metadata harvesting process.

Moreover, the model includes the enrichment of metadata with additional information from the authors. This information must comply with the FOAF specification, and it can come from the social networks of researchers or other information sources such as DBpedia. DBpedia is a collaborative effort to extract structured information from Wikipedia and Wikidata and make it available on the Web. DBpedia makes it possible to query Wikipedia in a more sophisticated way and link different data sets to the ones obtained from Wikipedia. This initiative makes these data available through bases of triples in different formats. The DBpedia data set uses a multi-domain ontology derived from Wikipedia.

2.2. OAI-PMH Metadata Harvesting

The process of harvesting or collecting metadata consists of sending HTTP requests to data providers that implement OAI-PMH for one or more records corresponding to information resources. A Java application, called OAI PMH Harvester, was developed for this process. OAI_PMH_Harvester works by receiving the response of the request sent to the repository:

```
[URLBase]?verb = ListRecords&metadataPrefix = oai_dc
```

Such a response is structured with the set of Dublin Core metadata and written to a file named (ListRecords.xml), which is managed as a tree structure using the Xerces Java Parser library. It is taken to a numerical sequence file of type ListRecords1.xml, ListRecords3.xml, ListRecords3.xml and so on, to save each batch of metadata sent by the data providers. There are some who send sets of 20, others sets of 100, and the configuration is particular to each repository. Thus, the harvester is able to receive varying amounts using the “resumptionToken” parameter. The harvester is also capable of performing harvests with optional parameters of the “ListRecords” verb—such as SET, FROM, and UNTIL—to run queries filtered by set or dates.

2.3. XML to RDF Conversion

XML_DC2RDF is the converter that takes the output of the harvester represented as XML files with Dublin Core to convert them to RDF. This conversion is necessary because, as confirmed in the Report on Metadata Models for Multimedia Content of the OMediaDis project (Open Platform for Content Management, Multichannel Distribution), many institutions provide access to their metadata

repositories through OAI-PMH but do not make their resources accessible through dereferenceable URIs, which restricts meaning and causes access to remain restricted to metadata [12].

There are some developments available that allow the exploitation of structured data, providing them with characteristics typical of semantic applications, such as the ones tasked with converting to RDF, also known as RDFizers. These tools convert data from relational databases, CSV files, XML, and text, among others. Three of them were compared to verify the feasibility of implementation within OntoOAI: OAI2LOD Server, a development designed to present OAI-PMH metadata as Linked Data [13]; D2R [14]; and OAI2RDF [15]. However, it was decided to develop a software application to perform this task because, in addition to overcoming some drawbacks, it was more suitable for incorporating the harvesting process with the rest of the implementation. This application is responsible for reading the metadata of XML files obtained from OAI-PMH harvesting using the SAXBuilder API.

SAXBuilder builds a JDOM document using a third-party SAX parser to handle the analysis tasks and uses an instance of SAXHandler to listen to SAX events and thus build a JDOM document content using JDOMFactory. JDOM is a Java representation of an XML document. JDOM provides a way to represent that document for easy and efficient reading, manipulation, and writing.

2.4. Integration of DC XML with FOAF

Author data are enriched to add co-author relationships and more information about the author. OntoOAI works with data sources that describe information about people when they are expressed with the FOAF specification. OntoOAI used DBpedia for this purpose, specifically the “PersonData” dataset. Thus, for each Dublin Core record obtained from the harvesting processes, information is added to the creator label corresponding to the author or co-authors of the resource.

The OntoOAI model expects that if an information resource has several authors, the creator label will be repeated such that the same procedure is carried out for each author. Then, for each individual, the RDF type is specified, such as <http://xmlns.com/foaf/0.1/Person>, and the name (foaf:name), surname (foaf:givenName, foaf:surname), people with whom the individual has a relationship, in this case co-authorship (foaf:knows), interest (foaf:topic_interest), and date of birth (onto:birthDate) are also included. Finally, the creator metadata are obtained as follows:

```
<dc:creator rdf:type = [http://xmlns.com/foaf/0.1/Person] foaf:name = [full name]
foaf:givenName = [name] foaf:surname = [last name] foaf:knows = [known person or
co-author] foaf:topic_interest = [topic of interest] dc:description = [occupation, grade, etc.]
onto:birthDate = [birthdate]/>
```

The process of recognizing an author as an entity was performed using the full name of the author and exact matching. Needless to say, the problem of homonymy or polysemy (two authors with the same name) and synonymy (an author who publishes under different names) among authors affects the results of this phase of the process.

Therefore, it is evident that there is an important area of opportunity in the unique identification of people on the web, particularly of researchers or authors in the academic–scientific content. No doubt, this model can be improved if the repositories that present metadata under the OAI-PMH make use of dereferenceable URIs, but this practice is not commonly included in OAI-PMH data providers. The use of ORCID or a similar unique identifier that guarantees an URI for an author is a desirable requirement for repositories and should be exposed through OAI-PMH.

2.5. Ontological Engineering

The model verifies the resultant knowledge through an ontology that results from combining two existing ones: Dublin Core and FOAF. This process is performed by merging ontologies. Next, each ontology and the technique used for merging are described, as well as the information validation process for which this new ontology is created.

Dublin Core: the resources that feed the knowledge base are structured with DC metadata under the OAI-PMH protocol. For this work, the <http://purl.org/dc/terms> namespace is used.

Friend of a Friend: the second ontology in OntoOAI is FOAF, used to incorporate additional data on the authors of articles, books, and other academic resources. In combination, their social relations based on co-authorship can even be expressed to determine related resources. The FOAF namespace is used to represent information about people, such as names, birth dates, photographs, and especially other people to whom they are related in some way. It is particularly useful for representing social network data.

FOAF works under the spirit of the AAA (Anyone can say Anything about Any topic) principle. In this case, the issues that anyone usually talks about are people. Other topics that are commonly related to what can be said about people and that are included in FOAF are organizations (to which they belong), projects (on which they work), documents (that they have created or that describe them), and many others [2].

The intention of merging the two ontologies, Dublin Core and FOAF, lies in generating a representation of the knowledge of the integration of data from the OAI-PMH repositories (Dublin Core) and other sources of information to enrich the author data (FOAF), with the goal of building a model to verify the consistency of this integration. The ontological model proposed is illustrated in the following graph. With it, one can model the properties of an information resource (e.g., book, article) with its author or authors; for example, a researcher is an author of a publication (dc:creator) but is also a person (foaf:person) with individual properties. Likewise, the relationship of authorship between a publication and a researcher is represented, along with the co-authorship of a researcher with one or more others with whom they write a publication (Figure 2).

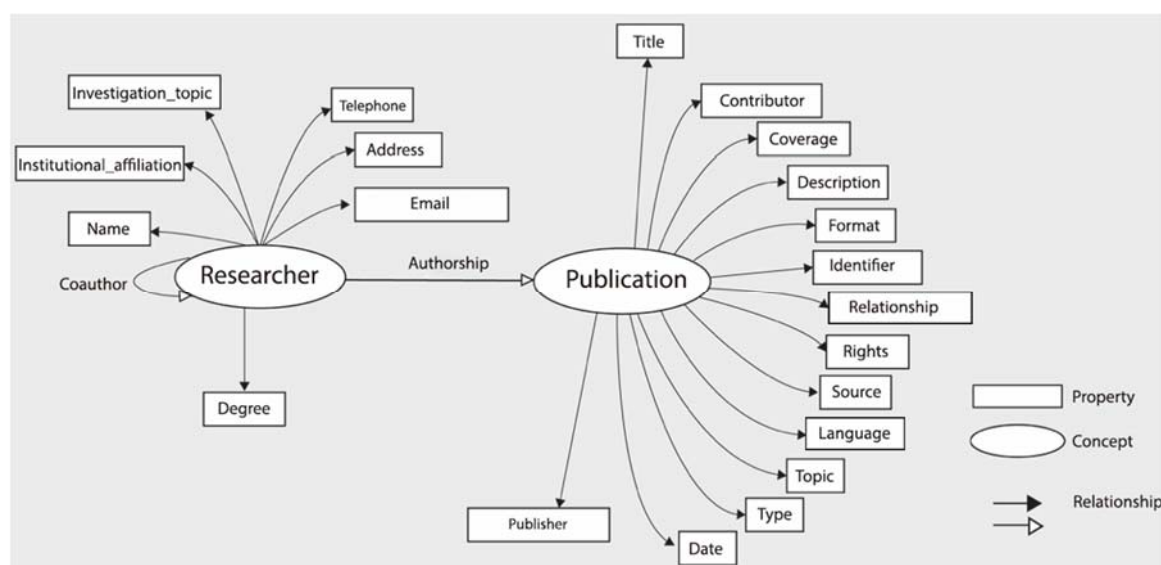


Figure 2. Authorship and co-authorship relations in merging ontologies.

In general, operations with ontologies are non-trivial tasks that cannot be fully resolved automatically due to a variety of factors, such as insufficient specification of an ontology for finding similarities with another ontology. Therefore, they are usually performed manually or semi-automatically, where a tool helps to find possible relationships between elements of different ontologies, but the final confirmation of the relationship is left to an expert, who will make decisions based on the description of the natural language of the elements of the ontology and common sense.

With the help of the Protégé tool and its “Refactor > Merge Ontologies” function, the merging of the two input ontologies was performed. As Ameen, Rahman, and Rani [16] mention, this automatic integration does not resolve the inconsistencies generated after the process. Therefore, the ontology

output was subjected to further refinement following the steps of the merging algorithm proposed by the same authors (Figure 3).

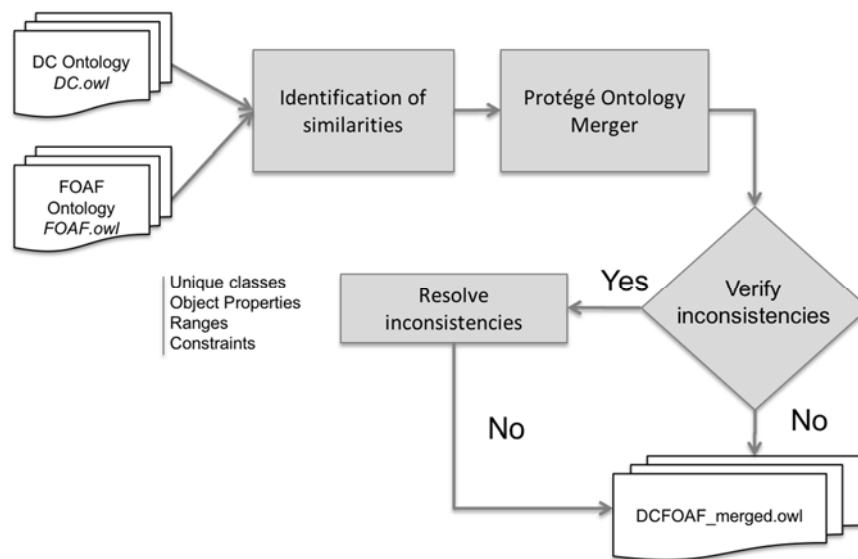


Figure 3. OntoOAI ontology merger algorithm based on [16].

The automatic process in Protégé identified classes that were identical in name (rdfs:Class and Agent) and performed the merger. However, for the case of “BibliographicResource” and “Document”, it did not perform a merger because the name did not match, and thus, the identification had to be made explicit manually, as it is shown in Figure 4.

The class hierarchy of the resulting ontology is shown in the Figure 5. In addition, it was necessary to identify other similarities to declare additional equivalencies and restrictions as well as to resolve inconsistencies.

Similarly, it was necessary to analyze the object properties to identify semantic matches, in which it is also important to check and match ranges and domains. Regarding these properties, the Dublin Core Creator (<http://purl.org/dc/terms/creator>) was defined as similar to FOAF maker (<http://xmlns.com/foaf/0.1/maker>), shown in Figure 6.



Figure 4. Declaration of “Document” and “BibliographicResource” equivalence.

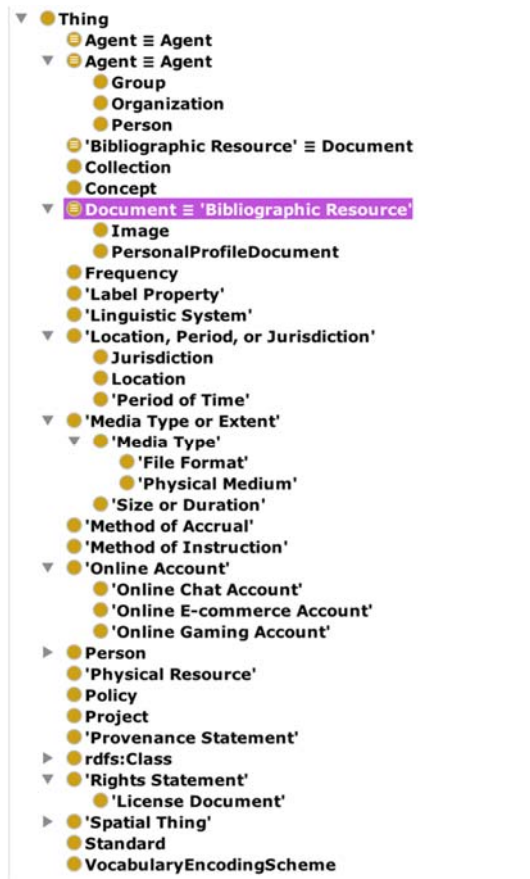


Figure 5. Class hierarchy of the merged ontology.

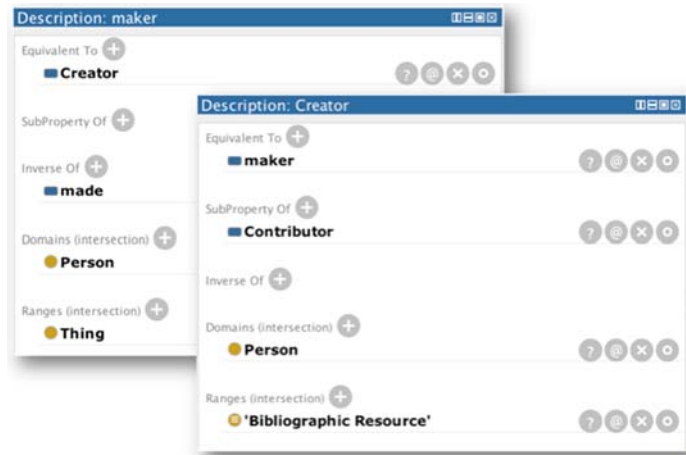


Figure 6. Equivalency of the “Creator” and “Maker” properties.

The resulting ontology, called “DCFOAF_merged.owl”, is verified using a reasoner; given that the stated and inferred models are similar, it is said that the ontology is consistent. The metrics of the resulting ontology are shown in the Table 1.

Thus, this ontology verifies the consistency of the RDF/XML files obtained in the previous phases.

Table 1. Metrics of the resulting ontology, “DCFOAF_merged.owl”.

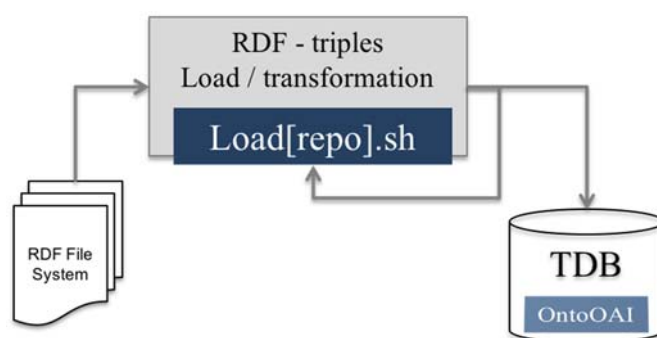
Classes	42
Object properties	66
Data properties	40

2.6. Storage

OntoOAI uses TDB of Jena framework as a triple store. Once the data are collected, processed, enriched, and validated, they pass to the triples store. Upon completion of these processes, a set of RDF graphs—serialized in RDF/XML—files are obtained. In each of them, a predetermined amount of information resources is described, which depends on the configuration of the source repository; as will be seen in the case study, each harvested repository provides a different number of resources per file.

Each of these resources is described by the 12 Dublin Core metadata, which, when repeatable, can represent at least 12 elements per resource, which make up the properties of each along with the properties included in the stages of conversion to RDF and the enrichment of authors. Each of these properties gives rise to a triple within TDB. Triples loading is performed using the Jena TDBLoader, which allows simultaneous bulk loading and the generation of an index. It performs bulk loading operations more efficiently than simple RDF reading in a TDB model.

To make this loading more efficient, shell scripts were programmed that run directly on TDB, a script for each repository harvested with the number of loading statements corresponding to the number of collected files, as discussed in the case study. The following diagram shows the process that is followed (Figure 7).

**Figure 7.** Triple storage process.

2.7. Context Awareness

The ability of OntoOAI to perceive information about the user environment and provide results based on this information. The fundamental concept is to infer non-explicit situations and thus manifest intelligent behaviour. Therefore, a profile is associated with a user that consists of a set of features, some more dynamic than others, to represent the circumstances of academic, personal, and even social space-time. Thus, a student profile contains identification data, curriculum, period, and registered courses, among others.

For this work, the student representation model based on ontologies for intelligent tutoring systems for the distance learning of Panagiotopoulos, Kalou, Pierrakeas, and Kameas [17] was adapted, using the following four classes:

- “Student”, which identifies any student.
- “StudentCourseInformation”, which includes information relevant to the educational process, such as modules for the program of study, school, homework, tests, and so on.

- “StudentCurrentActivity”, which refers to the details of the academic activity of the course year.
- “StudentPersonalInformation”, which is the static and permanent student information.

The ontology proposed by these authors was not located, and it did not exactly meet the needs of OntoOAI; accordingly, a new ontology called OntoOAIStudent was designed, based on the ontology proposed by those authors but with a new design.

The user context consists of the information known in advance and determined in a static manner based on the OntoOAIStudent model. The information comes from the school or school control offices, which could provide data on students for retrieving information relevant to their studies. The information is automatically imported from the school database to the OntoOAI database in an initial load through CSV files. This representation considers information independent of the application and does not include information result of the user interaction with the application.

To create the ontology that describes the student, the four major classes of [17] are adopted, but with the following hierarchy of classes (Figure 8).



Figure 8. Class hierarchy of OntoOAIStudent ontology.

The user profile is grouped into four main classes (Figure 9), as follows:

1. “Estudiante”, which represents a student.
2. “Cursos”, which describes the classes that the student takes, the school where the student is enrolled, and the program and program level (BA, MA, PhD).
3. “ActividadActual”, which describes the current period that the student is enrolled in, previous experience, course goals, module, and period.
4. “InformacionPersonal”, which represents accessibility, demographics, and student motivation.

Finally, the user profile represented in the ontology is among the inference engine inputs described below. While the OntoOAI model can be useful in discovering knowledge for teachers, the user profile used to test the model focuses on the student. However, it is easily expandable to teachers, academics, or researchers and only requires changing the representation of the user profile to one suitable for these types of users.

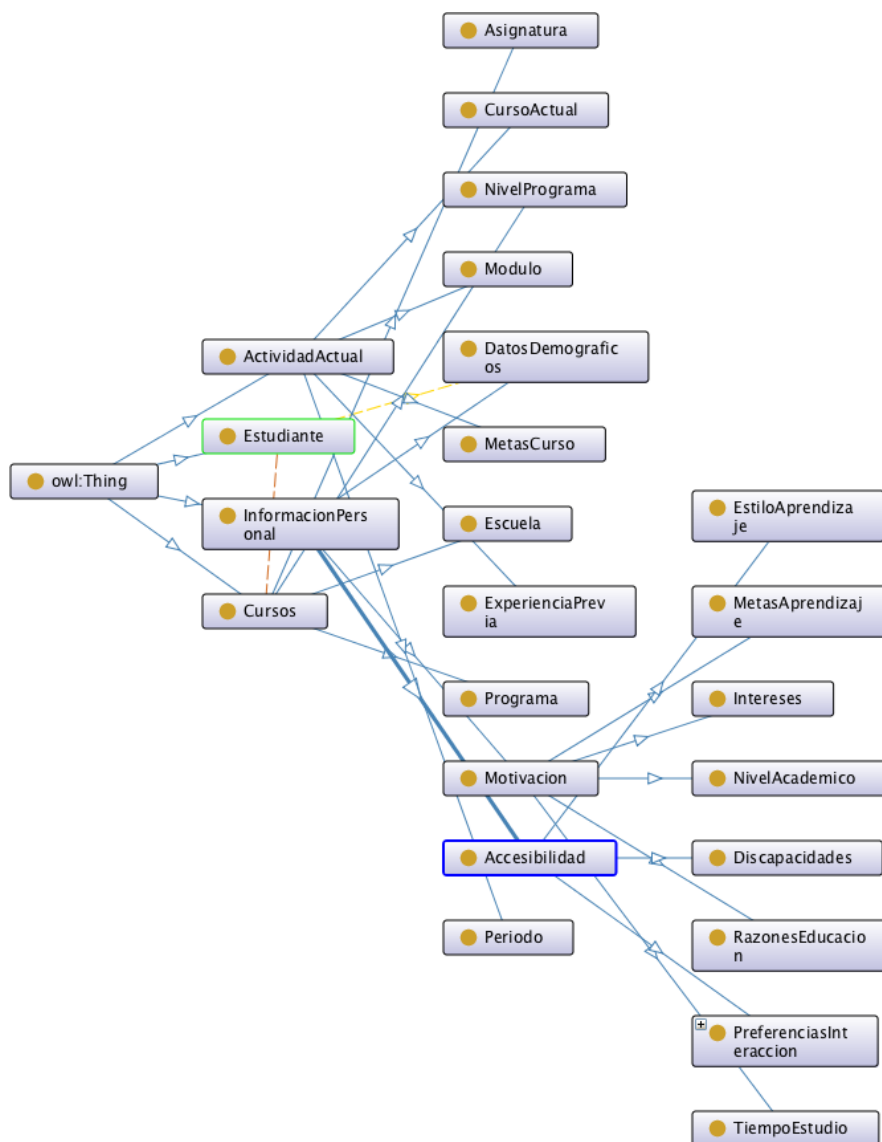


Figure 9. Graph of classes of the “OntoOAIStudent.owl” ontology.

2.8. Inference Engine and Knowledge Discovery

OntoOAI treats knowledge discovery as the nontrivial extraction of implicit, previously unknown, and potentially useful information, a concept defined by Frawley, Piatetsky-Shapiro, and Matheus [8]. Returning to the process of knowledge discovery defined by these authors but considering that it is conducted through logical reasoning using inference techniques, we have the following:

Given a set of facts F (knowledge base composed of triplets obtained from the harvesting and integration), a language L (OWL), and some degree of certainty C (considered “true” for the set of triplets inferred considering the previously defined user profile), a pattern is defined as a declaration S in L that describes relationships between a subset F_S of F with a certainty c such that S is simpler (in some sense) than the enumeration of all the facts in F_S . A pattern that is interesting (as it derives from facts) and sufficiently certain (because information is derived considering the user profile) is called knowledge.

Given the defined process of knowledge discovery, an application for performing it was designed. Its architecture and the interaction between the components are shown in the following figure.

It is important to remember that the required input information corresponds to the search terms of a user and to the context expressed in a profile, as explained in the previous section. Said input flows through a rule-based reasoner that processes a query on a database of triplets using the API of ontologies from the Jena Framework of the Apache Software Foundation (Figure 10).

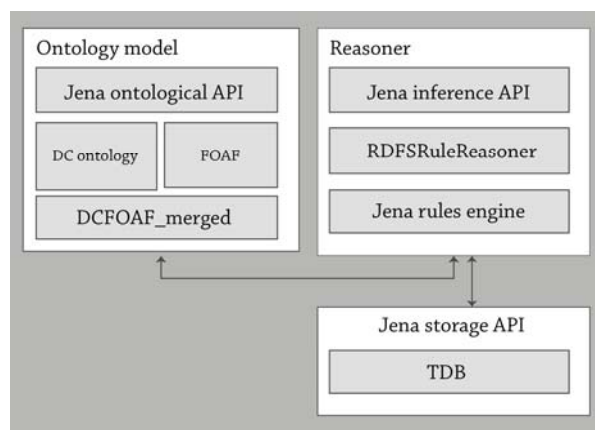


Figure 10. Architecture of the OntoOAI inference engine.

The Jena inference subsystem is designed to permit a range of inference engines or reasoners to be connected with Jena. Such engines are used to derive additional RDF declarations derived from an RDF base, together with optional ontological information and the axioms and rules associated with the reasoner [18]. The Jena engine infers logical consequences from a set of explicitly determined facts or axioms and typically provides automatic support for reasoning tasks such as classification, debugging, and query [19]. It is used to derive additional RDF declarations from the RDF base, together with optional ontological information and the axioms and rules from the reasoner.

For the purpose of experimenting with this development, the included OWL reasoner was used, which is an implementation based on the rules of OWL/Lite. The query process occurs in real-time on data previously collected from batch processing, giving as a result the set of inferences corresponding to information resources prepared as an input for applications that are responsible for end-user visualization.

The inference subsystem is designed to derive a set of declarations based on the database of facts from the OAI-PMH resource harvesting process, the context information from the user, and ontological information.

The inference engine is based on the assumption that if a resource of information is interesting for the user then the resource's author is also relevant, so it will be of the user's interest too other works whose author, subject, or title correspond to the mentioned author.

This way of linking resources was selected to run the inference engine due to the fact that this information is included in the Dublin Core specification.

Additional information as references or affiliations are not considered in the Dublin Core format of the OAI-PMH data providers. If they were, then this model could be expanded to include inferences based on more data.

The inference process, which follows a deductive reasoning modus ponens, or forward chaining, is based on the following syllogism:

Given the functions

$f(a) = Ra$, which defines the authorship or authorial contribution of a resource r written by an author a ,

$f(a) = Ta$, which defines the participation of an author a in a resource t , where participation means that the author is the subject, title, or resource author,

$f(b) = T_b$, which defines the participation of an author b in a resource t , where participation means that the author is the subject, title or resource author,

where:

$R = \{r \mid r \in R\}$, defined by the information resources contained in the knowledge base,

$A = \{a \mid a \in A\}$, defined by the authors or collaborators of information resources contained in the knowledge base,

$T = \{t \mid t \in T\}$, defined by the information resources contained in the knowledge base that were written by a certain group of authors,

$B = \{b \mid b \in B\}$, defined by the authors or collaborators of a given group of information resources contained in the knowledge base,

We have that:

If (the resource r is interesting for the user)

Then (the author a of resource r is interesting to the user).

If (the author a of resource r is interesting to the user)

Then (other works T_a whose author, subject, or title correspond to the author a are interesting to the user).

If (works T_a whose author, subject, or title correspond to the author a are interesting to the user)

Then (the set of authors or collaborators B of resources T_a is interesting to the user).

If (the set of authors or collaborators B of T_a resources is interesting to the user)

Then (the set of resources T_b written by authors B are interesting to the user)

The set of resources contained in sets T_a and T_b represent the knowledge discovered according to the search and context:

Knowledge discovered = $T_a \cup T_b$

$|T_a \cup T_b|$ = Total resources discovered for the user

The discovered knowledge forms the output for the user, in a table format with metadata and links to the full text of each resource obtained. This output can also be visualised in other ways, such as network type output and hierarchical visualisation, where one can see the relationships between resources, thereby enriching the user experience.

2.9. Batch Processing and Real-Time Processing

The nature of the operation of OAI-PMH requires collecting data repositories in batch processes for the following reasons:

- (a) Metadata harvest time depends on the response times of individual repositories. Additionally, the total metadata collection time is subject to the operation mode of the harvester, either sequential or parallel, for all repositories that are to be harvested. Thus, for the first mode, it will be the sum total of the response times of all repositories, and for the parallel mode, it depends on the number of repositories that are being harvested simultaneously and the slower response time of each processing thread.
- (b) The availability of repositories at the time of metadata collection processes can prevent a repository from being located. Thus, in a batch, it is possible to perform attempts to reconnect to the repository without impacting the end time.

Thus, batch processing and central storage of the results of the collection allow the handling of large volumes of data and the construction of applications and/or services based on integrated and available data. The query process is based on a set of parameters entering an inference engine to return a result, and it is performed in real time. It begins once the input is received and concludes with the sending of the output results. In this way, this information retrieval engine uses data

collected by the harvesting program that have been previously transformed, enriched, and stored in a centralized manner.

3. Experimental Results

Sources of Information

For the purposes of experimentation and validation of the proposed model, two repositories were located: Redalyc.org, the portal of the Network of Journals of Latin America and the Caribbean, Spain, and Portugal [20], and the institutional repository of Roskilde University (RUDAR—Roskilde University Digital Archive), Denmark.

Redalyc, available at <http://www.redalyc.org>, is a system of scientific information that currently maintains a collection of more than 500,000 full-text articles from more than one thousand peer-reviewed and open access scientific journals, edited by more than 600 institutions in Latin America. Redalyc also allows reading and downloading items and provides indicators for measuring scientific production for countries, institutions, journals, and authors.

Redalyc was selected first because it is one of the largest and most representative sources of scientific content in Latin America and second because the authors of this work have participated in this project since its inception, and thus, the experience and full knowledge of this project is gathered for the implementation of OntoOAI.

The RUDAR repository, at <http://rudar.ruc.dk/>, is the information system that collects, preserves, disseminates, and provides access to the intellectual and academic production of the University of Roskilde, a public university in Denmark founded in 1972. This repository mainly concentrates research products in digital form, including preprints, books, student reports, technical reports, articles, theses, conference proceedings, and images.

RUDAR was selected for the implementation of OntoOAI considering that this repository appears in OpenDOAR and therefore is displayed as a site that implements the OAI-PMH protocol. Thus, as RUDAR was selected, any other repository that complies with this minimum condition could be selected.

URLBaseRedalyc: <http://oai.redalyc.org/redalyc/oai>

URLBaseRudar: <http://diggy.ruc.dk:8080/dspace-oai/request>

Using the “OAI_PMH_Harvester” tool developed for OntoOAI, metadata collection processes were performed for each of the two repositories. Table 2 shows the results. From Redalyc, 19,027 XML files were obtained, corresponding to 380,540 recovered items, given that each XML file contains 20 articles. In the case of Rudar, 153 files were recovered with 100 resources each, for a total of 15,300, including books, theses, and other types of content.

Table 2. Results of the OAI-PMH harvesting process.

Repository	Total XML Files Obtained	Total Dublin Core Resources	Harvest Time	Harvest Size
Redalyc	19,027	380,540	45.6 h	762.9 MB
RUDAR	154	15,400	1.2 h	28.7 MB
Total	19,181	395,940	46.8 h	791.6 MB

One must remember that to raise the data to another semantic level, it is necessary to begin with the conversion of the XML files harvested from RUDAR and Redalyc to RDF. For this purpose, the converter developed for OntoOAI called XML_DC2RDF was used. Subsequently, the data from the authors are enriched to add more information that appears in data sources other than the harvested repositories (Figure 11). This enrichment is performed following the FOAF specification and by obtaining data from DBPedia, particularly the “PersonData” dataset, as described before.

An example of matches is presented below. As noted, the case of the author “Edgar Morin”, who has four items in Redalyc and whose data were enriched with information from DBpedia, corresponds to the data in the author’s Wikipedia biography.

```

<rdf:Description rdf:about="oai:redalyc.org:86019348002">
  <dc:rights>Signo y Pensamiento</dc:rights>
  <dc:subject>Comunicación</dc:subject>
  <dc:publisher>Pontificia Universidad Javeriana</dc:publisher>
  <dc:type>Artículo científico</dc:type>
  <dc:terms:modified rdf:resource="2012-01-13"/>
  <dc:terms:isPartOf rdf:resource="set:0120-4823"/>
  <dc:subject>Crisis cultural</dc:subject>
  <dc:identifier rdf:resource="http://www.redalyc.org/articulo.oa?id=86019348002"/>
  <dc:subject>Reforma universitaria</dc:subject>
  <dc:language>es</dc:language>
  <dc:source>Signo y Pensamiento (Colombia) Num.56 Vol.XXIX</dc:source>
  <dc:title>
    Anotaciones para un nuevo Emilio: transmisión sistémica del conocimiento
  </dc:title>
  <dc:creator rdf:type="http://xmlns.com/foaf/0.1/Person" foaf:name="Edgar Morin" foaf:givenName="Edgar" foaf:surname="Morin"
  <dc:description>"French philosopher and sociologist" onto:birthDate="1921-07-08"/>
  <dc:relation>http://www.redalyc.org/revista.oa?id=860</dc:relation>
  <dc:format>application_pdf</dc:format>
  <dc:subject>Formación transdisciplinar</dc:subject>
  <dc:subject>Saber especializado</dc:subject>
  <dc:description>
    La Universidad moderna nace de la reforma de Humboldt, quien introdujo las ciencias en plena expansión y contribuyó de manera decisiva a su desarrollo. Pero Humboldt era también consciente de que la Universidad debía proveer una cultura general, antes que una formación profesional especializada, para afrontar los problemas fundamentales y globales que se nos plantean en este siglo de manera cada vez más vital y urgente. Este principio cobra aún más vigencia ahora que vivimos una crisis cultural: la cultura de nuestro siglo necesita comunicación e intercambio entre el conocimiento científico y la cultura humanista. Sin embargo, en vez de irrigar como se debiera una competencia general, las disciplinas se encierran en sí mismas; la fragmentación del saber desintegra los problemas fundamentales y globales. Es entonces a esos problemas que se debería dedicar una nueva reforma de la Universidad, que no sería menos histórica que la de Humboldt. Ciertamente, cadauno de esos problemas requiere una competencia poli o trans-disciplinar (hay ejemplos), pero es justamente al afrontar esos problemas que se desarrolla esa competencia y se forman nuevos formadores.
  </dc:description>
  <dc:date>2010</dc:date>
</rdf:Description>

```

Figure 11. Example of enriched RDF.

If the participant OAI-PMH data providers exposed an URI for each author, this phase could be significantly improved. Harnessing initiatives such as the international researcher identifier provided by ORCID is important. It is noteworthy that ORCID and Redalyc are interoperable, given that ORCID currently does not display author data for the Semantic Web like Wikipedia does. All resources harvested by OAI-PMH from RUDAR and Redalyc were subjected to this process of author enrichment, with a total runtime of 36 h 34 min. After the enrichment phase, each RDF obtained was subjected to a verification process using an application in Jena to ensure compliance with the model data described by the “DCFOAF_merged.owl” ontology described above.

The final results of this stage of enrichment and verification show that 13.8% of the resources harvested from repositories had at least one author matching individuals in “PersonData”, for a total of 60,927 authors with matches (Table 3). This result shows that there are resources associated with more than one author. Some items could not be included in the resulting RDF files because they failed the verification with respect to ontology, as some resources omitted essential metadata, such as title or date, or had formation problems.

Table 3. Results of the process of author enrichment.

Repository	Resources Described	Resources with Authors in DBpedia	Authors Localised
Redalyc	380,534	52,251 (13.7%)	57,918
RUDAR	14,885	2550 (17.13%)	3009
Total	395,419	54,801 (13.8%)	60,927

The loading of triplets to TDB was performed in bulk and took 8.97 h in total. The results showed 7,917,081 triplets stored in the triplet store. This knowledge base uses 522.54 GB of disk space. The resulting composition of the OntoOAI base is shown in Table 4.

Table 4. Results of loading to TDB.

Property	Triplets
http://xmlns.com/foaf/0.1/Person	60,354
http://purl.org/dc/elements/1.1/identifier	394,776
http://purl.org/dc/elements/1.1/source	379,966
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	60,354
http://purl.org/dc/elements/1.1/date	394,775
http://xmlns.com/foaf/0.1/name	60,354
http://purl.org/dc/elements/1.1/format	385,975
http://purl.org/dc/elements/1.1/description	361,577
http://purl.org/dc/terms/modified	394,772
http://dbpedia.org/ontology/birthDate	60,354
http://purl.org/dc/elements/1.1/type	405,463
http://purl.org/dc/elements/1.1/title	394,776
http://purl.org/dc/elements/1.1/publisher	382,412
http://purl.org/dc/terms/isPartOf	394,780
http://purl.org/dc/elements/1.1/subject	1,600,540
http://purl.org/dc/elements/1.1/rights	379,966
http://xmlns.com/foaf/0.1/surname	60,354
http://xmlns.com/foaf/0.1/givenName	60,354
http://purl.org/dc/elements/1.1/relation	381,980
http://purl.org/dc/elements/1.1/creator	968,903
http://purl.org/dc/elements/1.1/language	394,650

The context awareness of OntoOAI is based on the profile of a user who describes their academic data according to the OntoOAIStudent ontology. To describe the context of users that was used in the case study, two profiles were defined: the first was identified with “itesm: A01210238” and an identifier consisting of the school initials and student registration number (Figure 12).

```

▼<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dcterms="http://purl.org/dc/terms/" xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:onto="http://dbpedia.org/ontology/"
  xmlns:student="http://www.semanticweb.org/arianna/ontologies/2016/3/OntoOAIEstudiante">
  ▼<rdf:Description rdf:about="itesm:A01210238">
    <student:name rdf:type="http://xmlns.com/foaf/0.1/Person" foaf:name="Rafael Gómez
    García" foaf:givenName="Rafael" foaf:surname="Gómez García"/>
    <student:programa>Sociología</student:programa>
    <student:nivelPrograma>Licenciatura</student:nivelPrograma>
    <student:escuela>ITESM</student:escuela>
    <student:cursoActual>Movimientos Sociales</student:cursoActual>
    <student:cursoActual>Epistemología</student:cursoActual>
    <student:cursoActual>Sociedades modernas</student:cursoActual>
    <student:intereses>Kant</student:intereses>
    <student:intereses>Teoría del conocimiento</student:intereses>
    <student:formatoPresentacion>Artículo</student:formatoPresentacion>
    <student:formatoPresentacion>Video</student:formatoPresentacion>
    <student:idiomaMaterno>es</student:idiomaMaterno>
    <student:idiomaPreferido>en</student:idiomaPreferido>
    <student:datosDemograficos>MX</student:datosDemograficos>
  </rdf:Description>
</rdf:RDF>

```

Figure 12. Profile of user itesm: A01210238.

The second profile (Figure 13) corresponds to a student of a German university identified with “univ: DL9510078”.

```

▼<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dcterms="http://purl.org/dc/terms/" xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:onto="http://dbpedia.org/ontology/"
  xmlns:student="http://www.semanticweb.org/arianna/ontologies/2016/3/OntoOAIEstudiante">
  ▼<rdf:Description rdf:about="univ:DL9510078">
    <student:name rdf:type="http://xmlns.com/foaf/0.1/Person" foaf:name="Margret Fintz"
      foaf:givenName="Margret" foaf:surname="Fintz"/>
    <student:programa>Pedagogy Educational Studies</student:programa>
    <student:nivelPrograma>Master</student:nivelPrograma>
    <student:escuela>Deutsche Universität</student:escuela>
    <student:cursoActual>Humanism and Pedagogy</student:cursoActual>
    <student:intereses>Social apprenticeship</student:intereses>
    <student:intereses>Johann Friedrich Herbart</student:intereses>
    <student:formatoPresentacion>article</student:formatoPresentacion>
    <student:formatoPresentacion>thesis</student:formatoPresentacion>
    <student:idiomaMaterno>de</student:idiomaMaterno>
    <student:idiomaPreferido>en</student:idiomaPreferido>
    <student:idiomaPreferido>es</student:idiomaPreferido>
    <student:datosDemograficos>DE</student:datosDemograficos>
  </rdf:Description>
</rdf:RDF>

```

Figure 13. Profile of user univ: DL9510078.

Once all previous steps were performed and the triplet store was ready for the reasoner, previously defined user profiles were used, and a pair of exercises in selective knowledge discovery were documented.

The experiment shown in this section was performed considering the user with the profile “item:A01210238”. A search for “complexity” was tested by introducing as such the term by the user, which is where OntoOAI began to be tested.

First, it is important to remember that the search is context-sensitive, which means that if a student of computer science or biology performs the same search, the results will be different because “complexity” in computing is approached from a different perspective from the sociological one.

Now, the process performed to obtain this result will be explained. First is a search based on matches that yields results in a scientific paper called “The Emergence of Sense from Non-Sense” by the author Edgar Morin, a French philosopher and sociologist (the latter information was obtained from Wikipedia along with other data about the author) published in the journal *Convergence* in 2007 and whose themes (described as dc:subject) are “sense of the world”, “non-sense”, “Sociology”, and “complexity”.

The diagram shown in Figure 14 shows the reasoning behind the output displayed to the user. As previously explained, the resource “:redalyc.org:oai:10504408”, referring to the article called “The Emergence of Sense from Non-Sense”, is located due to the similarities in the topics it treats, consistent with the theme of “complexity”, and the fit with the user program expressed in the profile, in this case “sociology”. The reasoning follows the syllogism explained below:

- If (the resource “oai:redalyc.org:10504408” is interesting to the user)
 - Then (the author “Edgar Morin” is interesting to the user).
- If (the author “Edgar Morin” is interesting to the user)
 - Then (other works whose author, subject, or title match “Edgar Morin” are interesting to the user).
- If (works Ta are interesting to the user)
 - Then (the set of authors or contributors B of resources Ta is interesting to the user).
- If (the set of authors B of resources Ta is interesting for the user)
 - Then (the set of resources Tb written by authors B is interesting to the user),

where:

r = “oai:redalyc.org:10504408”,
 a = “Edgar Morin”,

$T_a = \{ \text{"oai:redalyc.org:8601934800"}, \text{"oai:redalyc.org:27903809"}, \text{"oai:redalyc.org:49615201"}, \text{"oai:rudaruc.dk:1800/14983"}, \text{"oai:rudaruc.dk:1800/26454"} \},$
 $B = \{ \text{"Michel Olsen"}, \text{"Anders Sondergaard Skovbakke"}, \text{"Matias Find"}, \text{"Mads Christiansen"}, \text{"Allan Dreyer Hansen"} \}$
 $T_b = \{ \text{"oai:rudaruc.dk:1800/12794"}, \text{"oai:rudaruc.dk:1800/24877"}, \text{"oai:rudaruc.dk:1800/16477"}, \text{"oai:rudaruc.dk:1800/25805"}, \text{"oai:rudaruc.dk:1800/25386"}, \text{"oai:rudaruc.dk:1800/23930"}, \text{"oai:rudaruc.dk:1800/15047"}, \text{"oai:rudaruc.dk:1800/7734"}, \text{"oai:rudaruc.dk:1800/18163"}, \text{"oai:rudaruc.dk:1800/18070"}, \text{"oai:rudaruc.dk:1800/18892"}, \text{"oai:rudaruc.dk:1800/13812"}, \text{"oai:rudaruc.dk:1800/24240"}, \text{"oai:rudaruc.dk:1800/25873"}, \text{"oai:rudaruc.dk:1800/15987"}, \text{"oai:rudaruc.dk:1800/4230"}, \text{"oai:rudaruc.dk:1800/4231"}, \text{"oai:rudaruc.dk:1800/24002"}, \text{"oai:rudaruc.dk:1800/6878"}, \text{"oai:rudaruc.dk:1800/24581"}, \text{"oai:rudaruc.dk:1800/17313"}, \text{"oai:rudaruc.dk:1800/3831"}, \text{"oai:rudaruc.dk:1800/3830"}, \text{"oai:rudaruc.dk:1800/3970"}, \text{"oai:rudaruc.dk:1800/6613"}, \text{"oai:rudaruc.dk:1800/23967"}, \text{"oai:rudaruc.dk:1800/14741"}, \text{"oai:rudaruc.dk:1800/23613"}, \text{"oai:rudaruc.dk:1800/14129"}, \text{"oai:rudaruc.dk:1800/23282"}, \text{"oai:rudaruc.dk:1800/17645"}, \text{"oai:rudaruc.dk:1800/19221"}, \text{"oai:rudaruc.dk:1800/16704"}, \text{"oai:rudaruc.dk:1800/23015"} \}$

Thus, the set of resources contained in T_a and T_b comprise the output to the user, i.e., the knowledge found according to the search and context.

Discovered knowledge = $T_a \cup T_b$

For example, only resources discovered from the authors of the thesis "oai:rudaruc.dk:1800/26454" were considered, which totalled 34 after removing duplicates because several resources were co-written by an author of that same group. For display purposes, articles by the author Michel Olsen, author of "oai:rudaruc.dk:1800/14983", were omitted, equivalent to 25 additional resources.

While the exploration of the forms of visualisation of the obtained results is not proposed in this work, it should be noted that given the present associations between resources, it is possible to take advantage of graphs or hierarchical visualizations. Such forms of presenting information could allow the user to explore different levels of the results of the inferences and to discriminate information based on present relationships. An example is shown in Figure 15.

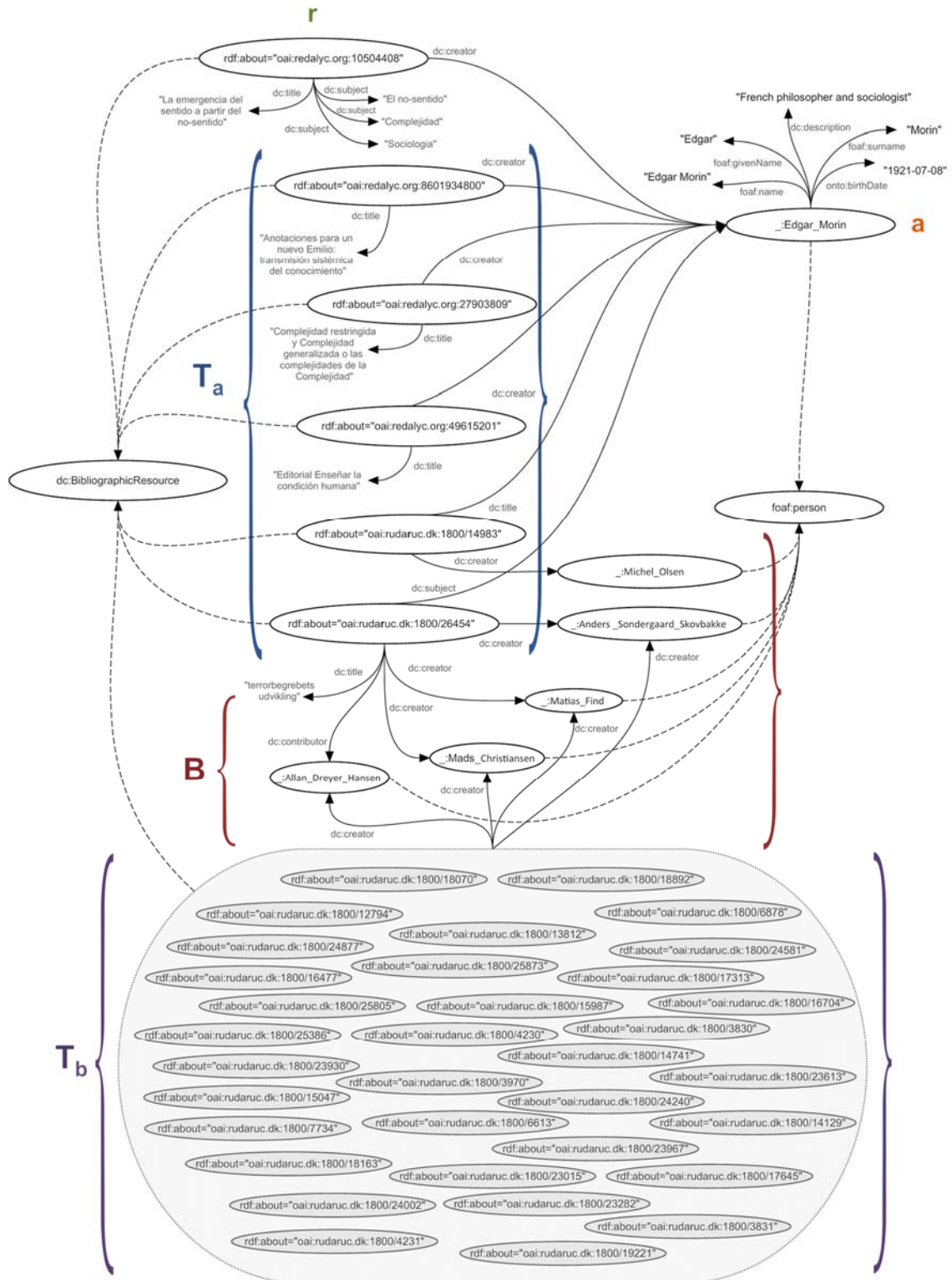


Figure 14. Selective knowledge discovery from the example.

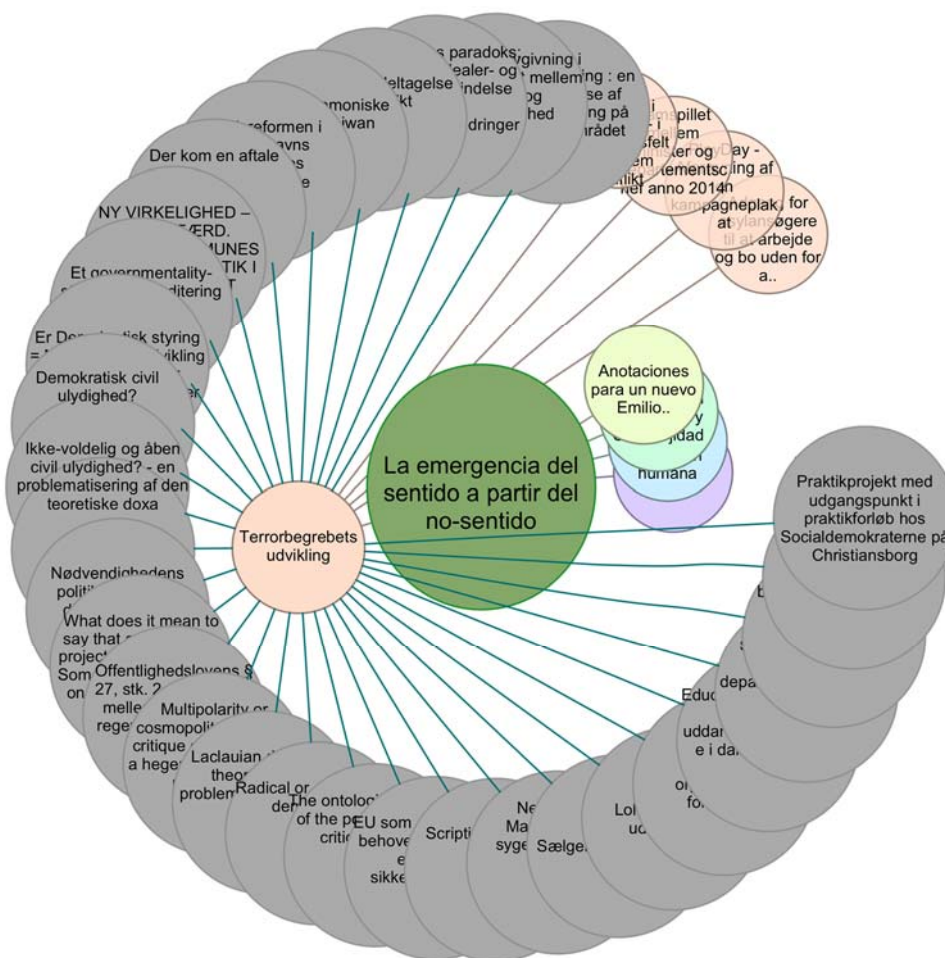


Figure 15. Graph-type visualization of the example results.

4. Conclusions and Future Work

On the path towards Semantic Web, sites are moving at very different rates. There are sites that do not incorporate any form of structure into the data that they present, even among the ones that use the most sophisticated and detailed formats to describe their information.

The protocol for harvesting OAI-PMH metadata is the result of a concern for making interoperable content available under the Open Access initiative. Thousands of repositories, which integrate intellectual, academic, and scientific production, participate in the Open Access movement and have adopted this specification to describe, display, and exchange information. However, the structuring of the OAI-PMH alone is not enough to take advantage of the benefits of the Semantic Web.

Thanks to the Semantic Web, software can process content, reason with it, combine it, and perform deductions logically to solve problems automatically. As Segaran, Evans, and Taylor [21] mention, building models of the world is not something new to system designers, but with semantic systems, these models must be expressed such that systems distributed throughout the Web can be read and the models understood and used with precision.

The main contribution is to propose a semantic approach to selective knowledge discovery, enabling data structured with OAI-PMH to be processed by semantic technologies, which allows the execution of inference mechanisms on the dataset in addition to the use of ontology merging techniques for the verification of knowledge.

Furthermore, this model suggests several lines of future work. OntoOAI could be extended to take advantage of sources of Linked Open Data information as input, in addition to OAI-PMH content,

and to apply knowledge discovery procedures, developing mechanisms for the filtering of academic or scientific information.

In addition, the proposal can be enhanced with the use of controlled vocabularies and/or multilingual ontologies to retrieve information in various languages.

Other inference engines—such as Pellet, Racer, or FaCT—should also be tested for full OWL DL reasoning.

Regarding awareness to context, the type of information stored in the user profile is static, determined during the design time, and cannot be updated or modified at runtime. In the sense that Eyharabide and Amandi (2012) highlight, next-generation user profiles should not just be a passive user data store but also an active component that can learn and update both the information and the type of information stored.

Finally, it is also considered as a future work to add more metadata to the inference engine in order to expand the model information retrieval capabilities that, in this moment, is based just on co-authorship.

Author Contributions: A.B.-G. designed the methodology, developed the software components, performed the experiment and wrote the article. E.A.-L. conceived the database, analyzed the data and interpreted the results.

Funding: This research received no external funding.

Acknowledgments: We like to show our gratitude to all Open Access journals that participate in Redalyc project. Without the invaluable information they publish this research would not have been possible.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Keßler, C.; d'Aquin, M.; Dietze, S. Linked Data for Science and Education. *Semant. Web J.* **2013**, *4*, 1–2.
2. Allemang, D.; Hendler, J. *Semantic Web for the Working Ontologist*, 3rd ed.; Morgan Kaufmann: Waltham, MA, USA, 2011; ISBN 978-0123859655.
3. Chan, L.; Cuplinskas, D.; Eisen, M.; Friend, F.; Genova, Y.; Guédon, J.-C.; Hagemann, M.; Harnad, S.; Johnson, R.; Kupryte, R.; et al. Budapest Open Access Initiative. 2002. Available online: <http://www.soros.org/openaccess/read.shtml> (accessed on 8 June 2018).
4. Max Planck Society. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. 2003. Available online: <http://openaccess.mpg.de/Berlin-Declaration> (accessed on 8 June 2018).
5. Brown, P.; Cabell, D.; Chakravarti, A.; Cohen, B.; Delamothe, T.; Eisen, M.; Grivell, L.; Guédon, J.-C.; Hawley, R.S.; Johnson, R.K.; et al. Declaración de Bethesda sobre Publicación de Acceso Abierto. 2003. Available online: http://ictlogy.net/articles/bethesda_es.html (accessed on 8 June 2018).
6. Albert, P.; Holmes, K.; Börner, K.; Conlon, M. Research Discovery through Linked Open Data. In Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries, Washington, DC, USA, 10–14 June 2012; ACM: New York, NY, USA, 2012; pp. 429–430.
7. Wiederhold, G. Foreword: On the barriers and future of knowledge discovery. In *Advances in Knowledge Discovery and Data Mining*; Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R., Eds.; AAAI Press: Menlo Park, CA, USA, 1996.
8. Frawley, W.J.; Piatetsky-Shapiro, G.; Matheus, C.J. Knowledge Discovery in Databases: An Overview. *AI Mag.* **1992**, *13*, 57–70.
9. IBM. Knowledge Discovery and Data Mining. Available online: http://researcher.watson.ibm.com/researcher/view_group.php?id=144 (accessed on 8 June 2018).
10. Becerril Garcia, A.; Lozano Espinosa, R.; Molina Espinosa, J.M. Semantic Approach to Context-Aware Resource Discovery over Scholarly Content Structured with OAI-PMH. *Comput. Sist.* **2016**, *20*, 127–142. [[CrossRef](#)]
11. Becerril Garcia, A.; Lozano Espinosa, R.; Molina Espinosa, J. Modelo para consultas semánticas sensibles al contexto sobre recursos educativos estructurados con OAI-PMH. In Proceedings of the Encuentro Nacional de Ciencias de la Computación, ENC 2014, Oaxaca, Mexico, 3–5 November 2014; pp. 1–15.

12. Ministerio para la Ciencia e Innovación de España. Informe Modelos de Metadatos para Contenidos Multimedia. Available online: http://omediadis.udl.cat/html/deliverables/215-Modelos_Metadatos_Contentidos_Multimedia/ (accessed on 10 November 2017).
13. Haslhofer, B.; Schandl, B. The OAI2LOD Server: Exposing OAI-PMH Metadata as Linked Data. In Proceedings of the International Workshop on Linked Data on the Web (LDOW2008), Beijing, China, 22 April 2008.
14. Bizer, C.; Cyganiak, R. D2R Server—Publishing Relational Databases on the Semantic Web. 2006. Available online: <http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/Bizer-Cyganiak-D2R-Server-ISWC2006.pdf> (accessed on 8 June 2018).
15. SIMILE. OAI2RDF. 2006. Available online: <http://simile.mit.edu/repository/RDFizers/oai2rdf/> (accessed on 8 June 2018).
16. Ameen, A.; Rahman Khan, K.; Rani, B. Semi-Automatic Merging of Ontologies using Protégé. *Int. J. Comput. Appl.* **2014**, *85*, 35–42.
17. Panagiotopoulos, I.; Kalou, A.; Pierrakeas, C.; Kameas, A. An Ontology-Based Model for Student Representation in Intelligent Tutoring Systems for Distance Learning. In *Artificial Intelligence Applications and Innovations*; Lazaros Iliadis, I.M., Ed.; Springer: Halkidiki, Greece, 2012.
18. Apache Software Foundation. Available online: http://jena.apache.org/about_jena/about.html (accessed on 8 June 2018).
19. Dentler, K.; Cornet, R.; Teije, A.; de Keizer, N. Comparison of Reasoners for large Ontologies in the OWL 2 EL Profile. *Semant. Web* **2011**, *2*, 71–78.
20. Becerril-García, A.; Aguado-López, E.; Rogel-Salazar, R.; Garduño-Oropeza, G.; Zúñiga-Roca, M. De un modelo centrado en la revista a un modelo centrado en entidades: La publicación y producción científica en la nueva plataforma Redalyc.org. *Aula Abierta* **2012**, *40*, 53–64. (In Spanish)
21. Segaran, T.; Evans, C.; Taylor, J. *Programming the Semantic Web*, 1st ed.; O'Reilly: Sebastopol, CA, USA, 2009; p. 302.



© 2018 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/>).