

An ontology-based approach to describe collaborative work by reusing and enriching data from an institutional repository

María-Auxilio Medina N.¹, Delia Arrieta D.², Jorge de la Calleja M.¹,
Laura Zacatzontetl H.¹, Marilú Zacatelco P.¹

¹ Departamento de Posgrado. Universidad Politécnica de Puebla. Tercer Carril del Ejido Serrano S/N. San Mateo Cuanalá. Juan C. Bonilla, Puebla, México. C. P. 72640

² Facultad de Economía, Contaduría y Administración.
Universidad Juárez del Estado de Durango. Fanny Anitúa y Priv. Loza S/N
Col. Los Ángeles Durango, Dgo. México. C.P. 34000
{maria.medina, jorge.delacalleja, marilu.zacatelco}@uppuebla.edu.mx,
{darrietad, laurita_z_h}@hotmail.com

Abstract. Besides tutoring and consultancies, the development of academic and scientific documents in universities evidenced collaborative work. This paper presents an ontology-based approach to describe different modes of collaboration by reusing and enriching data from an institutional repository, from a collection of posters. The approach uses an application ontology that makes explicit the relationships among authors and posters. The paper presents a list of competency questions that are answered in natural language and by the ontology terminology. The proposed approach is of value as this offers machine-readable data to support further analysis and inference mechanisms.

Keywords: Ontologies, semantic web, institutional repositories, document management.

1 Introduction

Besides tutoring and consultancies, the development of academic and scientific documents in universities are evidences of collaborative work that can be used for supporting management decisions. At present, the Universidad Politécnica de Puebla (UPPue) distributes open-access documents such as articles, master's thesis and posters by using the infrastructure of its institutional repository (IR), from now on, UPPue-IR.

Posters are documents written by graduate students of different academic programs where they report partial results of research activities; posters are often presented at symposiums or congresses. UPPue-IR is a documental database that allows users to retrieve validated documents frequently produced between teachers, students or both of them. From a technical point of view, this repository implements the Open Archives Initiative Protocol (OAI-PMH protocol) (Lagoze and Van de Sompel, 2001) to interoperate with the National Repository (RN, 2019); this protocol is also used to
Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)

export descriptive data of documents, commonly referred as metadata. The implementation of this protocol implies that documents are depicted by using the Dublin Core Metadata Element Set as the default metadata standard (DCMI, 2014). The elements of this standard related with collaborative work among authors of posters are creator and contributor, the first one stores the name of a student name, while the second one refers to his/her advisor; if there is a third or fourth author, their names are also stored in multiples instances of the contributor element. Unlike posters are retrieved by search engines, the order in a list of authors for posters or other types of academic documents neither their contribution are taken into account.

This paper presents an ontology-based approach to describe collaboration among authors of posters by reusing and enriching data from the UPPue-IR. The approach uses an application ontology that makes explicit the relationships among students, teachers and posters.

The paper is organized as follows. Section 2 presents user types and their competency questions (CQs). Section 3 describes the main ontology components. Section 4 contains the answer for CQs. Section 5 enumerates implicit information that is derived from the ontology. Finally, we conclude in Section 6 with a summary of the present work along with further research perspectives.

2 User types and their competency questions

According to (Gruber, 1995), an ontology is a “specification of a shared conceptualization”; in computer and information sciences, ontologies are formal definitions of types, properties and relationships between entities that exist in a particular domain of interest (Ecured, 2019). Ontologies are knowledge models composed by instances, concepts, rules and relationships that have a unique representation for a group of people or computers.

Table 1 shows the main user types of the posters’ collection, these users are highly likely to be found in other IR.

Table 1. User types of the posters’ collection.

User types	Description
Advisor	A person who directs the research work of a graduate student that is reported on a poster. The advisor is the second author in the authors’ list.
Manager	The manager of an IR in charged of exporting metadata
Student	The main author of a poster, the first in the authors’ list
Teacher	The third or fourth author of a poster, a person from the academic staff that reviews the content and structure of a poster

The scope of the ontology proposed is determined by the Competency Questions (CQs) of Table 2, more information about CQs can be found in (Noy and Hafner, 1997) and (Bezerra *et al.*, 2013). CQ1 to CQ4 support knowledge acquisition for IRs, CQ5 and CQ6 are related with the IR context while CQ7 and CQ8 have specific information about collaborative work between authors.

Table 2. Competency questions for by user types.

Number of CQ	Description of CQs in natural language
CQ1	What is a poster?
CQ2	What is a poster for?
CQ3	What kind of DC elements are used to describe a poster?
CQ4	Who use a poster?
CQ5	Which are mandatory metadata elements to deposit a poster into the UPPue-IR?
CQ6	How posters are introduced into the ontology?
CQ7	Who form the list of authors of a poster?

3 Main ontology components

The paper proposes an ontology to describe different modes of collaboration among authors of a posters' collection. The metadata for this collection are exported from the UPPue-IR and transform into ontology instances. Note that any other IR that implements the OAI-PMH protocol has also their own mechanisms to export metadata. The ontology is composed of a hierarchy of classes, a set of data properties (data property axioms), object properties (object property axioms) and instances (also known as individuals), this is edited by using the Protégé software tool version 5.2 (Musen, 2015). The following sections describe these components.

3.1 Main classes

The main class of the proposed ontology is called University, the purpose is to have a general concept that refers to the context of use for the proposed ontology. Table 3 shows the names and descriptions of three classes at the second level of the ontology, remaining concepts are obtained by generalization and specialization and distribute between the third or fourth level in the class hierarchy. By convention, class names starts with a capital letter.

Table 3. Classes at the second level of the proposed ontology.

Class	Description
Department	This class refers to the adscription of a student or teacher.
Poster	A document written by a student where he/she reports partial results of his/her research activities.
User	The User class integrates user types, (advisor, manager, student and teacher. An advisor is a type of teacher.

3.2 Data properties

The classes at the second level of the hierarchy are described by using data properties. For example, the name, last name or gender of a User, the title and date of a Poster are modeled as data properties. All interoperability aspects that correspond to the implementation of OAI-PMH protocol and the DC elements can be represented as data properties that link posters and users with data values from an XML Schema Datatype or an RDF literal (RDF, 2001).

3.3 Object properties

Collaborative work between authors to produce posters are modeled in the ontology as object properties, they are associated with domain and range restrictions as is illustrated in Table 4.

Table 4. Object properties for modeling collaborative work to produce posters.

Object property	Domain	Range
assignedTo	Teacher	Department
hasTeacher	Department	Teacher
wasProducedIn	Poster	Department
hasPoster	Department	Poster
hasStudent	Department	Student
studies	Student	Department
isAdvisorOf	Advisor	Student
isFirstAuthor	Student	Poster
isSecondAuthorOf	Advisor	Poster
isThirdAuthorOf	Teacher	Poster
isFourthAuthorOf	Teacher	Poster
isManagedBy	Poster	Manager

Table 5 shows the facets for the object properties, the notation is as follows: functional (F), inverse functional (IF), asymmetric (AS) and irreflexive (I). Object properties of Table 4, the facets in Table 5 and ontology instances form the ABox for the ontology, reasoners use this box to maintain logical consistency and to infer new knowledge. It is worth to mention that any of the object properties is considered symmetrical, transitive or reflexive.

Table 5. Facets for object properties.

Object properties	Facets
assignedTo	F, AS, IR
hasTeacher	AS, IR
wasProducedIn	F, AS, IR
hasPoster	AS, IR
hasStudent	AS, IR
studies	F, AS, IR
isAdvisorOf	AS, IR
isFirstAuthor	F, AS, IR
isSecondAuthorOf	F, AS, IR
isThirdAuthorOf	F, AS, IR
isFourthAuthorOf	F, AS, IR
isManagedBy	AS, IR

3.4 Ontology instances

Posters and user types are modeled as ontology instances. A semi-automatic process has been designed in order to transform metadata from UPPue-IR into ontology instances. As a way of illustration, Figure 1 shows how the 133 posters that form the posters' collection are distributed by year.

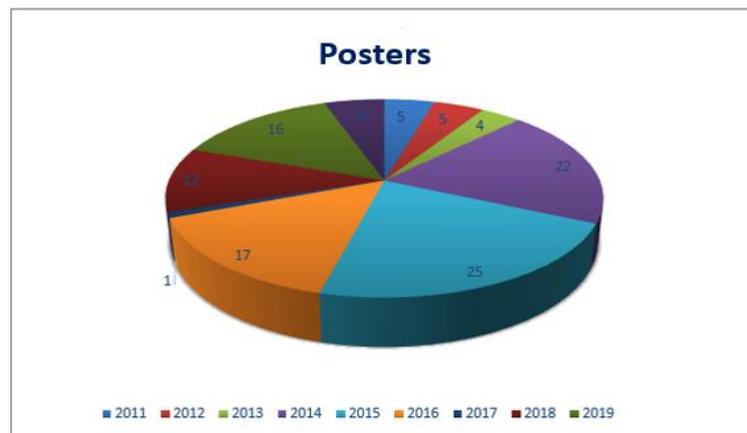


Fig. 1. Distribution of posters by year

Figure 2 shows information for a user in the Spanish version of the ontology. The translation of the Spanish terms is as follows:

- *apellidoMaterno*, second last name
- *nombreDePila*, name
- *Autor*, Author, a subclass of the User class
- *esAutorDe*, isAuthorOf
- *cartel1*, poster1
- *genero*, gener
- *apellidoPaterno*, first last name

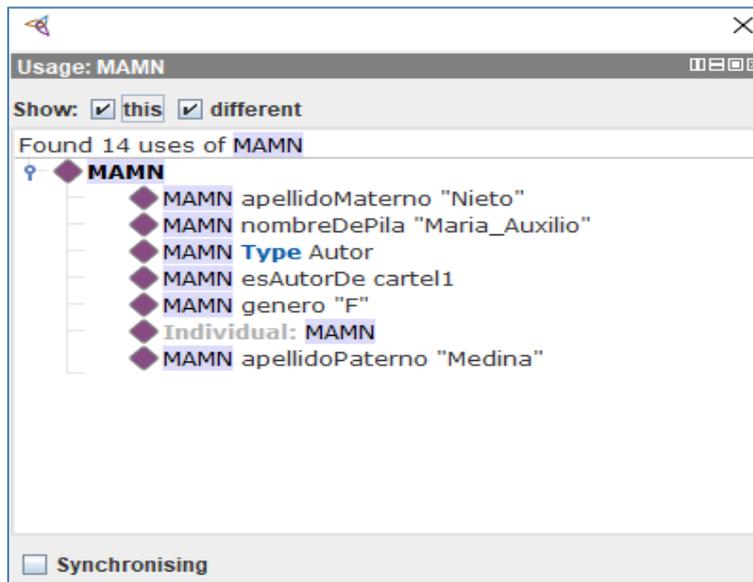


Fig. 2. Information about an ontology instance of the User class.

Figure 3 shows the information of usage of two different users. It is worth to notice that the role of these users is included in the ontology, (“tieneSinodal” is equivalent to “isThirdAuthorOf” and “ProfesorDeTiempoCompleto” is the Spanish term used for the “FullTimeTeacher” class).



Fig. 3. Information about collaborative work of two users.

Figure 4 shows the ontology metrics for the posters' collection. Note that the number of axioms is 2985 and that there are 396 ontology instances (individual account).

Ontology metrics:

Metrics	
Axiom	2985
Logical axiom count	2278
Class count	106
Object property count	23
Data property count	98
Individual count	396
DL expressivity	ALCROIF(D)

Class axioms	
SubClassOf axioms count	103
EquivalentClasses axioms count	5
DisjointClasses axioms count	9
GCI count	0
Hidden GCI Count	9

Fig. 4. Metrics of the ontology for posters.

4 Formal answers to competency questions

CQs are used as guidelines for ontology evaluation. This section presents the answers to CQs in natural language and using formal concepts. An excerpt of the usage information of the ontology elements are described in this section as formal answer.

CQ1: What is a poster?. A poster is a document written by a graduate student where he/she report partial results of his/her research activities.

Formal answer:

1. Annotation property: `rdf:isDefinedBy` for Poster
2. Data type property: `posterData` for Poster
3. Object property: `wasProducedIn`, `isManagedBy`, `isFirstAuthorOf`, `isSecondAuthorOf`, `isThirdAuthorFor`, `isFourthAuthorOf`

CQ2: What is a poster for?. A poster is a document to report advances or partial results of reserarch activities.

Formal answer:

1. Class: Poster
2. Poster SubClassOf University
3. Object properties: `wasProducedIn`, `hasPoster`

CQ3: What kind of DC elements are used to describe a poster?. Title, date, year, subject (for the department) and a list of authors (creator and contributor elements).

Formal answer:

1. Class: Poster
2. Poster SubclassOf University
3. Data property: `title`, (functional)
4. Data property: `year`, (functional)
5. Data property: `subject`, (functional)
6. Date property: `date`, (functional)

CQ4: Who use a poster?. UPPue-IR user types are advisor, manager, student and teacher.

Formal answer:

1. Class: User
2. (Advisor, Manager, Student, Teacher) SubClassOf User
3. Annotation property: `rdf:isDefinedBy` for Advisor, Manager, Student, Teacher

CQ5: Which are mandatory metadata to deposit a poster into the UPPUE-IR?.

Formal answer:

1. Data property: `title`, string or RDF literal
2. Data property: `year`, integer

3. Data property: subject, string or RDF literal
4. Date property: date, date

CQ6: How posters are introduced into the ontology?. Posters are introduced into the ontology as instances, the information about collaborative work of authors is represented in object properties.

Formal answer:

1. Class: Poster
2. Poster SubClassOf University
3. Object properties: see Table 4

CQ7: Who forms the list of authors in a poster?. A graduate student (the first author), an advisor (the second author) and two teachers (the third and fourth author).

Formal answer:

1. Object properties isFirstAuthorOf, isSecondAuthorOf, isThirdAuthorOf, isFourthAuthorOf.
2. isFirstAuthor, domain (Student)
3. isSecondAuthor, domain (Advisor)
4. isThirdAuthor, domain (Teacher)
5. isFourthAuthor, domain (Teacher)

In summary, although the ontology is simple in terms that this represents the addition of semantic information to a particular collection of data from an IR, this is able to represent CQs and their answers using its own terminology. All the inconsistencies were corrected before release. Hermit and Pellet reasoners were used for validation of logical consistency. The ontology can be exported to different semantic web languages such as RDF (RDF, 2001) or the Ontology Web Language (OWL 2004).

5 Implicit knowledge derived from the ontology

The formal features of the ontology enables to extract implicit knowledge as the following:

- If the second author of a poster is a teacher, then he/she is considered an advisor
- If a student is the first author of a poster, then he/she is a graduate student
- If a poster only has two authors, the first one is a graduate student and the second one his/her advisor
- A department has many teachers but a teacher is assigned only to a department
- The Poster and User are disjoint classes
- A user can not be a Student and a Teacher at the same time
- If a teacher is an advisor, that means that at least his/her name appears in the second place of an authors' list

The establishment of axioms, cardinality, domain and range restrictions as well as the definition of object properties, enables the formal representation of knowledge useful to discover possible data inconsistencies. For example, cardinality restrictions can be inserted into the ontology in order to establish a minimum, exactly or maximum number of authors for each poster. Ontologies as the described in this paper can be used to represent collaborative work of other types of documents according to the interests of potential users.

6 Conclusions

This paper presented an ontology-based approach to describe collaborative work by reusing and enriching data from an institutional repository. Ontology instances were obtained by exporting metadata of a posters' collection. The approach uses the ontology to formally represent relationships among users and posters.

The paper used a list of CQs that are answered by the ontology terminology in natural language and formal answers. The natural language answers are stored as definitions in the RDF language, while the formal answers are extracted from the usage dialogs from the Protégé ontology editor. Ontology information is used by reasoners to infer new knowledge as well as to discover possible data inconsistencies, the last feature add value to data from IRs.

The ontology itself and their instances form a machine-readable dataset that can be explored by semantic technologies. As future work, we plan to work in the design of an ontology assessment process to get feedback from the constructed ontologies.

References

1. Lagoze C., Van de Sompel H. The open archives initiative: building a low-barrier interoperability framework. In Proceedings of the first ACM/IEEE-CS Joint Conference on Digital Libraries JCDL'01. pp. 54-62. ISBN 1-58113-345-6. DOI:10.1145/379437. (2001).
2. National repository. Repositorio Nacional. Gobierno de México. Consejo Nacional de Ciencia y Tecnología (CONACYT). Retrieved from: <https://www.repositorionacionalcti.mx>. (2019).
3. DCMI Metadata Terms. Dublin Core Metadata Initiative. Retrieve from: <http://www.dublincore.org/specifications/dublin-core/dcmi-terms/>. (2014).
4. Gruber, T. R. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, Vol. 43 No. 4-5. pp. 907-928. (1995)
5. Ecured. Ontología. Retrieve from: <https://www.ecured.cu/Ontología>. (2019).
6. Noy, N. F., Hafner, C. D. The state of the art in ontology design: a survey and comparative review. In *AI Magazine*. Vol. 18. No. 1, pp. 53-74. (1997).
7. Bezerra, C., Freitas, F., Santana, F. Evaluating ontologies with competency questions. In Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), IEEE Computer Society Washington, DC, USA. Vol. 03. No. 1. pp. 284-285. ISBN: 978-0-7695-5145-6. DOI: 10.1109/WI-IAT.2013.199. (2013).

8. RDF1.1 XML syntax. Retrieved from: <http://www.w3.org/TR/rdf-syntax-grammar/> (2001).
9. Musen, M. A. The protégé project: a look back and a look forward. *AI Matters*. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, Vol. 1 No. 4, pp.4-12. DOI: 10.1145/2557001.25757003. (2015).
10. OWL Web Ontology Language Overview. Retrieved from: <http://www.w3.org/TR/owl-semantic/>. (2004).