OntoBib: an Ontology-Based System for the Management of a bibliography

Sahar Al-Sudani, Rim Al Hulou, Amedeo Napoli, Emmanuel Nauer

OntoBib: an Ontology-Based System for the Management of a bibliography

Sahar Al-Sudani* and **Rim Alhulou**** and **Amedeo Napoli**** and **Emmanuel Nauer****

Abstract. In this paper, we present our work aimed at improving retrieval and navigation services on bibliographic data. In this system, knowledge about the domain of documents (publications) is stored in an ontology, a formal model of conceptual knowledge. The ontology is employed to represent concepts and their relationships in order to carry out reasoning about documents. Reasoning helps researchers by providing more efficient document retrieval and navigation.

1 Introduction

It is becoming standard practice for researchers to publish their documents on the Internet (or intranets), via personal, institutional and discipline-specific servers. Today, we find on the web a large number of digital libraries allowing instant access to documents. Such libraries permit accessing documents by searching on metadata such as keyword, author, title, conference, series or journal name. They give access to full text or only to bibliographic entry of documents. Some digital libraries are discipline-specific like, for example, the Digital Bibliography and Library Project[2] (DBLP), a computer science bibliography database. Others are multi-disciplinary like for example, the full text archives of scholarly society serials publications[3].

Currently, digital libraries propose mainly search services based on metadata. In this case, the possible information retrieval consists generally in returning publications satisfying a set of properties represented by these metadata. However, it would be very useful to be able to query a scientific publications database with more intelligent queries, like *"return all the publications about a research topic"* (needing the definition of the links between the research topics), *"is a given publication a high level publication ?"* (needing the definition of high level publication), *"who are the senior researchers working on a given topic ?"* (needing the definition of senior researcher), etc. Answering these queries requires knowledge about the domain of publications and the domain of researchers as well, and needs to use relationships between the concepts within this knowledge (researchers, publications, conferences, research topics, keywords, etc.).

Such knowledge can be represented using ontologies, an organised set of concepts and relations. Ontologies are used in different domains for many types of applications, e.g. in medicine and bioinformatics. Currently, ontologies – associated to reasoning systems– are considered as the engine of the semantic web [21]. Domain knowledge is represented in ontologies; web pages are annotated using the concepts and relations of the corresponding ontology, and reasoning systems are in charge of using these semantic annotations to reason and answer queries. In the same spirit, ontologies can be used to represent knowledge about items of bibliographic data domain (authors, publications, keywords...), and reasoning systems can be used to carry out reasoning in order to answer complex queries or to enhance search services by using, for example, the available keywords which are more specific than or related to the submitted keywords.

In this paper we present an ontology-based system, called OntoBib, aimed at managing a bibliography in an "intelligent way". We show how knowledge on the domain of publications is used to improve information access. Reasoning is used by the system for proposing to the researchers a guided help, more efficient information retrieval and navigation activities.

The paper is organised as follows. In section 2, we present the general context of our work. The knowledge about publications is described in section 3. Section 4 presents the OntoBib system. Section 5 concludes the paper and presents the future work.

2 General Context

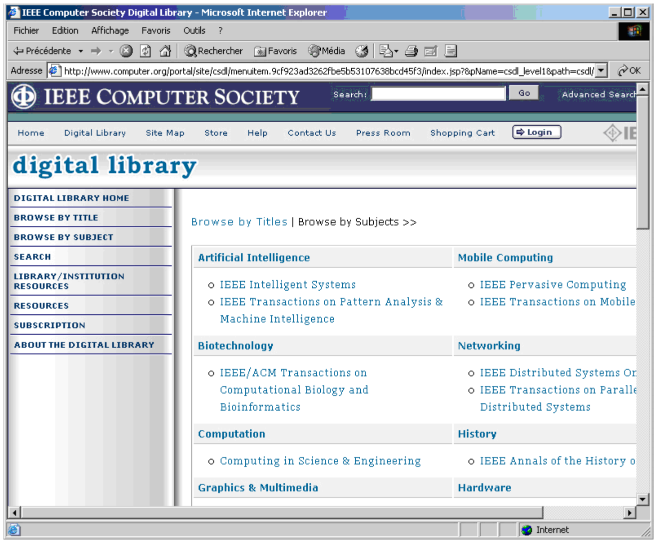

Figure 1. IEEE computer society user interface for navigation.

Most of existing digital libraries offer search as well as navigation services. Their user interfaces allows to formulate

1 * Informatics Institute for Post Graduate Studies, Baghdad, Iraq
Email: Sah_1970@yahoo.com
** LORIA – UMR 7503, B.P. 239, 54506 Vandœuvre-lès-Nancy CEDEX, France.
Email: {ahlulou,nauer,napoli}@loria.fr

2 http://www.informatik.uni-trier.de/~ley/db/

3 http://www.lib.uwaterloo.ca/society/full-text_soc.html

queries in order to search documents. Queries can contain one or more words to be searched in the full text document, or in one of the metadata fields. Users can also formulate multi-criteria queries, by defining more than one search condition concerning metadata (e.g. documents about "web" and/or/not authored by "Berners Lee") .

While search services are mainly based on metadata, navigation services supported by digital libraries are mainly based on catalogues of journals or series names, or on subject (topic) directories. A typical example of navigation services interface is, for example, the one provided by the IEEE computer society library[4], as given in figure 1.

2.1 The framework of accessing a bibliography

Bibliographic data are mainly available from local BibTex files or from bibliography servers like the DBLP database or CiteSeer[5]. In the real world, these data items (publications, authors, keywords, etc.) have their own properties and interrelations. It can then be helpful to use these relations and properties for search and navigation services.

Our work is motivated by the need to improve the existing search and navigation services using the knowledge available about bibliographic data. Some usage scenarios based on such relations and properties are as follows :

- Taking into account that "semantic annotation" is a subtopic of "semantic web", it will be useful to return documents about "semantic web" as well as "semantic annotation" to a beginner researcher who is searching for documents about "semantic web". This is not a standard functionality of available web search services.
- It will be interesting for a researcher to get information about people working in a specific topic, like for example information about senior researchers.
- Providing students with assisted and interactive search service. For example, after answering a user who is searching for publications about a given topic, say "OWL", it will be useful to ask him whether he is interested in publications on topics which are more specific than or related to the former.

Such search and navigation services have to rely on knowledge about the domain of documents, i.e knowledge about publications topics, authors and their relationships. This knowledge must be stored within a formal model associated to a reasoning system using knowledge for answering queries, and for suggesting further retrieval or navigation possibilities.

The design of an ontology-based system, able to take advantage of knowledge for answering queries, is mainly inspired by research works related to the semantic web field.

2.2 Semantic web

The semantic web was thought up by Tim Berners-Lee, creator of the web. There is a dedicated team of people at the World Wide Web consortium[6] (W3C) working to improve, extend and standardise the system, and many languages, publications, tools have already been developed. The idea of the semantic web is to annotate documents with *semantic annotations*, annotations that are not interpreted for display but rather as an expression of document content. This is often described as a *web for machines* as opposed to a *web for humans*. Such machine understandable web vision needs to establish a join terminology between members of a community of interest. These members can be human or automated agents. Ontologies are a solution proposed by the semantic web community to represent such terminology of a certain domain [6].

2.3 Ontologies

There are many definitions of what an ontology exactly is. A well-known proposal in the computer science domain is one proposed by Gruber [10], stating that *an ontology is an explicit specification of a conceptualisation*. For us, the meaning of *conceptualisation* refers to an abstract model materialised by a set of definitions of concepts and concept properties hierarchically organised. The *explicit specification* means that the model has to be implemented in a knowledge representation language (having a syntax and an associated semantics). Such a language allows the model to be exploited both by machines or by human beings [9].

An ontology includes machine-interpretable definitions of concepts in the domain and relations existing among them. So, ontologies are useful to share and reuse a common knowledge of the domain by people as well as applications. Ontologies can have different levels of expressiveness. The more detailed is the knowledge captured in the ontology the more expressive it is. One can have only a list of terms in controlled vocabulary, relations between terms (such as synonyms) in thesaurus or, in a very expressive ontology languages, a hierarchy of concepts and properties with inheritance of properties by subconcepts and a first order logic constraints between terms.

There are two kinds of advantages resulting of the existence of an ontology [21, 23]. From a *"communication and interoperability"* point of view, an ontology facilitates the information or the knowledge exchange (coming from heterogeneous data sources) between human and/or machines, because the ontology represents elements in a non-ambiguous way and allows to check for consistency. From a *"specification, integration and re-use"* point of view, an ontology can be helpful to specify problems in the domain being represented and can play the role of a bridge in order to facilitate the integration of an application into a platform. Finally, even if the construction of an ontology is linked to applications, as long as the structure of the ontology is not challenged, the ontology and the applications can evolve separately. In practice, having an ontology allows the resolution of a number of problems such as guiding a user for accessing data [8, 20], organising and/or comparing documents according to their content [1].

2.4 Related works

The practice of adding semantics to web resources using ontologies to build intelligent applications has been the focus of

[4] http://www.computer.org/portal/site/ieeecs/index.jsp
[5] http://citeseer.org/
[6] http://www.w3.org/

several authors and systems. SHOE [12] allows researchers to annotate their web resources with metadata, in order to build a distributed knowledge base. Like OntoBib, SHOE uses an ontology to declare the desired characteristics and relationships of a web resource and makes its properties explicit. The real potential behind SHOE is the ability to draw on the ontology to infer supplementary knowledge not directly stated within the facts describing the web resource.

Another system that is similar to OntoBib is ESKIMO [15] (The E-Scholar Knowledge Inference Model system) that demonstrates how ontologically modeled data can be used to infer new facts and resolve links based on analytical queries. ESKIMO uses XML (much less expressive language) for representing the ontology of publications and an embedded XML reasoner to infer new facts, while OntoBib system, presented in this paper, exploits description logics as an effective and expressive representation language for ontologies, supported by available reasoning tools like RACER. This separation between the representation of concepts through its ontology and the reasoning process through its reasoning capability make our prototype flexible and adaptable to new concepts. The publications ontology used in OntoBib is similar to the one used in ESKIMO only in the representation of the general relationships that are exists between the resources while an additional concepts definitions and relationships are introduced. Further the research topics are well-organized in a sub ontology of keywords that are related together to enhance deduction, and this sub ontology cover limited scientific community.

UNTANGLE [14], is another system that uses description logic for representing bibliographic information and help in information classification and retrieval. The Untangle project began as a exploratory research in using description logics for digital libraries, a web interface was then added and first web-based description logic system was born.

Another early project using DLs for the web was the FINDUR system [17], whose basic function was query expansion, i.e. is takes synonyms or hyponyms (more specific terms) and including them in input terms, thereby expanding the query, this function is already available in OntoBib system.

3 Representation of bibliographic data using an ontology

Digital libraries of publications generally manage a set of bibliographic entries (linked sometimes to their full text documents). A bibliographic entry describes a publication by a set of attribute/value pairs, like author, title, publisher, journal, keywords, etc. Queries on bibliographic data are formulated using these attributes. However, no further information, drawn by inference on the available knowledge, can be returned. This is because neither the semantics of the attributes nor relations existing between them are taken into account. Using an ontology representing the knowledge related to bibliographic data is a key feature in building a system being able to carry out inference (to reason) about such data. We refer to this ontology as $\mathcal{O}_{\mathcal{BIB}}$ in the following.

3.1 Ontology formalism

$\mathcal{O}_{\mathcal{BIB}}$ is coded in OWL-DL –an ontology representation language– based on the formalism of description logics [4]. This formalism involves three kinds of entities:

- *concepts* (or *classes*) represent set of individuals sharing a set of properties. The set of individuals is the *extension* of the concept. For example, the `InternationalConference` concept represents the notion of an international conference, with a location, a date, a organiser, a program chair, etc.
- *individuals* (or *instances* of concepts). Individual international conferences, e.g. `NLDB'06` and `ISWC'05`, may be represented as instances of the `InternationalConference` concept.
- *roles* (or *properties*) represent binary relations between concepts. For example, the role `hasPublisher` between the concepts `Journal` and `Publisher` states that a journal (instance of the concept `Journal`) is related to a publisher (instance of the concept `Publisher`). Each role has a *domain* (where the property is declared) and a *range* (where the relation is established). In the former example, the concept `Journal` (respectively `Publisher`) is the domain (respectively range) of the role `hasPublisher`.

Concepts and roles are built using a set of constructors to which a semantics is associated. Concepts (and possibly roles) are organised in a hierarchy by a *subsumption* (or *specialisation*) relation, noted $\sqsubseteq$, where $D \sqsubseteq C$ can be read like "C subsumes D" or "D is subsumed by C". A concept C subsumes a concept D if and only if C is more general than D (for every interpretation). All classes and properties are subsumed by $\top$, the root concept of the hierarchy.

Concepts can be either *primitive* or *defined*:

- primitive concepts are described by a set of *necessary* conditions. For example, `Conference` introduced by `Conference` $\sqsubseteq$ `Event` is a primitive concept.
- concepts defined by a set of *necessary* and *sufficient* conditions are called *defined* concepts. Whenever X is an instance of a defined concept C, X satisfies the properties attached to C (necessary conditions). Reciprocally, whenever X satisfies all properties attached to the concept C, then X is declared as instance of C (sufficient conditions). Defined concepts are introduced by an equivalence noted $\equiv$, where $C \equiv D$ means $C \sqsubseteq D$ and $D \sqsubseteq C$. For example, the defined concept `ConferenceProceedings` is introduced by `ConferenceProceedings` $\equiv$ `PublicationMedia` $\sqcap$ ($\exists$ `associatedEvent Conference`). If `X` is a publication media and is associated to a conference, then `X` is a conference proceedings ; conversely, a conference proceedings is a publication media which is associated to at least a conference.

Definitions of concepts use mainly three constructors:

- the *conjunction of concepts*, denoted $\sqcap$,
- the *disjunction of concepts*, denoted $\sqcup$,
- the *cardinality* which gives a minimal and maximal number of values associated to a role. The cardinality is written ($\leq$ `integer role`), ($\geq$ `integer role`) or (= `integer role`), respectively for maximal, minimal or exact cardinality. For example, ($\geq$`1 associatedEvent`) describes a

concept which is at least associated to one event by the `associatedEvent` role (the existential constructor, denoted $\exists$, is equivalent to a cardinality greater than or equal to 1, noted $\geq$1).

3.2 Ontology of publications

The $\mathcal{O}_{\mathcal{BIB}}$ ontology is used for expressing the contents of bibliographic entries. $\mathcal{O}_{\mathcal{BIB}}$ enables conceptual linking between the contents of publications, and semantic querying. Currently, there is no available ontology fully adapted to our needs, so we have to effectively construct $\mathcal{O}_{\mathcal{BIB}}$. $\mathcal{O}_{\mathcal{BIB}}$ is both a domain and an application ontology that is constructed from knowledge and experience of the research community. The main concepts and relationships of $\mathcal{O}_{\mathcal{BIB}}$ are shown in figure 2. Domain knowledge concerns `ResearchTopic` concepts, that are organised in a hierarchy. Application knowledge concerns a general representation of `Publication`, and links between a publication and other concepts like `Author`, `PublicationMedia`, `ResearchTopic`.

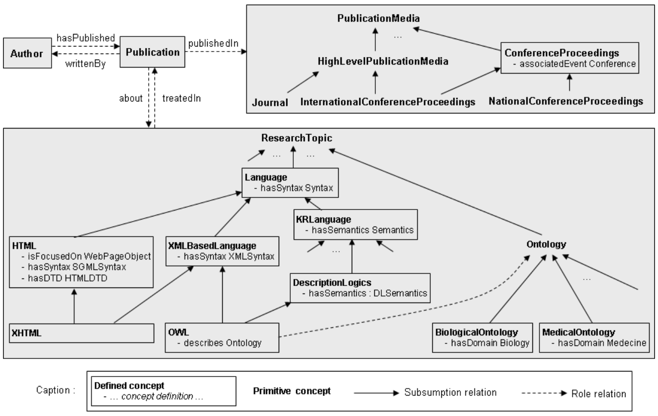

Figure 2. Ontology of publications

In order to establish scientific relationships between the publications on the base of their keywords, we introduce a specific (sub-)ontology of concepts for representing a large number of the keywords used in the publications. Each publication keyword is represented by a (defined or primitive) concept in the ontology; an ontology concept X represents the class of all keywords representing the concept X. So, the names of instances of a concept are the different keywords used in the publications for expressing the concept. For example, *OWL* and *Ontology Web Language* are two names of instances of the `OWL` concept. The concepts are automatically organised in a hierarchy according to the properties that they share (this will be detailed in the next section). The primitive concept `ResearchTopic` represents the super concept of all keyword concepts used in the publications (hereafter, we refer to keyword as a research topic). Section 4 shows how this hierarchy is exploited for improving the access to bibliographic information.

3.3 Ontology building

There is no unified method to build an ontology. However, the iterative process presented in figure 3, introduced in [21] [24], gives the majors steps required by an ontology construction.

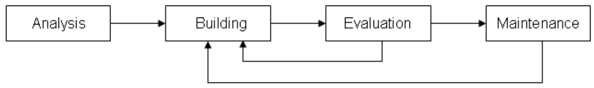

Figure 3. Steps required by an ontology construction.

The analysis step requires to define what the ontology conceptualises and in which use cases it will be used. The building step concerns the creation of the concepts and roles hierarchies, and the definition of the concepts. The objective of the evaluation step is to check the ontology consistency and the use cases performance in order to correct or adjust the necessary elements (by returning to the building step). The maintenance step concerns the effective use of the ontology: returning to the building step allows to take into account the necessary evolutions.

Semi-automatic methods, based on learning process can be used to build an ontology. Some of these methods use a classification process [2] or an association rule extraction process [19]. Other methods use syntactical analysis of textual data, followed by a distributional analysis [7, 3, 22].

As example of application, we choose to build $\mathcal{O}_{\mathcal{BIB}}$ on the base of the bibliographic data of our team, called ORPAILLEUR[7]. The main theme of these publications is *Knowledge management*, *Knowledge extraction* and *Semantic web*. The bibliographic data are used (1) to create the domain knowledge about research topics and (2) to instantiate the concepts of the ontology. As these textual data do not describe each research topic, but only mention it through the presence of a keyword, we could not use semi-automatic approaches. So, we choose to build $\mathcal{O}_{\mathcal{BIB}}$ manually, following the principles given in [18, 2], the main problem being the identification of the properties that have to be used for the definition of the domain concepts.

```
@INPROCEEDINGS{daquin05b,
       TEAM = {ORPAILLEUR},
       AUTHOR  = {d'Aquin, Mathieu and Lieber, Jean and Napoli, Amedeo},
       TITLE = {Decentralized Case-Based Reasoning for the semantic Web},
       BOOKTITLE = {{4th International Semantic Web Conference - ISWC 2005, Galway, Irlande}},
       YEAR ={ 2005},
       EDITOR = {Y. Gil et al.},
       VOLUME = {3729},
       SERIES = {Lecture Notes in Computer Science},
       PAGES = {142--155},
       PUBLISHER = {Springer},
       KEYWORDS  = {case-based reasoning, owl, c-owl, viewpoints, oncology},
       ABSTRACT  = {Decentralized case-based reasoning (DzCBR) is ... }
}
```

Figure 4. Example of entry in a BibTex file

Each bibliographic entry is translated into several instances of concepts in the ontology. For example, the BibTex entry given in figure 4 is translated into one instance of the concept `Publication` (because it describes a publication), three instances of the concept `Author` (one instance for each author of the publication), one instance of the concept `InternationalConferenceProceedings` (because it has

[7] http://www.loria.fr/equipes/orpailleur/Cadre.php?choix=Publications

been published in the proceedings of `ISWC 2005`, an international conference) and five instances of `ResearchTopic` concepts (one instance for each keyword of the publication). The `ResearchTopic` concepts to which the five previous concepts are attached need also to be created in the `ResearchTopic` hierarchy if they are not yet existing.

Creating the domain knowledge about research topics involves the creation of a concept to which keyword used in the publications are attached. Creating a defined concept implies to identify the set of roles (and so, the set of properties) of this concept. For example, we define the `OWL` concept as having a specific semantics, a specific syntax and that it describes on an ontology, with the three properties: $(\exists\ \texttt{hasSemantics DLSemantics})$, $(\exists\ \texttt{hasSyntax XMLSyntax})$ and $(\exists\ \texttt{describes Ontology})$.

The hierarchy of concepts is based on the subsumption relation. For defined concepts, the hierarchy is derived from the definitions. In this way: $\texttt{OWL} \sqsubseteq \texttt{DescriptionLogics}$ because

$$\texttt{DescriptionLogics} \equiv (\exists\ \texttt{hasSemantics DLSemantics})$$

and

$$\begin{aligned} \texttt{OWL} \equiv\ & (\exists\ \texttt{hasSemantics DLSemantics}) \\ \sqcap\ & (\exists\ \texttt{hasSyntax XMLSyntax}) \\ \sqcap\ & (\exists\ \texttt{describes Ontology}). \end{aligned}$$

For primitive concepts, the hierarchy is given by the expert working on the construction of the domain ontology. For example, the expert has stated that $\texttt{NationalConferenceProceedings} \sqsubseteq \texttt{ConferenceProceedings}$.

The readability of the ontology can be improved by creating abstract concepts, subsuming a set of more specific concepts. For example: $\texttt{XMLBasedLanguage} \equiv (\exists\ \texttt{hasSyntax XMLSyntax})$ subsumes all the concepts which are based on the XML syntax (the concepts of specific XML based languages have all an XML syntax).

$\mathcal{O}_{\mathcal{BIB}}$ has been implemented in OWL using the PROTÉGÉ[8] ontology editor [13]. The RACER system[9] has been used as a subsumption and classification engine [11]. $\mathcal{O}_{\mathcal{BIB}}$ is exploited within our ontology-based bibliographic data system, called OntoBib, in order to improve the access to bibliographic information.

4 The OntoBib system

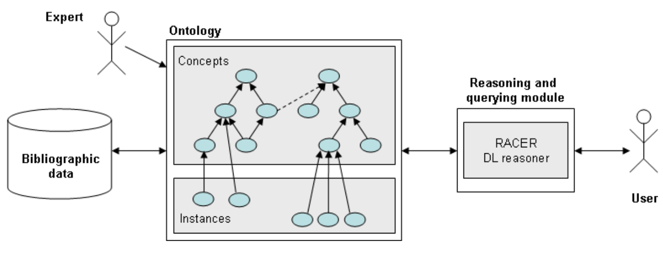

Figure 5. General structure of the OntoBib system

[8] http://protege.stanford.edu
[9] http://www.racer-systems.com/

This section describes the general structure of the OntoBib system. As shown in figure 5, the core of the system is the ontology model. Bibliographic data are used to build the ontology and bibliographic item (publication, author, keyword, etc.) are stored as individuals of this ontology.

Another module in the system tightly related to the ontology is the *Reasoning and querying module*. The main function of the reasoning module is to infer new facts from the existing ones using the description logic reasoner [11]. The reasoner inference services are used to answer queries. For example, the *instance checking* inference service is used to check whether a given individual is an instance of a specified concept. So, retrieving instances related to a concept will answer queries about this concept. We explain now, using examples, some functionalities of these modules.

4.1 Semantic Querying

One of the features implemented in the system relies on the inference of new facts from the existing ones. Based on the logical semantics of the representation language, different kinds of queries are defined as inference problems (hence, answering a query is called providing an inference service). The instance retrieval process is used to answer questions such as: *which are the high level publications*. OntoBib answers this question using the existing concept `HighLevelPublication` defined in the ontology as: $\texttt{HighLevelPublication} \equiv \texttt{Publication} \sqcap ((\exists\ \texttt{publishedIn Journal}) \sqcup (\exists\ \texttt{publishedIn InternationalConferenceProceedings}))$. This concept describes a high level publication as a publication published either in a journal or in the proceedings of an international conference. We agree that this definition is not consensual, but we choose it to simplify the example (a more complex definition can, of course, be used). Invoking the instance checking of RACER allows one to retrieve all instances of the concept `HighLevelPublication`.

The instance retrieval process is based on classification. A query is reifed as a concept, say the query concept, that is then classified in the ontology (searching for its most specific subsumers and its most general subsumees). Once the concept is classified, the instances that answer the query are instances of the subsumees of the query concept [5, 16]. For example, if one would like to know who are the senior researchers, then it needs to define what a senior researcher is, like for example an author having published at least 50 publications: $\texttt{SeniorResearcher} \equiv \texttt{Author} \sqcap (\geq 50\ \texttt{hasPublished})$. For this example, RACER classifies the query concept in the hierarchy, then the instance checking reasoning service is invoked again to retrieve all instances satisfying this query concept.

4.2 Improving the navigation process

The research topics hierarchy is a natural way to improve a navigation process, in which the user can explore the links between the topics in order to reach progressively the information he needs. Compared to a thesaurus in which the topics are linked without a precise semantics, $\mathcal{O}_{\mathcal{BIB}}$ provides a meaningful organisation of the research topics, according to their properties. The properties of the concepts and the subsumption relationships between concepts in $\mathcal{O}_{\mathcal{BIB}}$ can be exploited for an intelligent navigation:

- for example, suppose that a novice researcher is searching for publications about *knowledge representation language* without having more information about this topic. While a system returning only publications having the keyword *knowledge representation language* can miss a lot of interesting publications, OntoBib also recommends the researcher to search for more specific topics such as *Description Logics*, *OWL*, etc. (see figure 2). These more specific topics will help the researcher to learn more about the research topic. So, this is an example of providing the researcher with a *More Specific Readings* search option. This task is executing by the RACER reasoner, first by retrieving the concept corresponding to the searched research topic and second by retrieving all the subconcepts of this research topic concept.
- another search option that the system supports is *Related Readings*. This option allows retrieving the publications having research topics related with a given one. The ontology provides links between research topics on the base of their properties. So, someone interested by the *OWL* topic can be redirected on topics such as *XML*, *Description Logics*, *Ontology*, etc. The difference between the use of the ontology and a classical thesaurus *related-with* relation is that the links between concepts have a semantics directly deduced from their properties. For example, on figure 2, *OWL* is linked to *XML* because they have a common *XML syntax*, *OWL* is linked to *Description Logics* because they have a common *DL semantics*, or even *OWL* is linked to *Ontology* because the OWL concept describes an *ontology*.

Hence, the relationships between concepts in the ontology represent a method to improve the navigation process.

4.3 Synthesis and contributions

This work on the OntoBib system provides two kinds of contributions.

The first one concerns the creation of an ontology of bibliographic data, and especially about the research topics. This experience of knowledge domain specification shows how simple keywords can be represented by meaningful properties, on which a hierarchical organisation can automatically be derived. Comparing to the manual creation of a thesaurus, this ensures the correctness of the domain knowledge. Moreover, it is possible to know in which way different concepts are similar or different. Nevertheless, like a thesaurus construction, the construction of the ontology is a human time consuming task.

The second contribution concerns the use of the ontology. First, the instantiation step of the ontology building process is a way to add conceptual links between web resources, transforming in our case a weakly linked collection of publications, authors, topics into a set of richly interconnected resources. In this case, $\mathcal{O}_{\mathcal{BIB}}$ is used to store the bibliographic data through the instantiation of data items into concept instances. A problem is that this instantiation is actually also a manual time consuming task. Second, the use of a reasoner module on the $\mathcal{O}_{\mathcal{BIB}}$ ontology allows semantic querying about bibliographic data and improves navigation processes the bibliographic data, as shown before.

5 Conclusion and future work

In this paper, we have presented an experiment on the use of ontologies for search and navigation services in the domain of bibliographic data. Introducing a practical study we described an approach and architecture that remains an experimental system. Defining an ontology that represents research topics, authors, publications and their relationships, is the base of the OntoBib system which allows an intelligent management of bibliographic data, through a semantic query answering and through the exploitation of relationships existing between topics.

There is mainly two perspectives to this work. A first one concerns the extension of the actual $\mathcal{O}_{\mathcal{BIB}}$ in order to take into account the knowledge of other areas in computer science. A semi-automatic process, in which existing organised data of a domain (like thesaurus for example) will be used as starting elements for the ontology construction, is envisaged. A second and more difficult perspective is the building of an automatic parser that is able to convert the data in digital libraries to instances of the ontology.

REFERENCES

[1] R. al Hulou, A. Napoli, and E. Nauer, 'Une mesure de similarité sémantique pour la classification de documents par le contenu', in *Langages et Modèles à Objets (LMO'04)*, eds., J. Euzenat and B Carré, volume 10 of *RSTI - L'objet*, 217–230, Hermes, Paris, (2004).

[2] G. Antoniou and F. van Harmelen, *A Semantic Web Primer*, The MIT Press, Cambridge, Massachusetts, 2004.

[3] N. Aussenac-Gilles, B. Biébow B., and S. Szulman, 'D'une méthode à un guide pratique de modélisation des connaissances à partir de textes', in *Actes des 5èmes journées Terminologie et Intelligence Artificielle (TIA 2003)*, ed., F. Rousselot, pp. 41–53. ENSAIS, (2003).

[4] *The Description Logic Handbook*, eds., F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, Cambridge University Press, Cambridge, UK, 2003.

[5] H.W. Beck, S.K. Gala, and S.B. Navathe, 'Classification as a Query Processing Technique in the CANDIDE Semantic data Model', in *Proceedings of the Fifth International Conference on Data Engineering*, pp. 572–581, (1989).

[6] T. Berners-Lee. Semantic Web, 2000. http://www.w3.org/2000/Talks/1206-xml2k-tbl.

[7] D. Bourigault, N. Aussenac-Gilles, and J. Charlet, 'Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas', *Revue d'Intelligence Artificielle (RIA) , " Techniques Informatiques et structuration de terminologiques*, **18**(1/2004), 87–110, (2004).

[8] M. d'Aquin, C. Bouthier, S. Brachais, J. Lieber, and A. Napoli, 'Knowledge editing and maintenance tools for a semantic portal in oncology', *International Journal of Computer Human Studies*, **62**(5), 619–638, (2005).

[9] *Spinning the Semantic Web*, eds., D. Fensel, J. Hendler, H. Lieberman, and W. Wahlster, The MIT Press, Cambridge, Massachusetts, 2003.

[10] T. R. Gruber, *Formal Ontology in Conceptual Analysis and Knowledge Representation*, chapter Toward Principles for the Design of Ontologies Used for Knowledge Sharing, Kluwer Academic Publishers, 1993.

[11] V. Haarslev and R. Möller, 'RACER System Description', in *First International Joint Conference on Automated Reasoning, IJCAR'2001*, volume 2083 of *Lecture Notes in*

Computer Science, pp. 701–706. Springer-Verlag, (2001). http://www.racer-systems.com/.

[12] Jeffrey Douglas Heflin, *Towards the semantic web: knowledge representation in a dynamic, distributed environment*, Ph.D. dissertation, University of Maryland, 2001.

[13] M. Horridge, H. Knublauch, A. Rector, R. Stevens, and C. Wroe, *A Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools. Edition 1.0*, University Of Manchester, 2004.

[14] Ian Horrocks, Deborah L. McGuinness, and Christopher A. Welty, 'Digital libraries and web-based information systems', in *The description logic handbook: theory, implementation, and applications*, pp. 427–449, New York, NY, USA, (2003). Cambridge University Press.

[15] S. Kampa, T. Miles-Board, L. Carr, and W. Hall, 'Linking with meaning: Ontological hypertext for scholars', Technical Report 0-854327-37-1, University of Southampton, Southampton, UK, (2001).

[16] M. Lenzerini and A. Schaerf, 'Concept languages as Query Languages', in *Proceedings of* AAAI*'91, Anaheim, California*, pp. 471–476, (1991).

[17] D. L. McGuinness, 'Ontological Issues for Knowledge-Enhanced Search', in *Formal Ontology in Information Systems, Proceedings of the 1st International Conference (FOIS'98)*, ed., N. Guarino, pp. 302–316, Trento, Italy, (June 1998). IOS Press.

[18] Natalya F. Noy and Deborah L. McGuinness, 'Ontology development 101: A guide to creating your first ontology', Technical report, Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, (2001).

[19] B. Omelayenko, 'Learning of Ontologies for the Web: the Analysis of Existent Approaches', in *Proceedings of the International Workshop on Web Dynamics, 8th International Conference on Database Theory, ICDT'01*, (2001).

[20] B. Safar, H. Kefi, and C. Reynaud, 'Ontorefiner, a user query refinement interface usable for semantic web portals', in *Application of Semantic Web Technologies to Web Communities Workshop, 16th European Conference on Artificial Intelligence*, (2004).

[21] *Handbook on Ontologies*, eds., S. Staab and R. Studer, International Handbooks on Information Systems, Springer, 2004.

[22] S. Szulman, B. Biébow B., and N. Aussenac-Gilles, 'Structuration de terminologies à l'aide d'outils d'analyse de textes avec terminae', in *TAL*, volume 43, pp. 103–128, (2002).

[23] M. Uschold and M. Gruninger, 'Ontologies: Principles, methods and applications', *Knowledge Engineering Review*, **11**(2), 93–155, (1996).

[24] M. Uschold and M. King, 'Towards a Methodology for Building Ontologies', in *Workshop on Basic Ontological Issues in Knowledge Sharing, held in conduction with IJCAI-95*, (1995).