Ontologies and the Semantic Web: Problems and Perspectives for LIS Professionals

Ontologies and the semantic web: problems and perspectives for LIS professionals

Joseph T. Tennis (1) y Javier Calzada-Prado (2)

(1) School of Library, Archival and Information Studies, University of British Columbia, 6190 Agronomy Road, Vancouver, BC (Canada), jtennis@interchange.ubc.ca. (2) Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid, 28903 Getafe, Madrid (Spain), fcalzada@bib.uc3m.es

Resumen

En la actualidad, todavía hay profesionales del ámbito de la Biblioteconomía y Documentación que no conocen su papel en el desarrollo de la Web Semántica, en especial en lo concerniente a las ontologías. Este trabajo pretende contribuir a aclarar este asunto de dos maneras: en primer lugar, identificando las principales tendencias, temas y problemas tratados en la investigación sobre ontologías y, en segundo lugar, identificando las posibles contribuciones del área de la Biblioteconomía y Documentación al desarrollo de ontologías para la Web Semántica. Para ello, se presenta en primer lugar una revisión de literatura basada en búsquedas en bases de datos de cobertura internacional (LISA, SCI, SSCI, ACM Digital Library e IEEE Explore). A continuación, y a partir de dicha revisión, se presenta una discusión de las principales tendencias, temas y problemas identificados. Finalmente, se presentan recomendaciones sobre posibles contribuciones del área documental al desarrollo de ontologías para la Web Semántica.

Palabras clave: Ontologías. Web semántica. Biblioteconomía y Documentación.

Abstract

Nowadays, many Library & Information Science (LIS) professionals still do not know their precise role in the development of the Semantic Web, especially those aspects concerning ontologies. This paper intends to contribute clarifying this issue in a twofold manner: 1) identifying main trends, issues and problems concerning ontology research; and 2) identifying possible contributions from the LIS area to the development of ontologies for the semantic web. To do so, a thorough literature review based on searches in international databases (LISA, SCI, SSCI, ACM Digital Library and IEEE Explore) is carried out. Then, the paper presents a discussion of the main trends, issues and problems concerning ontology research. Recommendations of possible LIS contributions to the development of ontologies for the semantic web are finally presented.

Keywords: Ontologies. Semantic Web. Library & Information Science.

1. Introduction

Many years have passed since Berners-Lee envisioned the Web as it should be (1999), but still many information professionals do not know their precise role in its development, especially concerning ontologies –considered one of its main elements. Why? May it still be a lack of understanding between the different academic communities involved (namely, Computer Science, Linguistics and Library & Information Science), as reported by Soergel (1999)? The idea behind the Semantic Web is that of several technologies working together to get optimum information retrieval performance, which is based on proper resource description in a machine-understandable way, by means of metadata and vocabularies (Greenberg, Sutton & Campbell, 2003). This is obviously something that

Library & Information Science professionals can do very well, but, are we doing enough? When computer scientists put on stage the ontology paradigm they were asking for semantically richer vocabularies that could support logical inferences in artificial intelligence as a way to improve information retrieval systems. Which direction should vocabulary development take to contribute better to that common goal? The main objective of this paper is twofold: 1) to identify main trends, issues and problems concerning ontology research; and 2) to identify possible contributions from the Library & Information Science area to the development of ontologies for the semantic web. To do so, our paper has been structured in the following manner. First, the methodology followed in the paper is reported, which is based on a thorough literature re-

view, where main contributions are analysed. Then, the paper presents a discussion of the main trends, issues and problems concerning ontology research identified in the literature review. Recommendations of possible contributions from the Library & Information Science area to the development of ontologies for the semantic web are finally presented.

2. Methodology

According to the objectives of the present work, five of the most representative databases were selected: Library & Information Science Abstracts (LISA), Association for Computer Machinery (ACM) Digital Library, Institute of Electrical and Electronics Engineers (IEEE) Xplore and ISI's Science Citation Index Expanded (SCI) and Social Science Citation Index (SSCI). Taking the broader topic of our interest and the generic names of related disciplines as reference, four queries were executed in the before mentioned databases: "ontologies" AND "semantic web" (Q1), "ontologies" AND "computer science" (Q2), "ontologies" AND "library science" (Q3) and "ontologies" AND "linguistics" (Q4). Thesaurus or keyword searches (with truncation) were used where applicable, and no temporal limits were applied. The results of the corresponding queries are summarized in the following table:

	Q1	Q2	Q3	Q4
LISA	38	3	1	1
ACM	110	1	0	2
IEEE	517	7	1	63
SSCI	75	23	4	16
SCI	1060	66	15	24
Total recs	1800	100	21	106

Table I. *Results of database searches*

A total amount of 1800 records has been retrieved (overlap between databases not discounted) with Q1. Results from the rest of queries seem far from being representative. However, they do show a closer relationship between ontologies and Linguistics than that between ontologies and LIS.

Benefiting from the data analysis application of ISI's databases, the following yearly distribution has been observed for records retrieved with Q1 (Table II). As shown, the production of works dealing with ontologies and the semantic web has increased progressively during the last decade, and especially on the last five years.

Fifty-six ISI subject categories have been assigned to records produced during that time, showing certain interdisciplinarity in the topic. The category distribution observed is presented in the table III.

Year	Record Count	% of 1087	Bar Chart
1997	1	0.0920 %	
1998	1	0.0920 %	
1999	7	0.6440 %	
2000	5	0.4600 %	
2001	13	1.1960 %	
2002	62	5.7038 %	
2003	140	12.8795 %	
2004	219	20.1472 %	
2005	309	28.4269 %	
2006	286	26.3109 %	
2007	44	4.0478 %	

Table II. *Yearly distribution for Q1 ISI results*

Field: Subject Category	Record Count	% of 1087	Bar Chart
COMPUTER SCIENCE, THEORY & METHODS	625	57.4977 %	
COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	247	22.7231 %	
COMPUTER SCIENCE, INFORMATION SYSTEMS	140	12.8795 %	
COMPUTER SCIENCE, SOFTWARE ENGINEERING	77	7.0837 %	
ENGINEERING, ELECTRICAL & ELECTRONIC	63	5.7958 %	
COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	36	3.3119 %	
COMPUTER SCIENCE, HARDWARE & ARCHITECTURE	34	3.1279 %	
INFORMATION SCIENCE & LIBRARY SCIENCE	27	2.4839 %	
EDUCATION & EDUCATIONAL RESEARCH	15	1.3799 %	
OPERATIONS RESEARCH & MANAGEMENT SCIENCE	15	1.3799 %	
TELECOMMUNICATIONS	15	1.3799 %	
COMPUTER SCIENCE, CYBERNETICS	14	1.2879 %	
BIOTECHNOLOGY & APPLIED MICROBIOLOGY	11	1.0120 %	
BIOCHEMICAL RESEARCH METHODS	9	0.8280 %	
ERGONOMICS	9	0.8280 %	
MEDICAL INFORMATICS	9	0.8280 %	
PSYCHOLOGY, MULTIDISCIPLINARY	9	0.8280 %	
ENGINEERING, MANUFACTURING	7	0.6440 %	
MATHEMATICS, APPLIED	7	0.6440 %	
AUTOMATION & CONTROL SYSTEMS	6	0.5520 %	
BIOCHEMISTRY & MOLECULAR BIOLOGY	6	0.5520 %	
BUSINESS	6	0.5520 %	
ENGINEERING, CIVIL	5	0.4600 %	
ENGINEERING, MULTIDISCIPLINARY	5	0.4600 %	
GEOGRAPHY	5	0.4600 %	
GEOGRAPHY, PHYSICAL	5	0.4600 %	
ENGINEERING, BIOMEDICAL	4	0.3680 %	
ENGINEERING, ENVIRONMENTAL	4	0.3680 %	
ENGINEERING, INDUSTRIAL	4	0.3680 %	
HEALTH CARE SCIENCES & SERVICES	4	0.3680 %	
STATISTICS & PROBABILITY	4	0.3680 %	

(25 Subject Category value(s) outside display options.)

Table III. *Category distribution for Q1 ISI results*

As shown, the topic has been led by Computer Science, basically from the theoretical and methodological perspective. Artificial Intelligence is also well represented, along with Information Systems and Software Engineering. Library & Information Science can be found on the eighth position, while Linguistics (under "Applied Linguistics" and "Language & Linguistics theory" categories) is out of the chart in final positions. A more detailed analysis of works under the LIS category (table IV) shows that most of them also fall under the category devoted to the Information Systems perspective of Computer Science.

Field: Subject Category	Record Count	% of 27	Bar Chart
INFORMATION SCIENCE & LIBRARY SCIENCE	27	100.0000 %	
COMPUTER SCIENCE, INFORMATION SYSTEMS	20	74.0741 %	
GEOGRAPHY	3	11.1111 %	
GEOGRAPHY, PHYSICAL	3	11.1111 %	

Table IV. *Category distribution for LIS Q1 results*

Actually, author's affiliation data have revealed that 74% of works under the LIS category has been produced by authors from the CS field who have published in LIS journals.

3. Setting the stage: the Semantic Web

The purpose of this section is to provide a background on the Semantic Web (SW) and its requirements concerning vocabularies.

3.1. The vision

Although the idea was not completely new, having Paul Otlet (Rayward, 1994) and Vannevar Bush (1999) as its most renowned predecessors, it was Berners-Lee, current director of the World Wide Web Consortium (W3C), who succeeded in materializing the web and its most ambitious dimension, the semantic web, in the late 1990s. He defined it as "an extension of the current [web], in which information is given well-defined meaning, better enabling computers and people to work in cooperation" (Berners-Lee, 2001). Such vision is basically supported on three pillars (see figure I): XML, RDF and ontology-based agents. The purpose of these technical tools, whose development is being promoted by the W3C, is then to provide web resources with machine-understandable semantic annotations (metadata) in order to develop sophisticated services based on automatic information retrieval techniques.
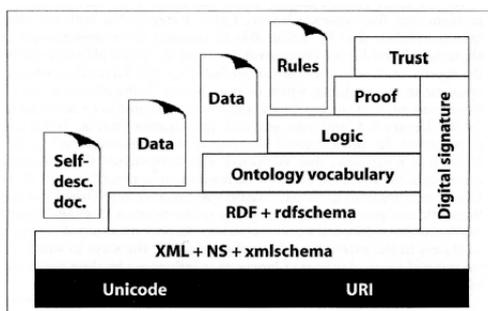

Figure I. *The Semantic Web layer-cake*

The preceding layer-cake (also called the 'Semantic Web Stack' by the W3C) can also be interpreted as different stages in the long way from proprietary, pre-XML documents to automated reasoning in the smart data continuum that will lead to the SW (Daconta, Obrst & Smith, 2003, p. 3).

3.2. The interoperability issue

Interoperability is the key issue for information retrieval in distributed information systems. It has been defined by IEEE as "the ability of two or more systems or components to exchange information and to use the information that has been exchanged" (IEEE, 1990). Making systems and information representations compatible implies a lot of agreement in creating and using a wide range of standards that will ensure syntactic and semantic interoperability.

3.2.1. Syntactic interoperability

HTML is a standard that provides an easy way to code resources, as well as embedding metadata, which is the reason why it has been so popular in the web. However, HTML is mainly presentation-oriented, and only a limited number of tags with scarce semantics can be used. Contrarily, XML is content-oriented, providing means to meaningfully and virtually unlimitedly describe the content of resources apart from their format, which constitutes the base for interoperability. Among other factors, content portability is what has made XML so successful. As shown in figure 1, XML is built upon the Unicode universal character set and Universal Resource Identifiers (URI), which makes coding and resources human interpretable and authorable as well as unambiguously identified.

3.2.2. Semantic interoperability

While XML provides syntactic interoperability, ontologies (in its widest sense, including metadata and different vocabularies) are intended for providing semantic interoperability. As mentioned before, both are essential elements to allow agents to make inferences from them and perform previously required tasks. RDF and RDF Schema are XML-based representational frameworks that intend to provide interoperable descriptions for web resources, defining ontologies or metadata schemas. Recently, SPARQL has completed the layer-cake of the SW architecture providing a query language for RDF.

3.3. The vocabulary spectrum

With the rise of the multidisciplinary approach to information organization, we have witnessed an increase in the diversity of types of vocabularies. The different vocabulary initiatives can be arranged on a spectrum, ranging from informal vocabularies to formal vocabularies.

3.3.1. Formal vocabularies

Vocabularies that are designed, maintained, and implemented for a particular purpose can be considered formal vocabularies. In LIS, this kind of vocabularies has been called controlled vocabularies or indexing languages. The latter makes explicit the pre-dominant purpose of these formal vocabularies: to index documents in information systems. The ultimate purpose of which is to retrieve them. The recently updated ANSI/NISO Z39.19-2005 (2005, p. 1-2) considers thesauri, taxonomies, and classification schemes to be indexing languages. Ontologies are also formal vocabularies. And though ontologies can be used for retrieval, they are not used solely for indexing with the ultimate purpose of retrieval. Ontologies, as specifications of conceptualizations that are machine-processable (Gruber, 1995) are formal because they are built, maintained, and implemented for a particular purpose.

3.3.2. Informal vocabularies

Informal vocabularies are not bound by a particular purpose, but rather grow from multiple purposes, many of them not designed, but emergent or ad hoc based on the immediate and context specific needs of the user. That is not to say that we cannot speak to the general nature of the purposes employed by these informal vocabularies. However, the distinction remains between these and formal vocabularies that the former are free from the constraints and therefore lacking the power of an explicit design imperative in their realization. Folksonomies, annotation services, and other Web 2.0 services like digg.com, del.icio.us and flickr can be considered informal vocabularies. They are often used for personal information management, indexing, bookmarking, and commentary (Tennis, 2006). Yet, because of their informal nature, none of them adheres to a strict design principle.

4. Positions of involved disciplines

In this section, we will revise key concepts and research trends associated with each of the main disciplines involved in ontology development for the SW, as identified in the literature review. We have intentionally let Philosophy apart, since it is not treated in current research but as theoretical background.

A first preliminary approach to the corresponding positions has been to observe the coverage of the topic in thesauri in order to draw a 'roadmap' of disciplines, in Soergel's (1999, p. 1119) terms. Vocabularies, in its attempt to represent existing knowledge, can reflect how different domains perceive each other. Thus, the ASIS Thesaurus of Information Science and Librarianship (1998) does not include the term "ontology", though it does include a related term, 'intelligent agents (software)' as narrower term of 'information retrieval software' and related term of 'artificial intelligence'. A more recent LIS thesaurus, the CINDOC's "Tesauro de Biblioteconomía y Documentación" (Mochón & Sorli, 2005) places ontologies as a narrower term of artificial intelligence, the latter being considered a narrower term itself of Computer Science. Curiously enough, the ACM Computing Classification System (2006) does not include the term 'ontologies', but it does include 'thesauruses' as narrower term of 'content analysis and indexing', itself a narrower term of 'information storage and retrieval' in 'information systems' (no direct reference at all to LIS).

An informal, brief analysis of titles and abstracts of the retrieved articles revealed the following topics in ontology research:

- Ontology-based indexing and querying for information retrieval.
- Software agent-based semantic applications and services for the semantic web.
- Ontology representation and query languages.
- Ontology software tools (annotation, editing, automatic generation, reasoners, etc.).
- Ontologies for information/document organization (knowledge management, web mining).
- Ontologies as key to semantic interoperability.
- Automatic ontology generation (automatic or semi-automatic ontology learning, population, etc.).
- Ontology validation, mapping, aligning, merging and translation for semantic interoperability.
- Thesauri to ontology transformation models.
- Ontology applications in e-commerce, e-government, e-learning and e-health.
- Ontology applications in wikis.
- Ontology applications for geospatial data.
- Ontology reusability.
- Ontology visualization and ontology-based semantic navigation.
- Semantic grid applications.
- Ontology-based information system architectures.

Representing some of the core topics in current research, the preceding list introduces some of

the main issues in positions of involved disciplines that we are about to detail.

4.1. Computer Science

In computer science, vocabularies are differentiated according to their 'semantic strength' (figure II), from weak semantics to strong semantics.
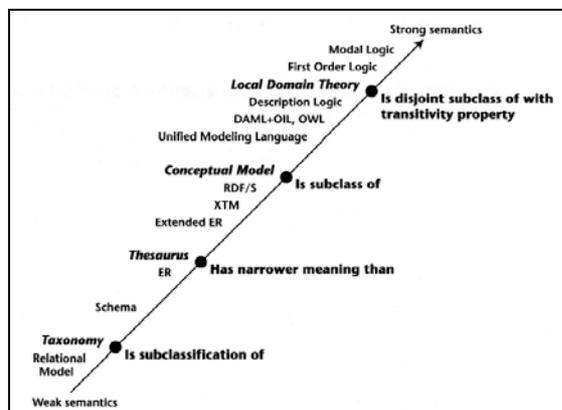

Figure II. *Ontology spectrum (Daconta, 2003)*

In the ascending line towards strong semantics, ontologies are situated above RDF Schema and under the logic rules level. Taxonomies and thesauri are placed a step behind, considered lightweight ontologies. In this field, the term 'ontology' is frequently used as an equivalent to 'vocabulary', 'taxonomy' or 'thesaurus'. But, what is an ontology for CS professionals, precisely?

The most influential definition from the CS field identified in the literature has been, without a doubt, that of Gruber (1995): "an ontology is an explicit specification of a conceptualization". As later explained by Studer, Benjamins and Fensel (1998), the idea of 'conceptualization' in ontologies refers to the abstract and simplified view of the world that is represented to serve a practical purpose; 'explicit' refers to the explicitness of the definition of concept types and use constraints; 'formal' refers to their machine-readability; and 'shared' means that it should constitute a set of agreed-upon terms.

Another relevant definition is that from Guarino (1998), who tried to clarify Gruber's definition by adding that the term is being used in AI referring to "an engineering artifact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words". Berners-Lee, Hendler & Lassila (2001), also relating the concept to AI, stated that an ontology is "a taxonomy and a set of inference rules". Clearly, the involvement of AI in ontology re-

search has been decisive for its development. AI, in its attempt to create machines able to perform intelligent actions, deals with reasoning about computationally-inferred world models, which is where it connects with ontologies. Typical IA methods applied to or related with ontologies (Pajares Martinsanz & Santos Peñas, 2005, p. 7-8) are: Petri networks, expert systems, fuzzy logic, neural networks and genetic algorithms.

Studer, Benjamins and Fensel (1998) distinguished four types of ontologies: domain ontologies (pertaining to particular domains), generic or core ontologies (valid for several domains), application ontologies (for domain modelling) and representational ontologies (which do not state what should be represented).

The general growing interest about ontologies has even determined the appearance of a new discipline in CS, *Ontology engineering*, which specifically addresses the development and management of ontologies (Sure, Tempich & Vrandecic, 2006).

Ontology development in CS involves dealing with organizing classes (concepts), properties (instances) and axioms relating them. Logical reasoning is, then, an integral part of ontology design. Representational ontology languages are either logic-based (first-order logic), frame-based (frame logic) or web-based (RDF, XML, HTML), as reported by Ding (2001). Usually, ontology construction is treated as a semi-automatic process with several techniques working at a time.

Main research trends concerning ontologies in CS are:

- *Agent systems*. As more web content is associated with explicit meaning and becomes more machine-readable, intelligent agents can retrieve and process information readily. Apart from providing resources with semantically machine-understandable information, machines that simulate human intelligent activities can make a valuable contribution. Intelligent software agents can be defined as "retrieval tools which routinely scan networks and systems for information meeting specified criteria and present the retrieved information to the user" (ASIS, 1998, p. 52). Autonomous software web agents make their decisions according to their knowledge (limited representation of the world), so it is essential that the shared ontology is consistent. Software Engineering modelling and verification techniques are applicable at many stages during the design, maintenance and deployment process of ontologies.

• *Semantic web mining*. This concept refers to the application of web mining techniques to ontology construction for web mining and the semantic web. Within web mining, three areas are usually distinguished: content mining, structure mining and usage mining (Stumme, Hotho & Berendt, 2006). When applied to ontology development, those areas provide several action lines (*ibidem*), from which we would like to highlight:

• *Ontology learning*. It refers to the automatic (unsupervised learning or semi-automatic extraction of semantics by means of machine-learning techniques. It is usually based on pre-existing resources such as dictionaries, thesauri, linguistic corpora and databases, from which the system 'learns' and populates the resulting ontology.

• *Ontology mediation: Mapping and merging ontologies*. The increasing number of overlapping ontologies available allows considering their reuse as a feasible way to build new domain-specific ontologies. The semi-automatic process of extracting and assembling ontology fragments based on syntactic and semantic matching techniques is called merging. On the other hand, mapping is oriented towards the achievement of semantic interoperability between information systems, and it refers to the process of finding equivalences between concepts from different ontologies, having one of them as reference and the others as target ontologies.

• *The semantic grid*. Originated in 2001 under the influence of the semantic web vision, this concept is related to what is called "grid computing", which refers to "coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations" (Sure & Roure, 2006). Research on the semantic grid is currently focused on ontology and agents-based applications for delivering intelligent, knowledge-based services.

4.2. Linguistics

As reported by Bontcheva, Cunningham, Kiryakov & Tablan (2006), there is a tension between "the increasingly rich semantic models in IT systems on the one hand, and the continuing prevalence of human language materials on the other". In this tension lies the reason for the increasingly closer relationship between Linguistics and CS, as suggested by database query results presented in section 2. New areas of joint research have matured over the years, such as HTL (Human Language Technology), which deals with the analysis, mining and production of

natural language; or a form of natural language analysis for information extraction, IE (Information Extraction), which "is becoming a central technology to link Semantic Web models with documents as part of the process of metadata extraction" (ibidem, p. 30).

4.3. Library & Information Science

Vickery (1997) defined ontologies as "a schedule, in some form that may involve the use of semantic categories, of concepts significant in a particular domain (that may be as wide as the universe of knowledge), together with a definition or scope note for each concept, and mechanisms for displaying its relationships to other concepts". He stated that the main difference between ontologies and documentary tools like bibliographic classifications and thesauri is their intended usage. Ontologies were to be used as knowledge bases in knowledge engineering as an application of artificial intelligence, while most documentary tools, like thesauri, were mainly designed for human consultation (Vickery & Vickery, 2004, p. 312). In his work, Vickery (1997) complains about the lack of interest of 'ontological engineers' in previous work in Information Science such as those of Chan (1985) and Lancaster (1986), which obviously denied the novelty of ontology construction.

Lancaster (2003) himself has complained about the gradual replacement of the traditional term in the library profession "classification" by newer, trendier terms like "taxonomy", "concept hierarchy" or "ontology", which he attributes to the multiplicity of professions currently contributing to information retrieval, and particularly to the influence of computer scientists'. As he reports, "computer scientists writing on information retrieval seem to recognize and cite only other computer scientists writing on information retrieval" (ibidem, p. xi).

Soergel (1999) has seen the ontology movement as the 'reinvention' of classification provoked by the lack of communication between disciplines. He has reported (ibidem) that the growing need for classification in other fields has led to the rise of ontologies, by largely ignoring the vast body of knowledge previously developed in Information Science: "classification by any other name is still a classification".

Garshol (2004) considers that ontologies, due to their open nature in opposition to controlled vocabularies, are the culmination in the progression of vocabularies towards the most powerful means of description. Contrarily, Gradmann (2004) argues that ontologies may be useful in what he calls "evident domains", that is, specia-

lized, restricted domains, but may not be appropriate for other complex domains where a higher degree of interpretation is required. That is something that even Gruber had in mind when defining ontologies and restricting them to particular areas of interest, as reported by Jacob (2003). Jacob also remarks that although ontologies have many similarities with documentary representational structures, their equation with any of them would diminish their potential in the Semantic Web. García Jiménez (2004) has agreed on the differentiation, but has also insisted on the special synergies that can be achieved between ontologies and thesauri. Moreira, Alvarenga and Paiva (2004) allege that both instruments possess different origins and purposes (as stated by Vickery & Vickery), but acknowledge that "the thesaurus accomplishes part of the objectives that computer science intends with ontology".

As for the existing types of ontologies, Fonseca (2005) identified two major types in Information Science research: ontologies of information systems and ontologies for information systems, both the result of a long-term research effort on conceptual modelling. Following McGuinness' (2003) classification of ontologies, Legg (2007) distinguishes three kinds of ontologies according to their expressivity: 'thesaurus ontologies" (among which she includes DCMES and Word-Net), ontologies with Description Logic expressivity (Topic Maps, RDF Schema, DAML+OIL/OWL and SHOE) and ontologies with first order logic expressivity (CYC, SUMO and SUO). She also distinguishes domain ontologies, as ontologies that belong to specific areas of knowledge.

Vocabulary development in LIS is typically under the form of controlled vocabularies, which involves domain analysis, relevant vocabulary selection and reuse, terminology control and establishment of term relationships. Their purpose is to avoid the ambiguity of natural language used in documents and achieve optimum information retrieval performance. In vocabularies, inference rules are usually understood as property inheritance among terms in a hierarchical organization.

Main research trends concerning ontologies in LIS are:

- *Automatic indexing and thesauri construction*. Research in this field started in the seventies. After demonstrating the limitations of statistical techniques, automatic thesauri construction begun to apply, among other techniques, those belonging to Natural Language Processing (NLP).

- *Thesauri to ontology transformation*. As suggested by Greenberg, Sutton & Campbell, (2003), "for existing and developing ontologies to be used and function fully in the Semantic Web environment, they need to adhere to standards supported by enabling technologies". An example application would be Topic maps, a standard (ISO/IEC 13250, 2000) that provides an ontology framework for information retrieval that has been discussed from the LIS perspective by Garshol (2004), among others.

5. Discussion

According to our observations, it seems evident that, first of all, there is a lack of terminological consensus (as reported by Vickery & Vickery, 2003, p. 312), not only among disciplines, but also within the respective disciplines. Both in CS and LIS, the term has become a buzzword applied to almost anything. However, LIS professionals look betrayed by other disciplines' use of their traditional terminology and techniques and seem reluctant to incorporate their advances to their own practice.

There is also a conceptual lack of understanding, which may have to do with the different educational background of the respective academic disciplines. But professional profiles are changing both for CS and LIS professionals, and blended education is beginning to be required. In the case of the latter, Rupp & Burke (2004) describes the evolution from cataloguers to ontologists in job advertisements published in professional journals.

Apart from that, we also find that there is a disagreement on ontology functions. A common position in CS is that unlike traditional, limited information environments where representation and retrieval based on controlled vocabularies were feasible, the open nature of the Web environment makes them inadequate. On the other hand, LIS professionals argue that controlled vocabularies are essential to give quality to ontology creation and consequently, to SW services. Things, of course, are not so black and white for everyone, and solution may be found on collaboration.

6. Recommendations for LIS professionals

Current research reviewed shows that the following recommendations can be made for LIS professionals:

- *Advancing on domain-specific vocabulary translation to semantic web syntax*. Examples of previous experiences on transforming vo-

cabularies into ontologies are, among many others, those of Qin & Paling (2001), Guzman Luna, Torres Pardo & López García (2006) and Marzal et al. (2006).

• *Advancing on vocabulary registration*. Providing universally available and authoritative reference on vocabularies can make a positive contribution to the SW vision. Exemplary is the case of the DCMI Vocabulary registry.

• *Advancing on enriching and expanding thesauri's semantic relationships*. An example of this is the Center for Bioinformatics of the US National Cancer Institute Metathesaurus (Golbeck, 2003). It is commonly understood that "the use of such techniques would greatly enhance the possibilities of using intelligent agents to handle information on the web" (Vickery & Vickery, 2004, p. 322).

• *Work on ontology-based information organization and visualization*. Conceptual navigation has promising applications for digital libraries, as demonstrated by, for instance, AquaBrowser.

• *Collaborate*. Participate and promote the development of multidisciplinary research groups.

References

ANSI/NISO (2005). Guidelines for the construction, format and management of monolingual controlled vocabularies. Z39.19-2005. Maryland: NISO Press.

ASIS, American Society for Information Science (1998). ASIS Thesaurus of Information Science and Librarianship. Medford: ASIS. ISBN 1573870501.

Berners-Lee, T.; Hendler, J.; Lassila, O. (2001). The semantic web. // Scientific American. May issue.

Berners-Lee, T. (1999). Weaving the web: the original design and ultimate destiny of the World Wide Web by its inventor. New York: Harper. ISBN 0062515861.

Bontcheva, K.; Cunningham, H.; Kiryakov, A.; Tablan, V. (2006) Semantic Annotation and Human Language Technology // J. Davies; R. Studer; P. Warren (eds.). Semantic web technologies: trends and research in ontology-based systems. England: Wiley. ISBN 0470025964.

Rupp, N.; Burke, D. (2004). From catalogers to ontologists: changing roles and opportunities for technical services librarians. // Serials Librarian. 46: 3-4 (2004) 221-226.

Bush, V. (1999). As we may think. // Library Computting. 18:3 (1999) 180-188.

Chan, L. M. (1985). Theory of subject analysis. Littleton: Libraries Unlimited. ISBN 0872874893

Daconta, M.C.; Obrst, L.J.; Smith, K.T. (2003). The semantic web: a guide to the future of XML, web services and knowledge management. Indianapolis: Wiley. ISBN 0471432571

Ding, Y. (2001). A review of ontologies with the semantic Web in view. // Journal of Information Science. 27:6 (2001) 377-384.

Fonseca, F. (2006). The double role of ontologies in Information Science research. // Journal of the American Society for Information Science and Technology. 58:6 (2006) 786-793.

García Jiménez, A. (2004). Instrumentos de representación del conocimiento: tesauros versus ontologías. // Anales de documentación. 7 (2004) 79-95.

Garshol, L.M. (2004). Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all. // Journal of Information Science. 30:4 (2004) 378-391.

Golbeck, J. (2003). The National Cancer Institute's Thésaurus and Ontology. // Journal of Web Semantics. 1:1 (2003) 75-80.

Gradmann, S. (2004). Going semantic? *Signum*. 8 (2004) 31-35.

Greenberg, J.; Sutton, S.; Campbell, D.G. (2003). Metadata: a fundamental component of the semantic web. // Bulletin of the American Society for Information Science and Technology. 29:4 (2003) 16-18.

Gruber, T.R. (1995). Toward principles for the design of ontologies used for knowledge sharing. // International Journal of Human-Computer Studies. 43:5-6 (1995) 907-928.

Guarino:(1998). Formal ontology in information systems. In: B. Bennett and C. Fellbaum (ed.) Frontiers in Artificial Intelligence and Applications, v. 150. ISBN 9781586036850.

Guzmán Luna, J.A.; Torres Pardo, D.; López García, A.N. (2006). Development of an ontology in the context of the semantic web: the migration of the traditional documentary thesaurus. // Revista Interamericana de Bibliotecología. 29:2 (2006) 79-94.

IEEE, Institute of Electrical and Electronics Engineers (1990). IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York: IEEE.

ISO/IEC (2000). Information Technology. Topic Maps. Available at: http://www.y12.doe.gov/sgml/sc34/document/0129.pdf

Jacob, E.K. (2003). Ontologies and the semantic web. // Bulletin of the American Society for Information Science and Technology. 29:4 (2003) 19-22.

Lancaster, F.W. (2003). Indexing and abstracting in theory and practice. London: Facet. ISBN 1856044823.

Lancaster, F.W. (1986). Vocabulary control for information retrieval. Arlington: Information Resources Press. ISBN 0878150536.

Legg, C. (2007). Ontologies on the semantic web. // Annual Review of Information Science and Technology. 47 (2007) 407-451.

Marzal, M.A.; et al. (2006). Modelado de un perfil de aplicación como ontología para la gestión de recursos en bibliotecas digitales educativas. // I

Conferencia Internacional sobre Ciencias y Tecnologías Multidisciplinares de la Información (InSciT2006), 25-28 October, Mérida, Spain.

McGuinness, D. (2003). Ontologies come of age. // D. Fensel, J. Hendler, H. Lieberman & W. Wahlster (eds.) Spinning the semantic web: bringing the world wide web to its full potential. Cambridge: MIT Press (2003) 171-194.

Mochón, G; Sorli, A. (2005). Tesauro de Biblioteconomía y Documentación. Madrid: CINDOC. ISBN 8400080912.

Moreira, A.; Alvarenga, L.; De Paiva Oliveira, A. (2004). "Thesaurus" and "ontology": a study of the definitions found in the computer and information science literature. // Knowledge Organization. 31:4 (2004) 231-244.

Pajares Martinsanz, G.; Santos Peñas, M. (2005). Inteligencia artificial e ingeniería del conocimiento. Madrid: Ra-Ma. ISBN 8478976760

Qin, J.; Paling, S. (2001). Converting a controlled vocabulary into an ontology: the case of GEM. // Information Research. 6:2.

Rayward, W. B. (1994). Visions of Xanadu: Paul Otlet (1868-1944) and hypertext. // Journal of the American Society for Information Science. 45:4 (1994) 235-250.

Soergel, D. (1999). The rise of ontologies or the reinvention of classification. // Journal of the American Society for Information Science. 50:12 (1999) 1119-1120.

Studer, R.; Benjamins, V. R. y Fensel, D. (1998). Knowledge engineering: principles and methods. // Data & Knowledge Engineering: Principles and Methods. 25:1-2 (1998) 161-197.

Stumme, G.; Hotho, A.; Berendt, B. (2006). Semantic web mining: state of the art and future directions. // Journal of Web Semantics. 4:2 (2006) 124-143.

Sure, Y.; Roure, D. (2006). Semantic grid: the convergence of technologies. // Journal of Web Semantics. 4:2 (2006) 82-83.

Sure, Y.; Tempich, C.; Vrandevic, D. (2006). Ontology Engineering methodologies. // J. Davies; R. Studer; P. Warren (eds.). Semantic web technologies: trends and research in ontology-based systems. England: Wiley. ISBN 0470025964

Tennis, J.T. (2006). Social tagging and the next steps for indexing. // J. Furner; J.T. Tennis (eds.) Proceedings of the 17th SIG/CR Research Workshop. Austin: ASIST.

Vickery, B.C. (1997). Ontologies. // Journal of Information Science. 23:4 (1997) 277-286.

Vickery, B.C.; Vickery, A. (2004). Information Science in theory and practice. München: K.G. Saur. ISBN 3598116586.