Thesauruses and ontologies

Silvia Arano

1. Introduction

During the past few years, the information representation and retrieval sector in the area of Documentation and Biblioteconomy has had to assume the important repercussions of the Internet and its associated technologies, and in particular, the World Wide Web (WWW). Technological modifications arising from these important changes are leading to the gradual digitalisation of the information representation and retrieval sector, affecting information artefacts, representation and retrieval tools and user requirements.

In the light of this growing context of digitalisation, diverse information representation and retrieval tools exist, which must be studied in addition to diverse fields of knowledge in which these tools have originated: Linguistics, Artificial Intelligence, Documentation, Linguistic Engineering... Hence, in specialised literature, analyses are performed on information representation and retrieval tools, taxonomies, classification systems, computational lexicons, lexical databases, thesauruses, titles lists, knowledge bases, conceptual maps, ontologies, synonym rings and semantic networks, among others. Among this wide spectrum of information representation and retrieval tools are thesauruses and ontologies, which are most often linked in bibliography, even though they come from completely different disciplinary areas. However, the conceptualisation applied by authors to the terms "thesaurus" and "ontology" is quite diverse, and sometimes authors confuse, oppose, complement or overlap both these concepts.

The overall objective of the present article [1] is to establish the relationship between the concepts of thesaurus and ontology in the Documentation and Biblioteconomy field. Two specific objectives have been established for this purpose. Firstly, to make an analysis of the thesaurus-based concept with a view to defining its most important characteristics and to verify the similarities and differences it shares with ontologies. And secondly, to establish a definition for the ontology concept, also for the purpose of verifying its characteristics and analysing the similarities and differences it has with thesauruses.

It is not, however, our intention to lay down a series of guidelines on these concepts, but to explain their main characteristics and describe their similarities and differences, in order to better understand what a thesaurus is and what an ontology is from the documentary standpoint. Or, in the words of Gilchrist (2002: 7):

" *It would not be sensible to pontificate on the "correct" meanings of these words, but in trying to delineate the central characteristics , it is hoped that the reader will gain a clearer understanding of their differences and similarities* "

2. Thesauruses

The term "thesaurus" has its etymological root in the Latin word *thesaurus* , which in turn comes from the Greek word *thesaurós* . In both cases, the meaning was treasure or repository of words.

In modern times, Lexicography was the first field in which the term "thesaurus" was used and in

which the first definitions were coined. In 1852 Peter Mark Roget Publisher his work *ROGET's THESAURUS of English Words and Phrases* (hereinafter, *ROGETs Thesaurus*), the purpose of which was to provide assistance and help in expression ideas and literary composition. *ROGETs Thesaurus* has a conceptually-based structure, where the concepts explained in the entries are used as a basis for relating and grouping together words that designate or name the concept in question, in different contexts. The meanings and uses of the term "thesaurus" started to be diversified, based on this conceptualisation proposed by Roget.

In 1957 the term "thesaurus" was used for the first time in the field of Biblioteconomy and Documentation, in a work presented to the *Dorking Conference on Classification* by Helen Brownson, member of the *American National Science Foundation* . In this work, the term "thesaurus" is used to analyse the problems of translating concepts and their relations, as expressed in documents, to a language with greater precision and without ambiguities, in order to facilitate information retrieval. Raising the issue of a historic evolution of the conception of thesauruses as a documentary tool in the field of Biblioteconomy and Documentation far exceeds the objectives of this article, and we shall therefore limit ourselves to selecting some definitions which, in our opinion, will enable us to define its main characteristics.

2.1. Establishing a conceptualisation of thesauruses

Within the regulatory context, *AENOR (Asociación Española de Normalización y Certificación) (Spanish Normalisation and Certification Association)* defines a thesaurus under the following terms:

"A controlled vocabulary of an indization language structured in a formal, a priori, manner, in order to make explicit the relations between concepts, (for example 'more generic than' or 'more specific than') '. (UNE 50-106-90, 1990: 5) [2]

This definition presents the thesaurus as an organised structure based on a series of conceptual relations that includes vocabulary control, which has indization as its function.

On the other hand, the *NISO (National Information Standards Organization)* [3] , guideline considers a thesaurus to be :

" *a controlled vocabulary of terms in natural language that are designed por postcoordination* " (ANSI/NISO Z39.19-2003, 2003: 1)

This definition includes the characteristics of postcoordination use of thesauruses; that is, the demand for their terms to be interrelated when searching for the information. At the present time, the *NISO is* about to terminate an updating of this guideline. Although the updated final text is still not available, in the preliminary documentation that is available for consultation in the *NISO* website, we can find another definition of thesaurus that stresses the vocabulary control work and structured organisation of this tool:

" *A set of word or phrases with equivalent terms explicitly identified and with ambiguous words or phrases (e.g.) homographs made unique. This set of terms also may include broader-narrower or other relations* ". (Z39.19 TAG Conference Call, 2003: 2)

If we consider the definitions proposed by the most widely-read manuals in the field of information representation and retrieval, the one given by Aitchinson and Gilchrist (1987) is worth mentioning. These authors define a thesaurus as a controlled vocabulary of an indization language that is organised formally in such a manner that the conceptual relations are established a priori, that can be used for information retrieval. Based on this definition, we can explicitly add yet another of its functions: that of information retrieval.

Slype (1991), for his part, considers that a thesaurus is:

a structured list of concepts, intended to represent in an univocal manner the contents of documents and consultations within a specific documentary system, and to assist users in the indization of documents and consultations ' (Slype, 1991: 24)

The above author, in addition to presenting the characteristics already referred to above (vocabulary control, conceptual, use in the indization and retrieval of information) enlarges the definition by introducing the user as a beneficiary of that tool.

Based on the characteristics we have inferred from the above definitions, a preliminary conceptualisation can be established of what a thesaurus is from ourstandpoint. a A thesaurus is a type of documentary language that represents the conceptual structure of a specific field of knowledge. A thesaurus offers a semantic structure, mainly through explaining the relations established between those concepts and finally, through a limited meaning of the terms that represent them. A thesaurus, in terms of Documentation and Biblioteconomy, is therefore a tool used for terminological control, since the thesaurus structure is based on concepts. Concepts are represented by selected terms which demonstrate this terminological control. This control aims to neutralise synonymy and polysemy, both of which are natural characteristics of language, that make it difficult to achieve precision in information indization and retrieval, which are the fundamental functions of a thesaurus. Thesauruses are tools created to help both information professionals and final users.

2.2. Thesauruses and the digital environment

As has already been said in the introduction, the technological repercussions of the digital environment has affected artefacts, tools and user behaviours alike in relation to information representation and retrieval. This effect opens up new possibilities in terms of design and elaboration, management and use of tools used for information representation and retrieval. Based on this standpoint López-Huertas (1997), De la Cueva Martín (2000), Shiri and Revie (2000) and Qin and Paling (2001) among others, have started to define the contributions made by the digital environments to thesauruses, and these can be summarised in the following elements:

- The first element to be considered is the enriching of the thesaurus structure functionality based on hypertexting. This leads to the establishing of hyperlinks among all the structural elements (descriptors, no descriptors, scope notes,etc.), and also among the different parts of the thesaurus.

- The second element is the reduction of updating and maintenance costs. Due to the growing digitalisation of thesaurus-construction processes and the gradual abandoning of paper formats in publishing these tools, cost-reduction is perfectly viable.

- The third element is user-integration into the process of creating, managing and optimising thesauruses, through usability tests, the use of user-modelling techniques, etc. This makes it possible to create tools that take user requirements into account and rules out their creation as simple theoretical structures.

- The fourth element is the possibility of applying methods of reuse and interoperability at the time of planning and creating the thesauruses. This makes it possible to use and make the most of the conceptual and linguistic information already generated for other artefacts.

From our point of view, this last element is a key factor in the new generation of digital thesauruses. The use of conceptual and linguistic information stored in other types of artefacts (e.g., in an ontology) enables advantage to be taken of the thesauruss structural elements and

makes enables the friendly nature of these tools to be increased for non-specialist, end-users.

3. Ontologies

Philosophy is the first field of knowledge in which the ontology concept is applied. The use of this concept has its roots in the Aristotelic notion *of first philosophy* then known as *metaphysics-,* which firstly studies the essence of living beings (*living beings as being)* , and secondly, the basic characteristics of reality as a whole (*the being* or principal *entity* upon which other entities depend). At the present time, it is considered a branch of Philosophy which has the objective of explaining existence in a systematic manner, dealing with the types and structures of objects, properties, events, processes and relations pertaining to each part of reality.

From the decade of the nineties on, ontologies started to become more important in the field of Artificial Intelligence, with special emphasis on Knowledge Engineering, PLN and Knowledge representation. As expressed by Studer (et al.), this is due to the fact that:

" *Artificial Intelligence (AI) deals with reasoning about models of the world. Therefore, it is not strange that AI researchers adopted the term ontology to describe what can be (computationally) represented of the world in a program* ". (Studer et al, 1998: 25)

Consequently, in the general sense of the word, for the purposes of Artificial Intelligence, ontologies are constructed artefacts that enable shared, common knowledge on something to be represented. This possibility of generating artefacts that can be shared and the natural consequence of exchanging the information stored inside them is what leads to a concept such as that of ontology (which hitherto only pertained to Artificial Intelligence) being filtered through into the working environments of other fields and especially, fields related to the management of artefacts and tools from the digital environment. Although different fields of knowledge exist [5] (Guarino 1998; McGuinness 2001) which are referred in research carried out into ontologies, our interest is focused on analysing ontologies as linguistic artefacts and determining their use in the Documentation and Biblioteconomy field.

The following chapter presents a series of definitions in the field of Artificial Intelligence that are used as a basis for extracting the most important characteristics of ontologies as such, for the purpose of proposing a proprietary definition of ontology that can be used as a guide in future considerations.

3.1. Establishing a conceptualisation of ontologies

With a view to obtaining a better understanding of the ontology concept, we have decided to carry out a search for and analysis of the different definitions of ontology in the field of Artificial Intelligence. Below is a series of definitions selected together with a brief comment on the most important characteristics of ontologies.

Neches et al. (1991) offer the following definition of ontology:

" *an ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary* " (Neches et al. quoted by Gómez-Pérez, 1999: 33)

This definition, which is brief and concise, gives a list of its component parts (terms, relations between terms and combination rules) and in addition, proposes elements that can be used to identify it: identifying the basic terminology and the relations arising from the terms; it then identifies the rules that permit them to be combined and finally, anticipates the definitions that

correspond to the terms and their relations. Ontology is seen as a dynamic artefact, insofar as it is formed by terms that have been created based on rules, in addition to those that are explicitly defined.

On the other hand, Gruber proposes the following definition: "an ontology is an explicit specification of a conceptualization" (Gruber, 1993b: 1). The author considers that a *conceptualization* is constituted by objects, concepts and other entities that exist within a specific area, and the relations existing between them. Although Grube's definition is more accurate and widely known in literature on ontologies, it is criticised due to the notion of *conceptualization* it uses, which is the same as the one used by Genesereth and Nilsson in their manual entitled *Logical foundation of Artificial Intelligence* published in 1987. To explain the meaning of *explicit specification* , it must be remembered that Artificial Intelligence considers that *all that which exists* is precisely all that which may be represented computationally. If the knowledge of a domain is represented through a declarative formalism, the series of objects that can be represented is referred to as the universe of discourse. This series of objects and the relations between them are shown in the vocabulary used to represent knowledge. In ontologies, the names of the entities in the universe of discourse are associated with each other by means of definitions (e.g categories, relations, functions or other objects) with the person describing them using a legible text and axioms that restrict interpretation and give those the terms the characteristic of being well formed. In formal terms, an ontology is a declaration of a logical theory.

In 1995, Guarino and Giaretta conducted a study in which they produced seven definitions in which the concept of ontology had different interpretations:

" *1. Ontology as a philosophical discipline; 2. Ontology as a an informal conceptual system; 3. Ontology as a formal semantic account; 4. Ontology as a specification of a conceptualization; 5. Ontology as a representation of a conceptual system via a logical theory: 5.1 characterized by specific formal properties, 5.2 characterized only by its specific purposes; 6. Ontology as the vocabulary used by a logical theory ; 7. Ontology as a (meta-level) specification of a logical theory* " (Guarino y Giaretta, 1995: 1)

These definitions (with the exception of the one regarding philosophical content) can be classified into two groups: a) those that conceive ontology as a conceptual framework on a semantic level (definitions 2 and 3); b) those that conceive it as a concrete object at syntactic level with a use guided by a specific purpose (definitions 4 to 7). In short, Guarinos opinion is that an ontology is:

" *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* ". (Guarino, 1998: 2)

It is therefore considered that an ontology is an engineering artefact constituted by a vocabulary with a specific meaning (through explicit assumptions), the purpose of which is to describe a part of reality.

In 1996, Bernaras et al proposed the following definition:

" *An ontology provides the means for describing explicitly the conceptualization behind the knowledge represented in a knowledge base* ". (Bernaras et al citado por Gómez-Pérez, 1999: 34)

Apart from the characteristics already mentioned in other definitions - *explicit description* of a *conceptualization* , that contributes *meaning-* , Bernaras et al. Again refer to ontology as an artefact used to represent knowledge in a knowledge base.

For Swartout et al. an ontology is:

" *a set of structured terms that describes some domain or topic. The idea is that an ontology provides a skeletal structure for a knowledge base* ". (Swartout et al 1996).

This definition does not provide any new information on the characteristics that have emerged from the definitions given by the other authors quoted above, and yet it is worthwhile separating the environment from which it arises: the concern for sharing knowledge among systems with a view to reducing costs and difficulties and the explicit connection as part of a knowledge database.

Uschold and Gruninger, on the other hand, take the view that:

" *Ontology is the term used to refer to the shared understanding o some domain of interest may be used as a unifying framework to solveproblems ... An ontology necessarily entails or embodies some sort of world view with respect to a given domain. The world view is often conceived as a set of concepts (e.g. entities, attributes, processes), their definitions and their inter-relations; this is referred to as a conceptualisation* ". (Uschold and Gruninger, 1996: p. 5)

The same authors also insist on the explicit nature of the representation provided by ontology to a conceptualisation.

In 1997, Borst reformulated the definition given by Gruber by specifying that " *Ontologies are defined as a formal specification of shared conceptualization*" (Borst quoted by Gomez-Perez, 1999: p. 33). The contribution made by this definition is to include the idea of *shared* in the notion of *conceptualisation* , when the very nature of the term *shared* contains one of the reasons for the existence of ontologies as artefacts. (Patil et al. 1992; Gruber 1993b; Guarino, Giaretta 1995).

Studer et al take the view that of the many definitions given for ontology, the ones given by Gruber (1993) and Borst (1997) are those that manage to capture the fundamental essence of this concept, combining both in the following definition: " *An ontology is a formal, explicit specification of a shared conceptualisation*" (Studer et al, 1998: 25). The elements constituting this definition are explained by the authors as follows: a) *Conceptualisation* : this refers to an abstract model of a phenomenon in the world arising from having identified the relevant concepts of that phenomenon, b) *Explicit* : this refers to the fact that the type of concept used and the restrictions governing its use are explicitly defined, c) *Formal* : this refers to the fact that an ontology must be able to be read by the computer and d) *Compartida* : this reflects the notion that an ontology captures consensual knowledge that is not the object of a single individual, but accepted by a group.

The above definitions show that a great many possible interpretations exist as regards the concept of ontology, which add a great variety of complementary points of view, even within the same area of knowledge. However, our opinion is that none of these fits in with all the parameters mentioned above, and consequently we propose the following as a synthetic definition: the explicit, formal representation of a shared conceptualisation that involves a perspective of a specific reality, and which is constituted in the conceptual structure of a knowledge base. Furthermore, its ultimate objective, as inferred from the definitions studied, is to share the knowledge it represents.

3.2. Ontologies and Linguistics

In paragraph 3.1 we proposed a definition of ontology for the field of Artificial Intelligence, which we believe is a reference for enabling it to be understood in other fields. Although based on this perspective, ontology is applicable to documentary work, our opinion is that regarding ontology as a linguistic artefact establishes an even closer link with the purposes and functions of the field of Biblioteconomy and Documentation.

Consequently, we base our theory on the fact that ontology offers a formal representation of knowledge, in which concepts, relations and conceptual restrictions are made explicit through

formalisms within a specific domain. If we consider its application in Linguistics, the most frequent function is that of providing support for Knowledge-Based Automatic Translation systems and in Terminology. In both cases, ontology is a formal, explicit representation of the conceptual structure of the field on which work is being performed.

As a result, in the Linguistics field, ontology is one of the modules associated to a knowledge base in which its function is to provide semantic support to words; i.e., words are described as linguistic objects in a lexical database and are related through a conceptual hierarchy located in the ontology. In the event of applying ontology as a semantic support for Terminology, the terms are described in a terminological database and then related to the conceptual structure of the ontology.

In that case, from the linguistics standpoint, ontology is an artefact that can be applied to documentary processing, since it facilitates the conceptual structuring of a specific field of knowledge which coincides in many respects with the knowledge representations used in Biblioteconomy and Documentation for information retrieval.

4. Relation between the concepts of thesaurus and ontology in the field of Biblioteconomy and Documentation

In paragraphs 2 and 3 we have outlined the most important characteristics of thesauruses and ontologies and proposed a conceptualisation for each of these information representation and retrieval tools. Based on the analysis made of the selected definitions, we can conclude that:

- A **thesaurus** is a documentary tool used in the field of information representation and retrieval that represents a field of specific knowledge through its conceptual structure. This conceptual structure provides a semantic organisation by making explicit the conceptual relations and restricting the meaning of the terms that represent them. The field of knowledge is structured based on hierarchical, associative equivalence-based conceptual relations. A thesaurus is used by both professional computer users and end-users.

- An **ontology** is a formal, explicit representation of the conceptual structure of a field of knowledge. Ontology is a semantic support for words that are described as linguistic objects in a lexical or terminological database. The conceptual relations represented in an ontology are extremely varied and depend on the field of knowledge to be structured. An ontology is constructed with the aim of sharing and reusing stored information, which, having been formalised, can be interpreted by both persons and computer programmes.

As we have also indicated above, if we consider ontology as a linguistic artefact, its close relation with the construction of documentary tools for information representation and retrieval is quite evident.

In the process of constructing documentary tools for information representation and retrieval, ontology contributes an explicit declaration of the conceptual relations within a particular field, through the semantic formalisation of that structure.

In turn, semantic formalisation makes it possible to obtain a logical, coherent representation of that conceptual structure, thereby generating a computer interpretation which is focused on interoperability and reuse by other artefacts or applications. Semantic formalisation, therefore, involves translating a message into a syntax that can be interpreted by the computer, facilitating the meaning of the type of relation existing between two or more concepts.

As an exercise of application, we propose formalising the hierarchical relations of the facet of a thesaurus. For this case, we shall take by way of example the *CELLS* facet from the *Tesauro ICYT*

de Biología Animal del CINDOC (Centre of Scientific Information and Documentation). Based on this facet, an analysis is made of the hierarchical relation between the facet title concept and the first-level specificity concepts.

Figure 1 *CELLS facet [4] : extraction of the facet title and subordinate first-level*

specificity concepts

```
CELULAS
  > ADIPOCITOS
  > ASTROCITOS
  > CELULAS DE KUPFFER
  > CELULAS ENDOTELIALES
  > CELULAS EPITELIALES
  > CELULAS GONADALES
  > CELULAS MAMOTROPAS
  > CELULAS MESODERMICAS
  > CELULAS NEOPLASICAS
  > CELULAS SECRETORAS
  > CITOPLASMA
  > ENTEROCITOS
  > ERITROBLASTOS
  > HEPATOCITOS
  > LINEAS CELULARES
  > MACROFAGOS
  > MASTOCITOS
  > MEGACARIOCITOS
  > MEMBRANA PLASMATICA
  > MIOBLASTOS
  > NEURONAS
  > NUCLEO CELULAR
  > PARED CELULAR
  > PINEOCITOS
  > SINOVIOCITOS
  > ZIGOTO
```

In the first place, let us observe the hierarchical relation between the concept "células" as a general term with respect to two of the concepts labelled as their specifics (>), "citoplasma" and "adipocitos". The formalisation of that relation would enable us to observe that they are not subordinated to the general concept based on the same characteristic. While "citoplasma" is considered *a part* of the cell (formally "citoplasma" <part of> "células"), "adipocitos" is really *a type of* cell. Therefore, the formalisation would correspond to "adipocitos" <is one> "cells". Although we are in the presence of two cases of hierarchical relations, they are not of the same type. The relation between "citoplasma" and "células" is a relation based on *all/part* , and the relation between "adipocitos" and "células" is a *gender/species* relation.

The confluence of both characteristics in constructing the same facet poses problems in terms of structural coherence, since in theory, a facet must be constructed based on the application of a single subdivision criterion applied to a specific concept. However, this not only affects the "purist" construction of thesauruses, but also, in more complex cases, it may be an erroneous guide in the search for and/or retrieval of information by the end-user.

For this reason, the semantic formalisation obtained through ontologies has a clear use in verifying the coherence of conceptual relations. It is even worthwhile asking oneself whether 'ontologising' conceptual relations (within the context of documentary languages) would also make it possible for the labelling of those relations to be transferred to the user interface, thereby making the conceptual structure more transparent when carrying out the information search and/or retrieval.

In short, the similarity between both information representation and retrieval tools lies in the fact that they structure determined parts of reality in conceptual terms.

On the contrary, the difference between both information representation and retrieval tools is rooted in the level of abstraction used to construct the conceptual organisation. Ontologies permit a deeper semantic development to be obtained, as they provide a logical, formal description of the information they store and therefore, this can be interpreted by both humans and computers. Thesauruses are tools with less semantic expressivity, and the information they organise can only be used by humans.

To make it easier to understand this difference of abstraction, let us reflect on what we should have to take into account when integrating both concepts, both into an ontological structure and a thesaurus-type structure. [5]

Firstly, integration into an ontological structure requires both concepts to be defined, their inclusion in the basic ontological conceptual categories to be established (e.g. objects, events, properties, etc.), their position in the conceptual hierarchy to be indicated, their conceptual relations to be indicated (which may vary, depending on the area of knowledge), and their properties, and all the information inherited from hierarchically-superior concepts to be specified. It should be mentioned that in an ontological structure, attributes and conceptual relations may be assigned in particular to the concept in question or they may be inherited from a hierarchically-superior concept. Ontology permits multiple inheritances, in other words, each concept is able to receive properties and conceptual relations from more than one concept that is higher up in the hierarchical scale. In addition, if ontology is associated with a terminological database, we would obtain information on linguistic variants and equivalents for the terms related to each concept. Furthermore, all these information parameters specified (definition, categorisation, hierarchy, properties and inheritance) are coded formally and explicitly; put another way, not only will it be possible for them to be understood by humans when reading but they can also be interpreted or decoded by a computer programme, and therefore rendered automatically 'processable'.

Secondly, the integration of both concepts into a thesaurus-type structure requires specifying the main conceptual relations (restricted to three types: hierarchical, associative or equivalence-based), in its optional form an end-note, and in its exceptional form, a definition. These information parameters are coded through words and they are intended to be read by humans. Similarly, establishing conceptual relations is to a certain extent opaque, due to the fact that the type of conceptual relation is indicated by means of a graphic symbol or abbreviation and no distinction is made between the subtypes within each type of conceptual relation (e.g., in hierarchical relations, one subtype would be the all/part relations and another would be the gender/species relations).

These differences of abstraction between ontological and thesaurus-based structures are, in our opinion, those which paradoxically reflect both the closeness and the distance between the concepts of ontology and thesaurus.

5. Conclusions

The growing consolidation and expansion of the digital environment makes the interrelation between different fields of knowledge increasingly more obvious. In the case of Biblioteconomy and Documentation, and more particularly, in the field of information representation and retrieval, few doubts can be harboured with respect to the importance of the contributions made by the fields of Linguistics and Artificial Intelligence.

In this area, in which Biblioteconomy and Documentation converge, Linguistics and Artificial Intelligence, feed-back between thesauruses and ontologies is extremely necessary.

Our thoughts should focus on the reuse of resources, in which the efforts put into gathering and systemising large volumes of information are not limited to the construction of isolated resources, but ensuring that this information is structured in such a way that it can be used for the generation or enriching of other resources. In this regard, we propose the use of an ontology, the structure of which will enable specific domains to be formalised and the necessary semantic support to be offered in proposing a thesaurus model. This link would make it possible to re-dimension thesauruses as tools in information representation and retrieval, with special emphasis on the coherence and relational enrichment of the conceptual structure.

In this article, we have merely presented as an example a small part of the semantic formalisation of hierarchical relations. One future field of work would be to study in depth the formalisation of hierarchical relations, and to conduct research into the formalisation of associative relations, equivalence-based relations or relations pertaining to a specific field of knowledge.

Another element into which research could be made would be the integration of the "scenario" concept, insofar as the semantic relations of a specific conceptual structure could be analysed, based on user-profiles.

6. References

ANSI/NISO Z39.19 - 2003 Guidelines for the Construction, Format, and Management of Monolingual Thesauri . [en línea] Bethesda: NISO. http://www.niso.org/standards/resources/Z39-19.pdf [Consulta: 19 marzo 2005]

A rano, S. (2004 *).* "La ontología: una zona de interacción entre la Lingüística y la Documentación ". [en línea]. EN: *Hipertext.net* (2). http://www.hipertext.net [Consulta: 19 marzo 2005]

Barité, M. (1997). *Glosario sobre Organización y Representación del Conocimiento. Clasificación, Indización y Terminología* . Montevideo: Escuela Universitaria de Bibliotecología y Ciencias Afines.

Bechhofer, S.; Goble, C. (2001). "Thesaurus construction through knowledge representation". [en línea]. EN: *Data & Knowledge Engineering* 37(2).25-45. http://www.sciencedirect.com/science [Consulta: 19 marzo 2005]

Cabré, M. T.; Feliu, J.; Vivaldi, J. (2004a). "Base de connaissances GENOMA: le rôle de lontologie". EN: *Journée détude Terminologie, Ontologie & Representation des Connaissances.* Lyon: Université Jean-Moulin Lyon 3. 19-25.

Cabré, M. T.; Bach, C.; Estopà, R.; Feliu, J.; Martínez, G.; Vivaldi, J. (2004b). "The GENOME-KB project: towards the integration of concepts, terms, textual corpora and entities". EN: *LREC 2004, Fourth international conference on language resources and evaluation: held in memory of Antonio Zampolli.* Paris: ELRA. 87-90.

García, A. (2004). "Instrumentos de representación del conocimiento: tesauros versus ontologías". [en línea]. EN: *Anales de Documentación* (7). 79-95. http://www.um.es/fccd/anales/ad07/ad0706.pdf [Consulta: 19 marzo 2005]

Gilchrist, A. (2002). "Thesauri, taxonomies and ontologies an etymological note". EN: *Journal of Documentation* 59 (1). 7-18.

Gómez-Pérez, A. (1999). *Ontological engineering: a state of the art.* [en línea] Pennsylvania: School of Information

Sciences and Technology(IST). Pennsylvania State University. http://citeseer.nj.nec.com/444416.html [Consulta: 19 marzo 2005]

Gruber, T. R. (1993a). *A translation approach to portable ontology specifications* . [en línea] http://gicl.mcs.drexel.edu/people/regli/Classes/KBA/Readings/KSL-92-71.pdf [Consulta: 19 marzo 2005]

Gruber, T. R. (1993b). *Toward principles for the design of ontologies used for knowledge sharing.* [en línea]. Pennsylvania: School of Information Sciences and Technology(IST). Pennsylvania State University. http://citeseer.nj.nec.com/gruber93toward.html [Consulta: 19 marzo 2005]

Gruber, T. R. (2001). *What is an Ontology?* [en línea]. Stanford: Knowledge Systems Laboratory. Stanford University. http://www-ksl.stanford.edu/kst/what-is-an-ontology.html [Consulta: 19 marzo 2005]

Guarino, N. (1998). *Formal ontology and information systems.* [en línea].Trento, Roma: Laboratory for Applied Ontology (LOA). http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/FOIS98.pdf [Consulta: 19 marzo 2005]

Guarino, N. (1995). *Formal ontology, conceptual analysis and knowledge representation.* [en línea] Trento, Roma: Laboratory for Applied Ontology. http://www.loa-cnr.it/Papers/FormOntKR.pdf [Consulta: 19 marzo 2005]

Guarino, N.; Giaretta, P. (1995). *Ontologies and knowledge bases: towards a terminological clarification.* [en línea]. Trento, Roma: Laboratory for Applied Ontology. http://www.loa-cnr.it/Papers/KBKS95.pdf [Consulta: 19 marzo 2005]

Guerrero, V.; Lozano, A. (1999). "Vínculos entre las Ontologías y la Biblioteconomía y Documentación". EN: *La Representación y la Organización del Conocimiento en sus distintas perspectivas: su influencia en la Recuperación de la Información (Actas del IV Congreso ISKO-España EOCONSID99, 22-24 de abril de 1999 en Granada), 1999.* Granada: ISKO-Facultad de Biblioteconomía y Documentación. 25-31.

ISO 2788:1986 (1986). *Documentation -- Guidelines for the establishment and development of monolingual thesauri.* 2 nd ed. Geneva: ISO.

López-Huertas, M. J. (1999). "Potencialidad evolutiva del tesauro: hacia una base de conocimiento experto". EN: *La Representación y la Organización del Conocimiento en sus distintas perspectivas: su influencia en la Recuperación de la Información (Actas del IV Congreso ISKO-España EOCONSID99, 22-24 de abril de 1999 en Granada), 1999.* Granada: ISKO-Facultad de Biblioteconomía y Documentación. 133-140.

López-Huertas, M. J. (1997). "Thesaurus structure design: a conceptual approach for improved interaction". EN: *Journal of Documentation* 53(2). 139-177.

McGuinness, D. L. (2002). *Ontologies come of age* . [en línea] Stanford: Stanford University. http://www.ksl.stanford.edu/people/dlm/papers/(continua) ontologies-come-of-age-mit-press-(with-citation).htm [Consulta: 19 marzo 2005]

Patil, R. et al. (1991). *The DARPA Knowledge sharing effort: progress report* . [en línea] Pennsylvania: School of Information Sciences and Technology(IST). Pennsylvania State University. http://citeseer.nj.nec.com/patil93darpa.html [Consulta: 19 marzo 2005]

Qin, J.; Paling, S. (2001). "Converting a controlled vocabulary into an ontology: the case of GEM". [en línea]. EN: *Information Research* 6(2). http://informationr.net/ir/6-2/paper94.html [Consulta: 19 marzo 2005]

Sánchez, E.; Martínez, R. (2002 *). Las ontologías y su aplicación en el ámbito de la documentación.* [en línea]. València: Universitat Politècnica de València. http://ttt.upv.es/ccarrasc/doc/2000-2001/Ontolog%EDas/Index.htm [Consulta: 19 marzo 2005]

Shiri, A. A.; Revie, C. (2000). "Thesauri on the Web: current developments and trends". EN: *Online Information Review* 24(4). 273-279. http://dlist.sir.arizona.edu/archive/00000163/01/thesauri.pdf [Consulta: 19 marzo 2005]

Slype, G. Van (1991). *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales* . Madrid, Salamanca: Fundación Germán Sánchez Ruipérez.

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

Studer, R. et al. (1998). *Knowledge engineering: principles and methods* . [en línea] Pennsylvania: School of Information Sciences and Technology(IST). Pennsylvania State University. http://citeseer.nj.nec.com/225099.html [Consulta: 19 marzo 2005]

Swartout, B. et al. (1996). *Toward distributes use of large-scale ontologies* . [en línea] Calgary: University of Calgary. http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html [Consulta: 19 marzo 2005]

Tudhope, D.; Alani, H.; Jones, Ch. (2001). "Augmenting thesaurus relationships: possibilities for retrieval". [en línea] EN: *Journal of Digital Information* 1(8). http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Tudhope/index.htm [Consulta: 19 marzo 2005]

Tramullas, J. (1999). "Agentes y ontologías para el tratamiento de la información: clasificación y recuperación en Internet". EN *: La Representación y la Organización del Conocimiento en sus distintas perspectivas: su influencia en la Recuperación de la Información (Actas del IV Congreso ISKO-España EOCONSID99, 22-24 de abril de 1999 en Granada), 1999.* Granada: ISKO-Facultad de Biblioteconomía y Documentación. 247-252

UNE 50106:1990. (1990). *Directrices para el establecimiento y desarrollo de tesauros monolingües.* Madrid: AENOR.

Uschold, M.; Gruninger, M. (1996). *Ontologies: principles, methods and applications* . [en línea] Pennsylvania: School of Information Sciences and Technology(IST). Pennsylvania State University. http://citeseer.nj.nec.com/uschold96ontologie.html [Consulta: 19 marzo 2005]

Z39.19 TAG Conference Call, June 30, 2003. (2003). En: NISO. *Developing the next generation of standards for controlled vocabularies and thesauri.* [en línea] Bethesda: NISO. http://www.niso.org/committees/MTinfo.html [Consulta: 28 febrero 2005]

Silvia Arano is member of the DigiDoc Group of the University Institute of Applied Linguistic of the University Pompeu Fabra . This article presents a part of the results of the HUM 2004-03162/filo project of National Plan I+D+I of the Ministry of Education and Science (Spain)

7. Notas

[1] Article published within the framework of the Project funded by the Ministry of Education and Science, ref. HUM2004-03162/FILO "Web Semántica y Sistemas de información documental", coordinated by Dr. Lluís Codina Bonilla (DIGIDOC - IULA UPF) [volver]

[2] The UNE guidelines, published in Spain by AENOR, are the translation into Spanish of the guidelines published by the ISO (International Organization for Standarization) [volver]

[3] The NISO (National Information Standards Organisation) is an institution that develops, maintains and publishes technical guidelines within the sphere of information management in the United States of America. [volver]

[4] The Italic and underlined text markers used to indicate the subordinated concepts have been added by us and are intended to highlight the conceptual differences between the subordinated concepts and the concept constituting the facet title. [volver]

[5] In both cases, a generic-type conceptual structure is used. [volver]