

Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives

Juan-Antonio Pastor-Sanchez, Francisco Javier Martínez Mendez and José Vicente Rodríguez-Muñoz.

Department of Information and Documentation. University of Murcia. Campus of Espinardo, Murcia, SPAIN.

Abstract

Introduction. This paper presents an analysis of the Simple Knowledge Organization System (SKOS) compared with other alternatives for thesaurus representation in the Semantic Web.

Method. Based on functional and structural changes of thesauri, provides an overview of the current context in which lexical paradigm is abandoned in favour of the conceptual paradigm. Likewise are described briefly various initiatives for the representation of thesauri using RDF vocabularies and hence for application to the Semantic Web, with particular attention to SKOS.

Analysis and Results. This brief comparison will allow us to know some highlights of these proposals to raise those transcendental aspects in a model representing thesauri. SKOS includes the main proposal of this model, which organizes the concepts into diagrams and libraries. The notions of descriptors and non-descriptors terms are replaced by the association to the concepts of preferred and alternative labels and may be defined hierarchical or associative relationships between concepts. SKOS also provides for the establishment of correspondence relations between concepts belonging to different conceptual schemes.

Conclusions. In accordance with the extent to which this series of requirements is fulfilled, we propose and defend the SKOS option as the best alternative for the development of multiple applications focusing on the use of thesauri in Web-based information services and systems, highlighting the advantages of this model over others proposed and in consideration of the user perspective in management, search and information query operations.

CHANGE FONT

Introduction

The concept of thesaurus has evolved from a list of conceptually interrelated words to today's controlled vocabularies, where terms form complex structures through semantic relationships. This term comes from the Latin and has turn been derived from the Greek "θησαυρός", which means *treasury* according to the Spanish Royal Academy, in whose dictionary it is also defined as: 'name given by its authors to certain dictionaries, catalogues and anthologies'. The increase in scientific communication and productivity made it essential to develop keyword indexing systems. At that time, Howerton spoke of controlled lists to refer to concepts that were heuristically or intuitively related. According to Roberts (1984), Mooers was the first to relate thesauri to information retrieval systems; Taube established the foundations of post-coordination, while Luhn dealt, at a basic level, with the creation of thesauri using automatic techniques. Brownson (1957) was the first to use the term to refer to the issue of translating concepts and their relationships expressed in documents into a more precise language free of ambiguities in order to facilitate information retrieval. The *ASTIA Thesaurus* was published in the early 1960s (Currás, 2005), already bearing the characteristics of today's thesauri and taking on the need for a tool to administer a controlled vocabulary in terms of indexing, thereby giving rise to the concept of documentary language.

Gilchrist defined a thesaurus as:

...a lexical authority list, without notation, which differs from an alphabetical subject heading list in that its lexical units, being smaller and more amenable, are used in coordinate indexing' (Gilchrist 1971: 11)

Almost simultaneously, another author, Wersig, gives another definition:

...lists of terms, previously prefixed, although extracted from the text of documents themselves and replicating concepts in simple units that are post-coordinated to avoid ambiguity. They are interrelated by hierarchical, associative and equivalence relationships. (Wersig 1971: 79)

Both authors agree in affirming the simplicity of the thesaurus elements and their coordination after indexing, although Wersig in particular emphasises the existence of semantic relationships among these units. Standardisation efforts increased in the 1980s with the appearance of the second edition of the standard ISO 2788:1986 (ISO, 1986) on monolingual thesauri, which defines thesaurus as:

...the vocabulary of a controlled indexing language, formally organized with the aim of state explicitly the existing relationships between concepts.

The following year, Aitchison and Gilchrist (1987) specifically introduced the role of thesaurus in information retrieval processes. The review performed by Miller (1997) as to the nature of thesauri as opposed to classification schemes is particularly interesting. It demonstrates that their functional nature conditions to a great extent their origin and actual experiences of application, focusing their evolution in one direction or another and making it at times impossible to draw a distinction between thesauri and conceptual schemes, meaning that it is not possible in all cases to restrict the field of application of the thesaurus solely to information retrieval (leaving knowledge organization to classification schemes). This aspect becomes even more important when it comes to the semantic Web, because it combines statistical information recovery techniques and the representation of information using metadata (Díaz Ortuño 2003) that are organized into structures defined by conceptual schemes (García Jiménez 2004). Nor should we overlook Shiriet *al.* (2002), whose study analyses the search strategies of different groups of users, demonstrating that the expansion of queries through thesauri and their integration in information retrieval is completely viable.

From a functional perspective, the thesaurus is a documentary language that uses a controlled vocabulary in order to solve the ambiguity issues of natural language when it comes to indexing and information retrieval processes. Thesauri are based on essentially lexicographical instruments and evolve towards systems focusing on the organization of information and the representation of the content of a documentary corpus, normally used for term extraction. Thesaurus-preferred terms are used in indexing processes and in the selection of terms for information searches, thereby increasing the communication capacity between the user and the information

retrieval system. They are a nexus of natural language and indexing. Usually, a general thesaurus offers a less specific domain knowledge representation than a specific thesaurus, where the focus is on the representation of the content of a particular documentary corpus. While exploring the thesaurus structure, the user may establish the search terms. This selection process is open to feedback, broadening, refining or expanding the constituent terms of the query.

Evolution of the thesaurus concept

Print editions of thesauri are no longer of use because of the increase in the volume of information in digital format, undermining the application of such static thesauri and making essential their adaptation to the Web environment. The process of integrating thesauri with information retrieval systems started in the early 1990s. The first projects dealt with the creation and maintenance of thesauri based on their representation using data models: the entity-relationship model ([Rodríguez Muñoz 1990](#) and [1992](#)) and the relational scheme ([Jones 1993](#)). Subsequently they were included in distributions of bibliographic databases (ERIC, INSPEC and MEDLINE, among others) as an assistance tool in the selection of terms for search queries. The relationship between hypertext and the thesaurus was soon identified, with Rada ([1991](#)) proposing a system of information prepared collaboratively within a hypertext environment, creating a network of links based on concepts contained in documents. Pastor and Saorín ([1993](#)) likewise had the productive idea of using hypertext as a framework for the development of applications to administer and query thesauri, subsequently proposing their '*documentary hypertext*' ([Pastor and Saorín 1995a](#) and [1995b](#)), with Rada's parallel structure taking the form of a semantic network represented by a thesaurus, within a working environment based on hypertext and subsequently expanded towards personal environments for the comprehensive management of information ([Pastor and Saorín 1998](#)).

Järvelin *et al.* ([1996](#)) produced a deductive model for the expansion of queries based on concepts using three levels of abstraction: conceptual, linguistic and occurrences. Concepts and their relationships are positioned at the conceptual level, while the linguistic level represents concepts through natural language, the expressions of which may have various representations at the level of occurrences. We are here faced with a vision of the thesaurus in which conceptual aspects prevail over lexical, giving them greater flexibility and adaptation ability, in line with the thesis upheld by López-Huertas ([1999](#)), for whom the thesaurus holds considerable potential for evolving from a mere lexical resource towards a powerful instrument for conceptual representation.

The appearance of the Web dramatically impacts the application of thesauri. Before its emergence, thesauri were effective in controlled environments with clearly defined structures and pathways for accessing information. On the Web, the heterogeneity of formats and structures, combined with the huge growth of resources and content, makes it impossible to apply this working model. The access pathways leading to elements of information are not delimited or established in advance and the constant updating of content makes it practically impossible to apply thesauri directly to the Web. It is also important to remember the practical non-existence of tools to represent thesauri, along with their management, usage and integration, in a distributed manner. Initiatives such as Web page indexes or directories (for example, Yahoo!, DMOZ and the Open Directory Project) have been overwhelmed by the evidence that it is practically impossible to exert control by applying a thesaurus or other type of tool based on manual indexing. The finally adopted solution has been to reduce the structure of the directory or otherwise its transformation into a search engine (as Yahoo! did in 2004). While thesauri have adapted to their publication and consultation in digital environments, their effective exploitation on the Web still leaves much to be desired for, among others, the following reasons ([Shiri and Revie 2000](#)), ([Greenberg 2004](#)), ([Roe and Thomas 2004](#)), ([Laguens García 2006](#)):

- Limited development and delimitation of the conceptual orientation of thesauri, with their lexical vision prevailing. The application of lexical thesauri within a dynamic environment such as the Web did not provide adequate results in comparison to the efforts devoted to indexing processes.
- Initial absence of adequate standards and models to represent different levels of abstraction on the Web (such as XML (the Extensible Markup Language), RDF (the Resource Description Framework) or SKOS (the Simple Knowledge Organization System))
- Access to Web development technologies (Java, PHP, MySQL) was sometimes costly or non-existent. Developments were confined to corporate environments with eminently pragmatic purposes, neglecting research, experimentation and the study of information retrieval dynamics using alternative tools.
- The initial and rapid success of search engines, which proved themselves effective by employing automatic Web crawling and index construction techniques, as a panacea which in many cases is difficult to criticise.
- The vision of the thesaurus as an obsolete tool or one of the limited usefulness, which need not necessarily have been the case.

Shiri and Revie ([2000](#)) advocate expanding the concept of the thesaurus, making it permeable to other tools and proposing new lines of work, such as the creation of navigation systems, and all this within an environment where the role of metadata will be essential for the re-use and exchange of thesauri. The function of thesauri within digital environments goes beyond a change of platforms, involving a genuine redefinition of the conceptual principles of such tools ([Arano and Codina 2004](#)). Tudhope ([2001](#)) suggested the possibility of enriching the specification and semantics of RT relationships (while maintaining compatibility with traditional thesauri) through a limited hierarchical extension of the associative and hierarchical relationships (expanding the information retrieval capabilities), in a solution closer to ontologies, which [Arano](#) (2005) sees as offering new possibilities for the redesign of documentary languages.

The expansion of the types of relationship would allow more adaptable and reusable developments from the perspective of different knowledge domains and over time ([García Torres *et al.* 2008](#)), as ontologies have a greater expressive capacity for modelling information systems. Inevitably, the future of the thesaurus on the semantic Web necessarily involves permeability with other proposals (ontologies and topic maps) or their combined usage, in establishing a paradigm shift in the creation of thesauri, focusing on conceptual thesauri ([Matthews *et al.* 2001](#)). They could also be applied in fields which appear alien to the idea of the thesaurus as a tool for terminological control (*folksonomies*), providing mechanisms for the creation of semi-controlled vocabularies the elements of which would be structured to facilitate query and usage in information retrieval tasks.

Background of thesaurus representation on the Web with the Resource Description Framework (RDF)

There have been several projects aimed at representing thesauri and conceptual schemes in XML format to be applied in organizations and in specific contexts. Some of these (Zthes (a specification for thesaurus representation), MeSH (Medical Subject Headings) and [Topic Maps](#)) have been long-lived. Nonetheless, the trend is to use the model proposed by RDF and its corresponding coding in XML for this purpose, as a logical evolution of the different study approaches for representing thesauri in the semantic Web. This shift is the result of the significant possibilities for integration, reuse and expansion offered by RDF and the Web Ontology Language (OWL) within the context of the semantic Web. An increasing number of initiatives are adopting RDF to represent information on the Web and using XML to code their information representation models (because a part of the potential of this model is based on the possibility of combining different specifications). The development of an RDF/XML vocabulary to represent thesauri means going beyond the model proposed by standards ISO 2788:1986 ([ISO 1985](#) and [1986](#)) and ANSI/NISO Z39.19 ([ANSI 2005](#)) for the construction and maintenance of thesauri. These standards describe the thesaurus as a set of terms of various kinds, among which semantic relationships are established, whereas the Web requires a broader vision which goes beyond the idea of a term as the central element of the thesaurus and expands the number, type and meaning of existing relationships. Many initiatives have therefore focused on the definition of RDF vocabularies to represent thesauri ([Matthews and Miles 2001](#)), some of which preserve at their core the idea of the term, while others introduce the notion of a concept to which the corresponding terms are assigned. Below we give a brief descriptive summary of some of these initiatives.

Language Independent Metadata Browsing of European Resources (LIMBER)

LIMBER is a proposal developed by Miller and Matthews ([2001](#)). Its aim is the use of RDF to build a comprehensive information system employing a thesaurus manager to index resources. As RDF is present throughout the system's elements, an RDF vocabulary was also developed to represent thesauri. In LIMBER, the elements of a thesaurus are *concepts*, *terms* and *scope notes*. Semantic equivalence relationships can be established among concepts and lexical relationships between concepts and terms. The semantic relationships can be hierarchical, associative or can indicate where a concept represents the starting point of a hierarchy (for which there is therefore a sub-class of concept). The equivalence relationships allow indicating whether they are exact equivalences, inexact, partial or one to many. There

are various types of scope notes and it is possible to develop multilingual thesauri by associating a language code both to concepts and to scope notes. The Institute for Learning & Research Technology developed an RDF vocabulary in the year 2000 with the aim of building conceptual schemes to help maintain information networks for research. The elements of this vocabulary are *concepts*, *terms*, *scope notes* and *term usage type*. The relationships between concepts may be associative or generic (specific relationships are inferred by the application). The terms, which may be assigned a language code, are classified as preferred or non-preferred and associated with concepts.

California Environmental Resources Evaluation System

The California Environmental Resources Evaluation System (CERES) is an information system developed by the California Resources Agency for access to electronic resources on the Environment. In partnership with the NBII (National Biological Information Infrastructure) a thesaurus was drawn up, designing a RDF specification for its publication on the Web. This specification is a fairly faithful expression of the concept of thesaurus under the ISO and ANSI/NISO standards. The vocabulary identifies terms which are in turn divided into categories, descriptors and entry terms (non-descriptor). There are appropriate properties to represent the relationships of each term type. For example, for a descriptor the available properties are SN (for scope notes), BT (for generic relationships), RT (for associative relationships) and UF (for non-descriptor relationships).

The Gateway to Educational Materials (GEM)

GEM is another RDF vocabulary, likewise based on the structure proposed by the ISO and ANSI/NISO standards, put forward by [the GEM Consortium](#). This vocabulary, created in 2001, limits itself to defining the different types of thesaurus relationship in the form of properties. There is no definition of any other type of element, leaving it open to the possible use of a complementary vocabulary.

Center for Army Lessons Learned (CALL) thesaurus ontology

[This vocabulary](#) was developed by Dynamics Research Corporation in 2002 within a project for the representation of a thesaurus of more than 18,000 terms for the Center for Army Lessons Learned, Fort Leavenworth, Kansas. This is a [DARPA Agent Markup Language](#) (DAML) ontology, defined with RDF. The thesauri applying this vocabulary are structured in terms that are associated with properties representing semantic relationships. The semantic restrictions of the thesauri are defined by making use of the DAML characteristics.

The Food and Agriculture Organization AGROVOC thesaurus

The KAON (Karlsruhe Ontology) model, developed by the [University of Karlsruhe, Information Technologies Research Centre](#) in 2002, served as the basis for development of an application for the representation of the AGROVOC thesaurus developed by the Food and Agriculture Organization and the Commission of the European Communities. Each term is modelled as an RDF class (`rdfs:Class`), with hierarchical relationships represented with the RDF subclass property (`rdfs:subClassOf`). Properties are defined to represent associative relationships, preferred terms and equivalent terms. The label associated with each term is the other element of this model which includes properties to represent the label text string, the associated concept and the language used.

[European Treasury Browser](#) Multilingual Educational Thesaurus (ETBT)

This is an RDF scheme based on the concept of labelled nodes, in other words both terms and relationships and the thesaurus itself, are nodes labelled differently to distinguish each individually. Although this is a multilingual thesaurus, in fact ETBT works with the sum of various monolingual thesauri. The types of node defined represent the thesaurus as a whole (Thes), each of the monolingual thesauri (Tmono), semantic fields (MT), term nodes (MTNode), nodes of equivalence between languages (ENode), non-preferred terms (UNode) and semantic relationships (RNode). Around each thesaurus node are organized the remaining nodes corresponding to each of the monolingual thesauri. Each monolingual thesaurus has associated a series of semantic field notes which in turn serve to structure the term and semantic relationship nodes.

Characteristics of a thesaurus representation model for the semantic Web

The RDF vocabularies proposed for representing thesauri provided various solutions to the same problem because of differences between the proposed conceptual models, being the main element a general predominance of term-based thesauri. However, this classical vision could not be entirely appropriate for the Semantic Web as this contingency involves a series of difficulties in thesaurus maintenance and document indexing processes. We will now develop on certain aspects that involve an evolution from the conceptual model of the thesaurus and offer a more novel approach that is more adaptable to the information retrieval possibilities offered by the Semantic Web.

Concept-based thesauri and term-based thesauri

This is perhaps the key aspect in the development of a thesaurus representation model and will be decisive in the corresponding development of the RDF/XML vocabulary. There are two approaches to the consideration of the central core of the thesaurus around which the remaining elements are structured. The first is to consider a thesaurus as a set of terms interconnected with a network of different types of semantic relationship. This is the classical representation given by ISO and ANSI/NISO standards, which describe a thesaurus as a set of terms related by different hierarchical relationships (generic term and generic term), associative relationships (related term) and equivalence relationships (equivalent term, preferred term). The second alternative conceives the thesaurus as a set of concepts related by semantic relationships similar to those of a term-based thesaurus. The concepts are tied to terms, in the form of lexical labels, which may or may not have lexical relationships established among them.

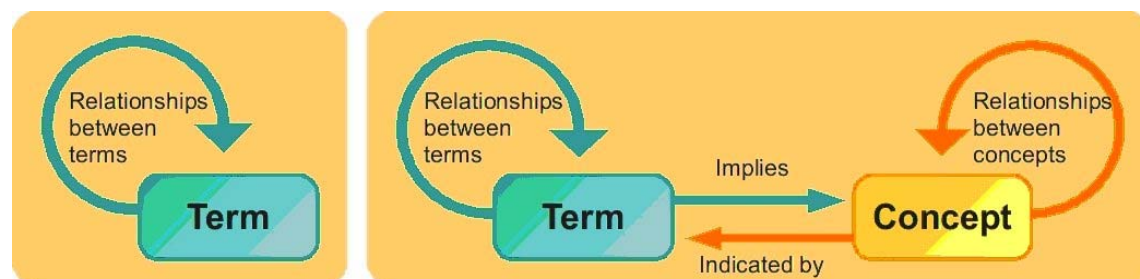


Figure 1: Term-based thesaurus compared with concept-based thesaurus.

In a term-based thesaurus the semantic relationships are established between the terms themselves. Meanwhile, in a concept-based thesaurus, certain semantic relationships are established between concepts and others between terms. The relationships between concepts generally include those which develop the hierarchical and associated structures, while the latter denote the relationships between terms associated with one single concept (abbreviations, common usage, technical language, etc.). A term-based thesaurus is thus structured at one single lexical-terminological level, whereas a concept-based thesaurus has a three-level structure: (a) conceptual level, where concepts are identified and their interrelationships established; (b) terminological correspondence level, where terms are associated (preferred or non-preferred) to their respective concepts and (c) lexical level where lexical relationships are defined to interconnect terms.

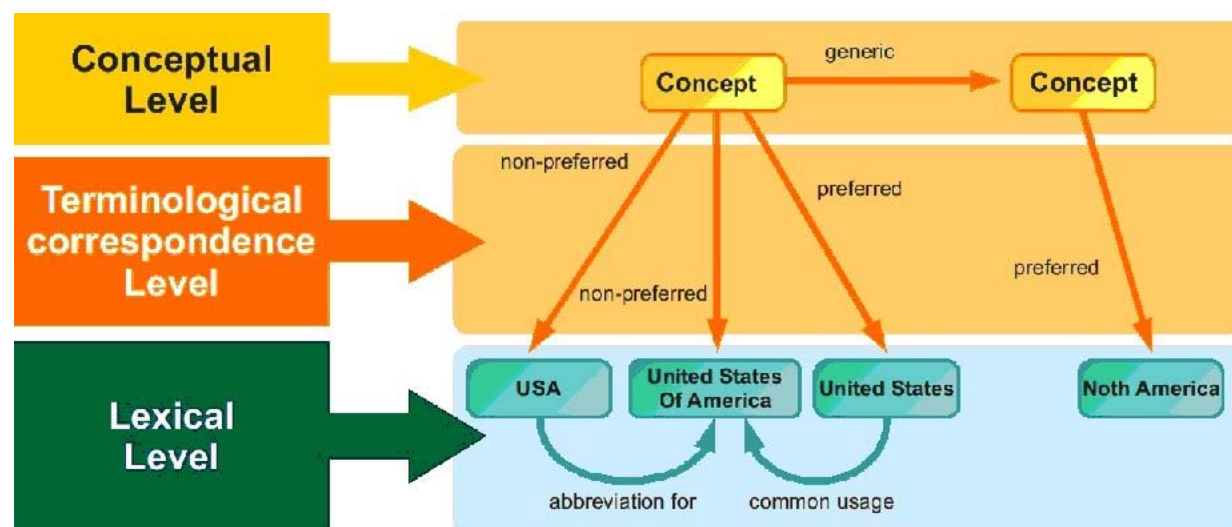


Figure 2: organizational levels of a concept-based thesaurus.

Term-based thesauri enable the representation of more complex schemes close to the traditional model. Concept-based thesauri allow more precise descriptions of the information structures, although they are not initially as intuitively comprehensible. This paradigm shift involves a considerable effort, compensated for many times over through the optimisation of the processes of preparation, maintenance and products derived from application of the thesaurus in the indexing of documents. Term-based thesauri involve problems of application in indexing when performing operations to change preferred and non-preferred terms. Thus, if a descriptor which has been used to index a document becomes a non-descriptor, the indexes linking that descriptor to the documents must be reorganized. This does not occur in the case of conceptual thesauri because the indexing processes are independent of terminological changes to the thesaurus. In this case, the documents are associated with concepts, not with terms; changes involving preferred and non-preferred terms do not impact on indexing.

Grouping structures

Occasionally we may wish to create groups or collections of concepts to establish concept families and provide the thesaurus with a greater level of semantic content, or allow complementary organizational criteria to be defined. This grouping option enables the organization of concepts by thematic areas that also act as a point of access for consulting the thesaurus. Concept groups may be labelled with expressions denoting the meaning of these structures. Another possibility is that of grouping concepts into facets, allowing them to be characterised according to multiple criteria while maintaining independent associative and hierarchical structures. In any case, it must be remembered that these groupings do not constitute elements of sufficient individual entity in order to establish semantic relationships with concepts or with other group structures. This is because of the difference in the granularity of the two elements, meaning that the establishment of semantic relationships between them could introduce distortions into the thesaurus structure.

Lexical relationships between labels

Labels and the lexical relationships between them are the elements which shape the lexical level of a thesaurus. Although the classical vision of a thesaurus does not include lexical relationships, they are of great interest because of the value-added information they provide. The establishment of lexical relationships will allow for a more precise description of the knowledge to be represented regarding the thesaurus domain. This precision will allow certain operations to be performed automatically, such as for example the disambiguation of the terms entered by users in queries during search processes.

Relationships as arcs or as nodes

Semantic or lexical relationships can be represented as arcs or as nodes. The arc-based solution is much more compact and intuitive than the node-based alternative. Meanwhile, node-based relationships involve an added difficulty in the development and maintenance of this type of thesaurus, as the tasks of creating and interpreting trees based on this approach are generally complex and confusing. Nonetheless, the relations expressed in arcs have the advantage of being able to be referenced using URIs. This type of relationship can be applied to those established at either a conceptual or a lexical level. Some RDF thesaurus representation vocabularies apply relationships as nodes, as in the case of the European Treasury Browser thesaurus.

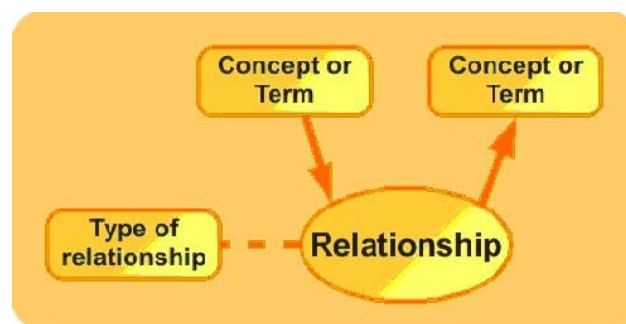


Figure 3: Term-based thesaurus compared with concept-based thesaurus.

Multilingual structures

Bearing in mind the scope of application of RDF-expressed thesauri (the semantic Web), it is essential to give consideration to mechanisms enabling to express their multilingual reality. It is therefore perfectly feasible to associate one concept with different preferred labels for each language. Another advantage of concept-based thesauri is the possibility of establishing lexical relationships adapted to the terminological reality of each language.

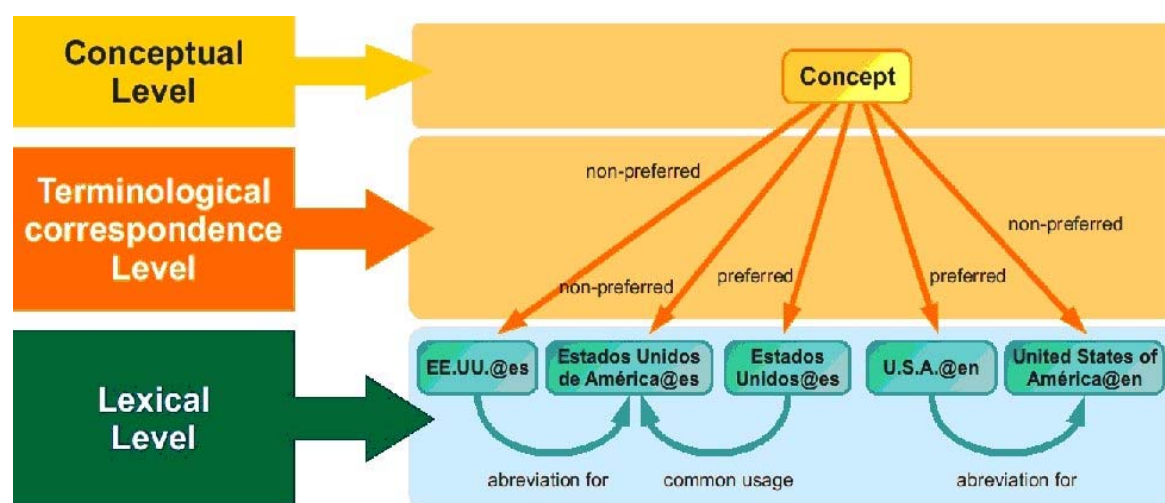


Figure 4: Independence of conceptual levels and adjustment of lexical levels to multilingual structures

Semantic restrictions

Semantic restrictions control aspects connected with the structural coherence of the thesaurus. Some are optional, while others are essential in order to guarantee proper development of the thesaurus-building process. This type of restriction is associated with processes and tasks for the control of poly-hierarchical structures, integrity at terminological correspondence level, the influence of symmetrical and inverse properties, the control of cycles in hierarchical structures or the control of the disjunctive property between associative and hierarchical relationships.

Expansion of relationship types

One of the advantages of applying RDF in the representation of thesauri is the possibility of expanding relationship types with OWL or simply with RDF/RDFS (Reynolds *et al.* 2005). This would allow for the creation of a new relationship by defining some of its properties such as transitivity, symmetry, inverse relationship or reflexivity, rigorously declaring the required semantic restrictions. This expansion would as well increase the likelihood of the successful application of thesauri for information retrieval on the Web. It furthermore represents a significant synergy in aspects related to the evolution and adaptation of the concept of thesaurus. The hierarchical and associative relationships can be expanded or their characteristics modified in the future, or customised for specific projects.

Equivalence among thesauri

It is highly desirable to be able to establish equivalences among concepts belonging to different thesauri. To do this we must define various types of equivalence, as the correspondence between concepts is not always precise, but may involve more specific or generic concepts. It is also possible to establish equivalences to associate certain aspects of the two concepts without necessarily meaning that one concept is more generic or specific than another. These relationships will be similar to those established between concepts belonging to a single thesaurus. This mapping allows the reuse or integration of external thesauri. The correspondence between concepts allows queries to be transformed from one system to another, or to be complemented, as one could consult information repositories indexed using a different conceptual scheme to that used at first instance by the user.

Definitions, explanations, scope notes and metadata in concepts and relationships

Classical scope notes are highly limited when it comes to express value-added information, which is particularly helpful in the use of thesauri on the Semantic Web. Such content may include definitions, notes of various kinds, information about changes in the thesaurus structure or the meaning of terms, usage examples, etc. This information, generally not included in printed thesauri, could be simply and agilely queried using a Web-based information system. Along with the types of content referred to earlier, we may wish to add metadata using Dublin Core (2008) to include data on authorship and edition associated with concepts, labels and relationships.

Simple Knowledge Organization System

The Simple Knowledge Organization System (SKOS) is a World Wide Web Consortium initiative in the form of an RDF application providing a model to represent the basic structure and content of conceptual schemes as header lists of subject matter, taxonomies, classification schemes, thesauri and any type of controlled vocabulary. The origins of the project date back to the preparation of a thesaurus of activities within the SWAD-Europe project:

SKOS Core was developed as draft of an RDF Schema for thesauri compatible with relevant ISO standards. Further work extended it to multilingual thesauri and mappings between thesauri and developed some pilot tools ([Alistair et al. 2005](#)).

The first version of SKOS Core was presented in 2003. In the System, concepts are identified with Uniform Resource Indicator references; these concepts can be labelled in text strings in one or more languages, documented and then structured using various types of semantic relationship. The model is capable of mapping concepts of different schemes and defining ordered collections and concept groupings. It can also establish relationships between the labels related to the concepts.

The use of RDF in developing SKOS allows it to provide documents in a format that is legible in computer applications, as well as their exchange and publication on the Web. SKOS was designed to create new organizational systems or to migrate those in existence to the semantic Web in a quick and easy manner. It provides a simple vocabulary and an intuitive model which can be used together with the Web Ontology Language (OWL) or independently. SKOS represents a mid-way step between the low level of structuring seen presently on the Web and the rigorous descriptive formalism of ontologies defined with OWL. The structure of SKOS is described in a series of documents including SKOS Primer ([W3C 2008b](#)), SKOS Reference ([W3C, 2008a](#)) and SKOS Use Cases and Requirements ([W3C 2007](#)). While SKOS is at an initial stage of development, the essential core work is already established, along with its basic vocabulary.

The SKOS data model

This model is an ontology defined with OWL Full. Being based on RDF, SKOS structures the data in the form of triples which can be coded in any syntax valid for RDF. SKOS can be used together with OWL to give formal expression to structures of knowledge regarding a specific domain, as SKOS cannot perform this function because it is not a language for formal knowledge representation ([Alistair et al. 2005](#)). Knowledge described explicitly as a formal ontology is expressed as a set of axioms and facts. However, a thesaurus or any type of classification scheme does not include this form of affirmation, but rather identifies and describes (with natural language or non-formal expressions) ideas or meanings which we refer to as concepts. These concepts can be organized into structures which lack formal semantics and cannot be considered as axioms or facts. In other words, a thesaurus only provides an intuitive map of how themes are organized within the processes of classification and the search for objects (generally documents) relevant to a specific domain.

In order to convert a thesaurus into formal knowledge, it must be transformed into ontology ([Van Assem et al. 2006](#)), a highly costly process because ontology does not provide a data model which can easily be applied. This occurs because thesauri have been developed without formal semantics, essentially as tools to help in navigation or information retrieval processes. Nonetheless, OWL can be applied in building a data model (in our case SKOS) appropriate to the level of formalisation required by a thesaurus. Thus, the concepts of a thesaurus represent entities in the SKOS data model and the relationships between concepts are facts about those entities.

For SKOS, a knowledge organization system is expressed in terms of concepts structured into relationships to shape concept schemes. Both the concepts and the concept schemes are identified using URIs. The concept can be labelled in any language. A concept can be related to multiple labels, but only one of these for each language can be related as a preferred label. The remaining labels related to the concept are referred to as alternative labels. Hidden labels can also be defined in order to assign to a concept labels which would only be applicable in the search and indexing processes, but which would not be visible to users. Concepts may be assigned to classification or identification codes within a specific conceptual scheme. These notations are not expressed in natural language, but using mnemonic or similar codes. Concepts can also be documented using different types of notes, such as definitions, scope notes or edition notes, among others. The SKOS model covers the establishment of links between concepts known as semantic relationships. These relationships may be hierarchical or associative, although this typology could also be expanded. Concepts can also be grouped into collections, which may in turn be labelled in order. SKOS is supplemented by the possibility of intermapping concepts from different schemes using hierarchical, associative or precise equivalence relationships.

Concepts and concept schemes

A concept represents an idea, a notion or a unit of thought. SKOS requires a degree of flexibility, given that it is intended to represent a semi-formal knowledge organization system. It is therefore clearly an abstract entity independent of the term or terms which may be used to label it. Concepts are linked to a URI or an RDF identifier for reuse and reference. Concepts may also be linked to conceptual schemes. A conceptual scheme is the aggregation of one or more SKOS concepts, generally used to represent and identify thesauri or classification schemes. A concept scheme may have one or more *top concepts*, which head the hierarchical structures within the concept scheme itself. These are usually the start points for search and navigation tasks for users.

Element	Vocabulary	Descriptive Definition
Concept	skos:Concept	Instance of owl:class
Concept scheme	skos:ConceptScheme	Instance of owl:class Disjoint with skos:Concept
Inclusion into a concept scheme	skos:inScheme	Instance of owl:ObjectProperty Domain: skos:Concept Range: skos:ConceptScheme
Top Concept	skos:hasTopConcept	Instance of owl:ObjectProperty Domain: skos:ConceptScheme Range: valores skos:Concept

Table 1: Classes and properties for concepts, schemes, associations with schemes and top concepts

Two examples are set out below. The first example illustrates a basic definition of various concepts and their link to concept schemes. The second example contains the definition of a concept scheme and its top concept.

```

1. <rdf:RDF xmlns:skos="http://www.w3.org/2004/02/skos/core">
2.   <skos:Concept rdf:about="http://www.example.org/conceptos#america">
3.     <skos:inScheme rdf:resource="http://www.example.org/esquema"/>
4.   </skos:Concept>
5.   <skos:Concept rdf:about="http://www.example.org/conceptos#asia">
6.     <skos:inScheme rdf:resource="http://www.example.org/esquema"/>
7.   </skos:Concept>

```

```

8. <skos:Concept rdf:about="http://www.example.org/conceptos#americanorte">
9. <skos:inScheme rdf:resource="http://www.example.org/esquema"/>
10. </skos:Concept>
11. </rdf:RDF>

```

Example 1: Definition of concepts and links to concept schemes.

```

1. <rdf:RDF xmlns:skos="http://www.w3.org/2004/02/skos/core">
2. <skos:ConceptScheme rdf:about="http://www.example.org/esquema">
3. <skos:hasTopConcept rdf:resource="http://www.example.org/conceptos#america"/>
4. <skos:hasTopConcept rdf:resource="http://www.example.org/conceptos#asia"/>
5. </skos:ConceptScheme>
6. </rdf:RDF>

```

Example 2: Definition of a concept scheme and top concepts.

Lexical labels

A lexical label is a string of characters representing a natural language expression. SKOS provides preferred, alternative and hidden labels. The preferred labels are associated with concepts to represent descriptor terms. Only one preferred label for each language can be related to a concept. The presence of identical lexical labels to represent different concepts is not recommended. Alternative labels can be used to assign multiple non-preferred expressions to a concept. This is helpful in representing equivalent terms, non-descriptors, synonyms and acronyms. These expressions enrich the vocabulary available within the system, offering a greater number of possible access routes to a concept and increasing the chances of success in indexing and search processes. The hidden labels are generally assigned to concepts, not to be accessible to users but to be processed by computer applications. One example would be the variant spelling errors for other labels, either preferred or alternative. In SKOS a preferred label can be assigned on an individual basis for each language using UNICODE characters, which would allow different writing systems to be used.

Element	Vocabulary	Descriptive Definition
Preferred label	skos:prefLabel	Instance of owl:DatatypeProperty Range: rdf:PlainLiteral Domain: Cualquier recurso A resource has no more than one value of skos:prefLabel per language Pairwise disjoint with skos:altLabel and skos:hiddenLabel
Alternative label	skos:altLabel	Instance of owl:DatatypeProperty Range: rdf:PlainLiteral Domain: Cualquier recurso Pairwise disjoint with skos:prefLabel and skos:hiddenLabel
Hidden label	skos:hiddenLabel	Instance of owl:DatatypeProperty Range: rdf:PlainLiteral Domain: Cualquier recurso Pairwise disjoint with skos:prefLabel and skos:altLabel

Table 2: Vocabulary and definitions of classes and properties for lexical labels

```

1. <rdf:RDF xmlns:skos="http://www.w3.org/2004/02/skos/core">
2. <skos:Concept rdf:about="http://www.example.org/conceptos#americanorte">
3. <skos:prefLabel xml:lang="es">América del Norte</skos:prefLabel>
4. <skos:altLabel xml:lang="es">Norte América</skos:altLabel>
5. <skos:hiddenLabel xml:lang="es">América Norte</skos:hiddenLabel>
6. <skos:prefLabel xml:lang="en">North America</skos:prefLabel>
7. </skos:Concept>
8. </rdf:RDF>

```

Example 3: Definition of concept and associated lexical labels

Types of relationship between labels can be defined using an extension known as SKOS-XL4 ([W3C 2008c](http://www.w3.org/TR/2008/REC-skos-xl4-20081126/)), with the need to define labels as resources, as with concepts, schemes and collections. This defines a special type of lexical entity which is assigned a literal chain which can be repeated for various units.

Element	Vocabulary	Descriptive Definition
XL label	skosxl:Label	Instance of owl:Class Cardinality exactly 1 with skosxl:literalForm Class disjoint with skos:Collection, skos:Concept and skos:ConceptScheme
Literal form	skosxl:literalForm	Instance of owl:DatatypeProperty Domain: skosxl:Label Range: literal string
XL preferred label	skosxl:prefLabel	Instance of owl:DatatypeProperty Range: rdf:PlainLiteral Domain: Cualquier recurso Pairwise disjoint with skos:prefLabel and skos:altLabel
XL alternative label	skosxl:altLabel	Instancia de owl:ObjectProperty Dominio skos:Concept Rango de valores skosxl:Label The property chain (skosxl:altLabel+skosxl:literalForm) is a sub-property of skos:prefLabel
XL hidden label	skosxl:hiddenLabel	Instance of owl:ObjectProperty Domain: skos:Concept Range: skosxl:Label

		The property chain (skosxl:prefLabel+skosxl:literalForm) is a sub-property of skos:prefLabel
Relation between Labels	skosxl:labelRelation	Instance of owl:ObjectProperty y owl:SymmetricProperty Domain and range: skosxl:Label

Table 3: Classes and properties for concepts, schemes

Lexical units are defined as resources of the class *skosxl:Label* and associated with concepts with the corresponding properties of *skosxl:prefLabel*, *skosxl:altLabel* and *skosxl:hiddenLabel*. As resources defined with *skosxl:Label* always have associated a literal form with *skosxl:literalForm*, the combination of this property with any other of the types of *skosxl* lexical unit is defined as a sub-property of the corresponding type of SKOS lexical unit. For example, if the concept is related to a resource of the class *skosxl:Label* and this in turn with its literal form by means of *skosxl:prefLabel*, the existence of the property *skosxl:prefLabel* associated with that literal may be inferred. The relationships among lexical entities are applied using the property *skosxl:labelRelation*. This property can be used directly, but the most effective approach is to define a customised relationship type as a sub-property of *skosxl:labelRelation*. The example below illustrates the SKOS XL mechanism for expressing this type of relationship.

```

1. <!DOCTYPE rdf:RDF [
2. <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#">
3. <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#">
4. <!ENTITY ex "http://www.cine.org/">
5. <!ENTITY skos "http://www.w3.org/2004/02/skos/core#"> ]>
6. <!ENTITY skosxl "http://www.w3.org/2008/05/skos-xl"> ]>
7. <rdf:RDF xmlns:skos="&skos;" xmlns:rdf="&rdf;" xmlns:rdfs="&rdfs;"
  xmlns:ex="&ex;" xmlns:skosxl="&skosxl;" xml:base="&ex;">
8. <rdf:Property rdf:ID="fullForm"/>
9. <rdfs:subPropertyOf rdf:resource="&skosxl;labelRelation"/>
10. <owl:inverseOf rdf:resource="#acronymForm"/>
11. </rdf:Property>
12. <rdf:Property rdf:ID="acronymForm"/>
13. <rdfs:SubPropertyOf rdf:resource="&skos;LabelRelated"/>
14. </rdf:Property>
15. <skosxl:Label rdf:ID="ONUFull">
16. <skosxl:literalForm xml:lang="es">Organización Naciones
  Unidas</skosxl:literalForm />
17. <ex:acronymForm rdf:ID="ONUAbrev" />
18. </skosxl:Label>
19. <skosxl:Label rdf:ID="ONUAbrev">
20. <skosxl:literalForm xml:lang="es">ONU</skosxl:literalForm />
21. <ex:fullForm rdf:ID="ONUFull" />
22. </skosxl:Label>
23. <skos:Concept rdf:about="http://www.example.org/conceptos#onu">
24. <skosxl:prefLabel rdf:ID="ONUAbrev" />
25. <skosxl:altLabel rdf:ID="ONUFull" />
26. </skos:Concept>
27. </rdf:RDF>

```

Example 4: Definition of of a relationship between lexical labels.

Two properties are defined: *ex:fullForm* and *ex:acronymForm*, in order subsequently to establish a relationship to indicate that one lexical unit is an acronym of another. They are additionally declared using OWL as inverse properties. We then create the *skosxl:Label* resources corresponding to the complete form and the acronym of a concept, defining its properties *skosxl:literalForm*. Lastly, the relationships between the two resources are established with labels *ex:fullForm* and *ex:acronymForm*, along with the relationships *skosxl:prefLabel* and *skosxl:altLabel* between the concept and the lexical units. The resulting tree is slightly more complex than that produced on the basis of a situation where there is no need to apply any type of relationship between lexical units, as illustrated in the figure below.

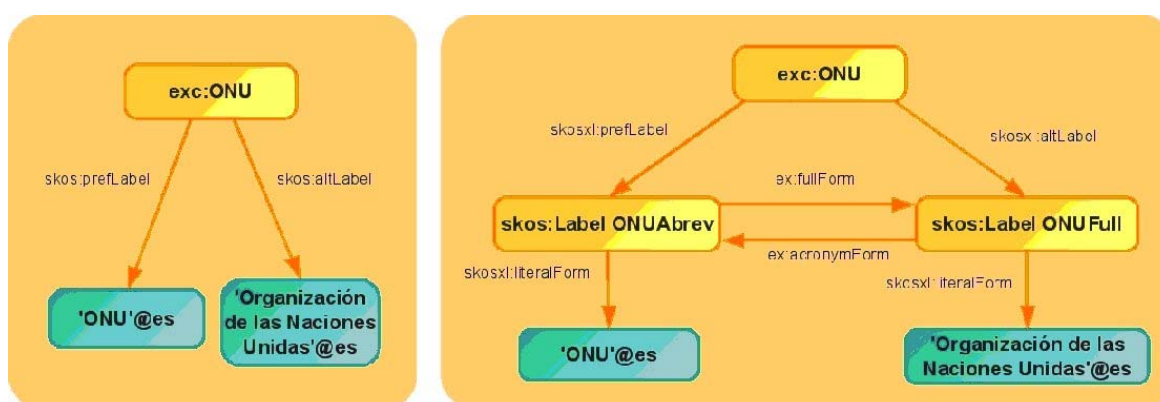


Figure 5: Tree corresponding to the coding of the previous example using SKOS Excel and the corresponding tree with no relationships between lexical entities established

Semantic relationships

Semantic relationships in SKOS are links between concepts inherent in their meaning. SKOS distinguishes between two basic types: hierarchical and associative. A hierarchical relationship between two concepts indicates that one is more general than the other. Meanwhile, an associative relationship indicates that both concepts are connected in some manner, although neither is more generic or specific than the other. This type of relationship is almost identical to that applied in thesaurus building. In SKOS semantic relationships are essential in defining concepts beyond the associated lexical labels. In fact, the meaning of a concept is also complemented through its semantic relationships with other concepts. The basic hierarchical relationships are defined without the transitive property, simply to create

declarations of concept structures. Transitive hierarchical relationships are also covered, in order to establish influences and implement expanded search algorithms in search applications. This is why higher classes of hierarchical relationship are defined, associated with the transitive property, with basic usage relationships, which do not include this property, as a sub-class of the former.

Element	Vocabulary	Descriptive Definition
Semantic relationship	skos:semanticRelation	Instance of owl:ObjectProperty Domain: skos:Concept Range: skos:Concept
Broader transitive	skos:broaderTransitive	Instance of owl:ObjectProperty Instance of owl:TransitiveProperty and Sub-property of skos:semanticRelation
Narrower transitive	skos:narrowerTransitive	Instance of owl:ObjectProperty Instance of owl:TransitiveProperty Defined as owl:inverseOf de skos:broaderTransitive Sub-property of skos:semanticRelation
Generic broader	skos:broader	Instance of owl:ObjectProperty Not instance of owl:TransitiveProperty Sub-property of skos:broaderTransitive
Generic narrower	skos:narrower	Instance of owl:ObjectProperty Not instance of owl:TransitiveProperty Defined as owl:inverseOf de skos:broader Sub-property of skos:narrowerTransitive
Related	skos:related	Instance of owl:ObjectProperty Instance of owl:SymmetricProperty Not instance of de owl:TransitiveProperty Sub-property of skos:semanticRelation Disjoint with skos:broaderTransitive

Table 4: Vocabulary and definitions of classes and properties for elements of semantic relationships

For its part, the associative relationship is defined with the symmetric property. In other words, if a concept A has an associative relationship with B, it is deduced that B has an associative relationship with A. Likewise, it should be pointed out that the associative relationship does not have the transitive property, meaning that if A has an associative relationship with B and B in turn with C, one cannot deduce that an associative relationship can be established between A and C.

```

1. <rdf:RDF xmlns:skos="http://www.w3.org/2004/02/skos/core">
2. <skos:Concept rdf:about="http://www.example.org/conceptos#america">
3. <skos:prefLabel xml:lang="es">América</skos:prefLabel>
4. <skos:narrower rdf:resource="http://www.example.org/conceptos#americasur"/>
5. </skos:Concept>
6. <skos:Concept rdf:about="http://www.example.org/conceptos#americanorte">
7. <skos:prefLabel xml:lang="es">América del Norte</skos:prefLabel>
8. <skos:broader rdf:resource="http://www.example.org/conceptos#america"/>
9. </skos:Concept>
10. <skos:Concept rdf:about="http://www.example.org/conceptos#americasur">
11. <skos:prefLabel xml:lang="es">América del Sur</skos:prefLabel>
12. <skos:broader rdf:resource="http://www.example.org/conceptos#america"/>
13. <skos:related rdf:resource="http://www.example.org/conceptos#cultivocafe"/>
14. </skos:Concept>
15. <skos:Concept rdf:about="http://www.example.org/conceptos#cultivocafe">
16. <skos:prefLabel xml:lang="es">Cultivo de Café</skos:prefLabel>
17. <skos:related rdf:resource="http://www.example.org/conceptos#americasur"/>
18. </skos:Concept>
19. </rdf:RDF>

```

Example 5: Definition of concepts and associated semantic relationships

We will now look at an example based on the entry in the UNESCO thesaurus (2009) for the descriptor *thesaurus*. This first includes a description of the entry we are using in this example.

Tesouro
English term: Thesauri
Terme français: Thésaurus
Русский термин : Тезаурусы
NA Lenguaje documental controlado y dinámico que contiene términos relacionados semántica y genéricamente que abarcan de manera exhaustiva una esfera concreta del conocimiento.
MT 5.05 Ciencias de la información
UP Descriptores
UP Tesouro monolingüe
UP Tesouro multilingüe
UP Tesouro
TG Lenguaje de indexación [32]
TE Compilación de tesouro [70]
TR Control terminológico [19]
TR Encabezamiento por materia [185]
TR Material de referencia [123]
TR Terminología [404]

Example 6: Entry for the descriptor *thesaurus* in the UNESCO thesaurus

We now detail the coding which SKOS would perform for this descriptor, without including the coding which would be used for the descriptors with which the semantic relationships are established.

```

1. <rdf:RDF xmlns:skos="http://www.w3.org/2004/02/skos/core">
2. <skos:Concept rdf:about="http://www.example.org/conceptos#tesauro">
3. <skos:prefLabel xml:lang="es">Tesauro</skos:prefLabel>
4. <skos:prefLabel xml:lang="en">Thesauri</skos:prefLabel>
5. <skos:prefLabel xml:lang="fr">Thésaurus</skos:prefLabel>
6. <skos:prefLabel xml:lang="ru">Тезауры</skos:prefLabel>
7. <skos:scopeNote xml:lang="es"> Lenguaje documental controlado y dinámico
8. que contiene términos relacionados semántica y genéricamente que abarcan
9. de manera exhaustiva una esfera concreta del conocimiento.
10. </skos:scopeNote>
11. <skos:altLabel xml:lang="es">Descriptores</skos:prefLabel>
12. <skos:altLabel xml:lang="es">Tesauro monolingüe</skos:prefLabel>
13. <skos:altLabel xml:lang="es">Tesauro multilingüe</skos:prefLabel>
14. <skos:altLabel xml:lang="es">Thesauri</skos:prefLabel>
15. <skos:broader rdf:resource="http://www.example.org/conceptos#lengindex"/>
16. <skos:narrower rdf:resource="http://www.example.org/conceptos#compes"/>
17. <skos:related rdf:resource="http://www.example.org/conceptos#controlterm"/>
18. <skos:related rdf:resource="http://www.example.org/conceptos#encmateria"/>
19. <skos:related rdf:resource="http://www.example.org/conceptos#matref"/>
20. <skos:related rdf:resource="http://www.example.org/conceptos#terminologia"/>
21. </skos:Concept>
22. </rdf:RDF>

```

Example 7: Coding with SKOS of the descriptor *thesaurus* according to the corresponding entry in the UNESCO thesaurus

The above example can be represented graphically in the following manner:

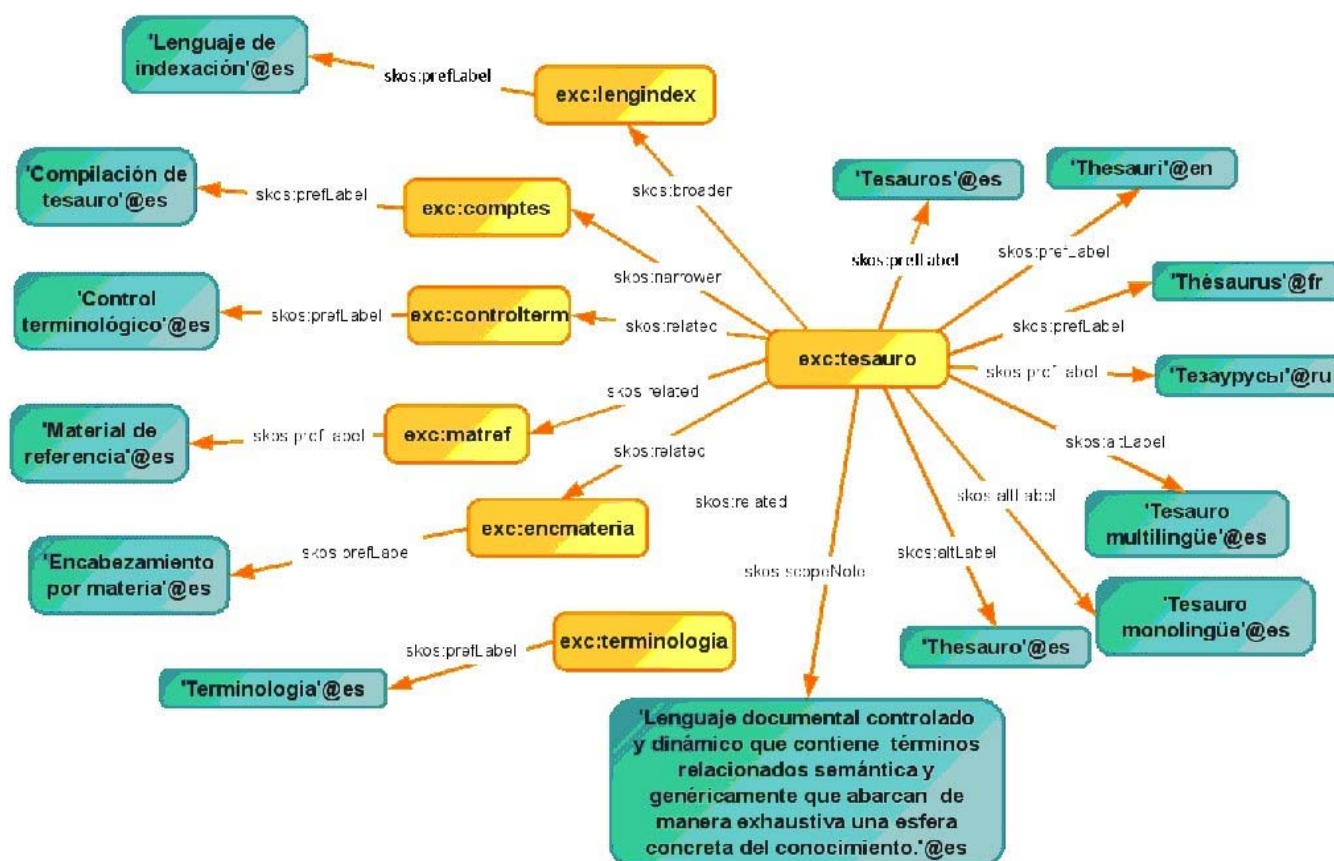


Figure 6: Tree corresponding to the coding with SKOS of the descriptor *thesaurus* according to the corresponding entry in the UNESCO thesaurus

One of the characteristics of SKOS is its flexibility (W3C 2009) which is perhaps why semantic restrictions have not explicitly been defined, preventing the creation of cycles in the hierarchical structures or the establishment of associative relationships between concepts belonging to the same hierarchy. Nonetheless, the SKOS model does cover the semantic restriction involved in the disjunction between *skos:related* and *skos:narrowerTransitive*, which prevents the establishment of associative relationships between concepts structured within the same hierarchical line. The applications themselves will have the task of ensuring that these and other possible restrictions not previously mentioned (such as the reflexivity of semantic relationships) are applied correctly, at all times in accordance with the requirements of the conceptual scheme defined by the system users.

Notations

With SKOS a notation can be associated with a concept, allowing it to be tied to its corresponding entry within a thesaurus, classification or other organization system

which identifies elements with decimal signatures, identification codes or similar. A concept is thus associated with the specific field of a conceptual scheme. The values associated with notations are generally defined as a typed literal, normally constructed on the basis of data types defined with an XML scheme. The value which can be associated with a notation is thus a combination of a typed literal and a URI defining the format of that literal. One single concept may be assigned various notations, although one notation should only be assigned to a single concept. This last point is a convention, but is not formally defined in the SKOS data model, meaning that there are no integrity conditions for this element.

Element	Vocabulary	Descriptive Definition
Notation	skos:notation	Instance of owl:DatatypeProperty Domain: skos:Concept Range: Typed literal

Table 5: Vocabulary and definitions of classes and properties for notations

Documentation

Documentary elements allow for the inclusion of non-formal information concerning the meaning of relationships, concepts and labels, along with their evolution over time, publication notes, examples, usage scope notes and other types of data essentially addressing users. These elements allow certain aspects of the conceptual scheme to be documented. SKOS offers a typology allowing the type of documentation applied to be distinguished, meaning that such content can be formalised, up to a point. The typology of documentary elements covers a wide range of notes: scope, history, changes, definition, edition and example.

Element	Vocabulary	Descriptive Definition
Note	skos:note	Instance of owl:ObjectProperty Domain: rdfs:Resource
Scope note	skos:scopeNote	Instance of owl:ObjectProperty Sub-property of skos:note Domain: rdfs:Resource
History note	skos:historyNote	Instance of owl:ObjectProperty Sub-property of skos:note Domain: rdfs:Resource
Change note	skos:changeNote	Instance of owl:ObjectProperty Sub-property of skos:note Domain: rdfs:Resource
Definition	skos:definition	Instance of owl:ObjectProperty Sub-property of skos:note Domain: rdfs:Resource
Editorial note	skos:editorialNote	Instance of owl:ObjectProperty Sub-property of skos:note Domain: rdfs:Resource
Example	skos:example	Instance of owl:ObjectProperty Sub-property of skos:note Domain: rdfs:Resource

Table 6: Vocabulary and definitions of classes and properties for documentation elements

Another advantage of SKOS is that it allows elements of other vocabularies, such as Dublin Core (2008), to be exploited. Example 9 includes an illustration of this in line 7, which makes use of the element "source". The SKOS documentation elements can be employed without any type of restriction in documents where the main content uses another vocabulary.

Concept collections

Collections of concepts in SKOS allow groupings to be defined, thereby enriching the structure without going so far as to establish explicit semantic relationships which would distort the hierarchical or associative structures of the conceptual scheme. The concepts and concept collections are disjoint in SKOS, making it impossible to establish semantic relationships of any kind between the two. Collections can be linked to RDF resources identified by a URI, although it is more typical for them to be declared as empty nodes. Collections can be associated with lexical labels and as with RDF collections, can be ordered using lists of elements. The collection elements can likewise be inferred on the basis of the elements of an ordered collection. Collections can be included within other collections, as in the case of an additional element.

Element	Vocabulary	Descriptive Definition
Collection	skos:Collection	Instance of owl:Class Disjoint with skos:Concept y skos:ConceptScheme
Ordered collection	skos:OrderedCollection	Instance of owl:ObjectProperty Instance of owl:FunctionalProperty Domain: skos:OrderedCollection Range: rdf:List Allowed more than one value of skos:member
Member of a collection	skos:member	Instance of owl:ObjectProperty Instance of owl:FunctionalProperty Domain: skos:OrderedCollection Range: rdf:List Allowed more than one value of skos:member
Member of a list	skos:memberList	Instance of owl:ObjectProperty Instance of owl:FunctionalProperty Domain: skos:OrderedCollection Range: rdf:List Allowed more than one value of skos:member

Table 7: Vocabulary and definitions of classes and properties for elements of collections

Mapping properties

It is clearly possible to include concepts in various conceptual schemes, thereby allowing their reuse. Additionally, SKOS offers the functionality of mapping between concepts included in different conceptual schemes. This makes it possible to establish a correspondence between such concepts. It may occur that the correspondence between the two concepts is not precise, in which case SKOS distinguishes between different types of mapping. It can be declared that an exact correspondence exists between two concepts, that one is more generic or specific than the other, or otherwise an association correspondence may be established. The correspondence properties are sub-classes of their corresponding semantic relationships, as may be seen in the following figure.

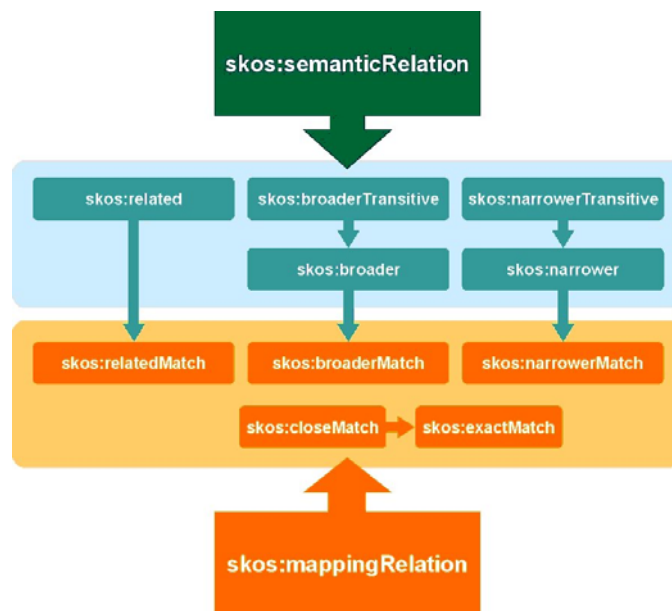


Figure 7: Organizational diagram of semantic and correspondence relationships

The table below describes the elements, associate vocabulary and definitions of the classes and properties.

Element	Vocabulary	Descriptive Definition
Mapping relationship	skos:mappingRelation	Instance of owl:ObjectProperty Domain: skos:Concept Range: skos:Concept
Exact match	skos:exactMatch	Instance of owl:ObjectProperty Instance of owl:SymmetricProperty Sub-property of skos:mappingRelation
Broad match	skos:broadMatch	Instance of owl:ObjectProperty Sub-property of skos:mappingRelation Sub-property of skos:broader
Narrower match	skos:narrowerMatch	Instance of owl:ObjectProperty Sub-property of skos:mappingRelation Sub-property of skos:narrower Defined as owl:inverseOf de skos:broader //identical situation as for skos:narrowerTransitive and skos:narrower//

Table 8: Vocabulary and definitions of classes and properties for elements of collections

SKOS applications

The traditional problems of Information Retrieval (relevance of response, precision, comprehensiveness, etc.) combine with others inherent in the nature of the Web (Martínez and Rodríguez 2003), progressively aggravated through the emergence of the Web 2.0, among other reasons, because the quality, structuring and originality of content have not evolved in parallel with the ease of Web publication. There are large numbers of duplicate pages, as many users prefer to copy content rather than to reference it using hypertext links. Many pages make improper use of HTML metadata and some include all kinds of terms in order to confuse indexing bots and it is impossible to distinguish the type of object retrieved in the search. On such a scenario, Web search engines are on occasion unable to offer useful results. The situation would improve with the use of standardised metadata models, such as Dublin Core (2008), along with the application of conceptual schemes. The indexing of Web pages with a thesaurus allows us to present queries without users having to perform a predictive selection of terms. Search possibilities are expanded as the user would be consulting a network of terms, which could be combined with a new query language. This is a search based on the exploration of a network of concepts which would guide the user in selecting the query terms, enabling an increase in effectiveness by establishing filters in accordance with the content of certain metadata.

The first obvious application of SKOS is therefore the representation of conceptual thesauri in a manner which moulds itself perfectly to the requirements of the Semantic Web. Although this phase is still in development, many thesaurus management or document indexing applications already use this model (a compilation of these applications may be found in the W3C Website, given the fact that its essential properties and elements have been fully defined and represent a solid development core. SKOS may evolve, incorporating new types of relationship and adapting with RDF and OWL. Even so, the model for structuring concepts, labels

and relationships will be maintained over time and therefore had therefore come into use even before W3C had drawn up a first definitive recommendation.

The thesauri in particular and the conceptual schemes in general have been used in information retrieval and organization tasks, although only partial solutions have been devised, with a limited scope of operation. On the Web, most such initiatives have created XML or RDF/XML vocabularies to implement functional systems (Greenberg 2004). Nonetheless, until the emergence of SKOS, these developments were not coordinated, nor did they have a global vision of an application valid for any type of conceptual scheme. They likewise failed to share a common model, meaning that the exchange of data between systems required the application of mapping mechanisms, which on occasion could not operate completely.

The interoperability of systems is another of the positive aspects involved in SKOS, in particular if one remembers that metadata represent the essential core of the Semantic Web. It is much more beneficial to define rules and restrictions regarding a system's information, outside the programming code and to represent these in documents drawn up in a standard language. This is what metadata and ontologies offer through RDF and OWL and is therefore a further contribution of SKOS.

The exchange of conceptual schemes represented with a standard vocabulary and models such as SKOS offers a number of advantages tied to the incorporation and consequent reuse of external resources. One possible application would be to incorporate them as an element integrated into information systems, for organization and search processes. This is a technique similar to the syndication of Web content. The usefulness of conceptual schemes is not limited to information search processes, but could also be employed in Content Management Systems. The main difficulty of these systems is that of creating an information organization tool integrated with the automatic generation processes of the corresponding Website navigation systems. A thesaurus could be used to devise an organizational scheme, while an ontology could define the correspondence rules to create the navigation system dynamically. SKOS could be used to describe how the content of a Website is organized and that would be involved, using an ontology, in the design of Website navigation systems.

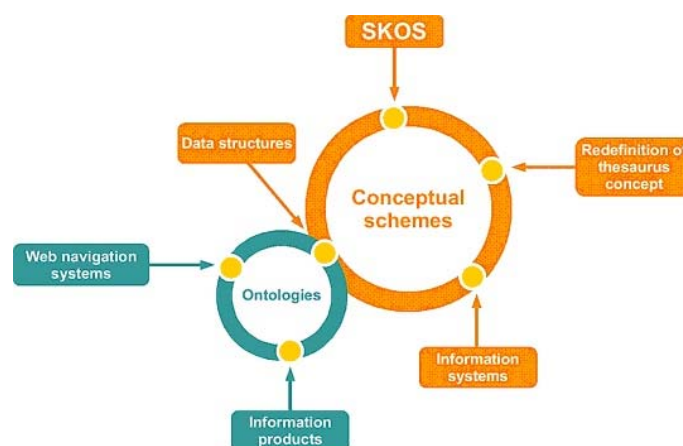


Figure 8: SKOS applications in Information Systems, complemented by ontologies

Given all the above, we believe that SKOS, in addition to its potential employment at a general level on the Web, could also be used initially within corporate information systems, associated with their operation over Intranets. The joint use of SKOS and OWL could allow the conceptualisation of an information system as a sphere of structured data, which could serve to define products and services in the form of specific views of the content managed by the system.

Conclusions

Over the course of this work we have presented various solutions focusing the resolution of one single problem: the representation of conceptual schemes, such as thesauri, in a manner appropriate to their use for information retrieval and the organization of knowledge on the Web and more specifically on the semantic Web. We have established how initiatives to represent thesauri using XML, RDF and even using alternatives with a different focus, such as Topic Maps, can work. Initiatives using XML, such as Zthes and MeSH, have not been developed on the basis of the common descriptive framework offered by RDF and RDF Schema. Their integration in the descriptive processes of the semantic Web must therefore take place through processes to convert the representation formats employed. Something similar occurs with the proposal made by Topic Maps, which provides the XML Topic Maps (XTM) specification, an XML vocabulary appropriate for the representation of this type of tool within information technology environments. It would, however, be reasonable to believe that the field of application of Topic Maps, more general than that of SKOS, could represent an advantage over the latter solution. Nonetheless, SKOS can be expanded and take part in the much more general logical processes of the Semantic Web, used jointly with OWL. This would then offer a high level of flexibility at the descriptive formalisation level.

The alternatives employing RDF to represent thesauri, such as LIMBER and CERES, referred to earlier, involve their own developments which are not integrated within the W3C initiatives and which on occasion have lexical units as their central elements. The adoption of SKOS as a common model to represent thesauri allows conceptual thesauri to be represented in a standardised manner. To an extent, OWL offers greater possibilities of representation and potential application than SKOS. OWL could be used directly to develop ontologies with which to represent thesauri. However, the direct use of ontologies raises the drawback of the complexity of thesaurus management tasks. This job is simplified with SKOS, while maintaining and expanding the scope of application, as this is a specialised OWL ontology and can be expanded in the future. To an extent this guarantees the viability of SKOS and its evolution over time. SKOS is not therefore a closed development, but rather its very nature gives access to the mechanisms required for its adaptation.

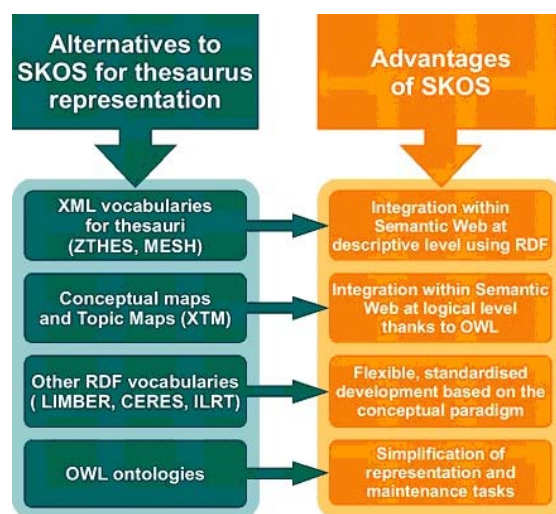


Figure 9: Advantages of SKOS over other alternatives

From the user perspective, the use of thesauri developed with the SKOS model affects the managers of information services and systems, along with those who use them through query operations. For information managers, SKOS offers a closer approach to knowledge organization and management, complementing the automatic extraction of textual content from documents with its indexing through conceptual entities. Furthermore, thesauri can be mutually combined through correspondence relationships, enabling their reuse and a more effective use of the tasks of indexing and the construction of navigation systems for Web information services and systems.

SKOS would allow us to devise multi-level information query and search tools. This would then give us a model of the Web structured on various levels: conceptual, navigational and documentary structure. Thesauri would organize the conceptual level, constituting the solution employed for information retrieval procedures. The navigability provided by hyperlinks on the Webpages consulted would be maintained. The documentary structure level could be integrated with other tools such as Topic Maps or ontologies which would allow one to design feedback processes for the structure of the thesaurus or thesauri employed.

Thesauri could also allow searches to be expanded and redefined, or show references to documents with content related to that of those directly retrieved in the search, or suggest new search terms. One could even develop an interface in which the user is not required to enter keywords but instead navigates around a network of concepts, selecting those of interest to the search being performed. In this manner, the reuse of conceptual schemes would offer the end user a more coherent and integrated vision of information search systems and services.

About the authors

Juan-Antonio Pastor-Sanchez has a BA and PhD in Information Science from the University of Murcia. He has worked since 1994 as Web Manager in the University of Murcia and, currently, he is Senior Lecturer of Digital Information Services in the Department of Information and Documentation of the same university. His research has always been related to Thesauri and Information Technology. He is a member of the editorial board of the Spanish journal *Scire: Representación y Organización del Conocimiento* and reviewer of papers in Spanish for *Information Research: an electronic journal*. He can be contacted at pastor@um.es

Francisco Javier Martínez Méndez teaches information technology in the Department of Information and Documentation of the University of Murcia, of which he is currently Head. A Graduate and PhD in Information and Documentation. Professor of the UNITWIN Chair in 'Information Management in Organizations', sponsored by UNESCO. Member of the Information Technology Research Group of the University of Murcia. He is also the author of the blog, [recuperación de información en la web](http://recuperación.de.información.es). He can be contacted at javima@um.es

José Vicente Rodríguez-Muñoz is Professor in the Information and Documentation Department, Faculty of Information Science, University of Murcia, Spain. Graduate in Chemistry, PhD in Computing Science. Guest Professor in the University of La Habana, Cuba. Co-ordinator of the UNITWIN Chair in 'Information Management in Organizations', sponsored by UNESCO. Member of the 50th Committee, Documentation, of AENOR (Spanish Association of Normalization and Certification). Scientific Consultant to several projects of the Regional Government of Murcia. Chief of the Information Technology Research Group of the University of Murcia. His research area, projects and publications, covers information management, information retrieval and evaluation of Web searching. Currently Dean of the Faculty of Information Science. He can be contacted at jovi@um.es

References

- Aitchinson, J. & Gilchrist, A. (1987). *Thesaurus construction. A Practical manual*. 2nd edition. London: ASLIB.
- Alistair, M., Matthews, B., Beckett, D., Brickley, D., Wilson, M. & Rogers, N. (2005). *SKOS: a language to describe simple knowledge structures for the Web*. In *XTech 2005: XML, the Web and beyond*. Retrieved 14 December, 2009 from <http://epubs.cclrc.ac.uk/bitstream/685/SKOS-XTech2005.pdf> (Archived by WebCite® at <http://www.webcitation.org/5m2gMRJih>)
- American National Standards Institute and National Information Standards Organization. (2005). *Z39.19-2005: Guidelines for the construction, format and management of monolingual controlled vocabularies*. New York, NY: National Information Standards Organization. Retrieved 14 December, 2009 from <http://bit.ly/6VMJ9H>
- Arano, S. & Codina, L. (2004). *La estructura conceptual de los tesauros en el entorno digital: ¿Nuevas esperanzas para viejos problemas?* [The conceptual structure of thesauri in the digital environment: new hopes for old problems?] *Jornades Catalanes d'Informació i Documentació* (No. IX), 41-58. Retrieved 14 December, 2009 from <http://www.lluiscodina.com/ontotesauros.doc> (Archived by WebCite® at <http://www.webcitation.org/5m2hE8Mec>)
- Arano, S. (2005). *Los tesauros y las ontologías en la biblioteconomía y la documentación*. [Thesauri and ontologies in librarianship and documentation.] *Hipertext.net*, (No. 3). Retrieved 14 December, 2009 from <http://www.hipertext.net/Web/pag260.htm> (Archived by WebCite® at <http://www.webcitation.org/5m2hu3Mn3>)
- Brownson, H. (1957). Current United States research on topics related to the conference. In *Proceedings of the International Study Conference on Classification for Information Retrieval 13-17 May 1957, Dorking, England*. (pp. 99-100). London: ASLIB.

- Currás, E. (2005). *Ontología, taxonomía y tesauros: manual de construcción y usos*. [Ontologies, taxonomies and thesauri: construction manual.] Gijón, Spain: Trea.
- Díaz Ortuño, P.M. (2003). [Problemática y tendencias en la arquitectura de metadatos Web](#). [Problems and trends in the architecture of Web metadata.] *Anales de Documentación*, (No. 6), 35-58. Retrieved 14 December, 2009 from <http://www.um.es/ojs/index.php/analesdoc/article/viewFile/1691/1741> (Archived by WebCite® at <http://www.webcitation.org/5m2jooEt4>)
- Dublin Core Metadata Initiative. (2008). [Metadata Element Set, Version 1.1](#). Retrieved 11 September 2009 from <http://dublincore.org/documents/dces/> (Archived by WebCite® at <http://www.webcitation.org/5m2jyDQli>)
- García Jiménez, A. (2004). [Instrumentos de representación del conocimiento: tesauros versus ontologías](#). [Tools of knowledge representation: thesauri versus ontologies.] *Anales de Documentación*, (No. 7), 79-95. Retrieved 14 December, 2009 from <http://www.um.es/ojs/index.php/analesdoc/article/viewFile/1691/1741> (Archived by WebCite® at <http://www.webcitation.org/5m2kKQP8h>)
- García Torres, A., Pareja Lora, A. & Pradana López, D. (2008). Reutilización de tesauros: el documentalista frente al reto de la Web semántica. [Reuse of thesauri: the documentalist and the challenge of the semantic Web.] *El profesional de la información*, **17**(1), 8-21.
- Gilchrist, A. (1971). *The thesaurus in retrieval*. London: ASLIB.
- Greenberg, J. (2004). [User comprehension and searching with information retrieval thesauri](#). *Cataloging & Classification Quarterly*, **37**(3/4), 103-120. Retrieved 14 December, 2009 from <http://ils.unc.edu/mrc/pdf/greenberg04user.pdf> (Archived by WebCite® at <http://www.webcitation.org/5m2khE3NQ>)
- International Organization for Standardization. (1985). *ISO 5964:1986 Guidelines for the establishment and development of multilingual thesauri*. Geneva, International Organization for Standardization.
- International Organization for Standardization. (1986). *ISO 2788:1986 Guidelines for the establishment and development of monolingual thesauri*. Geneva, International Organization for Standardization.
- Jarvelin, K., Kristensen, J., Niemi, T., Sormunen, E. & Keskustalo, H. (1996). [A deductive data model for query expansion](#). In *Proceedings of the 19th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval* (pp. 235-243). New York, NY: ACM Press. Retrieved 15 August 2008, from http://www.info.uta.fi/tutkimus/fire/archive/ExpansionTool_Sigir96.pdf (Archived by WebCite® at <http://www.webcitation.org/5m2kxJw2E>)
- Jones, S. (1993). A thesaurus data model for an intelligent retrieval system. *Journal of Information Science* **19**(3) 167-178.
- Laguens García, J.L. (2006). Tesauros y lenguajes controlados en Internet. [Thesauri and controlled vocabularies in the Internet.] *Anales de Documentación*, (No. 9), 105-121.
- López-Huertas Pérez, M.J. (1999). Potencialidad evolutiva del tesoro: hacia una base de conocimiento experto. [The evolutionary potential of the thesaurus: building a base of expertise]. In *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 EOCOSID'99, 22-24 April 1999* (pp. 133-140). Granada: ISKO, Facultad de Biblioteconomía y Documentación.
- Martínez Méndez, F. J. & Rodríguez Muñoz, J.V. (2003). [Síntesis y crítica de las evaluaciones de la efectividad de los motores de búsqueda en la Web](#). [Synthesis and critique of the evaluation of effectiveness of Web search engines.] *Information Research*, **8**(2), paper no. 148 Retrieved 14 December, 2009 from <http://InformationR.net/ir/8-2/paper148.html> (Archived by WebCite® at <http://www.webcitation.org/5m2llgKqk>)
- Matthews, B. & Miles, A. (2001). [Review of RDF thesaurus work: a review and discussion of RDF schemes for thesauri](#). (Workpackage 8 Deliverable 8.2 SWAD-Europe). Retrieved 14 December, 2009 from <http://www.w3.org/2001/sw/Europe/reports/thes/8.2/> (Archived by WebCite® at <http://www.webcitation.org/5m2owkpNr>)
- Matthews, B., Alistair, M. and Wilson, M. (2001). [Modelling thesauri for the semantic Web](#). (Workpackage 8 Deliverable 8.1/8.2 SWAD-Europe). Retrieved 14 December, 2009 from <http://www.w3c.rl.ac.uk/SWAD/thesaurus/tif/deliv81/final.html> (Archived by WebCite® at <http://www.webcitation.org/5m2lmCyQY>)
- Miller, U. (1997). Thesaurus construction: problems and their roots. *Information Processing & Management*, **33**(4), 481-493.
- Miller, K. & Matthews, B. (2001). [Having the right connections: the LIMBER Project](#). *Journal of Digital Information*, **1**(8), Article 37. Retrieved 14 December, 2009 from <http://journals.tdl.org/jodi/article/view/34/35> (Archived by WebCite® at <http://www.webcitation.org/5m2m5xfqV>)
- Pastor Sánchez, J.A. & Saorín Pérez, T. (1993). [Un interfaz hipertexto para un sistema de gestión de tesauros](#). [A hypertext interface for a thesaurus management system.] *Scire: representación y organización del conocimiento* **2**(1), 51-62. Retrieved 2 January 2009 from <http://ibersid.eu/ojs/index.php/scire/article/view/1049/1031> (Archived by WebCite® at <http://www.webcitation.org/5m2mNjDPO>)
- Pastor Sánchez, J.A. & Saorín Pérez, T. (1995a). [El hipertexto documental como solución a la crisis conceptual del hipertexto. El reto de los documentos cooperativos en redes](#). [The hypertext document as a solution to the conceptual crisis hypertext. The challenge of cooperative documents on networks.] *Cuadernos de Documentación Multimedia*, (No. 4), 41-56. Retrieved 14 September, 2009 from <http://www.uctm.es/info/multidoc/multidoc/revista/cuadern4/hiperdoc.htm> (Archived by WebCite® at <http://www.webcitation.org/5m2mdwKnd>)
- Pastor Sánchez, J.A. & Saorín Pérez, T. (1995b). Exposición conceptual y aplicación de un modelo para la organización del conocimiento en Internet: el hipertexto documental. [Conceptual exposition and application of a model for the organization of knowledge on the Internet: the hypertext document.] In *Organización del Conocimiento en Sistemas de Información y Documentación: Actas del II Encuentro de ISKO-España, Getafe 16 y 17 November 1995*. (pp. 247-269). Zaragoza, Spain: Universidad de Zaragoza.
- Pastor Sánchez, J.A. & Saorín Pérez, T. (1998). [La escritura hipermedia](#). [Hypermedia writing.] *Cuadernos de Documentación Hipermedia*, (No. 6-7), 221-238. Retrieved 14 December, 2009 from <http://www.uctm.es/info/multidoc/multidoc/revista/cuad6-7/saorin.htm> (Archived by WebCite® at <http://www.webcitation.org/5m2mlV0OK>)
- Rada, R., Zeb, A., You, G.-N., Michailidis, A. & Mhashi, M. (1991). Collaborative hypertext and the MUCH system. *Journal of Information Science*, **17**(4), 191-196.
- Reynolds, D., Thompson, C., Mukerji, J. & Coleman, D. (2005). [An assessment of RDF/OWL modelling](#). Bristol, UK: Hewlett Packard. Retrieved 12 September 2009 from <http://www.hpl.hp.com/techreports/2005/HPL-2005-189.pdf> (Archived by WebCite® at <http://www.webcitation.org/5m2nEoV2w>)
- Roberts, N. (1984). Historical studies in documentation. The pre-history of the information retrieval thesaurus. *Journal of documentation*, **40**(4), 271-285.
- Roe, Sandra K & Alan R. Thomas, (Eds.) (2004). *The thesaurus: review, renaissance and revision*. New York, NY: Haworth Press.
- Rodríguez Muñoz, J.V., Martínez Méndez, F.J. & Díaz Ortuño, P. (1990). Los modelos de datos como alternativa en la construcción de tesauros. [Data models as alternatives to the construction of thesauri.] En *Actas de las III Jornadas Españolas de Documentación Automatizada*. Palma de Mallorca: FESABID.
- Rodríguez Muñoz, J. V. (1992). *Construcción del esquema conceptual del tesoro mediante un modelo de datos*. [Construction of the conceptual scheme of the thesaurus via a data model.] Unpublished doctoral thesis, Universidad de Murcia, Murcia, Spain
- Shiri, A.A. & Revie, C. (2000). Thesauri on the Web: current developments and trends. *Online Information Review*, **24**(4), 273-279. Retrieved 14 December, 2009 from <http://dlist.sir.arizona.edu/archive/00000163/01/thesauri.pdf>
- Shiri, A.A., Revie, C. & Chowdhury, G. (2002). Thesaurus-assisted search term selection and query expansion: a review of user-centred studies. *Knowledge organization* **29**(1), 1-19.
- Tudhope, D., Harith, A. & Jones, C. (2001). [Augmenting thesaurus relationships: possibilities for retrieval](#). *Journal of Digital Information*, **1**(8), Article 41. Retrieved 14 December, 2009 from <http://journals.tdl.org/jodi/article/view/181/160> (Archived by WebCite® at <http://www.webcitation.org/5m2nn4dLN>)
- UNESCO. (2009). [UNESCO thesaurus](#). Paris: Unesco. Retrieved 14 December, 2009 from <http://databases.unesco.org/thesaurus/> (Archived by WebCite® at <http://www.webcitation.org/5m2nyx9xG>)
- Van Assem, M., Malaisé, V., Miles, A. & Schreiber, G. (2006). [A method to convert thesauri to SKOS](#). In *The Semantic Web: Research and*

Applications. (pp. 95-109). Berlin: Springer. Retrieved 11 September, 2009 from <http://www.cs.vu.nl/~mark/papers/Assem06b.pdf> (Archived by WebCite® at <http://www.webcitation.org/5m2oBIbhg>)

- World Wide Web Consortium (W3C). (2004). *OWL Web Ontology Language guide. W3C Recommendation 10 Feb 2004*. Edited by M.K. Smith, C. Welty and D.L. McGuinness. Retrieved 14 December, 2009 from <http://www.w3.org/TR/owl-guide/> (Archived by WebCite® at <http://www.webcitation.org/5m2pEK5Qj>)
- World Wide Web Consortium (W3C). (2007). *SKOS Use Cases and Requirements. W3C Working Draft 16 May 2007*. Edited by: Antoine Isaac, Jon Phipps and Daniel Rubin. Retrieved 15 June, 2008 from <http://www.w3.org/TR/2007/WD-skos-ucr-20070516/> (Archived by WebCite® at <http://www.webcitation.org/5m2qW4Gfk>)
- World Wide Web Consortium (W3C). (2008a). *SKOS Simple Knowledge Organization System reference. W3C Working Draft 29 August 2008*. Edited by Alistair Miles and Sean Bechhofer. Retrieved 1 September 2008 from <http://www.w3.org/TR/2008/WD-skos-reference-20080829/> (Archived by WebCite® at <http://www.webcitation.org/5m2qNGvya>)
- World Wide Web Consortium (W3C). (2008b). *SKOS Simple Knowledge Organization System primer. W3C Working Draft 29 August 2008*. Edited by Antoine Isaac and Ed Summers. Retrieved 1 September 2008 from <http://www.w3.org/TR/2008/WD-skos-primer-20080829/> (Archived by WebCite® at <http://www.webcitation.org/5m2qE1rS1>)
- World Wide Web Consortium (W3C). (2008c). *SKOS Simple Knowledge Organization System eXtension for Labels (XL) RDF schema 20th August 2008 'Last call' edition*. Edited by: Alistair Miles and Sean Bechhofer. Retrieved 29 September 2008 from <http://www.w3.org/TR/2008/WD-skos-reference-20080829/skos-xl.html> (Archived by WebCite® at <http://www.webcitation.org/5m2q1XRaN>)
- World Wide Web Consortium (W3C). (2009). *SKOS Simple Knowledge Organization System reference. W3C proposed recommendation 15 June 2009*. Edited by Alisair Miles and Sean Bechhofer. Retrieved on 14 December, 2009 from <http://www.w3.org/TR/2009/PR-skos-reference-20090615/> (Archived by WebCite® at <http://www.webcitation.org/5m2phfs3S>)
- Wersig, G. (1971). *Das Krankenhaus-Informationssystem (KIS): Überlegungen zu Strukturen und Realisierungsmöglichkeiten Integrierter Krankenhaus-Informationssysteme*. [The hospital information system (HIS): considerations of structure and possibilities of realization of integrated hospital information systems.] Munich-Pullach, Germany: Verlag Dokumentation.

How to cite this paper

Pastor, J.A., Martinez, F.J. & Rodriguez J.V. (2009). "Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives" *Information Research*, **14**(4) paper 422. [Available at <http://InformationR.net/ir/14-4/paper422.html>]

Find other papers on this subject

[Scholar Search](#)

[Google Search](#)

[Bing](#)

■ [Bookmark This Page](#)

88

© the authors, 2009.
Last updated: 14 December, 2009

W3C
OneStat.com

[Contents](#) | [Author index](#) | [Subject index](#) | [Search](#) | [Home](#)