Advantages of SKOS for providing semantic stability to to the information resources: an example of use in a collaborative institutional web portal.

Pastor Sánchez, J.A.; López Carreño, R. y Martínez Méndez, F.J. Departamento de Información y Documentación, Universidad de Murcia. SPAIN

Resumen.

SKOS has emerged as the W3C initiative to provide semantic stability to the Web. Although this proposal is still in development, many thesaurus management or document indexing applications already use this model. We present a proposal of use of SKOS into a collaborative information system dedicated to the Cultural management in order to improve the description level of the documents of the portal and the information retrieval effectiveness.

Palabras clave.

SKOS, Websites, Conceptual Schemes, Thesauri, Information Systems, Information Retrieval.

Introduction.

In the origin of the Web, the sites contained large sets of pages that, because of technological developments, have become to portals where to offer services along with interactive information. Their degree of development is diverse and not always meet the expectations created, particularly affected by the downgrade of the contents and information architecture problems that difficult navigation and thus, the information retrieval. The use of content management systems[1] (CMS) for managing websites is becoming more common. This facilitates their management and contribute to their dynamism and update, releasing managers of design documents and increasing their productivity. These systems are typical applications of Web 2.0 (or *Social Web*) environment where is changing the paradigm editor-reader into a new concept, the *collaborative editing*. Now, portals provides content management and enables a range of tools for collaborative production, offering these services to the community of interested users.

Portals and Web 2.0

The Web 2.0 seeks to establish networks or users communities with similar interests, working with content production applications (blogs, multimedia content editors, wikis, etc.) and share comments on social networks, finding information by searching based on folksonomies. In this new environment is easy to find institutions that have enabled a space, while maintaining their traditional portals, trying to expand their level of impact in Internet. Content syndication technology (RSS and Atom mainly) has helped to facilitate the monitoring of the current report, allowing the reuse of digital content as expression of the theoretical conception of the primeval hypertext (Conklin, 1991). In

---

[1] Drupal, Joomla, phpnuke, etc

parallel, users retrieval information from free labeling systems with shared metadata, assigning labels or keywords (tags) to information resources (a page, a photo, a link, a post, etc.). Tags are shared by the community of users: *social tagging or folksonomies[2]*. The implementation of these new applications and the adoption of these collaborative habits is changing the portals, 'is pervasive in the use of information, loosening control over it and handing it to citizens who are proactive in getting customised information and using it to suit their needs' (Peña López, 2008).

SKOS.

SKOS (Simple Knowledge Organization System) is a W3C 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. In SKOS concepts are identified with URI references and can be labelled in text strings. 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. SKOS structure is described in a series of documents including *SKOS Primer* (W3C, 2008a2), *SKOS Reference* (W3C, 2008a1) 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.

SKOS Applications.

Common problems of Information Retrieval (relevance, precision, comprehensiveness, etc.) combine with others inherent in the nature of the Web 2,0, among other reasons because the quality, structuring and originality of content have not evolved in parallel with the increase of Web publication. The situation would improve with the use of standardised metadata models 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.

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[3]. SKOS may evolve, incorporating new types of relationship and adapting with RDF and OWL. The thesauri in particular and the conceptual schemes in general have been used in information retrieval and organisation tasks, although only partial solutions have been devised. SKOS could be used to describe how the content of a website is organised, and that would be involved, using an ontology, in the design of Website navigation systems.

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. Using SKOS into a website, like

---

[2] Although this phase is still in development, many thesaurus management or document indexing applications - Del.icio.us o Flickr - already use this model.
[3] *Dublin Core* for example.

a terminological control tool, would help to standardize the description of documents while enhancing the effectiveness of information retrieval

An example of use: a collaborative portal for the Cultural Industries management.

We present an example of use within an intranet portal for the cultural management. This portal integrates databases of cultural industries along with information about cultural events that take place throughout the year. The databases management and the adding of information about the events is conducted in a collaborative form. Cultural enterprises are registered in the portal, they fill cards introducing their description. If one of these companies wants to spread an event, the portal provides a space for the information and an editing tool for maintaining accurate information. All these pages are also labeled directly by the users of the portal. At the same time, administrators can use the portal as a communication way for their activities, also labeling the documents generated. It is easy to foresee a problem in the indexing consistency by the lack of labeling control, that can be overcome through the use of SKOS within the website to provide greater stability in the semantic description of this content.
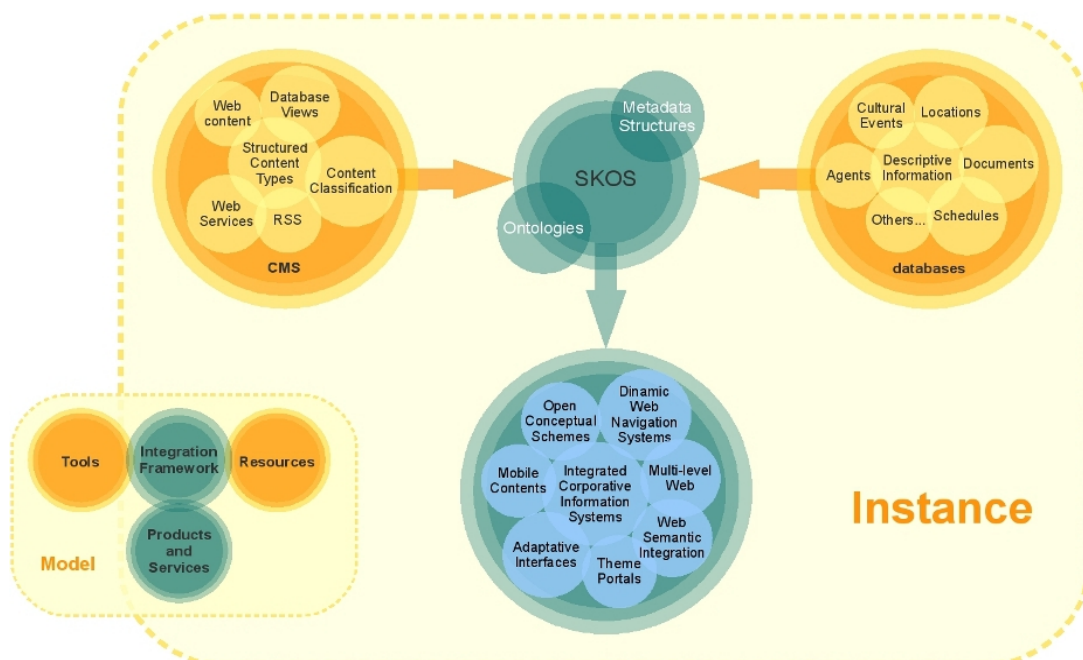

Figure 1: SKOS applications in Information Systems, complemented by ontologies

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 and in the indexing of their pages. For information managers. SKOS offers a closer approach to knowledge organisation 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.

The application of SKOS is based on a development model that consider four elements:

- *Information resources*: Includes all the information items related in the corporative system. Therefore, the system can obtain information from databases, repositories, document files or any type of content unit.
- *Tools*: This element covers the applications, those with which it is possible manage and process the information resources.
- *Services and Products*: Considers the results of applying Tools and Information resources by means of a Integration Framework
- *Integration Framework*: It makes possible the description of semantic structures. Basically this element brings together different specifications, mainly of the Semantic, as OWL, RDF/RDFS, SKOS among others.

The core of the information system are basically the databases: *agents* (institutions, organizations and enterprises) and the *cultural events* and their descriptive information (program, locations, dates, tickets, etc.). The term 'database' implies highly structured elements. However, it is possible to include managed objects for collaborative document management systems. This would provide greater capacity for integration and interrelation between the content.

The proposed integration framework consists of SKOS for the creation of frameworks for the design of controlled vocabularies to be used in indexing. It is essential to use metadata serialized in RDF/RDFS for the provision of data by web services. Ontologies are also part of the integration framework for defining the processes and structures needed for the conversion of the conceptual schemes developed with SKOS. The information stored in databases is integrated into the content managed by the CMS, together in a variety of products and services. This model envisages the CMS as an integral part of corporate information systems and the use of metadata and conceptual schemes combined with SKOS ontologies to define thematic portals. The contents of these websites are designed using dynamic navigation systems as a result of adapting frameworks through ontologies. The user interface and content development for all types of devices (like mobile) is developed from the same sort of adjustment, using the processes of indexing and description of these contents. These websites are structured in three levels: conceptual, navigational and structural (documentary). Moreover, the conceptual schemes could be used by external applications through Web services, resulting in their integration in the Semantic Web.

Conclusion.

The main conclusion is simple, SKOS would improve the description level of the documents of the portal. This better level will contribute to expand the searching capabilities and to do more suitable the user experience. Obviously, it improves the semantic description of the information resources and the information retrieval effectiveness.

References.

Conklin, Jeff. "Hypertext: an introduction and survey," *IEEE Computer*, vol. 20 no. 9 (1987), 17-41.

Peña López, I. 'Towards e-Government 2.0: Review of the IV Internet, Law and Politics Congress – Political Track". IDP Issue 7 (2008) http://www.uoc.edu/idp/7/dt/eng/pe-na.pdf

W3C (2007). SKOS Use Cases and Requirements. W3C Working Draft 16 May 2007. Edited by: A. Isaac, J. Phipps and D. Rubin. Retrieved on 15 June 2008, from http://www.w3.org/TR/2007/WD-skos-ucr-20070516/

W3C (2008a1). SKOS Reference. W3C Working Draft 29 August 2008. Edited by: A. Miles y S. Bechhofer. Retrieved on 1 September 2008, from http://www.w3.org/TR/2008/WD-skos-reference-20080829/

W3C (2008a2). SKOS Primer. W3C Working Draft 29 August 2008. Edited by: A. Isaac and E. Summers. Retrieved on 1 September 2008, from http://www.w3.org/TR/2008/WD-skos-primer-20080829/