

Procedures for extracting keywords from web pages, based on search engine optimization

Authors: Mari Vallez (Universitat Oberta de Catalunya, Universitat Pompeu Fabra) , Cristòfol Rovira (Universitat Pompeu Fabra) , Lluís Codina (Universitat Pompeu Fabra) y Rafael Pedraza (Universitat Pompeu Fabra)

Citation: Vallez, Mari, Rovira, Cristòfol, Codina, Lluís, Pedraza, Rafael (2010). "Procedures for extracting keywords from web pages, based on search engine optimization". *Hipertext.net*, 8, http://www.upf.edu/hipertextnet/en/numero-8/keywords_extraction.html



Abstract: We present a research project whose main objective is the development of a tool that facilitates the assignment of keywords to web documents. We synthesize the main characteristics and features of the tool we are constructing and analyze its theoretical basis. This line of research is motivated by the current interest in semantic technologies as a mechanism to facilitate and optimize access to information.

Keywords: Keywords extraction, Metadata, Semantic Web, Meta tags, Semantic Annotation, Taxonomies, Thesaurus, Ontologies, Search Engine Optimization (SEO), Semantic technologies, Controlled language.

Table of contents

1. Introduction
2. The tool
 - 2.1. Initial limitations
3. Theoretical basis
 - 3.1. Semantic Web
 - 3.2. Metadata and semantic annotation
 - 3.3. Controlled languages: taxonomy, thesaurus and ontology
 - 3.3.1. Taxonomy
 - 3.3.2. Thesaurus
 - 3.3.3. Ontology
 - 3.4. Search Engine Optimization (SEO)
4. Conclusions
5. Bibliography

1. Introduction

The World Wide Web offers a universe of information and knowledge within which it can often be difficult to locate the pertinent information needed in any particular case. Algorithms based on links analysis have greatly improved the ranking search engine results, although there is still a long way to go, especially if we consider using intelligent search engines to automate a broader portion of the information retrieval process.

The proposed Semantic Web (Berners-Lee, 2001) could represent important progress in this domain because it offers a paradigm shift, transforming the current web, based almost exclusively on natural language, into a structured, organized web where natural language content is semantically tagged in an explicit manner so that machines can interpret it, facilitating automatic processing of web content. One of these processes is information retrieval (Ding, 2005).

Tagging and metadata are, therefore, among the basic elements of the Semantic Web project, with implications for all aspects of the creation and distribution of web content. The new paradigm implies a new form of content creation, in which the creators must assume the task of tagging if they want the content to be semantically interpreted by new search engines and user applications. In this context, we see the need for tools that facilitate the automatic or semi-automatic creation of this metadata and ensure its quality.

This article presents a research project with the main objective of developing and exploring the possibilities of a tool that facilitates semi-automatic assignment of keywords to web documents. This tool will extract keywords based on their occurrence in the text of the document and on a taxonomy that will be predefined although it can always be edited and modified. Candidate keywords generated by this procedure will be ordered according to the relevance criteria of the algorithms used in ranking search engine results.

We synthesize here the main characteristics and features of the tool we are constructing and analyze its theoretical basis.

This line of investigation is motivated by the current interest in semantic technologies as a mechanism to facilitate and optimize access to information (Davies, 2009; Kiryakov, 04), the W3C Semantic Web project also must be placed in this context.

2. The tool

The objective of the tool we propose to develop is to facilitate the semi-automatic assignment of metadata using keywords to represent the thematic content of web documents.

In general terms, it works as follows:

1. The text content of a web page is processed, comparing the terms used in the document with those of the taxonomy previously developed and for the topic of the page in question.
2. The terms of the page that also appear in the taxonomy are selected as candidate keywords to represent the content of that document.
3. Each candidate keyword is assigned a score based on the relevance criteria normally used in algorithms for ranking search engine results, e.g., number of occurrences, the location of the term in main areas of the web page (e.g., <title>, headers (<h1, h2> ...), or anchors), the url, or tagging with emphasis markers such as bold type.
4. Candidate keywords are prioritized by the relevance score obtained in each case. The system permits automatic assignment of keywords to the document in question, based on a predetermined threshold score or the manual selection of the best keywords from the list of candidates.
5. The group of keywords selected can become part of the document's metadata in some of the standard web formats, such as Dublin Core metadata as part of the document's source code, as an external RDF file, etc.

Assigning a group of keywords that pertain to a web document has three important consequences:

1. Facilitation of both the presentation of and access to information. Coding the group of keywords in a metadata format synthesizes the document in a way that offers great semantic capacity. It helps provide access to the information because it facilitates searching by concepts (Douglas, 2006). The keywords obtained from controlled language, such as a taxonomy, are a means of bringing the author of the content and the recipient, in this case the person searching for information, closer together. It is a proposed solution to a major part of the problem presented by linguistic variation (synonymy and polysemy) of the natural language (Bontcheva, 2006). At present, there is no evidence that search engines use keyword metadata found in web documents in any generalized manner. Nonetheless, keywords are an element that can be helpful in page ranking if they are in fact related to the content of the document. In addition, they are a retrieval tool that adds value to internal search engines, not only with respect to Intranets but also internal searches of open web sites with a large volume of information.
2. Improvement of rankings. We must emphasize that the tool we plan to develop will also be of interest from the perspective of search engine rankings. The candidate keywords with the highest score will be the terms that correspond with the probable topics where the page will be well ranked. Therefore, the authors will have information that allows them to determine whether it is necessary to revise the content for more efficient processing by search engines to meet their objectives.
3. Preparation for the new intelligent search tools. The page will be much better prepared for the semantic web and for future processing by intelligent agents.

We would highlight the importance of our research with respect to social communication, especially the two primary directions in which it is intended to make a contribution: information retrieval and search engine ranking. The current information society has brought with it new communication channels, a large volume of information sources, and powerful information processing tools (Castells, 1997). The project we propose would be an effort to optimize communication processes within this context.

2.1. Initial limitations

Even though the proposed tool has a polyvalent goal, in the first phase we will explore its efficiency and effectiveness in a limited context, defined by these four elements:

1. Limitation by type of document. Scientific or academic documents with a large amount of textual information will be processed.
2. Limitation by topic. Documents related to topics connected to the sciences of the web will be processed.
3. Limitation by type of processing. The efficiency of assigning keywords based on the results obtained by searches using the primary search engines: Google and Yahoo!.
4. Limitation by results. This exploratory research has the ultimate goal of evaluating the efficacy and efficiency of the proposed tool.

Limitation by topic is especially important. An important part of successful extraction and assignment of keywords will depend on the quality of the controlled language (taxonomy) employed. As would be expected in the field of knowledge representation, limitation to one domain (in this case, Sciences of the Web) allows us to develop a more complete taxonomy that will have greater potential for success.

3. Theoretical basis

Various disciplines and areas related to semantic technologies converge in our proposal. In recent times, this type of technologies is awakening expectations. Nonetheless, reports from different international information technology research and advisory companies that investigate technology trends (e.g., Gartner: <http://www.gartner.com> and Forrester: <http://www.forrester.com>) demonstrate the low introduction of this technologies in organizations. Semantic technologies are treated with caution, since after years of development they are not yet considered sufficiently established, even though their great potential is acknowledged (Gartner, 07A / b), and the Semantic Web, as defined by W3C, is considered as an emerging technology with 1-5% market penetration that is still more than 10 years from large-scale development.

Semantic technologies are integrated into various topic areas, techniques and disciplines with very diverse origins, all of which are connected, such as for example: information retrieval, natural language processing, information extraction, controlled languages, semantic annotation, ontology population , etc.

The tool presented is framed in the context of semantic technologies, and is therefore related to all of the relevant topic areas. Nonetheless, there are three areas of interest that have a more direct impact on the research that is underway:

- metadata and semantic annotation
- controlled languages: taxonomies, thesauri and ontologies
- search engine optimization

3.1. Semantic Web

One of the final objectives of the Semantic Web is the creation of a system of intelligent agents that would be able to carry out inferences (in an automated format) with the information published to the web. This objective is more utopic than real, even in the mid-term (Codina 2006). Nonetheless, many of the developments that occurred recently thanks to the new paradigm have given rise to new services that are experiencing great success in the current web environment. One of the successes to the Semantic Web is the implementation of various standards for the representation and processing of information in a more sophisticated manner. These standards permit the expression of metadata in a logical format; at the same time they represent controlled languages (for example, thesauri or ontologies) so they can be processed by computer programs. These formats (including XML, RDF, SKOS-Core and OWL) are being utilized in a generalized manner.

After associating semantic values with available resources, it is necessary to have tools that facilitate efficient information localization. These instruments would be the so-called intelligent agents, which they could interpret and comprehend the information. Later, they can provide the processed information to the users. Nonetheless, these tools are still far from being reality.

These technologies will enable the conversion of the web into a globally described infrastructure where it will be possible to share and "recycle" data and documents between different types of users. This should allow users to retrieve the information they need in a more precise manner, in accordance with the content described.

Our research is situated in the context of migration toward the Semantic Web (Pedraza-Jimenez, 2008). This development will help the first phase, with the objective of facilitating the assignment of keywords to web documents. Fortunately, it is not necessary to wait for better development of the Semantic Web to begin to enjoy the advantages of this tagging. As we will see, assigning keywords generates immediate improvements, both in information retrieval using current search engines and in rankings in their lists of results.

3.2. Metadata and semantic annotation

As we have seen, metadata is one of the fundamental elements of the Semantic Web, i.e., information (data) describing the content of the associated documents and explicitly representing their meaning (Aguado de Cea, 2002).

The semantic annotation achieved with metadata gives semantic content to documents and allows machines to interpret the information of a specific domain.

Assigning metadata is a complicated, slow and costly process. One of the tasks that would help in this phase is the construction of tools to automatically extract information and then convert it to metadata (Cunningham, 2005).

Information Extraction is the term utilized for the activity of automatically extracting specific information from natural language texts. Different approaches to realizing this process that can be grouped in two principal categories: machine learning and systems based on rules and patterns (Flynn, 2007).

Machine learning techniques are primarily based on calculation of probability, based on training collections. They adapt very well to different environments, although we also must cite some of their drawbacks: they require many examples, selection of proper sources is complicated, they consume considerable amounts of time before results are obtained, the productivity declines with increased heterogeneity of the documents, and the adaptation or inclusion of new fields for extraction is complex.

Systems based on rules and patterns are grounded in the experience of the individual who develops them and therefore specialists in each domain are needed to define the rules for data extraction. The process of definition requires much time and the introduction of changes to the systems is complicated because in some cases this can mean a return to redefining the system.

Annotation tools convert the semantic content extracted from web pages to metadata (Ureña, 2006). These applications can be classified into two large groups: external tools and authoring tools. The first group allow the association of metadata to web pages, but does not store the data on the page itself; instead, the information is saved in an external repository. Annotation tools intended for authors assist with the incorporation of metadata either within the web page or elsewhere, according to the applicable standards (e.g., xml, rdf). Our project falls within the latter group.

There are various approaches to semantic annotation, which can be grouped into three large categories. The first model is based on linguistic annotation, the discipline that originated the concept of annotation is corpus linguistics. The objective is to tag a text based on different levels of language, beginning with the lemma and proceeding through the morphosyntactic, syntactic, semantic, and discursive levels (Buitelaar, 2003). The definition of terms and their interrelationship is interesting because this information can affect the value of a term as a keyword. Nonetheless, this system is not well established in the context under study because it is computationally expensive.

The second approach is based on ontologies, which are used as the central resource for extracting connections between terms, thereby demonstrating their meaning (Niremburg, 01). Although this system is having great impact at present, it is not consolidated because the process of ontology creation is not yet well established (Maedche, 2001; Pedraza-Jimenez, 2007).

The third approach proposes the use of a controlled language, such as a thesaurus or taxonomy, to facilitate the assignment of metadata. This is the model traditionally used in Library and Information Science to manually index information. This model is directly linked to the assignment of metadata and semantic annotation (Guyot, 2006). The project presented here is based on this last model, with the addition of a layer of automatic processing based on the presence in the taxonomy of a document's terms and their validation with respect to the criteria applied by a ranking algorithm.

3.3. Controlled languages: taxonomy, thesaurus and ontology

Controlled languages are mechanisms for the presentation and organization of knowledge, with the objective of controlling and normalizing the assignment of keywords to a document. Therefore, they are among the essential elements for effective use of metadata, whether we apply them in a manual or automated context. Taxonomies, like thesauri and ontologies, are tools that permit the structuring of information and provide a minimum of semantics (Gilchrist, 2003). The current growth of information available on the web has generated new possibilities for the design and development of controlled languages.

The main characteristics of controlled languages that are most related to the project described here are presented below.

3.3.1. Taxonomy

Taxonomies are a form of hierarchical classification of content. The concept originated in systematic biology, which studies the relationship between organisms and their evolutionary history. They are used to establish classification criteria, which enable diverse organisms to be grouped according to their shared characteristics.

This idea has been extended to other contexts and a taxonomy has come to be understood as a semantic hierarchy in which units of information are related at the level of classes and subclasses as a way of organizing knowledge (Chris, 2007). Taxonomies are the backbone of the ontologies we will see below.



Figure 1. Example of Taxonomy

3.3.2. Thesaurus

Thesauri are lists of words or terms used to organize knowledge within a domain, with the objective of controlling the thematic description of a document. A thesaurus is a type of language for documentation that consists of standard terms, descriptors, and the semantic and functional relationships that are established between these terms. The semantic relationships used are equivalence, association and hierarchy (López-Huertas, 99).

Thesauri have highly controlled terminology and a great capacity for specialization. They are very useful for optimizing the information retrieval in closed systems, since they help to remove ambiguity and support

semantic standardization in the expression of document content. They are standard tools in libraries, documentation centres, image banks and scientific databases, but are not as widely used in settings related to information retrieval.

Lenguaje de indexación [32]	
English term:	Indexing languages
Terme français:	Langage d'indexation
Русский термин :	Языки индексирования
NA <i>Lenguaje artificial empleado por los sistemas de</i>	
MT	5.05 Ciencias de la información
UP	Lenguaje de búsqueda documental
UP	Lenguaje de indización
UP	Lenguaje documental
TE	Control terminológico [18]
TE	Lenguajes controlados [1]
....TE2	Sistema de clasificación [408]
.....TE3	Encabezamiento por materia [193]
TE	Lista oficial de autoridades [29]
TE	Tesaurus [270]
....TE2	Compilación de tesaurus [69]
TR	Lingüística [468]
TR	Recuperación de información [336]
TR	Terminología [409]

Figure 2. Example of a Thesaurus

3.3.3. Ontology

Ontologies have their origins in metaphysics, a branch of philosophy that focuses on the nature of reality.

Ontology was devised to describe the existence of being and basic relationships, and to define entities and their typologies (Echeverría, 1998).

Since the '80s, ontologies have been used in artificial intelligence to represent knowledge in a particular area. Ontologies are formal, explicit specifications that represent concepts and relationships in a particular domain (Gruber, 1993).

The spectacular evolution of the web and the great interest that exists in the development and implementation of the Semantic Web have led to a very important role for ontologies, even though they are more symbolic than real. In theory, they are one of the key pieces in communication between organizations, individuals and software applications and as such facilitate interoperability between systems. Thanks to the knowledge stored in ontologies, intelligent agents could directly extract data from web pages, process them and make inferences. Nonetheless, this functionality is not yet available outside of certain restricted domains.

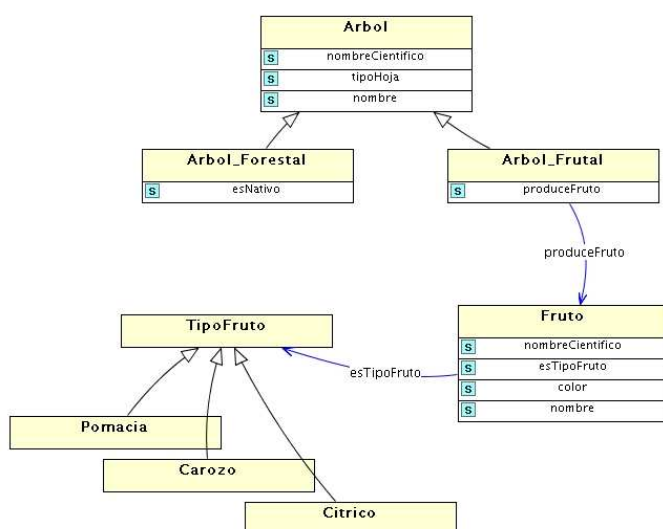


Figure 3. Example of Ontology

Knowledge standardization in ontologies is one of the initial barriers to implementation of the Semantic Web, since the construction of ontologies is an extremely slow, costly and error-prone process. It requires great effort and a degree of specialization that many organizations do not have within their reach.

Various methods and tools exist that can help in the semi-automatic creation and development of ontologies. Ontology engineering is the discipline charged with the study and construction of tools that have as their objective the design of mechanisms for a more agile process of constructing ontologies for a particular domain. Nonetheless, there is no consensus in the scientific community concerning the different specific phases that will be involved in this development.

As stated above, the tool we propose to develop will be based on a manually constructed taxonomy, a very specific topic, and with the majority of the relationships typical of a thesaurus, such as equivalency, association and hierarchy.

3.4. Search Engine Optimization (SEO)

In search engines, ranking is done by a group of techniques used to make a web page appear near the top of lists of results when users execute particular search equations.

For the authors of web content, well-ranked web pages are primordial, given that the proportion of traffic from search engines is constantly on the rise. Recent studies have found that between 50 and 70 percent of the total traffic to a web site could be driven by search engines (Valentine, 2007) and cases where the percentage reaches 90% are not unusual.

One of the principal phases in improving the ranking of a particular page is determining the keywords that will achieve the best page ranking. Keywords must be selected on the basis of content, objectives and the intended audience. In this context, it is useful to identify three or four primary keywords, taking into account the following aspects (Gonzalo, 04):

- Relationship with the content. The selected keywords must reflect the content of the web page and coincide with those users would use to locate the web page to be ranked.
- Popularity and competency. The most frequently used individual terms tend to have great competency and therefore it becomes difficult to position a page among the top results using these terms. The solution is normally to select "key phrases" consisting of two or three words that are not very popular and using them to optimize the web pages.

The efficacy of the selected keywords can be evaluated using the Keyword Effectiveness Index. This indicator calculates the potential of a particular word, based on its popularity, number of searches per month using the term, and its competitiveness, or the number of results obtained when this word is used in a search.

On the other hand, to improve ranking we must take into account how a search engine's ranking algorithms behave with respect to the text on web pages. It is known that the principal search engines post a page in which the words used in searches are placed in areas of special relevance, such as in the title (<title>), headings (<h1>, <h2> ...), the anchors (<href>), boldface, or graphics titles (<title>), or even in the text at the beginning of the document or the anchors of the links on other pages that point to the page we want to position.

Search Engine Optimization has two key implications that have been considered as important in this study:

1. Creation of the taxonomy. In the process of selecting taxonomy terms, it is a priority to consider the Keyword Effectiveness Index and prioritize the terminology most often employed by web users and facilitate its retrieval.
2. Prioritization of candidate keywords. As indicated above, our tool provides users with a list of candidate keywords based on the taxonomy and the content of the page being analyzed. This list will be prioritized based on the greater or lesser presence of terms in relevant areas of a page that apply the algorithms for ranking search engine results. As a consequence of this prioritization, the user knows which terms are most important in describing the content in agreement with a group of criteria widely used by search engines. In addition, it will be important to explore whether automatic assignment based on a particular store is effective.

4. Conclusions

Even though the Semantic Web remains utopic, migration in that direction is a definite reality. Metadata tagging of web content using standard formats is becoming generalized and the available content on the Internet is slowly preparing for a future we are not sure will be exactly what W3C has promised. However, the final product is not as important as the interesting steps being taken toward a more functional web.

This project is within this phase of migration toward the Semantic Web, with a certain amount of scepticism about the final result and great enthusiasm for the preliminary results. Semi-automatic extraction of keywords from web pages could be a new aspect with immediate and interesting repercussions that, at the same time, could be one more step toward the final objectives. The research we propose explores the possibilities of automating keyword assignment, using relatively simple procedures based on models that are widely used in the disciplines of Information Retrieval, Search Engine Optimization and Library and Information Science.

5. Bibliography

Aguado de cea, Guadalupe; Álvarez de mon, Inmaculada; Pareja, Antonio (2002). «Primeras aproximaciones a la anotación lingüístico-ontológica de documentos de web semántica: OntoTag». IN: *Revista Iberoamericana de Inteligencia Artificial*, num. 17, p. 37-49.

Berners-Lee, Tim; Hendler, James; Lassila, Ora (2001). "The Semantic Web". *Scientific American*, vol. 284, num. 5 (May), p. 34-43.

Buitelaar, Paul; Declerck, Thierry (2003). "Linguistic Annotation for the Semantic Web". IN: Siegfried Handschuh

- and Steffen Staab: *Annotation for the Semantic Web, Frontiers in Artificial Intelligence and Applications Series*, vol. 96, IOS Press, 2003.
- Castells, Manuel (1997). *La era de la información. Economía, sociedad y cultura*. (3 vols.). Madrid: Alianza.
- Chris, M. (2007). "Taxonomy development: assessing the merits of contextual classification". IN: *Records Management Journal*, 17, p. 7-16.
- Codina, Lluís; Rovira, Cristòfol (2006). "Web Semántica: visión global y análisis comparativo". IN: *Tendencias en documentación digital*. Gijón: Trea.
<http://eprints.rclis.org/archive/00008637/http://eprints.rclis.org/archive/00008637/>.
- Codina, Lluís; Marcos, Mari-Carmen; Pedraza, Rafael (coords.). *Web semántica y sistemas de información documental*. Gijón: Trea, 2009.
- Cunningham, Hamish; Bontcheva, Kalina; Li, Yaoyong (2005). "Knowledge management and human language: crossing the chams". IN: *Journal of Knowledge Management*, vol. 9, num. 5, p. 108-131.
- Davies, John; Grobelnik, Marko; Mladenčić, Dunja (2009). "Challenges of Semantic Knowledge Management". IN: *Semantic Knowledge Management*, p.245-247.
- Ding, Li, et al. (2005). "Search on the Semantic Web". IN: *IEEE Computer*, vol. 38, num. 10 (October), p. 62-69.
- Douglas, T.; Ceri, B.; Dorothee, B.; Daniel, C. (2006). "Query expansion via conceptual distance in thesaurus indexed collections". IN: *Journal of Documentation*, 62, p. 509-533.
- Echeverría, Rafael (1998). *La ontología del lenguaje*, Editorial Dolmen, 5ª Ed, Palma de Mallorca.
- Flynn, P.; Zhou, L.; Maly, K.; Zeil, S.; Zubair, M. (2007). "Automated template-based metadata extraction architecture". IN: *Lecture notes in computer science*, 4822, p. 327-336.
- Gartner Research Group (2007a). "Taking stands on the Semantic Web". Id Number: G00148696.
- Gartner Research Group (2007b). "Finding and exploiting value in Semantic Technologies on the web". Id Number: G00148725.
- Gilchrist, Alan (2003). "Thesauri, taxonomies and ontologies: an etymological note". IN: *Journal of documentation*, Volum 59, Num. 1, p. 7-18.
- Gonzalo Penela, Carlos (2004). "La selección de palabras clave para el posicionamiento en buscadores". [on line]. IN: *Hipertext.net*, núm. 2, 2004. <http://www.hipertext.net>.
- Gruber, Thomas (1993). "A translation approach to portable ontologies". IN: *Knowledge Acquisition*, vol. 5, num. 2, p. 199-220.
- Guyot, Jacques; Radhouani, Saïd; Falquet, Gilles (2006). "Conceptual Indexing for Multilingual Information Retrieval". IN: *Accessing Multilingual Information Repositories, Lecture Notes in Computer Science*, vol. 4022, p. 102-112.
- Kiryakov, Atanas; Popov, Borislav; Terziev, Ivan; Manov, Dimitar; Ognyanoff, Damyan (2004). "Semantic annotation, indexing, and retrieval". IN: *Journal of Web Semantics*, 2 (1), p. 49-79.
- López-Huertas, M. J. (1999). "Potencialidad evolutiva del tesoro: Hacia una base de conocimiento experto". IN: *La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información*, IV Congreso ISKO, p. 133-140.
- Maedche, A.; Staab, S. (2001). "Ontology Learning for the Semantic Web". IN: *IEEE Intelligent Systems, Special Issue on the Semantic Web*, vol. 16, num. 2, p. 72-79.
- Miller, George A. (1995). *WordNet: a lexical database for English*, Communications of the ACM, Vol. 38, Num. 11, p. 39-41.
- Nirenburg, Sergei; Raskin, Victor (2001). "Ontological semantics, formal ontology, and ambiguity". IN: *Proceedings of the international Conference on Formal ontology in information Systems*, ACM, vol. 2001, p. 151-161.
- Pedraza-Jimenez, Rafael; Codina, Lluís; Rovira, Cristòfol (2007). "Web semántica y ontologías en el procesamiento de la información documental". IN: *El profesional de la información*, 2007, noviembre-diciembre, v. 16, n. 6, p. 569-578.
- Pedraza-Jimenez, Rafael; Codina, Lluís; Rovira, Cristòfol (2008). Semantic Web adoption: online tools for web evaluation and metadata extraction. IN: Da Ruan et al. (ed) *Computational Intelligence In Decision And Control*. Proceedings Of The 8Th International Flins Conference. New Jersey: World Scientific Publishing Co Pte Ltd, 2008. ISI Document Delivery No.: BIF16.
- Rovira, Cristòfol; Marcos Mari-Carmen (2006). "Metadatos en revistas-e de Documentación de libre acceso". IN: *El profesional de la información*; 15(2): p.136-143.
- Rovira, Cristòfol; Marcos Mari-Carmen (2007). "Repositorios de publicaciones digitales de libre acceso en Europa: análisis y valoración de la accesibilidad, posicionamiento web y calidad del código digital". IN: *El profesional de la información*; 16(1): p. 24-38.
- Uren, Victoria, et al. (2006). "Semantic annotation for knowledge management: Requirements and a survey of

the state of the art". IN: *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 4, num. 1 (January 2006), p. 14-28.

Valentine, Mike (2007). Search Engine Optimism.

http://searchengineoptimism.com/Google_refers_70_percent.htmlhttp://searchengineoptimism.com/Google_refers_70_percent.html

Vossen, Piek (2004). "EuroWordNet: A multilingual database of autonomous and language-specific wordnets connected via an inter-lingual-index". IN: *International Journal of Lexicography*, Vol. 17, Num. 2, p. 161-173.



Last updated 08-11-2010
© Universitat Pompeu Fabra, Barcelona