The Semantic Dimension in Information Retrieval, from Document Indexing to Query Reformulation

Abdelkrim Bouramoul

Computer Science Department, Misc Laboratory, University of Constantine,
B.P. 325, Constantine, Algeria, <a.bouramoul@yahoo.fr>

Abdelkrim Bouramoul is an Assistant Professor in the Computer Sciences Department, Faculty of Engineering, University of Mentouri, Constantine, Algeria, since 2006. He is also researcher in OSIG research group of the MISC Laboratory at the same University. He obtained a magister degree in software engineering in 2006, a post-graduate diploma from the Paris-Sud University, France, in 2004 and engineer degree from the Constantine University in 2002. Now, he is preparing his PhD in Computer Sciences at the Constantine University. Abdelkrim Bouramoul has published several articles in international journals and conferences. His research interests concern semantic and contextual information retrieval.

Bouramoul, Abdelkrim. **The Semantic Dimension in Information Retrieval, from Document Indexing to Query Reformulation.** *Knowledge Organization, 38(5)*, 425-437. 11 references.

ABSTRACT: In the context of this research, we present an approach for representing the semantic content of documents and guiding the automatic query reformulation using a domain ontology. The aim is to improve the performance of information retrieval systems. In order to operationalize our proposal, the development of a set of external resources was needed, so we have constructed the 'AnimOnto' domain ontology relating to the animals domain and a document base that covers the same field. They were used to test and validate our proposal. Specifically, we propose in this paper a general architecture based on three complementary processes, this architecture uses the ontology during the semantic indexing stage and in the query reformulation phase. We also describe the 'AnimeSe Finder' tool (***Anim***al ***Se***mantic ***Finder***). The latter has the advantage of being generic and adaptable to other search types. It is possible to use another ontology with another document base for a new domain, to exploit the general functionalities offered by the 'AnimeSe Finder' tool.

Received 5 May 2011; revised 16 June 2011; accepted 20 June 2011.

1.0 Introduction

One of the current challenges of Information Retrieval Systems (IRS) is to develop tools able to integrate more semantics in their treatment. The aim is twofold: 'understanding the content of documents' and 'understanding the user needs' and to be able to link them. It is possible, according to Boughanem et al. (2009), to regroup the knowledge reflected in an IRS into three main classes:

– Knowledge about users (profiles or user models);
– Knowledge about the documents (index); and,
– Knowledge about the concepts of the application field.

The first point has already been explored in our research (Bouramoul et al. 2010) when we proposed a system for taking into account the user context via their profile for the query reformulation. The obtained results using this technique have been convincing in the sense that the documents returned by the contextual reformulation were more relevant than those obtained with the original query. On the other

hand, the satisfaction of the users who participated in the experiment was remarkably better.

A perspective of our previous contribution was to use the other two aspects described by Boughanem, namely the knowledge about documents and the concepts of the application field as a reference to reformulate the user queries. In this paper, we substantiate this idea by using an external resource (ontology) to improve the relevance of Information Retrieval Systems (IRS), the same ontology will also be used to index and describe the content of documents. Our work falls within the area of semantics consideration via ontology in IRS, particularly that of using a domain ontology to index documents and reformulate the user queries by exploiting relations between concepts. This idea is not new, but the value of our contribution lies in its completeness. This research here deals autonomously with the search process, starting from the ontology creation, via the establishment of the system architecture and its implementation, and ending at the validation of results.

This paper is organized into three main parts: The first presents the areas on which our contribution is based, namely the use of ontology in the Information Retrieval (IR) field and the different approaches for query reformulation. In the second part we present our own contribution. Here we first determine the choices that we have adopted for the system parameters, then we present the architecture of the tool that we propose, a description of its implementation, an evaluation of its performance, and the discussion of the obtained results. The third part consists of the conclusions and perspectives.

2.0 Ontology, a crucial need in IR

Several definitions of ontology have emerged in the last twenty years, but the most referenced definition and also the most synthetic one is probably that given by Gruber (1993, 199): "An ontology is an explicit specification of a conceptualization." Starting from this definition, ontology is used in the IR field to represent shared descriptions related to a more or less formal domain in order to add a semantic aspect to IRS. It is therefore natural that research focused on the integration of ontology in IRS. A first solution aimed to build ontologies from corpora on which IR tasks will be carried out (Saias and Quaresma 2003; Koo et al. 2003). A second solution aims at reusing existing resources. In this case, ontologies are usually chosen only from the knowledge domain that they address (Vallet et al. 2005; Baziz et al. 2005). In general, the knowledge represented by ontology can be used at three different levels in the IR process. It can help to index documents, and then called semantic indexing. It can also assist the formulation of the user need and access to documents. Finally, the ontology can be used in the model itself to achieve matching between the need and the document particles.

In our case, we use domain ontology in the indexing process and query reformulation to interrogate a document base using the 'AnimSe Finder' tool that we propose. The idea is to guide the user's query through the conceptual network of the 'AnimOnto' ontology, developed for this purpose, to enrich it with new words coming from the vocabulary of this ontology. The interest is twofold:

- *Increase the recall* by the query expansion taking into account the terms that are not present in its initial form. This can be done by adding synonyms recovered from the ontology (related concepts) to the query terms that are in relation with the chosen domain (animals); and,
- *Increase precision* thanks to the semantic indexing of documents using concepts which are recovered from the ontology instead of ambiguous terms. This can be done by extracting terms guided by the chosen domain ontology, then by a weighting of concepts using the relationships between them.

3.0 Query reformulation, approaches and limitations

The query reformulation aims to modify the user's query by adding significant terms. The idea of query refinement is not new; several approaches use different techniques for selecting terms to be added to the initial query. We distinguish three types of query reformulation approaches and the difference between them lies on the one hand in the source of terms used in the reformulation. On the other hand, it lies in the method used for selecting terms to be added to the initial query.

The first type of the approaches is based on a global analysis of the considered collection of documents and is most commonly based on a statistical analysis of document corpus (Cui et al. 2002). The objective is to increase the frequency of words appearing together in one document and to select the terms with the highest coefficient. The information thus obtained is used to reformulate query automatically by adding terms related to the terms already used in the query. The terms added from the docu-

ments give a better correspondence between information need and the document collection.

The second type of approaches based on the principle of relevance feedback aims to reformulate the initial query such that it corresponds better to the content of the documents collection. The principle is as follows; the user submits his initial query and the system returns an initial set of documents that the user has to evaluate (relevant, irrelevant). The information on the relevance of initially returned documents is used for selecting terms to be added to the initial query. We quote in this category the work of (Lin et al. 2006) in which the system offers, based on the first query, a set of documents and according to those viewed by the user, the system updates its terms index in concordance with automatic learning methods.

The last type of the approaches, described in the literature, uses external resources of terms such as thesauri or ontology that contain the vocabulary used in the query enrichment. Such approaches use ontology with equivalence and subsumption relations (Navigli and Velardi. 2003), in order to extract the terms to be added to the initial query.

4.0 System parameters

The parameters of the system that we propose concern the following elements:

4.1 The concept source

To recover terms that will be added to the initial query, we use the 'AnimOnto' ontology as a source of concept. This is a domain ontology related to the domain of animals created under Protégé. The latter can produce two code formats for describing an ontology (OWL and XML). For implementation reasons we chose the XML format. The XML describing the 'AnimOnto' ontology characteristics is subsequently used as input for the reformulation process.

4.2 The concept selection method

The concepts are recovered from the ontology as follows: if the concept (C1) is present as a term in the initial query, it will be replaced by the concept (C2) recovered from the 'AnimOnto' ontology. The choice of the concept (C2) is done after parsing the XML file describing the ontology. Finally, the enlargement of concepts is based on the presence of a semantic link between (C1) and (C2) (synonym or broader concept).

4.3 The user role

We opted for an automatic query reformulation; the user's role is therefore passive. The latter does not intervene in the reformulation process and it is the system that supports the entire operation. The idea is to increase the weight of words found in documents considered as relevant and conversely decrease the weights of terms considered as irrelevant.

5.0 The animal semantic finder tool

5.1 Construction of the 'AnimOnto' ontology

Although so far no general methodology had been able to be generally admitted, many principles and criteria for building ontologies have been proposed. These methods concern the whole process guiding the construction steps. This is the case of the Kactus method (modelling Knowledge About Complex Technical systems for multiple USe) proposed by Schreiber (Schreiber et al. 1995). We use this method to construct the domain ontology that will serve as a source for terms used in the reformulation and indexing process.

Basing on the idea that our goal is not to build the ontology in itself, but rather its application to guide our approach and thus validate our proposal. The choice of this method is justified by two main reasons: first, by the fact that its use is not expensive, and unlike other methods, it does not require collaboration between domain experts, engineers and future ontology users. Secondly, it is appropriate for the development of small ontologies that are generally used during test and validation steps. The Kactus method is based on three main stages, namely:

1) Specification of the ontology based on the chosen field. It consists in particular of determing the terms to be collected and the tasks to be performed using this ontology;
2) Organization of terms using the meta-categories such as: concepts, terms, relationships, attributes, etc.; and,
3) Refinement of the ontology by its structuring according to principles of modularization and hierarchical organization.

5.1.1 Step 1: domain specification

The 'AminOnto' ontology will be constructed to provide a conceptual vocabulary, which allows the annotation of documents relating to the animal domain.

The choice of the animal domain is justified by the availability of data to construct the ontology on the one hand, and the possibility of building a scalable document base whose terms are semantically related to the concepts of ontology on the other hand. We note in this context that our system is open insofar as it can be used for an information search in any other document base whose content is belonging to another domain.

5.1.2 Steps 2 and 3: from the terms collection to the ontology refinement

In practice we cannot separate the ontology construction steps, because it is a nonlinear process. Several iterations were made during the construction of the 'AminOnto' ontology for the following reasons:

– It was not possible to know from the outset, whether the collected terms would be sufficient to meet the purpose for which the ontology was constructed. We had added new terms when it was necessary, in the same way we have removed terms that we judged unnecessary.
– It was not always easy to predict whether a term would play the role of a class or an attribute. Repeated corrections had to be made in this respect. Figure 1 presents a part of the 'AminOnto' ontology that we have created.

5.2 Architecture of the 'AnimSe Finder' tool

Animal Semantic Finder is a tool for semantic information searching in a document base whose content is related to the animal domain. The semantics is taken into account by the 'AnimOnto' domain ontology when selecting terms that are used in the document indexing and the query reformulation. In addition to a search guided by ontology, our system allows for a classical search using the query just as it was formulated by the user. The aim is to compare the results given in both cases, and thus measure the contribution of the semantics inclusion in the information retrieval process.

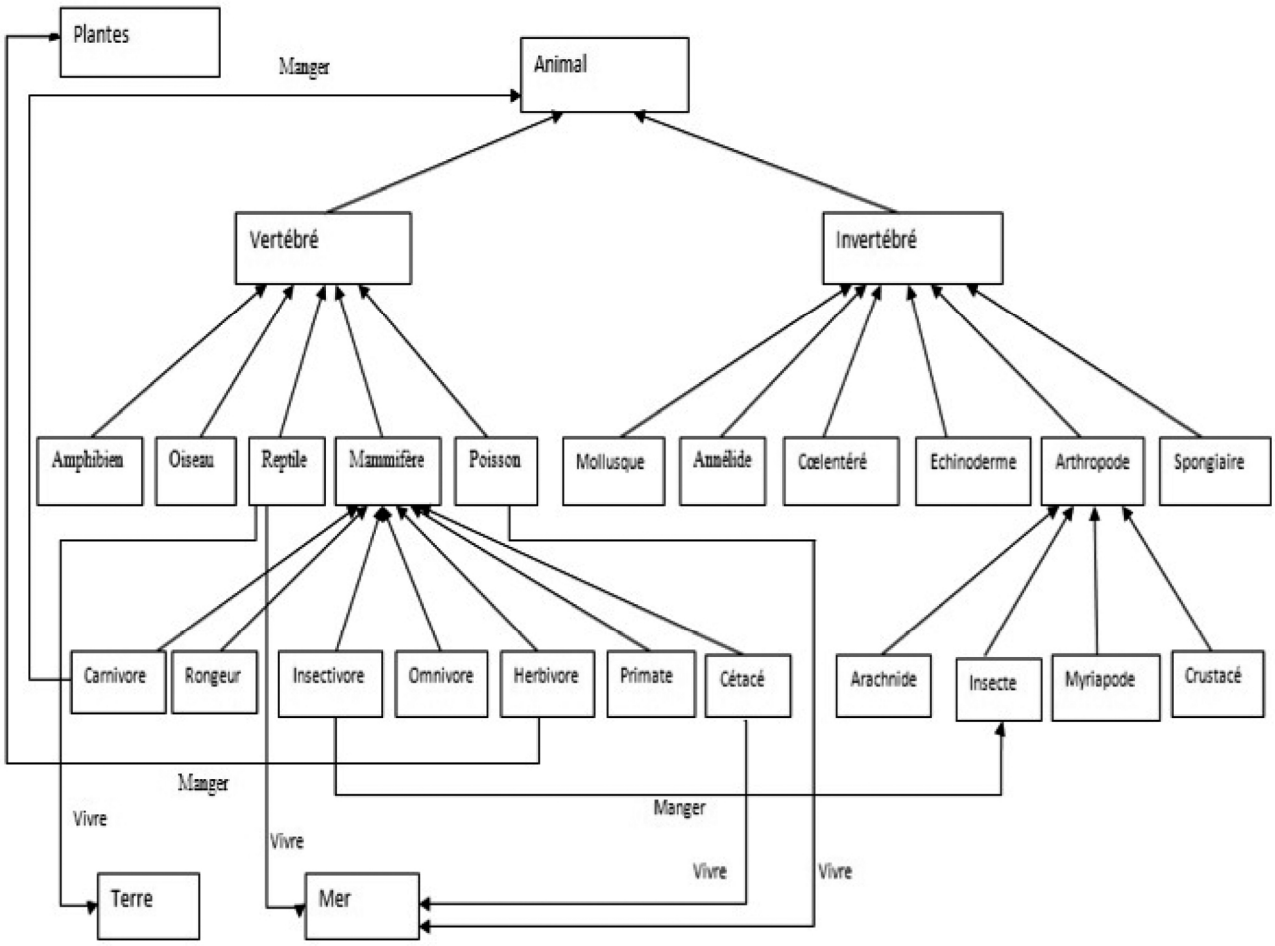

Figure 1. A part of the 'AminOnto' ontology

The different modules included in our system communicate by sending messages. A message can be a user query, a query reformulated by the system, a concept of the ontology or a document. In this context, and in order to ensure a better structuring of its modules, the 'AnimSe Finder' tool is based on three complementary processes: indexing process, reformulation process and search process. We present in the following each of these processes, the different elements that compose them, and the manner in which they operate.

5.2.1 Indexing process

The direct use of the document base during a search is a time-consuming operation, according to the number of documents and the size of each document. For this reason a special treatment for identifying the relevant elements to be used by the search process is necessary, i.e. the indexing operation. Two types of indexing have been considered: the semantic indexing guided by ontology, and the classical indexing which consists of constructing a set of terms to characterize the content of a document. In this second type of indexing, weighting is done before creating the index by calculating the occurrence frequency of each term in the concerned document. A step common to both types of indexing consists in cleaning the document; this can be done by eliminating words with zero informational value using an anti-dictionary. These stop words appear in most documents and are not discriminating. They can be articles, prepositions, conjunctions, or even verbs.

The semantic indexing process constructs a set of terms representing the informational content of each document; these keywords are derived from the concept ontology. The indexing process is composed of two stages: The document indexing and then their weighting. Once the weighting is complete, the indexing module transmits the set of triples (concept, weight, document location) associated to each document to the storage module. The latter updates the index base for an eventual use in future searches sessions. Figure 2 shows the operating principle of the indexing process and illustrates the interaction between its different elements.

The indexing module is the heart of this process. It retrieves the concepts from the ontology and calculates their frequency in the documents. It proceeds as follows:

A) Concept extraction: The aim of this step is to extract all the terms of the document that may represent concepts in the ontology. These terms corre-

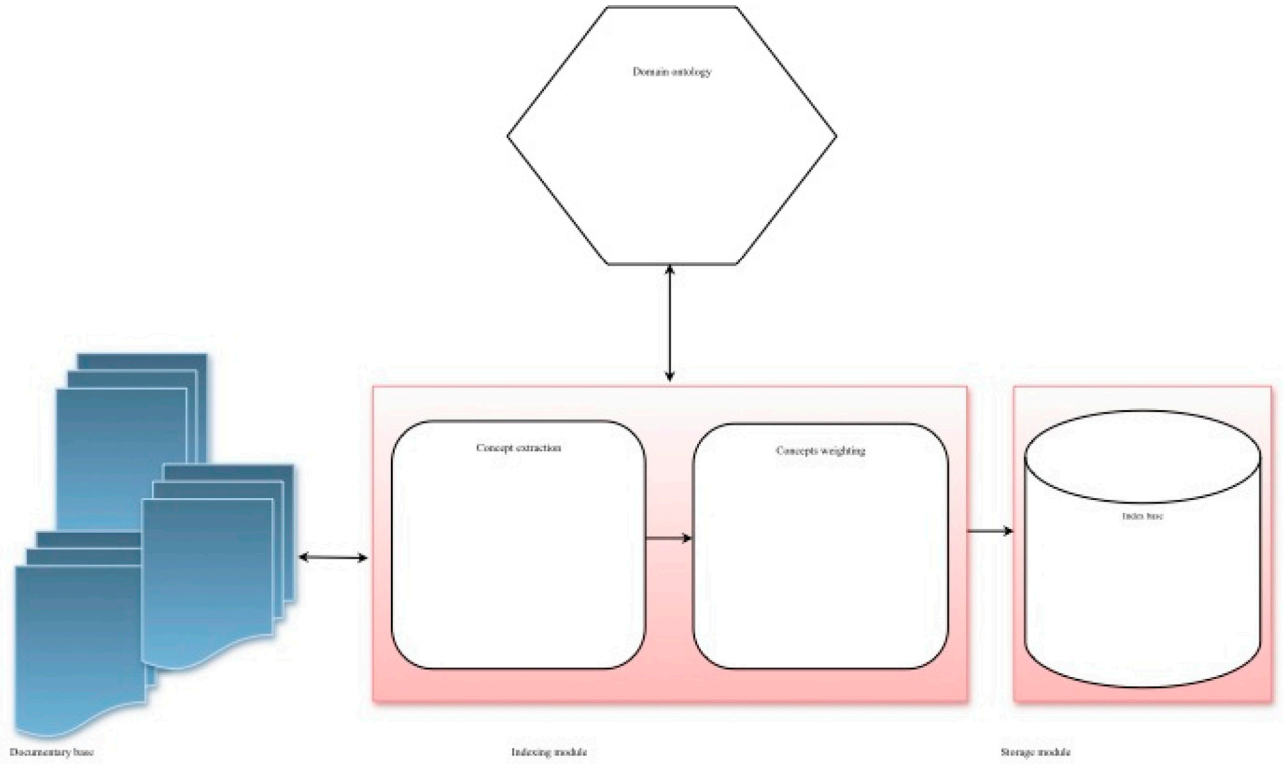

Figure 2. Architecture of the indexing process

spond to different entries (or nodes) in the ontology. For this purpose, we use a technique that consists to project the ontology on the document. This is done by parsing the ontology using to identify the concepts that occurrent as terms in the document.

B) *Concepts Weighting:* The weight of a word reflects its importance in the document. The weighting phase has a major impact on the quality of the search process itself, it allows to associate with each concept found in the document a weight (appearance frequency). We used a statistical weighting method to calculate the frequency of each term extracted according the number of its occurrences.

5.2.2 Reformulation process

The user often has difficulties to express his or her exact information needs. Therefore, among the documents that are returned by search tools, some of them interest him less than others. In so far, an improvement of the way how the user expresses his information needs is a supplement that can enhance the quality of the returned documents. In the context of our research, the reformulation process supports the generation of a new query using the 'AnimOnto' ontology and the initial query in order to return documents which are more relevant than those provided by a non-reformulated query.

The structure of the reformulation process, the different modules that compose it and the interactions between them are illustrated in Figure 3.

This process is based essentially on the reformulation module that supports the expansion of the user query to better reflect his information need. This module is conducted in two steps acting on different aspects of the initial query. These steps are:

– *Concept extraction:* In this step the reformulation module retrieves the query terms present in the ontology, then it goes through the ontology using these words as an entry point to extract the concepts directly related to each term in different hierarchical levels of the ontology.
– *Query reformulation:* In this step, the reformulation module takes as input the concepts retrieved in the previous step; it uses them in generating the new query that will be forwarded later to the research process.

5.2.3 Search process

This process is mainly responsible for the fundamental decision that allows an association to a query, such that all relevant documents can be returned. It is based on the Boolean search model, in which a document (d) is represented by a set of its terms (t_i), and a query (q) as a logical term expression. A document does correspond to a query if the implication ($d \Rightarrow q$) is valid. This correspondence C (d, q) is determined as follows:

– $C(d, ti) = 1$ if $ti \in d$; 0 otherwise
– $C(d, q1 \wedge q2) = 1$ if $C(d, q1) = 1$ and $C(d, q2) = 1$; 0 otherwise
– $C(d, q1 \vee q2) = 1$ if $C(d, q1) = 1$ or $C(d, q2) = 1$; 0 otherwise
– $C(d, \neg q) = 1$ if $C(d, q) = 0$; 0 otherwise

To clarify the operating principle of the search process, we present two search scenarios according to the two search modes provided by our system:

– *Scenario 1 'classical search':* Suppose the user wishes to have information about the domain of cats. He then formulates a query containing the word 'cat'. The search process takes as input this query and searches all the documents in the document base through the index base and retrieves the relevant ones. These are documents $\{d_1, d_2, d_3, \ldots .d_n\}$ in which we will find the word 'cat'. The search module then ranks the documents from the highest

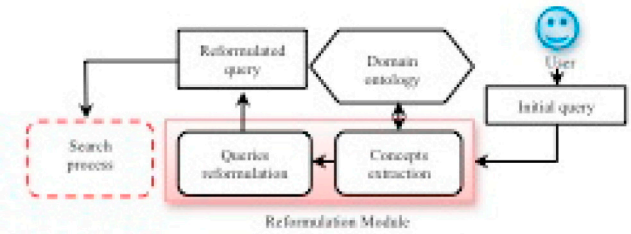

Figure 3. Architecture of the reformulation process

weight to the lower ones $\{P_1,P_2,P_3,\ldots P_n\}$. The result is finally displayed to the user.

– *Scenario 2 'semantic search':* For the same information need, the user query is sent now to the reformulation process that automatically generates a new query using the 'AnimOnto' ontology. The new query will be transmitted to the search process to be used later in order to return the documents reflecting the user need. In the chosen example, the result of the query reformulation is: "carnivorous + cat", the word 'carnivore' was recovered from the ontology using a parser developed for this purpose. Finally, the search process retrieves the set of documents $\{d'_1,d'_2,d'_3,\ldots,d'_n\}$ containing the word 'cat' and/or the word 'carnivore', it calculates thereafter the sum of the weights $\{P'_1+P''_1, P'_2+P''_2, P'_3+P''_3,\ldots, P'_n+P''_n\}$ in order to classify the resulting documents according to their relevance. The result is then displayed to the user.

Figure 4 shows the manner in which the search process operates and the interaction between its different elements.

Finally, the combination of the three processes will allow us to define the general architecture of our system; it is presented in Figure 5.

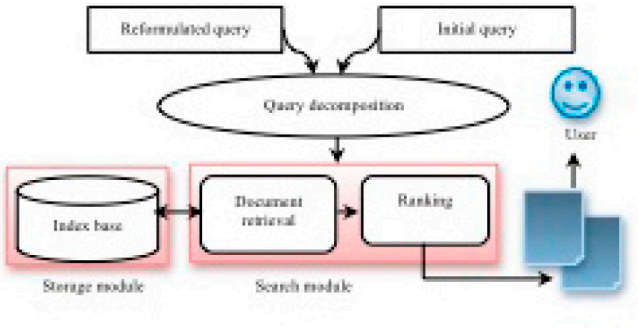

Figure 4. Search process architecture

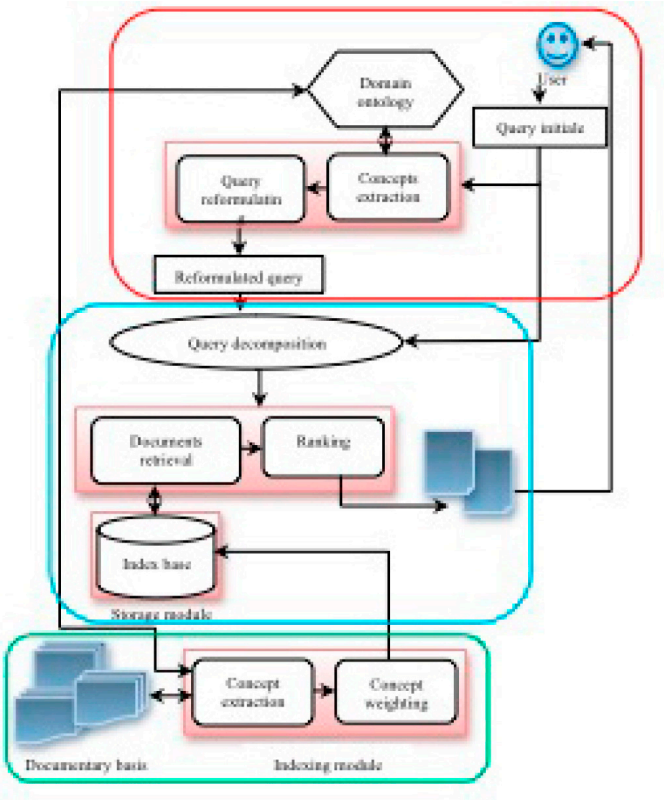

Figure 5. Architecture of the semantic information search system

5.3 Description of the developed tool

The Animal Semantic Finder tool has been developed to operationalize the main objective of our contribution, i.e. to consider the semantics based on domain ontology to improve information retrieval system. Our application offers to users the ability to perform two search types:

- *Classical search:* In this search mode, the user formulates a query relating to the animals domain in order to get answers for frequently asked questions in this field such as: the animal lifetime, food, categories ... etc.. The system processes the user query and provides all documents that meet the request.
- *Semantic search:* Among the documents returned in the first search type, some do not meet the exact needs of the user. To this end, we offer the opportunity to expand the search by adding other terms to the original query in order to return documents closer to those the user wants. These terms are retrieved from the domain ontology associated with our system.

5.3.1 Techniques used for the application development

For the development of 'AnimSe Finder', our choice in terms of tools and implementation techniques was:

- **Protégé:** A freeware allowing ontology editing. It is also a knowledge base structure that produces an ontology written in XML, RDF, OWL etc ... To edit the 'AnimOnto' ontology, we have opted for the XML format. The XML format is well adapted to small size ontologies as in our case. It also permits the use of the DOM Api for the ontology parsing and exploring.
- **DOM API:** The Document Object Model defines a standard command set that parsers should incorporate in order to access the contents of XML documents. The XML parser that supports DOM extracts data from an XML document and describes them using a set of objects. In our case the DOM will be used by the XML parser to extract ontology concepts associated with representative terms of a document. The integration of the DOM API with the VB.Net language is ensured by establishing a reference to the MSXML library types provided in Msxml.dll.
- **VB.Net:** Visual Basic.Net is an object-oriented programming language allowing developments with .Net technology using Visual Studio. This language makes available to applications that run in the NET, a set of classes that allows the user to interact with the system. We chose this language for its qualities in terms of database managing and the facilities that it offers for manipulating external documents (XML in our case). It allows more precisely, a very simple and polyvalent data access and it includes a rich object library using the Windows API.
- **Access:** We use the 'Access' DataBase Management System to create the database used to store indexes and references associated to the different documents of the document base. Access is based on the relational model and operates on the following principle: the information is stored in tables that are linked by relationships. The interrogation of the database is done by queries written in SQL language (Structured Query Language).

5.3.2 Presentation of AnimSe Finder

A) **Main interface:** The search tool that we propose offers all the functionality described in the proposed architecture. More exactly, it allows to edit a query and provide documents that meet this request by two types of research (classical and semantical). It also gives the possibility of expanding the document base by adding new documents, so they will be taken into account when responding to future requests. Finally, it offers the possibility to visualize the 'AnimOnto' ontology by three display types: XML code, tree form and graph form. The different areas of the main interface are shown in Figure 6.

B) **Indexing:** Once the document is added to the document base, it must be indexed. The activation of the indexing operation will lead the triggering of a series of operations, starting with a pretreatment to remove stop words of each document using an anti-dictionary. Thereafter, one has to extract the representative concepts of the document. Finally, the concept weighting consists in assigning a weight for each concept found and in updating the index base. These operations are performed automatically when a new document is added, and they are fully supported by the system. Figure 7 shows the indexing interface.

C) **Ontology Visualization:** The 'AnimSe Finder' tool gives the possibility to visualize the 'AnimOnto' ontology with the aim to guide the indexing and searching operation. Three display types are offered: XML code, tree form and graph form. Figure 8 shows how the ontology is displayed.

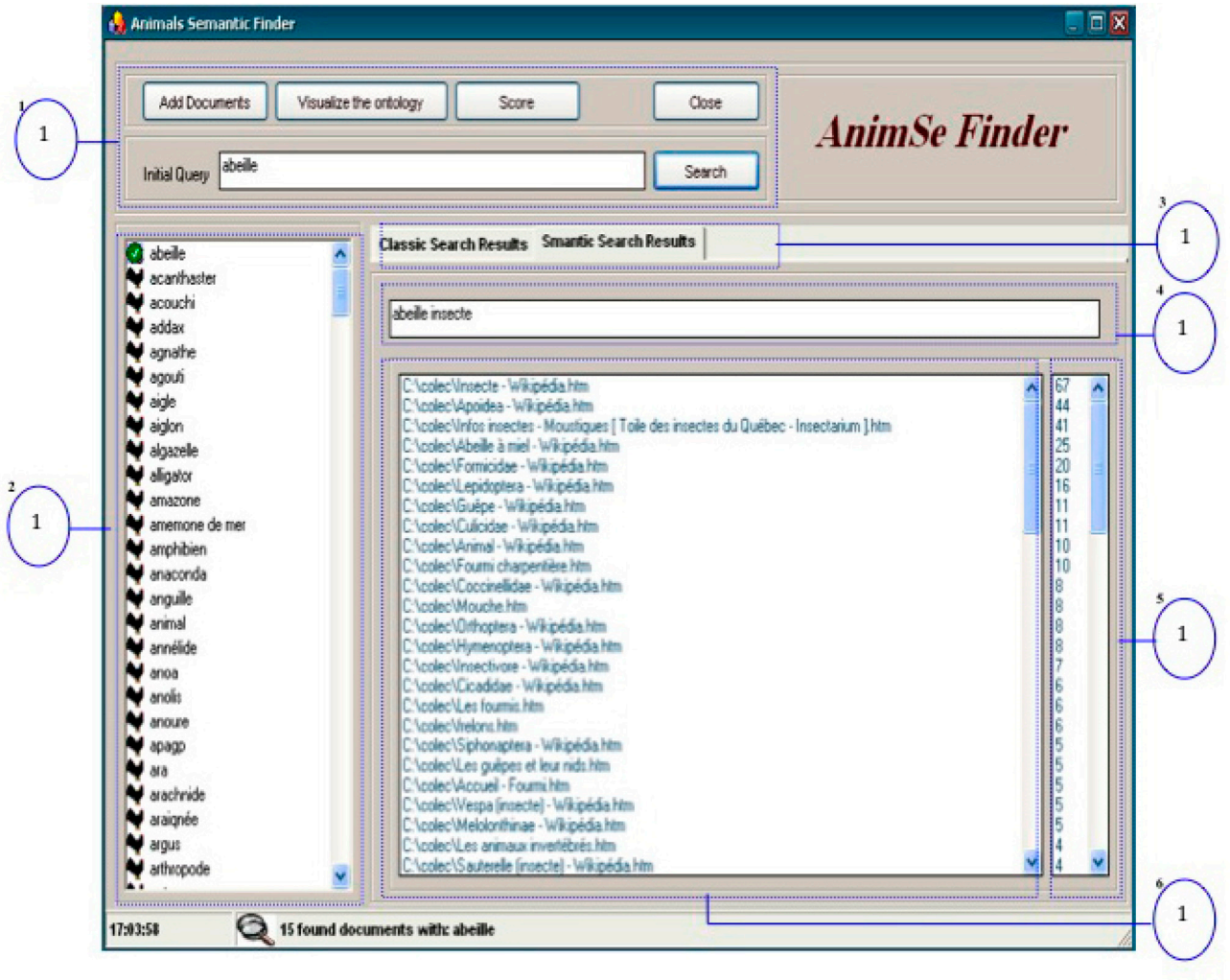

Figure 6. Main interface of the AnimSe Finder tool (1. Initial query; 2. 'AnimOnto' ontology concepts; 3. Offered search type; 4. Reformulated query; 5. Relevance scores; 6. Returned documents)

6.0 Evaluation of the 'AnimSe Finder' performance

To assess the 'AnimSe Finder' tool performance, we conducted an experiment that aims to understand and measure the contribution of the semantic treatment in the indexing and the reformulating process. The evaluation was based on the metrics generally used in the information retrieval field. These metrics are precision, recall and F-measure, they are calculated as in Figure 9.

6.1 Test collection characteristics and evaluation method

The evaluation protocol that we used is based on four elements: a corpus of documents (document base), a corpus of queries, relevance judgments (reference value) indicating that a certain document is relevant for a given query), and evaluation metrics (recall, precision and F-measure).

- **Document base:** The used test collection has been created from the web and encyclopedias by choosing arbitrarily documents that are related to the animal domain. These documents are selected according to two criteria: their accessibility and diversity (writing styles, vocabulary, text length, etc.). The document database includes 13,000 text documents. The advantage of this documents corpus lies in the fact that it is consists of reliable resources.
- **Reference value:** Reference values of the correct answers are made up as follows: For each query, all potentially relevant documents are gathered in a 'relevance group'. We have, to this end, controlled the relevance of each document according to the

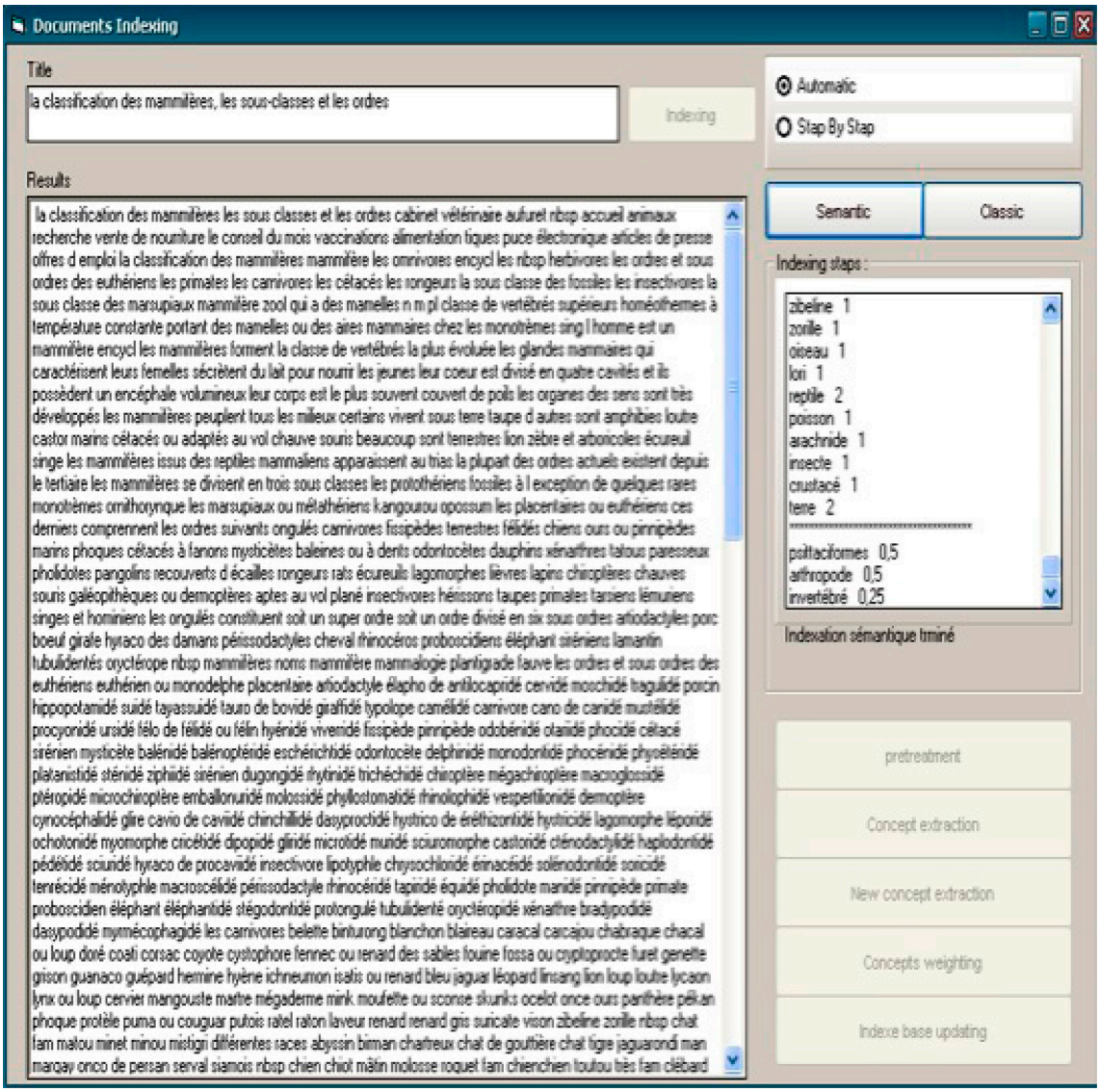

Figure 7. Semantic indexing mechanism

query during the construction of a reference value. Table 1 summarizes the characteristics of the test collection used for the performance evaluation of the 'AnimSe Finder' tool.

Category	Number
Total number of documents	13,000
Number of queries	25
Average number of relevant documents per query	220

Table 1. Characteristics of the test collection

– **Operation principle:** Using the document base previously described, the system performs the semantic indexing of all documents. It takes as input a set of 25 queries expressed by users in natural language. These users have the role of judging whether the documents found in both searches (semantical and classical) are relevant or not according to the query. The documents returned in both search cases have been listed by their relevance order according to each query. They were finally subjected to evaluation using the previously described protocol.

6.2 Results and discussion

The obtained results are presented in Table 2.

Search type	Recall (%)	Precision (%)	F-Measure
Semantic	91.18	88.37	0.792
Classical	78.12	72.24	0.687

Table 2. Evaluation results

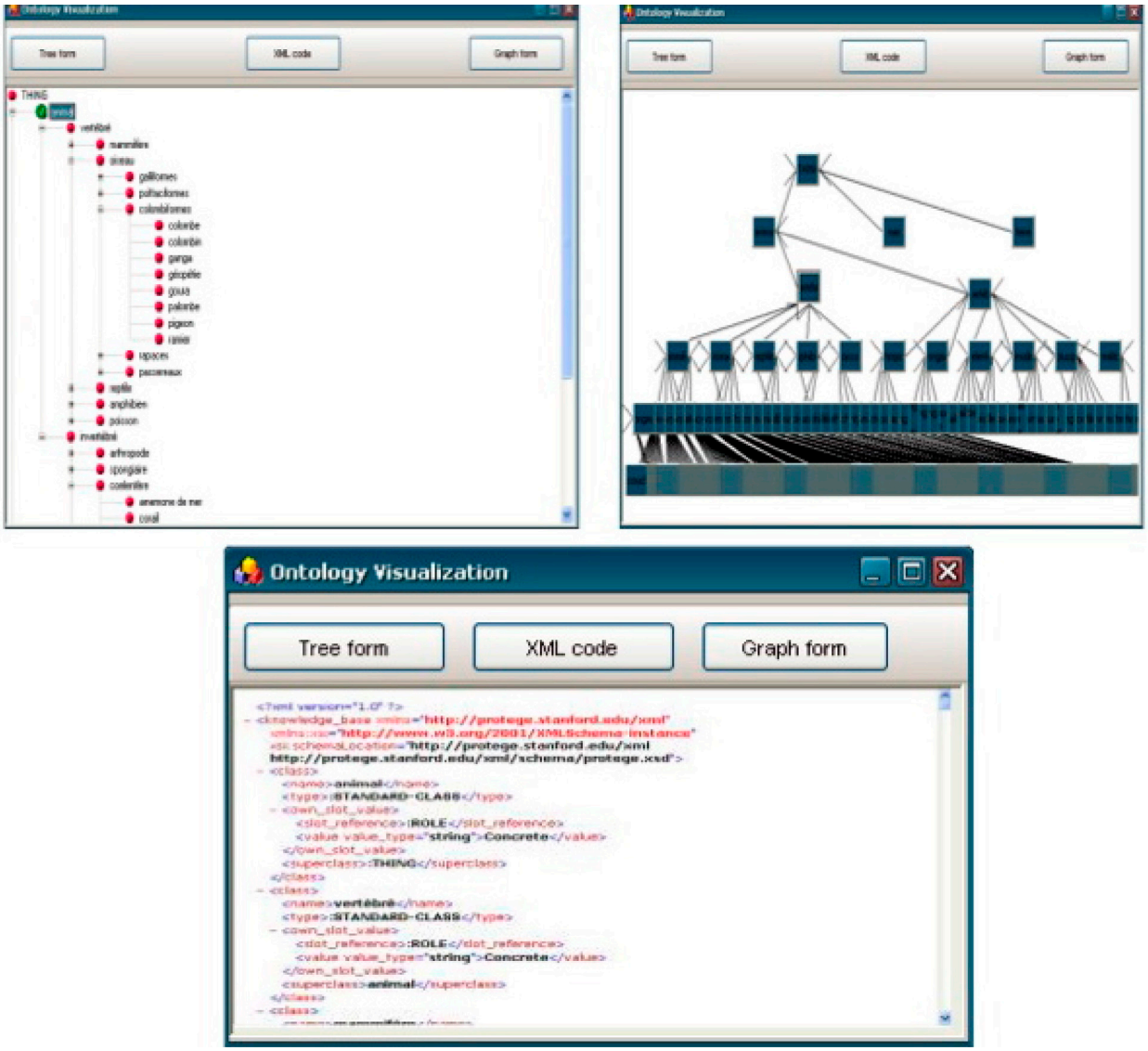

Figure 8. Visualization of the 'AnimOnto' ontology

The obtained scores show that the recall rate is higher in the case of a search guided by the ontology. The addition of the semantic dimension during indexing and reformulation produced a gain of 13.06 percent compared to the classical search. This rate is considered as a 'silent' corresponding to the relevant documents that have not been returned in the case of a classic research. This gap in terms of recall is also explained by the strong match between the words that were added to the query and by the information content of the returned documents. With respect to precision, scores show that the semantic search presents an improvement of 16.13 compared to classical search. This high rate of 88.37 means that only a few nonrelevant documents are provided by 'AnimSe Finder' and that the latter may be considered as "precise". In contrast the loss of 16.13 percent in the case of classical search represents documents that are returned although they are insignificant or irrelevant; they constitute the 'noise'. Finally, these results are confirmed by the F-measure that combines precision and recall and their weighting.

7.0 Conclusion

In this paper we presented two complementary proposals reflecting two viewpoints of ontology used in information retrieval. The first proposal, semantic indexing, is based on the hypothesis that a document can be viewed as a set of concepts, where the importance of a concept depends on the number of links with other concepts that share the same document. In this proposal, the document concepts are selected by a projection of the 'AnimOnto' ontology on the differ-

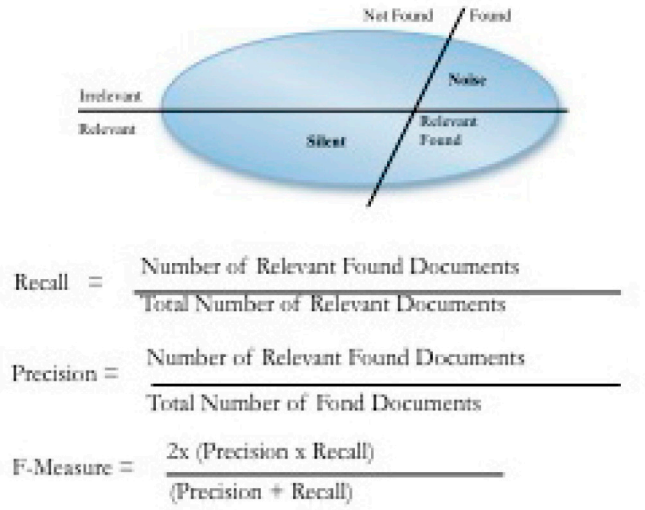

Figure 9. Precision, recall and F-measure

ent documents. The second proposal concerns the query reformulation by exploiting the semantic relationships between their terms in order to improve the performance of the search tool.

To validate these proposals, we implemented the 'AnimSe Finder' tool. In this application the different phases of document representation has been implemented, namely the term extraction, the new concepts extraction and weighting. 'AnimSe Finder' materialized also the query reformulation phase by expansion using the 'AnimOnto' ontology. This is done by taking into account the semantic relationships between concepts presented in the queries and those of the ontology.

Finally, the developed tool has the advantage of being generic and adaptable to other search types. It is only necessary to use a different ontology and another document base corresponding to the desired domain in order to extend the functionality offered by the 'AnimSe Finder' tool. The comparison of scores obtained by a classical search and those of a search guided by the 'AnimOnto' ontology, in the case of the present document collection, has shown that the use of terms generated by the ontology and the query reformulation provides a remarkable improvement in terms of the relevance of the returned results. As perspective, it would be interesting to deepen the performed research with the idea of using a query with word combinations, as well as to make a lexical query analysis to correct the users spelling errors. It would also be interesting to widen the application domain and extend the 'AnimOnto' ontology by adding new concepts and semantic relationships, or to apply the search in other fields than that of animals.

References

Baziz, Mustapha, Boughanem, Mohand, Aussenac-Gilles, Nathalie, and Chrisment, Claude. 2005. Semantic cores for representing documents in IR. In Liebrock, Lorie M. ed. *Proceedings of the 2005 ACM symposium on applied computing, Santa Fe, New Mexico, March 13-17, 2005*. New York: ACM, pp. 1011-17.

Boughanem, Mohand, Berrut, Catherine, Mothe, Josiane and Dupuy, Chantal. 2009. Advances in Information Retrieval, 31th European Conference on IR Research, ECIR 2009, Toulouse, France, April 6-9, 2009, Proceedings. New York: Springer.

Bouramoul, Abdelkrim, Kholladi, Mohamed-Khireddine, and Doan, Bich-Lien. 2010. PRESY: A context based query reformulation tool for information retrieval on the Web. *Journal of computer science* 6: 470-77.

Cui, Hang, Wen, Ji-Rong, Nie, Jian-Yun and Ma, Wei-Ying. 2002. Probabilistic query expansion using query logs. *Proceeding of the 11th International Conference on World Wide Web, May 07-11*. New York: ACM, pp. 325-32.

Gruber, Tom. 1993. A translation approach to portable ontology specifications. *Knowledge acquisition* 5: 199-220.

Koo, Sang Ok, Lim, Soo Yeon, and Lee, Sang Jo. 2003. Building an ontology based on hub words for informational retrieval. In Liu, Jiming ed. *Proceedings of the IEEE/WIC International Conference on Web Intelligence (WI 2003) Halifax, Canada, October 13-17, 2003*. Los Alamitos, CA: IEEE, pp. 466-69.

Lin, Hsi-ching, Wang, Li-hui and Chen, Shyi-Ming. 2006. Query expansion for document retrieval based on fuzzy rules and user relevance feedback techniques. *Expert systems with applications* 31: 97-405. DOI: 10.1016/j.eswa.2005.09.078.

Navigli, Roberto and Velardi, Paola. 2003. An analysis of ontology-based query expansion strategies. In Ciravegna, Fabio, and Kushmerick, Nicholas eds., *Proceedings of the 2003 workshop on Adaptive Text Extraction and Mining, Monday 22 September 2003, Cavtat-Dubrovnik (Croatia)*, pp. 42-49.

Saias, José, and Quaresma, Paulo. 2003. A methodology to create ontology-based information retrieval systems. In Moura-Pires, Fernando, and Abreu, Salvador eds. *Progress in artificial intelligence, 11th Portuguese Conference on Artificial Intelligence, EPIA 2003, Beja, Portugal, December 4-7, 2003, Proceedings*. New York: Springer, pp. 424-34.

Schreiber, Guus. Wielinga, Bob and Jansweijer, Wouter. 1995. The kactus view of the 'o' word. In *IJCAI'1995, Workshop on Basic Ontological Issues in Knowledge Sharing, Montréal, 1995*. Berlin: Springer. Also in Bioch, J.C. and Tan, Y.-H. (eds.), *Proceedings 7th Dutch National Conference on Artificial Intelligence NAIC'95, EURIDIS, Erasmus University Rotterdam, The Netherlands,* [s.l.: s.n.], pp. 159-68.

Vallet, David, Fernández, Miriam, and Castells, Pablo. 2005. An ontology-based information retrieval model. In Gómez-Pérez, Asunción, and Euzenat, Jérôme eds. *The Semantic Web: Research and applications.* Berlin: Springer, pp. 103-10.