Ontology Learning by using text clustering techniques

Ana B. Rios-Alvarado

Information Technology Laboratory
CINVESTAV - Tamaulipas

Thesis advisor:
Ivan Lopez-Arevalo

December 9th, 2010
Abstract

Ontologies provide a structural organizational knowledge, they support the exchange and sharing of information. Ontology learning techniques from text have emerged as a set of techniques to get ontologies from unstructured information. An important task in ontology learning is get a taxonomy. This report presents a little revision of the state of the art in ontology learning and the first proposal to get taxonomic (is a) relationships between relevant terms using the Web search and WordNet. This activity corresponds to advances on the first year of the doctoral studies.
Chapter 1

Introduction

At the beginning of the 21st century the access to digital information resources has been increase doing that the unstructured information is growing rapidly. Not only is the unstructured information growing on the Web, it also is growing into organizations, institutions, and companies. In an organization, for example, documents represent a significant source of collective expertise (know how). In order to store, retrieve, or infer knowledge from this information, it is necessary represent its knowledge by a conceptual structure. This can be achieved by means of an ontology.

Background

According to Studer et al. [1], an ontology is defined as a formal, explicit specification of a shared conceptualization. Conceptualization refers to an abstract model of some phenomenon in the world. Explicit makes reference to define the type of concepts used and the constraints of their use. Formal involves the fact that the ontology should be machine-readable. Shared shows that an ontology captures consensual knowledge, that is, it is accepted by a group of experts in the domain. Neches et al. [2] describe an ontology as an element that it defines the basic terms and relations contained into the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions of a conceptualization. Recently, Staab et al. [3] defines an ontology as a formal description of concepts and relationships that can exist for a community of human and/or machine agents. The notion of ontologies is crucial for the purpose of enabling knowledge sharing and reuse. In WordNet (lexical database for English) appears the follow definition: an ontology (in Computer Science) is a a rigorous and exhaustive organization of some knowledge domain that is usually hierarchical and contains all the relevant entities and their relations. Thus, an ontology should: 1) capture a shared understanding and 2) enable logical inference on facts through axioms.

According to Noy and McGuinness [4] the main reasons to develop an on-
ontology are:

- To share common understanding of the structure of information
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To separate domain knowledge from the operational knowledge

The common components of an ontology are:

- Concepts: represent the ideas to formalize
- Relations: correspond to a type of association between concepts, these relations can be taxonomic (is a) or non-taxonomic (part-of, related-to)
- Instances: illustrate given elements or individuals
- Axioms: sentences which are always true

An ontology can be built in a manual manner through the knowledge engineers and domain experts resulting in long and tedious development stages, which can results in a knowledge acquisition bottleneck [5]. As a consequence, an important research area is **Ontology Learning**.

Ontology Learning is defined as the set of methods used for building from scratch, enriching or adapting an existing ontology in a semi-automatic fashion using heterogeneous information resources [6]. The ontology learning deals with discovery entities and how such entities can be grouped, related, and subdivided according to its similarities and differences.

There are different points of view in the ontology learning techniques. Maedche and Staab [5] considered the cycle of life in the building of an ontology and claimed four parts in the ontology learning process: extract, prune, refine, import or reuse. Weng et al. [7] emphasized in the extraction methods considering four categories: dictionary-based, text clustering, association rules, and knowledge base. Particularly, text clustering techniques can found relations between words that appears in different places into the text.

In ontology learning, an unsupervised manner to build conceptual structures is using text (terms) clustering techniques. However, these approaches do not considerer that words are ambiguous and sharing a semantic context. Pantel and Lin [8] provides a soft clustering algorithm called *Clustering by Committee* (CBC) which can assign words to different clusters. This method only creates clusters of terms but it not create a hierarchical structure. Cicurel et al. [9] evaluated CBC concluding that was a good technique to identify senses of words. Its disadvantage is that it requires four parameters: $\theta_1$, for ensure that the committees are far each other, $\theta_2$ that makes sure that words are far from the committees, $\theta_3$ refers to similarity between committees, and $\theta_4$ refers to similarity word to committee that has been already assigned.
On the other hand, many works currently focus on inexpressive ontologies. Ontologies can describe more than just terms, attributes, and the relations between terms. Ontologies can enable logical inference on facts. Only a few approaches as [10, 11] present methods for enriching inexpressive ontologies with axioms which is based on a syntactic analysis. In this research, ontology learning techniques are proposed and developed to automatically discover terms, concepts, relations, and axioms from documents.

Goals

The main goals of the research are:

• To get a model for ontology learning from textual resources using a text clustering technique, Web search to get relationships, and representing discovered axioms by using description logic.

• To get general topics/concepts from textual resources using a text clustering algorithm.

• To obtain a technique to get a taxonomic relationships between concepts.

• To obtain a technique to get a non-taxonomic relationships between concepts.

• To develop a technique to discover axioms using textual resources and the created taxonomy.

Particularly, this report focuses on present an overview of the state of the art and present a method for discovering taxonomies using Web search and WordNet.

Overview of this document

The rest of the document is organized as follows: Chapter 2 presents the state of the art, where relevant approaches on ontology learning using clustering techniques have been proposed. In Chapter 3 is introduced a method for discovering taxonomies by using Web search and WordNet. Finally, in the Chapter 4 some conclusions are presented.
Chapter 2

State of the art

Ontology learning systems have different purposes, they mainly extract concepts and relationships from a collection of documents (corpora) related to specific domain in order to construct an ontology. Cimiano et al. [12] presented the different methods applied to learn certain ontology primitives according to the follow tasks:

- Extracting the relevant domain terminology and synonyms from a text collection.
- Discovering concepts which can be regarded as abstractions of human thought.
- Deriving a concept hierarchy organizing these concepts.
- Extending an existing concept hierarchy with new concepts.
- Learning non-taxonomic relations between concepts.
- Populating the ontology with instances of relations and concepts.
- Discovering other axiomatic relationships or rules involving concepts and relations.

Gomez-Perez et al. [13] presented a survey on the methods, techniques, and software used for ontology learning. Their work emphasizes in proposals based on texts, dictionaries, knowledge bases, semi-structured schemas, and relational schemas.

Buitelaar et al. [14] described the process to build an ontology based on the named cake model. The cake model considers building an ontology as overlay, where each layer corresponds to a task that allows to get a component of the ontology. From the bottom to top layer is organized as: terms, synonyms, concepts, hierarchies, relationships, and rules. The methods that can perform these tasks are classified into four groups based on [15]: lexical-syntactic patterns, information extraction, machine learning, and co-occurrence analysis.

In general, ontology learning from text involves tasks as [16]:
- Natural Language Processing: part of speech tagging, phrase chunking, and stemming.

- Information Extraction: identify terms and relations.

These tasks are able to get:

- the vocabulary for a domain
- the relationships between concepts

The text mining is the process that allows to discover patterns and new knowledge for collection of text. Techniques of Natural Proccesing Language are tipically used for recognizing relevant terms and their relationships. The text requires a processing phase, where tasks as: 1) extraction of plain text, 2) splitting of text into sentences, 3) elimination of stopwords, 4) tagging of sentences, and 5) parsing of the sentences are applied.

Some works [17, 18, 19, 20] used the Web as knowledge base for discovering relations. Through Web search and queries on the Web it is possible to determine how much two concepts are related to a particular relation. The Web scales statistical measures are used to determine the probability of co-occurrence of words into a context.

On the other hand, the text clustering techniques allow to get related terms. Clustering is to split data into groups with similar characteristics. Clustering algorithms are divided on partitioning methods (k-means, k-medoids) and hierarchical clustering (agglomerative or divisive algorithms). Text clustering algorithms have been used in ontology learning [21, 22, 23, 24] but such works do not consider the ambiguous words or their semantic relationships.

Some characteristics in works are presented in Table 2.1.

<table>
<thead>
<tr>
<th>Author</th>
<th>Goal</th>
<th>Techniques</th>
<th>Expressiveness</th>
<th>Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beil et al., 2002 [21]</td>
<td>Term extraction and build taxonomy</td>
<td>Hierarchical clustering</td>
<td>Taxonomy relations</td>
<td>Text documents</td>
</tr>
<tr>
<td>Karoui et al., 2006 [22]</td>
<td>Term extraction and build taxonomy</td>
<td>Hierarchical clustering</td>
<td>Taxonomy relations</td>
<td>HTML documents</td>
</tr>
<tr>
<td>Yeh and Yang, 2008 [24]</td>
<td>Topic extraction and build taxonomy</td>
<td>LDA</td>
<td>Taxonomy relations</td>
<td>Digital library</td>
</tr>
<tr>
<td>Sung et al., 2008 [23]</td>
<td>Build taxonomy</td>
<td>Hierarchical clustering</td>
<td>Taxonomy relations</td>
<td>“postings” of news</td>
</tr>
</tbody>
</table>

Table 2.1: Some proposals using clustering techniques in ontology learning

According to review, some of the major limitations of previous proposals are: 1) some clustering techniques give poor results, 2) level expressiveness is limited to non-taxonomic relationships, and 3) dependency to linguistic databases and the domain.
Chapter 3
Discovering taxonomic relationships

The discovery of hyperonyms\(^1\) has recently been a subject covered in many works, mainly for its importance in the construction of taxonomies used as organizational and classification models. Previous works have used specific lexical patterns or have focused on identifying new patterns, recently the use of the Web as source of collective knowledge seem a good option for finding appropriate hyperonyms. In present work, an approach to find hyperonyms relations between terms into a knowledge domain is presented. This approach combines WordNet synsets and context information for building a query set. This query set will be execute a Web search for recovery the most representative hyperonym for a term.

Introduction

According to Gruber [25], “ontologies are often equated with taxonomic hierarchies of classes”; thus, it can be said that the key component in an ontology is the taxonomy. Such taxonomies, as the main component of an ontology provide an organizational model for a domain (domain ontology), or a model suitable for specific tasks (task ontologies) [26]. Nevertheless, learning taxonomy is a hard task.

For building a taxonomy, the identification of hyperonomy/hyponomy relations between terms (in the present work terms are nouns) is mandatory. Hyponymy can be defined as: an expression A is a hyponym of an expression B if the meaning of B is part of the meaning of A and A is a subordinate of B. By contrast, an expresion B is a hyperonym of A if B includes the meaning of A and B is a superior to A. For example, *Mercury*, *Jupiter*, and *Mars* are

---

\(^1\)In linguistics, a hyperonym is a word or phrase whose semantic range includes a set of another words.
hyponyms of Planets, in the contrast Planet are hyperonym of Mercury, Jupiter, and Mars. Other names for the hyponym relationship are is-a, parent-child, or broader-narrower relationships [27]. Caraballo [28] claimed that, “a word A is said to be a hyperonym of a word B if native speakers of English accept the sentence B is a (kind of) A”.

Several works, to identify hyperonyms, have used lexical patterns but, recently, some works are using the Web as source of collective knowledge for this purpose. In addition to using Web and lexical patterns, some works [29, 30] identify new lexical patterns, and so it is possible to obtain more specific hyponyms; but its necessary rely on the known hyponym relations for training a classifier, which is not always possible to have them.

In this chapter, a unsupervised method to find hypernyms relations between the terms into a knowledge domain is proposed. This is, given a corpus of text and the group of related terms (called topics), a combination of lexical patterns with supervised information and context information is applied for building a query set. This query set will be execute a Web search for recover the most representative hyperonym for a term.

The rest of the chapter is structured as follow. In Section 2 a brief description of the related work in automatically discovering hyperonyms is given. In Section 3 the approach and the method for find hyperonyms are described. Later, in the Section 4, the experiments and preliminary results are presented. Finally, Section 5 gives some conclusions and the further work.

Related Work

One of the first works in automatic discovering hyperonyms from the text was proposed by Hearst [31]. He proposed a method to identify a set of lexico-syntactic patterns occurring frequently in the text. Caraballo [28] proposed to build a noun hierarchy automatically from text using data on conjunctions and appositives appearing in the Wall Street Journal corpus. Both methods are limited by the number of patterns used. Pantel et al. [32] showed how to learn syntactic patterns for identifying hypernym relations and binding these with clusters built from co-occurrence information. Blohm and Cimiano [33] proposed a procedure to find lexico-syntactic patterns indicating hypernym relations from the Web. From this work, Ortega-Mendoza et al. [29] and Sang [34] developed a method to extract hyponyms and hyperonyms using lexical patterns respectively. Snow et al. [30] generated hyperonym patterns and combined them with noun clusters to generate high-precision suggestions for unknown noun insertion into WordNet. Also, they described a variant of their classifier including evidence from coordinate terms (terms with common ancestor classes) increasing precision. Ryu and Choi [35] proposed a method to build a taxonomy which relies on term specificity and similarity. Recently, McNamee et al. [36] detected hypernyms in named entities (i.e. proper nouns) in order to improve the performance of Question Answering Systems. Ritter et al. [37] presented a method based on lexical patterns that find hypernyms of arbitrary noun phrases. They
used a SVM (Support Vector Machine) classifier to found the correct hypernyms from matches to the Hearst patterns.

Most of these studies are limited due the hand selection of pairs of terms that have hyperonomy relationship, which are initial seed for the discovery of new patterns.

On the other hand, Cimiano and Staab [38] showed that a potential way to avoid the knowledge acquisition bottleneck is acquiring collective knowledge from the World Wide Web using a search engine. This idea was used by Sánchez [39] using the Web for acquiring taxonomic and non-taxonomic relationships.

The representation model

Typically the text is represented using the Bag of Words model. This model assumes that the order of words has no significance, however, current applications consider that a semantic representation focus on Natural Processing Language has a major potential for new developments. Thus, word-context matrices and pair pattern matrices are most suited to measuring the semantic similarity of word pairs and patterns [40]. In the present work, the propose is using a syntactic parser which extracts the grammatical context where each word occurs. Using the Minipar\(^2\) is possible to get different features, two of them are the dependency relationships \(<\text{subject}, \text{verb}>\) and \(<\text{verb}, \text{object}>\). With these relationships, representative pairs of words into a context (topic) cluster are identified.

A pair-term matrix is used as representation model:

<table>
<thead>
<tr>
<th>Verbs</th>
<th>Nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>visit</td>
<td>Porto Novo</td>
</tr>
<tr>
<td>go</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>like</td>
<td></td>
</tr>
<tr>
<td>cup of tea</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.1: Example of pair verb-noun matrix

The mutual information is the measure used for the association strength between two words (verb-noun) and is defined as follows (Equation 3.1):

\[
mi_{xy} = \log \frac{P(x,y)}{P(x)P(y)}
\]

(3.1)

For each pair verb-noun its mutual information is calculated, so the pair verb-noun is mapped to numerical values as shows in Figure 3.2

\(^2\)http://webdocs.cs.ualberta.ca/~lindek/minipar.htm
The method: Querying the Web

The discovering hyperonyms follows the next steps (see Fig. 3.3):

- Pre-processing: To identify dependencies between nouns sharing a verb into the same context. These dependencies are obtained using the Minipar parser. A pair-pattern matrix [41] is used as representation model. In the pair pattern matrix, a row represents a pair terms and the column corresponds to the pattern in which the pair occur. The pairs corresponds to the terms appearing in a triple term structure $<\text{subject}>\text{verb}<\text{object}>$, then the pair subject-object is taken as a representative pair term. This representation reduces the vocabulary of the corpus that will be get into groups.
- Topics: The topics from the corpus are inferred using an adaptation of
the CBC algorithm proposed by Pantel [42].

- Discovering hyperonyms: For each topic, a taxonomy is constructed. For each pair of nouns in the topic, a set of queries is generated; this considering the following:

  1. The Hearst patterns (see Table 3.1) have shown good evidence identifying that entity (noun) \( A \) is a hyponym of \( B \). But also, Snow et al. [30] identified other possible patterns as result of their method for discovering hyperonyms (see Table 3.2).

  2. A general query like \textit{such as}< hyperonym >\ is insufficient for obtaining interesting and precise information. In order to obtain useful information, the query needs to be more specific [34]. Then, related information to the query is added: 1) contextual information and 2) supervised information. The contextual information is given to the terms with the highest frequency in the corpus (without stopwords and after of a lemmatization process). The supervised information is given to the more representative terms in the WordNet\(^3\) synset where the term appears.

  3. For each query in the hyperonym query set, the \( n \) first pages are retrieved. The text for each \( n \) page is cleaned (eliminating images, videos, banners, etc.) and parsed. Each sentence is POS-tagged using the Stanford tagger\(^4\) and the lexical pattern of the query and their candidate hyperonym are identified. A term is selected as hyperonym if it is a noun or adjective but is not a stopword.

  4. The list of candidate hyperonyms is evaluated using a new query set, where each possible hyperonym will be replaced in the lexical pattern. Each candidate hyperonym is evaluated using its query set and the number of hits by means of the following measure [38]:

  \[
  \text{Score}_{\text{CandHyperonym}} = \frac{\text{hits(LexicalPattern(term, CandHyperonym))}}{\text{hits(CandHyperonym)} \quad (3.2)}
  \]

  5. The \textit{LexicalPattern(}term, \textit{CandHyperonym)}\ corresponds to build a query like: \textit{<term>,+and+other+<CandidateHyperonym>} where \textit{and} and \textit{other} correspond to some lexical pattern.

  6. The hyperonym with the highest score in the results for the query will be the hyperonym associated to the term.

**Preliminary Results**

A sample of the Lonely Planet\(^5\) corpus was used in the experiments, this corpus includes a taxonomy constructed by hand.

\(^3\)http://wordnet.princeton.edu/
\(^4\)http://nlp.stanford.edu/software/tagger.shtml
\(^5\)http://olc.ijs.si/lpReadme.html
To illustrate the experiment, the term museum was considered. The terms with the greater frequency in the sample corpus were: cash, travel, and product. The words whose WordNet synset were extracted were: collection, object, and display; they lexical pattern query set is shown in Table 3.3.

Using the query set with only lexical patterns, the list of candidate hyperonyms was: <site, place, attraction, department of history>. Using a query with added information, the new candidate hyperonyms were: <depository, institution>. A new lexical pattern query set was created using each one. Then using the number of obtained hits in the web search, the corresponding score was computed for each candidate hyperonym. For example, for term attraction the obtained hits are shown in Table 3.4.

In Table 3.5 can be seen that the best hyperonym to museum is attraction and the tourist context could be a good option, but it is important to note that the second best candidate is institution and according to different authors the definition of museum are:

...a museum is a building or institution which houses and cares for a collection of artifacts and other objects of scientific, artistic, or historical importance and makes them available for public viewing through exhibits that may be permanent or temporary...\(^6\)

Museums enable people to explore collections for inspiration, learning and enjoyment. They are institutions that collect, safeguard and make accessible artefacts and specimens, which they hold in trust for society...\(^7\)

The museum is an empowering institution, mean to incorporate all who

\(^7\)http://www.museumsassociation.org/about/frequently-asked-questions
Lexical Patterns with the highest frequency terms

<table>
<thead>
<tr>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>museum, + and + other + cash + travel + product</td>
</tr>
<tr>
<td>museum, + or + other + cash + travel + product</td>
</tr>
<tr>
<td>museum + is + a + cash + travel + product</td>
</tr>
<tr>
<td>such as + museum + cash + travel + product</td>
</tr>
<tr>
<td>including + museum + cash + travel + product</td>
</tr>
<tr>
<td>especially + museum + cash + travel + product</td>
</tr>
<tr>
<td>called + museum + cash + travel + product</td>
</tr>
<tr>
<td>particularly + museum + cash + travel + product</td>
</tr>
<tr>
<td>for example + museum + cash + travel + product</td>
</tr>
<tr>
<td>among which + museum + cash + travel + product</td>
</tr>
</tbody>
</table>

Lexical Patterns with the terms in WordNet synsets

<table>
<thead>
<tr>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>museum, + and + other + collection + object + display</td>
</tr>
<tr>
<td>museum, + or + other + collection + object + display</td>
</tr>
<tr>
<td>museum + is + a + collection + object + display</td>
</tr>
<tr>
<td>such as + museum + collection + object + display</td>
</tr>
<tr>
<td>including + museum + collection + object + display</td>
</tr>
<tr>
<td>especially + museum + collection + object + display</td>
</tr>
<tr>
<td>called + museum + collection + object + display</td>
</tr>
<tr>
<td>particularly + museum + collection + object + display</td>
</tr>
<tr>
<td>for example + museum + collection + object + display</td>
</tr>
<tr>
<td>among which + museum + collection + object + display</td>
</tr>
</tbody>
</table>

Table 3.3: Example Web query set

Thus, according to the added information to queries, the term institution is a good candidate hyperonym to museum.

Following the experiments, in Table 3.6 can be seen the hyperonyms obtained for each term and their appropriate WordNet hyperonym. The query set was used to find the hierarchical structure shown in Fig. 3.4 which corresponds to the terms: <plant, vegetation, park, garden, region, safari, environment>.

Conclusions and Further Work

An approach to discover hyperonyms have been presented in this chapter. The use of the related information in the queries on the Web seems to be a good approximation to narrow the results of the search; which allows the method can be applied to any domain of knowledge. This kind of queries are the most concrete and indicates 1) that the relation between terms is taxonomic and 2) the terms and their hyperonym are in the same context. However, WordNet is limited on nouns with more one terms, and only includes some proper nouns.

---

### Table 3.4: Example web query set for evaluating the term attraction

<table>
<thead>
<tr>
<th>Query Set</th>
<th>Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>museum, and other attraction</td>
<td>12300000</td>
</tr>
<tr>
<td>museum, or other attraction</td>
<td>12300000</td>
</tr>
<tr>
<td>museum is a attraction</td>
<td>26900000</td>
</tr>
<tr>
<td>attraction such as museum</td>
<td>26900000</td>
</tr>
<tr>
<td>attraction including museum</td>
<td>26900000</td>
</tr>
<tr>
<td>attraction especially museum</td>
<td>26900000</td>
</tr>
<tr>
<td>attraction called museum</td>
<td>11600000</td>
</tr>
<tr>
<td>attraction particularly museum</td>
<td>26800000</td>
</tr>
<tr>
<td>attraction for example museum</td>
<td>3780000</td>
</tr>
<tr>
<td>attraction among which museum</td>
<td>12500000</td>
</tr>
</tbody>
</table>

### Table 3.5: Score of candidate hyperonyms

<table>
<thead>
<tr>
<th>Candidate Hyperonym</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>site</td>
<td>0.09794</td>
</tr>
<tr>
<td>place</td>
<td>0.21463</td>
</tr>
<tr>
<td>attraction</td>
<td>3.74220</td>
</tr>
<tr>
<td>department of history</td>
<td>0.82055</td>
</tr>
<tr>
<td>depository</td>
<td>1.50125</td>
</tr>
<tr>
<td>institution</td>
<td>3.65833</td>
</tr>
</tbody>
</table>

The obtained results can be improved resolving ambiguous terms that are not in WordNet. Also, add new lexical patterns to questions and extending the search on Frequently Questions Blogs and Wikipedia are a good options to explore.
Figure 3.4: Taxonomy found

<table>
<thead>
<tr>
<th>Term</th>
<th>Hyperonym obtained</th>
<th>WordNet Hyper</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 plant</td>
<td>organism</td>
<td>organism, being</td>
</tr>
<tr>
<td>2 park</td>
<td>plant</td>
<td>tract, piece of land</td>
</tr>
<tr>
<td>3 garden</td>
<td>park</td>
<td>vegetation</td>
</tr>
<tr>
<td>4 region</td>
<td>park</td>
<td>location</td>
</tr>
<tr>
<td>5 safari</td>
<td>garden</td>
<td>expedition, travel</td>
</tr>
<tr>
<td>6 environment</td>
<td>garden</td>
<td>geographical area</td>
</tr>
<tr>
<td>7 vegetation</td>
<td>plant</td>
<td>collection, aggregation</td>
</tr>
</tbody>
</table>

Table 3.6: Score of candidate hyperonyms
Chapter 4

Conclusions

The main goal in this research is to obtain a model for ontology learning from textual resources about a specific domain. This implies several challenges:

- obtain representative concepts
- find relevant taxonomic and non-taxonomic relationships
- achieve a higher level of expressiveness (axioms)

The text clustering seems to be a good technique to obtain terms and web search to obtain relationships (hyperonym and hyponym). The lexical-patterns show a good evidence for identifying taxonomic and non-taxonomic relationships.

According to the schedule of activities in the first phase, the main achieved activities were: searching and collecting textual resources, obtaining the representation model, integration with the clustering algorithm, and implementation of selected techniques to get relevant vocabulary and taxonomic relationships.

The next phase in the doctoral studies considers to study and select a technique to get non-taxonomic relationships.
Bibliography


