An interactive pattern based approach for extracting non-taxonomic relations from texts

Marie Chagnoux and **Nathalie Hernandez** and **Nathalie Aussenac-Gilles** [1]

Abstract. In this paper, we propose a methodological approach based on pattern design and acquisition from texts in order to enrich lightweight ontologies with non-taxonomic relations. Since learning approaches require constrained domains and corpora with strong regularities, an alternative method is needed to locate sharp relations in versatile corpora. Our approach relies on an existing ontology, a pattern bank that can be enriched and a tagged corpus resulting from a morpho-syntactic analysis. The objective is twofold : (i) the morpho-syntactic patterns stored in the bank are used to identify new relations between the concepts from the ontology (ii) new patterns identifying new kinds of relations are extracted from the context of cooccurring concept labels; these patterns enrich the pattern bank and can be matched to look for new semantic relations. At the end of the process, the original ontology is enriched with the relations extracted from texts.

1 Introduction

Relation extraction from texts contributes in two complementary ways to Ontology Engineering: automatically identifying relation instances between concept instances results in Ontology Population ; extracting relations or properties between concept classes is part of Ontology Building, and it is automated for Ontology Learning.

On the one hand, recent works in Ontology Population rely on learning technics based on statistics combined with works in linguistics [14] [20] [15] [8]. Their rather good results arise from the fact that the task is highly constrained by the existing material: concepts and relations are already defined in the ontology; sets of manually tagged documents are available as well as label lists of concept instances; the searched relations can be expressed with a reduced number of variant phrases that are listed *a priori*. Brewster and Ciravegna [5] experimented machine learning algorithms to support pattern-based relation extraction. Their system is able to define one or several patterns from a set of linguistic expressions that have been tagged as revealing a relation between concept instances. Learning is particularly relevant in fields where texts are little disposed to variations. For instance, many works [7] have been done in the biomedical field where texts offer the advantage of redundant structures which can be easily fixed in reproductive patterns.

On the other hand, Ontology Building deals with a more complex task since neither concepts nor relations are known: both elements must be discovered. Many works have been dedicated to the extraction of terms as concept labels and to the extraction of taxonomic relations [4], but the extraction of non taxonomic relations is still a challenging task [11]. This lack of initial anchorage makes it more difficult to carry out learning approaches. In this context, matching linguistic patterns with texts offers a robust solution under a human supervised process. Based on Hearst's ideas, we developed a first tool, Caméléon, that implements pattern matching in corpora to identify relations and concepts for ontology engineering [16].

Our aim is to build on our experience in pattern-based relation extraction and ontology building with Caméléon in order to preserve the sturdiness of extraction and improve the assistance given to the user during the critical steps of the ontology building process. We propose a methodological framework that better support the user during the pattern and relation identification tasks in the case where a hierarchy of concepts and their related terms are already available. In keeping with the options made in Caméléon, this framework promotes pattern reuse and adaptation. Its first novelty is to guide more efficiently the identification of related terms in the sentences matched with known patterns. The major change is to automatically suggest corpus-specific patterns.

In the first part of this paper, we introduce the main issues related to using pattern for ontology enrichment. In the second part, we present our framework by explaining our algorithm and its implementation.

2 Patterns for ontology enrichment

Pattern-based approaches for semantic relations discovery are based on the premise that semantic relations between terms such as hypernymy or part-whole can be *a priori* inferred from corpus investigation and abstracted in a pattern which embeds all the potentially related linguistic sequences.

These approaches suppose three different steps : during the acquisition step, the patterns are designed; during the matching step, the patterns are used to extract lexical relationships between pairs of terms from texts; during the modelling step, each lexical relation may lead to defining pairs of concepts connected with a semantic relation in the ontology.

A given relation REL can be discovered by means of different patterns. We call this set $\mathcal{P}^{REL} = \{P^{REL_1}, .., P^{REL_n}\}$. We consider that a pattern P^{REL_i} is a lexico-syntactic construction which is defined for capturing the semantic relation REL between two entities X and Y. This capture assumes an implicit intermediary lexical relation between two terms: t^1 related to X and t^2 related to Y. A relation REL can be designated by several labels or terms noted $Label(REL)$. When matched on the texts, P^{REL_i} allows to extract pairs of terms $REL^t(t^1, t^2)$ linked by the relation REL.

Since we intend to use pattern matching to enrich an existing ontology, patterns should help to extract relations between concepts of this ontology or new concepts to be added to the ontology. This goal implies focusing on an extra step which consists in mapping pairs

[1] IRIT - University of Toulouse, France, email: {aussenac, chagnoux, hernand}@irit.fr

of terms to pairs of appropriate concepts, in identifying the right semantic relation and in enriching the ontology with this new relation. According to the proposed notation, the pairs $REL^t(t^1, t^2)$ extracted from texts thanks to patterns must be transformed into pairs $REL^c(c^1, c^2)$ where c^1 and c^2 are two new or existing ontology concepts. The main problem concerns the interpretation process.

In the next section, we discuss the major issues involved in this step.

2.1 Main issues

2.1.1 Identifying concepts from terms

The identification of the concepts referenced by the terms of a pair is mainly complicated by three linguistic issues : polysemy and metonymy at local level and anaphora at global level.

Polysemy Polysemy is simpler, from an ontological point of view, than from a general perspective since ontologies are more generally dedicated to a specific domain. Polysemy occurs when a term corresponds to a label shared by different concepts. The term can thus refer to different concepts cj or ck. For instance, "*plant*" is ambiguous since it could refer to the concept "*factory*" or to the concept "*plant*" in the biological sense.

Metonymy Metonymy is a recurrent issue in Ontology Building. In metonymy, the term *t* must not be associated with the concept cj to which it seems to refer but with ck which is closely associated to cj. For instance, in the sentence "*the pilot climbed to height 600 m*" must be understood as "*the aircraft climbed*" as it is the aircraft which has the property of going upward.

Anaphora Anaphora is probably the main issue since it is the most frequent phenomenon in texts. Pattern design is spatially constrained and so far no method has been proposed to overlap sentence segmentation. Consequently, valuable information is frequently lost since patterns are matched to each sentence independently. When related terms appear in two consecutive (or worse non-consecutive) sentences, entity *X* or entity *Y* -or even both- can be missing. For instance, in 1, the semantic content of the second sentence will not be taken into account if the matched pattern is *N VB[BE] [ADJ]? NP * [to manufacture]*.

(1) "*The Airbus A310 is a medium to long-range widebody airliner manufactured by Airbus SAS.* ***It*** *was the second model to be introduced by Airbus after the A300.*"

2.1.2 Matching semantic relations

Once the concepts referenced by the pair of terms have been identified, a problem remains when enriching the ontology with a new relation. It is often difficult to determine at which level of the ontology the relation has to be added. The relation can be true only for the concepts identified from the terms or can be generalized to one of their parent concepts.

For instance, let's assume that an ontology includes the following two hierarchies: *CF6-50 engine* `IS-A` *CFM engine* `IS-A` *engine* `IS-A` *Aircraftcomponent* ; *A300* and *A318* `IS-A` *AirbusAircraft* `IS-A` *Aircraft*. Then, the lexical relations $POWER^t(American\ \ General\ Electric\ CF6-50\ engine, A300)$ and $POWER^t(The\ Airbus\ A318\ Elite, CFM\ engine)$ extracted for the relation $POWER^t$ from the following two examples 2 and 3 can lead to several potential conceptual relations listed in 4.

(2) "*American General Electric CF6-50 engines powered the A300*"

(3) "*The Airbus A318 Elite will be powered by CFM engines*"

(4) a. *POWER*^{c}*(CF6-50 engine,A300)*
    b. *POWER*^{c}*(CFM engine,Airbus A318 Elite)*
    c. *POWER*^{c}*(engine,AirbusAircraft)*
    d. *POWER*^{c}*(engine,Aircraft)*
    e. *POWER*^{c}*(Aircraftcomponent,A300)*
    f. *POWER*^{c}*(Aircraftcomponent,Aircraft)*

Since it seems to be difficult to automate relation conceptualization at the right level, the only option is to ask the user for validation.

2.2 Caméléon : benefits and drawback of past experiments

The method described in this paper is based on Caméléon, a system that supports ontology enrichment with semantic relations according to pattern-based relation extraction from texts. Caméléon provides assistance to reuse, adapt or design patterns for syntactically tagged texts. Then it supports pattern matching, human validation of the sentences found with the patterns that leads to defining terms and lexical relations, and finally it proposes an ontology editor where conceptual relations can be added.

Caméléon's ability to extract non-taxonomic relations from texts has been recently evaluated in [1] and [2]. These papers have shown both benefits and drawbacks of the approach. The undeniable asset of the pattern-based approach is its accuracy in discovering relations. The advantage of applying the method to a varied set of corpora is to make it possible to formulate some domain and corpus independent thresholds.

- Firstly, the experiment confirmed the high cost of pattern design, and the gain brought by pattern reuse.
- Secondly, when building an ontology, a major problem comes from the difficulty in identifying the terms involved in the relation and matching them with ontology concepts.
- Thirdly, reuse is successful in identifying cross-domain relations like hypernymy but it may lead to few conceptual relations depending on the textual genre of the corpora. For instance, only didactic handbooks contain definition patterns. When defining domain-specific patterns, this method is tedious and severely limited by the small number of patterns typically employed.
- Finally, the evaluation of the method on disparate corpora has shown a great heterogeneity of results from one text to another as well as from one relation to another. These results will be true for any pattern-based approach for relation extraction.

Because we assume that a taxonomy of concepts and their related terms form the kernel of an ontology to be enriched with conceptual relations, we identified two possible means to improve this process:

- During pattern matching, the system will propose all known ontological terms occurring in the sentence as being potential related terms. From these lexical relations, potentially conceptual relations can be defined. The validation could be delayed and could bear only on candidate conceptual relations.

- The system will learn patterns from the contexts of co-occurring terms or concepts from the ontology. We will first enrich texts with semantic annotations (the concepts corresponding to term occurrences). Learning will take into account POS tags and semantic annotations to define high level patterns. This step should reduce the cost of pattern definition and improve concept identification, because the process fixes terms and concepts to determine the relation.

3 An interactive approach to define and reuse linguistic patterns

Most of the approaches in the literature are concerned with concept identification. Patterns are used to extract relevant lexical unit pairs. The approach we present in this paper differs in that the elements we attempt to find are relations and not concepts in relations : we suppose that, in our patterns, X and Y are identified since they match ontology-defined concepts.

Moreover, seeing that most relation extraction systems aim at discovering taxonomic relations [19], part-whole relations [11] and hypernyms [17], we have focused our work on transversal relations. Other works have already dealt with this question: [10] is interested in causal relations, [12] compares causal relations in English and French, [18] explores Semantic Web possibilities, etc. But most of the time, the works narrow their typology to a very small number of relations. We believe that the pattern-based approach is able to capture more accurate relations provided that the resources are extended. This belief was already one of the fundamental options in Caméléon where the list of searched relation types is open and can be adapted to each new studied corpus. It is also the option chosen by [13] to identify lexical patterns which represent semantic relations between WordNet concepts. In our approach the user lies at the heart of the process since he validates all suggestions, especially during pattern acquisition. Our work can also be compared to the dynamic iterative approach of [6] since it is founded on the same three kinds of resources.

Thanks to the ongoing interest in the semantic web, existing ontologies are now accessible through the Web. For example, gateways such as `Watson`[2] or `Swoogle`[3] make it possible to retrieve and access ontologies and ontology entities. In our approach, we propose to reuse relations or properties found in existing ontologies.

3.1 Required resources

Our approach takes as input a list of existing patterns, a lightweight ontology and a corpus.

3.1.1 Pattern base

The first step of our methodological proposal is the reuse of already existing patterns: we adapted Caméléon patterns[4] as well as those from other works[5] in a base of patterns where they are organized according to the REL relations and their labels.

[2] `Watson` [9] for the Semantic Web aims "to provide an efficient access point to the online ontologies and semantic data." See http://watson.kmi.open.ac.uk/Overview.html

[3] `Swoogle`, http://swoogle.umbc.edu/, crawls the World Wide Web for Semantic Web documents written in RDF

[4] Since Caméléon was built for French, patterns were rebuilt to fit English

[5] We notably reused [3] and [12] among others

3.1.2 Ontology

As our goal is to identify the concepts referenced by pairs of terms extracted thanks to patterns, our proposal is highly dependent on the ontology lexical component. For this reason, we consider lightweight ontologies which are composed of two semiotic levels [15]. The lexical level (L) covers all the terms or labels defined to designate concepts or relations. The conceptual level defined in the structure (S) of the ontology represents the concepts and the semantics defined from the conceptual relations that link them.

The structure of an ontology is a tuple $S := \{C, R, \leq, \partial_R\}$ where:

- C, R are disjoint sets containing concepts, non taxonomic relations
- $\leq$: $C \times C$ is a partial order on C, it defines the hierarchy of concepts
- $\partial_R : R \rightarrow C \times C$ is the signature of an associative (or non-taxonomic) relation.

The lexicon of a lightweight ontology is a tuple L : $\{L^C, L^R, F, G\}$

- L^C, L^R are disjoint sets containing labels (or terms) referencing concepts and relations
- F, G are two relations called reference, they enable access to the labels (or terms) referencing respectively the concepts and the relations.

Note that a concept can be defined by several terms, and a term, when it is ambiguous, can reference several concepts.

In our approach, a lightweight ontology is given as input to our system. This ontology is at least a hierarchy of concepts defined with a lexicon $L^0 := \{L^C, F\}$ and a structure $S^0 := \{C, \leq\}$. The aim of our approach is to enrich both S^0 by identifying R and ∂_R from texts and L^0 by defining G.

3.1.3 Corpus

The choice of the corpus is a deciding factor in the ontology enrichment process. The corpus must describe the items of knowledge that will be integrated into the ontology. In our approach, the corpus is extracted from existing corpora and experts must guarantee that it covers the whole domain in a representative period.

3.2 The Webcontent project: a case study

Our work takes place in the context of WebContent[6], a project which aims at building a computing environment to explore and use the Semantic Web technologies for applications like, for instance, Technological Watch. One of the application fields is the economic watch in the aviation industry. Consequently, all the resources cited as examples in this paper are related to our work in the project for this application. The pattern base presented in the above section has been manually enriched with non-taxonomic relations discovered from the corpora. Table 1 illustrates some examples of relations discovered by patterns and associated with the linguistic segments.

In the framework of the WebContent project, we propose to enrich a handmade hierarchy of concepts of the aeronautical domain. A sample is presented in figure 1. Each concept of this ontology is associated with a list of labels.

We propose to extract relations between concepts from two different kinds of texts. The corpus we used to illustrate our approach is

[6] http://www.webcontent.fr, June 2008

Table 1. Examples of transversal relations and related patterns

Relation	examples of pattern
Communicat.	X instruct (DT)? Y
Build	X (MD)? build Y
Build	X (MD)? assemble (DT)? Y
Own	X belonging to Y

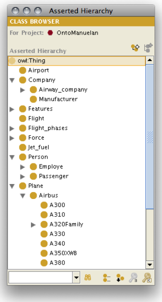

Figure 1. Extract of the ontology used in WebContent

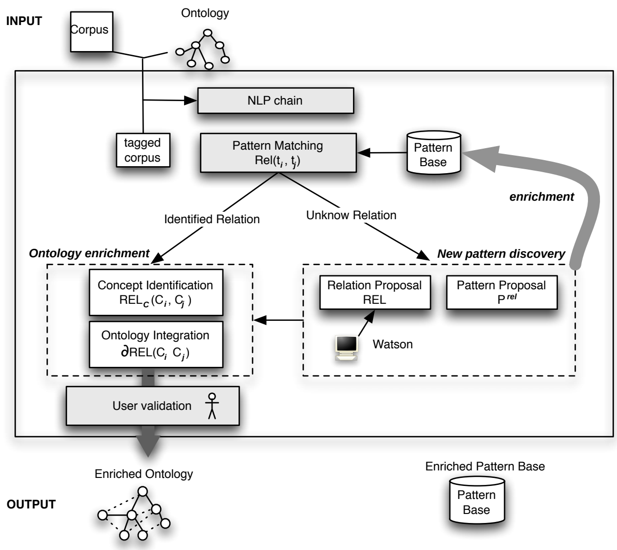

Figure 2. Overview on our system

in English. It is built up from 513 short news releases from AFP [7] and 14 *Wikipedia* articles about particular aircrafts such as A300 or B737. This double nature situates our work both in the field of economics with the news from AFP and in the field of aeronautics with the *Wikipedia* articles.

3.3 Algorithms and implementation

The aim of our proposal is twofold: using and discovering relation extraction patterns and enriching an existing ontology with new relations.

In this section, we give an overview of our system by presenting our general algorithm. We then detail its different steps.

3.3.1 Overview

Figure 2 presents an overview on our system.

The first part of the algorithm is dedicated to processing - relations and pattern discovery - and the second one to validation by the user. The validation phase is carried out in a second time in order to propose to the user all the pairs extracted for a relation. We believe that this way the user's work is facilitated.

Processing For each pair of distinct ontology concepts (c^i, c^j), we look for all the sentences that contain t^i, t^j where t^i, t^j belong respectively to the set of labels associated with concepts c^i, c^j. If one of the base patterns $P^{REL}i$ can be matched on the sentence s, we store the relation REL extracted by the pattern, the pair of concepts $REL^c(c^i, c^j)$ and s. If not, we search for a relation that could be defined in an existing ontology. If such a relation is found, we store the new relation noted REL_{new}, the pair of terms $REL_{new}{}^c(c^i, c^j)$ and s.

Pattern proposition and Validation For each new relation REL_{new} detected thanks to existing ontologies, we display to the user the relation and all the pairs of concepts $REL_{new}{}^c(c^i, c^j)$. The user is asked to validate the relevance of the relation. If he validates the relation, the system proposes a set of patterns that can be generated according to the sentences where the pairs have been identified. The user validates the relevance of the patterns. Validated patterns are added to the base.

Relation validation For each relation detected (REL and REL_{new}), the system displays to the user the labels of the relation and the pair of related concepts. The user then decides where to add the relation in the ontology. He can either decide to

- add the relation between c^i and c^j
- add the relation between an ancestor of c^i and an ancestor of c^j
- add the relation between a concept linked to c^i in the ontology and a concept linked to c^i
- reject the relation for the pair

In order to facilitate his choice, for each pair the user can have access to the sentences where the pair have been found in the text and to the context of each concept in the ontology (labels of the concepts and related concepts).

The implementation of the different steps is detailed in the following.

3.3.2 Ontology enrichment

Pre-treatment Our system has been implemented with `GATE` which provides a Language Engineering software architecture to develop and deploy NLP chains. It has the advantage of offering finite state techniques to extract or annotate information. Each text submitted to our system is tokenised, split into sentences, tagged with `TreeTagger` and annotated with ontology information.

[7] Agence France-Presse, a French news agency

Pattern matching Once the text has been processed through this sequence of tools, relations are discovered with `JAPE`, the *Java Annotation Patterns Engine*, which provides finite state transduction over annotations based on regular expressions. For each pair of terms that correspond to concept labels, we use `GATE` to discover potential relations. Each type of relation is associated to a JAPE grammar: i.e. a set of phases running sequentially in order to constitute a cascade of finite state transducers over annotations defined at the start of each grammar. TreeTager results will be exploited in the pattern design as they provide morphosyntactic information and lemma.

Table 2 presents examples of relations identified between concepts according to existing patterns after the disambiguation process. It also presents sentences which illustrate these relations.

Table 2. Examples of transversal relations between concepts.

Relation	Concept c^1	Concept c^2	Examples of textual segment
Comm.	Pilote	Controller	"*the controller instructed the pilot*"
Build	Aircraft Manufacturer	Plant	"*Airbus is expected to begin to build plant*"
Build	Factory	Plane	"*the plant will have the capacity to assemble four" aircrafts per month*"
Own	Airway Company	Plane	"*the Airbus A320 belonging to Armenian Airlines*"

Concept Identification In order to identify the concepts involved in pairs $REL^t(t^1,t^2)$, gazetteers annotate text words according to the concept labels that are defined in the lexicon of the ontology. For each concept c, a gazetteer is constructed according to $F(c)$. For example, the gazetteer for the concept *plane* will be composed of the terms *aircraft, plane, aeroplane, airplane*. This step is crucial as it will determine the X and the Y used in our patterns and the anchorage of the pattern proposal. To ensure that the most specific concepts are identified, gazetteers are implemented in order to favor the detection of labels that correspond to the longest possible phrases (i.e. composed of the greatest number of words). For example, in the sentence "*The heads report to the vice president*", the labels of two concepts are present. We will favor the extraction of the label "*vice president*" rather than "*president*".

As mentioned in a previous subsection, a label may refer to different concepts. We thus need to disambiguate the terms of the pairs: when the label found in a pair refers to a single concept, the corresponding concept is identified. For instance, in $BUILT^t(plant, aircraft)$, the term "*aircraft*" is not ambiguous as it refers to the concept *plane*. On the other hand, when the label corresponds to several concepts, the user will identify the relevant concept during the validating step.

Enriching the ontology with new relations Once the concepts involved in the relations are identified, we propose to enrich the ontology with the relations extracted for the most general concepts. For example, our approach extracts the relation "*BUILD*" for the pair of concepts "*factory*", "*plane*" (as shown previously) and for the pair "*assemblyLine*, "*plane*". As the concept "*assemblyLine*" is defined as a subconcept of "*factory*", the system only proposes to add the relation in the ontology between "*factory*", "*plane*". Before adding relations to the ontology, the user will be asked to validate each proposal.

3.4 Discovering and validating new patterns

Sometimes, no relation can be identified between two concepts whose instances co-occur in a sentence — we thus have pairs of "orphan" concepts. In this case, we implicitly assume that already-existing ontologies may include already-established relation(s) linking these two concepts. We thus enrich our strategy with relation and pattern proposals to the user.

Relation Proposal First, such pairs of orphan concepts are discovered from texts and submitted to `Watson`. This step yields all available relations between such two concepts, which are present in the database. For instance, none of the patterns previously stored in our base corresponds to a relation between the concepts "*Person*" and "*Company*", although they frequently occur within the same context, as illustrated below:

1. "*Airbus Names New Chief Managers for A380, A320 Programs*"
2. " *Aircraft maker Airbus has named Mario Heinen senior vice president and chief manager of the A380 aircraft program*"
3. "*Laurence Barron, the vice senior president of Airbus*"

`Watson` proposes the "*Work_for*" relation.

Pattern Proposal Meanwhile, the system suggests a set of patterns which could be associated with the new relation. Patterns combine all kinds of elements defined by the `Gate` resources in the pattern set: for instance, the lemma "name" is treated either as a syntactic category (verb), as a semantic category (a verb referring to a choice), or simply as a string[8]. For each item in the sentence, the system proposes either its grammatical category, its semantic category or its word (lemma). The result is a list of combinations mixing heterogeneous elements and matching the linguistic configuration.

From sentences such as examples 1 and 2, our system will propose several combinations illustrated by examples 1a–1d. If Named Entities or any other semantic knowledge is available in particular gazetteers, they will appear in patterns as in 1e.

1.(a) X name Y
  (b) X name (NP)? (NP)? Y
  (c) X (VB=choose) Y
  (d) X (MD)? (VB=choose) Y
  (e) X name NE_Person Y (with NE_Person semantic class of Named Person)

Contrary to learning approaches, the entire control of the user on the pattern construction process guarantees the semantic significance of the patterns and the relevance of the identified conceptual relations. Meanwhile, the interactive pattern acquisition accelerates the manual validation process.

4 Conclusion

Pattern-based relation extraction from a corpus can be an efficient means to enrich an ontology provided patterns can be collected, accumulated, adapted and semi-automatically acquired, and provided

[8] If the semantic category is chosen, the user has to create a gazetteer from instances of relations proposed in the thesaurus. In our case, among the sixteen selected synonyms, he could retain { *select; pick; decide on; nominate; designate*} and ignore {*give a name; baptize*}.

related terms can be easily identified in matched sentences. To carry out this process, we propose a tool which extends the Caméléon relation extraction tool by integrating learning principles. This new tool embeds (i) term identification in phrases matching already-written patterns and (ii) a pattern-creation assistant based on automatic proposals but not on machine learning.

Our tool is still being implemented as we now endeavor to improve two aspects of our approach. The first hindrance is induced by the problem of defining the context window where concept instances are to appear. The default window in Caméléon is the sentence. Indeed, longer windows lead to prohibitively complex combinations, while shorter windows often prove to be deficient in discovering relations. The second issue relates to the way patterns are proposed. Currently, all possible combinations of linguistic elements are proposed to the user. The next step will be to reduce the number of patterns when there are numerous occurrences of the same relation thanks to a learning or clustering stage of the abstracted patterns. If learning is not possible, at least, patterns will be ordered with respect to an automatically-measured accuracy. This measure will compare the structure of new patterns to that of those already present in the base. Evaluations will have to be carried in order to compare our proposal with existing approaches.

REFERENCES

[1] Nathalie Aussenac-Gilles and Marie-Paule Jacques, 'Designing and evaluating patterns for ontology enrichment from texts', in *EKAW 2006, 15th International Conference on Knowledge Engineering and Knowledge Management , Prague, oct. 2006*, ed., Springer Verlag, pp. 158 – 165, Prague, (2006).

[2] Nathalie Aussenac-Gilles and Marie-Paule Jacques, 'Designing and evaluating patterns for relation acquisition from texts with caméléon', *Terminology, special issue on Pattern-Based approaches to Semantic Relations*, **14**(1), 45 – 73, (2008).

[3] Caroline Barrière and Akakpo Agbago, 'Terminoweb: A software environment for term study in rich contexts', in *International Conference on Terminology, Standardisation and Technology Transfer (TSTT 2006)*, Beijing, Chine, (25 et 26 août 2006).

[4] Christopher Brewster, 'Techniques for automated taxonomy building: Towards ontologies for knowledge management', in *Proc. of the 5th Annual CLUK Research Colloquium, Leeds*, (2002).

[5] Christopher Brewster, Fabio Ciravegna, and Yorick Wilks, 'User-centred ontology learning for knowledge management', in *NLDB*, eds., Birger Andersson, Maria Bergholtz, and Paul Johannesson, volume 2553 of *Lecture Notes in Computer Science*, pp. 203–207. Springer, (2002).

[6] Christopher Brewster, Jose Iria, Ziqi Zhang, Fabio Ciravegna, Louise Guthrie, and Yorick Wilks, 'Dynamic iterative ontology learning', in *Proceedings of the Recent Advances in Natural Language Processing Conference (RANLP)*, (Septembre 2007).

[7] Massimiliano Ciaramita, Aldo Gangemi, Esther Ratsch, Jasmin Saric, and Isabel Rojas, 'Unsupervised learning of semantic relations between concepts of a molecular biology ontology', in *IJCAI*, pp. 659–664, (2005).

[8] Philipp Cimiano, *Ontology Learning and Population from Text. Algorithms, evaluation and applications*, Springer, Berlin, 2007.

[9] Mathieu d'Aquin, Claudio Baldassarre, Laurian Gridinoc, Sofia Angeletou, Marta Sabou, and Enrico Motta, 'Watson: A gateway for next generation semantic web applications', in *ISWC, Poster session of the International Semantic Web Conference*, (2007).

[10] Roxana Girju and Dan Moldovan, 'Text mining for causal relations', in *FLAIRS 2002*, pp. 360 – 364, Pensacola Beach, Florida, (14-16 mai 2002 2002).

[11] Roxana Girju Girju, Adriana Badulescu, and Dan Moldovan, 'Learning semantic constraints for the automatic discovery of part-whole relations', in *HLT-2003*, (2003).

[12] Elizabeth Marshman, 'Towards strategies for processing relationships between multiple relation participants in knowledge patterns: An analysis in english and french', in *Terminology*, ed., John Benjamins, volume 13:1, pp. 1 – 34, (2007).

[13] Maria Ruiz-Casado, Enrique Alfonseca, and Pablo Castells, 'Automatising the learning of lexical patterns: An application to the enrichment of wordnet by extracting semantic relationships from wikipedia', *Data Knowl. Eng.*, **61**(3), 484–499, (2007).

[14] Alexander Schutz and Paul Buitelaar, 'Relext: A tool for relation extraction from text in ontology extension', in *ISWC 2005, LNCS 3729*, ed., Y. Gil et al., pp. 593 — 606, (2005).

[15] Alexander Schutz and Paul Buitelaar, 'Relext: A tool for relation extraction in ontology extension', in *Proceedings of the 4th International Semantic Web Conference*, Galway, Ireland, (November 2005).

[16] Patrick Séguéla, *Construction de modèles de connaissances par analyse linguistique de relations lexicales dans les documents techniques*, Ph.D. dissertation, Thèse de Doctorat en Informatique, Toulouse 3 Paul Sabatier, 2001.

[17] Rion Snow, Daniel Jurafsky, and Andrew Y. Ng, 'Learning syntactic patterns for automatic hypernym discovery', in *Advances in Neural Information Processing Systems 17*, (2005).

[18] Maria Tchalakova, Borislav Popov, and Milena Yankova, 'Methodology for bootstrapping relation extraction for the semantic web', in *AIMSA 2006*, Varna, Bulgaria, (13-15 sept. 2006).

[19] Paola Velardi, Alessandro Cucchiarelli, and Michael Petit, 'A taxonomy learning method and its application to characterize a scientific web community', *IEEE Transactions on Knowledge and Data Engineering*, **19**(2), 180 – 191, (2007).

[20] Dominic Widdows, 'Unsupervised methods for developing taxonomies by combining syntactic and statistical informations', in *HLT/NAACL*, pp. 276 –283, (2003).