Document Annotation and Ontology Population from Linguistic Extractions

Florence Amardeilh
Mondeca, R&D Lab.
Paris, France.
Florence.amardeilh@mondeca.com

Philippe Laublet
Laboratoire LaLICC,
Université Paris IV, France.
philippe.laublet@paris4.sorbonne.fr

Jean-Luc Minel
Laboratoire LaLICC,
Université Paris IV, France.
jean-luc.minel@paris4.sorbonne.fr

ABSTRACT
In this paper, we present a platform for semi-automatic ontology population from textual documents. Our platform provides an environment for mapping the linguistic extractions with the domain ontology thanks to knowledge acquisition rules. Those rules are activated when a pertinent linguistic tag is reached. Those linguistic tags are then mapped into a concept, one of its attributes or even a semantic relation between several concepts. The rules instantiate these concepts, attributes and relations in the knowledge base constrained by the domain ontology. This paper deals with the underlying knowledge capture process and presents the first experimentations realized on a real client application from the legal publishing domain.

Categories and Subject Descriptors
I.2.4 Knowledge Representation Formalisms and Methods – *Representation languages.*
I.2.7 Natural Language Processing – *Language parsing and understanding.*

Keywords
Knowledge acquisition tool, Knowledge extraction from text, Knowledge capture for the Semantic Web, Knowledge capture using natural language processing, Method for ontology population, Semantic Web.

INTRODUCTION
From publishing industry to business competitive intelligence, important volumes of data coming from various documentary sources are processed and analysed by professional users. Firstly, taken the documents released each day in their fields, they must select the most relevant ones wrt to their applications. Secondly, for every resource, they must manually extract the pertinent information. That information is used to annotate the document by a set of descriptors (terms from the thesaurus like 'divorce' and named entities like 'Mr. Bouscharain') and to enrich their knowledge base (with named entities, attributes of these named entities and semantic relations between these named entities).

In the Semantic Web context, the content of a document can be described and annotated using knowledge representation languages such as RDF, XTM and OWL. RDF, the Resource Description Framework (Ora, 1999), is a formalism of knowledge representation originated from the semantic networks field and whose syntax is XML-based. It is mainly used to describe resources, such as an electronic web document, by a set of documentary metadatas (author, date, source, etc.) and of descriptors. Those metadatas are composed of triplets : (subject, verb, object) or (object 1, relation, object 2) or (resource, property, value) according to the needed description type.

The Topic Maps are another formalism of knowledge representation which is also based on an XML syntax (Park, 2003). The Topic Maps define a set of topics relatively to a same domain with interactions between them constituting a semantic map of the knowledge. A topic represents everything that can be described or thought by a human. It can participate to one or many relations, called associations, in which it plays a specific role. The topics have at least a name and intrinsic properties, called occurrences. This language allows a great flexibility in knowledge representation, especially with regards to the modelling of complex n-ary semantic relations.

OWL, Ontology Web Language (Hendler, 2004), is used to formalise an ontology (Gruber, 1993), or more generally some ontological and terminological resources (Bourrigault, 2003), by the definition of concepts used to represent a domain of knowledge. This language describes those concepts by a set of properties, relations and constraints. The formalism used in OWL is taken from some formalisms of description logic.

In our projects, we use RDF to describe the content of a documentary resource, OWL to model the ontology which will represent an applicative or a functional vision of the concerned domain and the Topic Maps to implement the knowledge base that will contain the instances of the concepts, properties and relations described in the domain ontology. The pertinent information of the domain, contained in the documentary resources will be captured to instantiate the knowledge base and to create documentary annotations. These semantic annotations can then be interpretable by the machine to be later shared, published, queried or more generally used (Laublet, 2002).

Semantic annotation and ontology population are greatly dependent of the information extracted from the documentary resources by the professional users. This manual process of the documents is extremely expensive in time and resources. The entire process involves productivity and quality issues. For all those reasons, the companies are more and more looking for implementing solutions based on the use of linguistic tools that semi-automatically capture the pertinent information from textual documents.

Those natural language processing technologies should be tightly integrated into the future Semantic web applications et shall even become essential to the development, acceptation and use of the Semantic Web (Bontcheva, 2003). Thanks to the functionalities offered by the natural language processing technologies, et mainly those of Information Extraction, solutions adapted to the specific needs of Semantic Web might be developed such as:

- The semi-automatic construction of terminologies/vocabularies of a domain from a representative documentary corpus as well as their maintenance (Bourrigault, 2003).
- The semi-automatic enrichment of knowledge bases by named entities and semantic relationships extracted from textual documents (Kyriakov, 2003).
- Semantic annotation of documentary resources (Kahan, 2001) (Handschuh, 2002) (Vargas-Vera, 2002).

We noticed in our own projects that the linguistic tools and the ontology of the client domain are implemented independently from one another contrary to the research works previously cited. That's why we decided to implement a gateway between the concepts of the ontology and the semantic tags resulting from the linguistic tools that will capture the pertinent knowledge of the domain.

In this article, we present an innovative platform for document annotation and knowledge acquisition. In the next section of this paper, we will talk about the origin of actual problematic and we will describe the implementation of our solution. Then, we will present the results of the first experimentations issued from a project realized in the legal publishing field. This project will be used to illustrate our work all along the article. At last, we will sum up the results in order to develop new hypothesis and conclude on the future perspectives of our research.

LINGUISTIC TOOLS INTEGRATION IN A WEB SEMANTIC PORTAL

The tools used: ITM™ & IDE™

Our solution is based on the Intelligent Topic Manager™ (ITM) tool from the company Mondeca. ITM is a software engineering platform for knowledge management. ITM integrates a semantic portal (Amardeilh, 2004) providing four key functions : Edition, Search, Navigation and Publication. The domain ontology, formalized in OWL, constraints the knowledge base model (implemented in Topic Maps), the user interfaces as well as every functionalities of the portal. The knowledge base elements point out towards their relative documents, accessible by URL on the Internet or in a content management system.

The linguistic analysis is done by the Insight Discoverer™ Extractor (IDE) developed by the company Temis. This tool implements a finite-state transducer method (Grivel, 2001) that relies on a pre-treatment involving document segmentation in textual units (usually sentences), lemmatisation and morpho-syntaxic analysis of those textual units. IDE™ produces a tagged conceptual tree. Each node of that tree is named by the semantic tag attributed to the textual unit extracted depending of the domain analysed (cf. **Erreur ! Source du renvoi introuvable.**).

On the one hand, the ITM™ portal doesn't allow the (semi-)automatic enrichment of its knowledge base. On the other hand, the IDE™ information extractor, once the information extracted on a textual documentary corpus, simply presents these information to the user through an html interface without recording in a knowledge base or even a database for later reuse. Both companies decided to collaborate on several projects (documentation, publishing, competitive intelligence, etc.). However, the customisation of their tools for a client application is always done independently from one another, each having its own constraints.

Actually, Mondeca builds the domain ontology, if it doesn't pre-exist already, according to the client needs and its existing data, whilst Temis develops specific linguistic resources for each application domain, reusing existing resources when possible (as the named entities recognition tool). That's why the linguistic tags of the conceptual tree produced by IDE™ have different names from the concepts defined in the ontology even if they are describing the same subject. As a consequence, we must find a way to map one with another in order to be able to instantiate the right concepts from the linguistic extractions.

ITM/IDE Integration in the Semantic Portal

The integration between the linguistic extractions from IDE™ and the ontological concepts of the domain defined in ITM™ must be achieved according to the following steps: 1) parsing the conceptual tree resulting from the linguistic analysis; 2) defining the acquisition rules between linguistic tags and ontological concepts; 3) automatic processing of these rules on a corpus of documents (cf. Figure 1).

This process is applied to each of our client projects. We will illustrate this implementation through an example taken from one of our client projects. It is about the domain of legal publishing: the author of legal articles must be informed of every legal text and court decisions. Thus, for every published document, a reference is recorded in the knowledge base with all it properties as well as the other textual references cited.

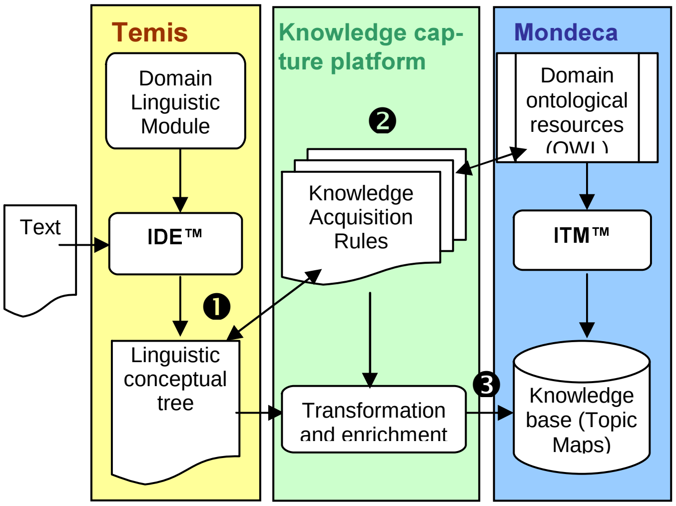

Figure 1. The process of Ontology population.

The corpus used in our example is only composed of report of legal decision issued by cassation court about divorces or employment contracts. The reports, cf. Figure 2, are divided in two parts: firstly a semi-structured header representing the information linked to this decision (date, supreme Court of appeal, decision number, appeal number, etc.) and then the unstructured document body describing the implied parties, the motifs, the argumentation with the references to the legal coded texts (called « TC », e.g. « common Law ») and non coded texts (called « TNC », e.g. « Decree of the 30th of September 1953 »).

CIV. 1 D.S

COUR DE CASSATION

Audience publique du **23 mars 2004** Cassation partielle

M. BOUSCHARAIN, président Arrêt n° 510 F-D

Arrêt n° 510 F-D

Pourvoi n° F 02-19.839

(…) **REPUBLIQUE FRANCAISE**

AU NOM DU PEUPLE FRANCAIS

LA COUR DE CASSATION, PREMIÈRE CHAMBRE CIVILE, a rendu l'arrêt suivant :

Sur le pourvoi formé par Mme H, épouse Y, demeurant xxxxx, 75019 Paris, (…)

Sur le rapport de Mme G-L, conseiller référendaire, les observations de Me B H, avocat de Mme H, de la SCP V, avocat de la société P, les conclusions de Mme P, avocat général, et après en avoir délibéré conformément à la loi ;

Sur le moyen unique, pris en sa seconde branche :

Vu l'article L. 311-37 du Code de la consommation, dans sa rédaction antérieure à la loi n°2001-1168 du 11 décembre 2001 ; (…)

Figure 2. Extract from a cassation court legal decision report.

Conceptual tree resulting from the linguistic analysis

As stated above, the IDE™ produces a conceptual tree from each linguistic analysis of a legal decision report (cf. Figure 3). At each node of this tree corresponds a linguistic tag and its textual value taken from the legal document, recalled between parenthesis. Our solution must parse this valued tagged tree in order to extract the pertinent information and to map it with an existing concept of the domain ontology, which can be a topic, an attribute, an association or a role in the knowledge base. To do this, we model the knowledge acquisitions rules that will produce the creation of an instance of an ontological concept at each corresponding node of the conceptual tree.

```
/REFERENCE DECISION(cassation 10400510)
   /FORMATION(CIV . 1)
      /Chambre civile(CIV . 1)
   /JURIDICTION(COUR DE CASSATION)
   /DATE SEANCE(Audience publique du 23 mars 2004)
      /DATE(23 mars 2004)
         /MonthDayNumber(23)
         /month(mars)
         /YearNumber(2004)
   /Noms-de-personnes(M. BOUSCHARAIN , président)
      Nom(M. BOUSCHARAIN)
      role(président)
         /Role/Juridique(président)
   /DECISION/ARRET/ARRET DIFFUSE(Arrêt n° 510 F-D)
         num(510 F-D)
   /POURVOI(Pourvoi n° F 02-19.839)
      num(F 02-19.839)
   …
   /REFERENCE(article L. 311-37 du Code de la
consommation)
      ref(article L. 311-37 du Code de la consom-
mation)
         /ARTICLE unique(article L. 311-37)
            art num(L. 311-37)
```

Figure 3. Extract from a conceptual tree dealing with a legal decision.

The tree parsing is ruled by some basic principles:

- A tree has necessarily a root, representing here the document or the main subject of the document (in our example, the decision itself).
- The tree parsing is a top-down parsing by a prefixed order : starting from the root, the algorithm parses first the left child before parsing the right child et so on recursively.
- Two parsing are indeed necessary: the first one to get the topics with their attributes and the second one to get the associations with the different roles played by the topics.

These two parsing are essential because not every topic necessarily play a role in an association. Therefore, they wouldn't be instantiated if the parsing of the tree was only considering the associations, then their roles and finally the corresponding topics. In our example, it is especially the case of the topics « Person » having attributes such as « Name » and « Role » whereas not participating to any association as modelled in the client ontology.

In order to process the conceptual tree, we choose, in a first step, to implement the Knowledge acquisition rules in the Xpath language[1]. Indeed, this language allows to parse a tree (XML document, conceptual tree, etc.), to directly reach any of its nodes and from any node to select any of its ancestors, descendants or siblings.

Definition of the Knowledge Acquisition Rules

Table 1. Examples of mappings between semantic tags and concepts.

Name of the linguistic tag	Name of the concept in the ontology	Type in the kb	Context
/nom lex	Person	Topic	
/noms lex	Person	Topic	
/COURT MEMBERS	Legal personality	Topic	∃ Descendant = /Juridique
	Political personality	Topic	∃ Descendant = /Politique
/REFERENCE	Ref Editorial Legislative TNC	Topic	∃! Child = /article
	Ref Editorial Legislative TNC Article	Topic	∃ Child = /article
	Simple Reference	Association	∃ Parent = /REFERENCE DECISION
	Targeted link	Role	∃ Parent = /REFERENCE DECISION
/art num	Num Article	Attribute	
/MOTIF			
	Originated link	Role	

Each node from the conceptual tree must be manually mapped with a concept of the domain ontology, whatever is type is (topic, attribute, association and role)[2]. To do this, we define knowledge acquisition rules that will set off the creation of an instance of the ontological concept at each corresponding node of the conceptual tree. **Table 1** sums up the various possible cases:

- A linguistic tag can be mapped with only one concept: « /art num » with the attribute « Num Article ».
- Many linguistic tags can be mapped to the same concept: « /Nom lex » and « /Noms lex » with the topic « Person ».
- A linguistic tag can be mapped with several concepts of the same type: « / COURT MEMBERS » with the topics « Legal personality » and « Political personality ».
- A linguistic tag can be mapped with several concepts of different types: « /REFERENCE » with the topics « Ref Editorial Legislative TNC » and « Ref Editorial Legislative TNC Article », with the association « Simple reference » and with the role « Targeted link ».
- A linguistic tag can't be mapped with any ontology concept: « /MOTIF ».
- A concept can't be mapped with any linguistic tags: the role « Originated Link ».

When a linguistic tag may instantiate more than one concept, the context of the ancestor, descendant or sibling nodes can help in solving the ambiguities. For instance, if the node « /REFERENCE » has a child node « /ARTICLE », the topic « Ref Editorial Legislative TNC Article » will be instantiated, otherwise it will be the topic « Ref Editorial Legislative TNC ».

The first part of a report, and therefore the linguistic extractions, deals with the cassation court decision. It contains all the attributes of the topic representing this decision, i.e. "Ref Editorial Case Law" marked by the tag « /REFERENCE DECISION ». It is then possible to put in relation each of the nodes in this first part with the corresponding attributes, such as the tag « /FORMATION » with the attribute « formation » in Figure 4.

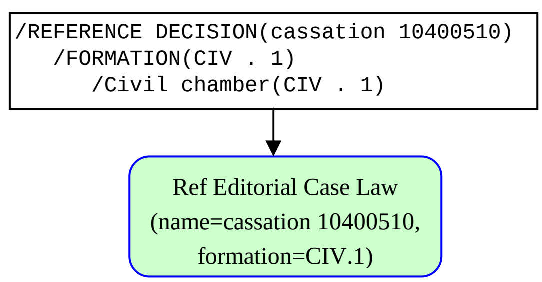

Figure 4. Linguistic extraction of the topic "Ref Editorial Case Law".

The second part of the document deals with other types of concept instances, such as the persons, being the different parties or the legal personalities (lawyers, presidents, counsellors, etc.), and the references to the legal documents on which is based the argumentation of the different parties. Those references will be instantiated according to their concept corresponding to a coded text or not, with their attributes (date, type of text, etc.) then related with the decision through the association named « Simple Reference » and the specification of their role , i.e. « Targeted link », cf. Figure 5.

1 Site web du W3C : http://www.w3.org/TR/xpath

2 The vocabulary used here is the one of the Topic Maps.

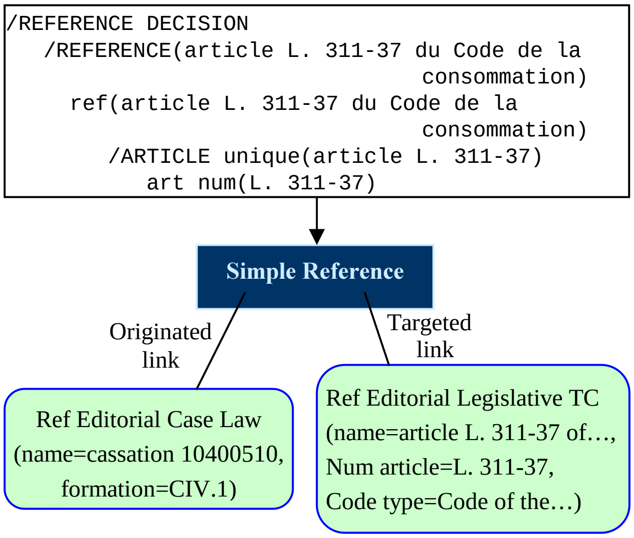

Figure 5. Linguistic extraction modelled as an Association of type « Simple Reference ».

Once the mapping is done, each acquisition rule will be formalised using Xpath and added in the domain ontology as an attribute of the concept it will instantiate. For example, the concept "Ref Editorial Legislative TC Article" will have the following knowledge acquisition rule :

"/REFERENCE DECISION/REFERENCE/ref[ARTICLE and TEXTE]".

Processing an Acquisition Rule

After linguistic analysis, the conceptual tree of the document selected by the user is automatically parsed by the entire set of the acquisition rules. At each pertinent node, the corresponding instantiation action of the knowledge base associated with every acquisition rule is processed. However, in order to avoid the multi-creation of the same instance in the knowledge base, a check is done before the concept creation to verify its existence in the knowledge base. Once the tree parsing terminated, the used can visualised every new instances added to the knowledge base through a validation interface. From this interface, the user can modify and/or delete any created instance, and add another ones. Thanks to this interface, the final user can control the quality of the underlying knowledge base.

EXPERIMENTATIONS AND RESULTS

Our corpus of experimentation is composed of 36 reports from supreme Courts of appeal. On these 36 documents, given by the legal publishing company, 4 only were used to manually define the acquisition rules. The other 32 documents were used as the test corpus. After reception of the extraction patterns produced and compiled by the linguists from Temis, we processed the whole test corpus and get for each document its corresponding conceptual tree. We compared the linguistic tags with each detected concept and we noticed which ones were correctly created, incorrectly created or even no created at all in the knowledge base.

In order to evaluate quantitatively the results of this process, we used the precision and rappel measures, defined to measure either information retrieval results (cf. TREC conferences), or information extraction results (cf. MUC conferences). In our case, we applied those measures to the tagged linguistic extractions with regards to the instantiated concepts in the knowledge base. Hence, we obtained the two following adapted measures:

- **Precision** measures the number of instances correctly acquired by the number of instances acquired
- **Rappel** measures the number of instances correctly acquired by the number of instances existing in the conceptual tree

Following the analysis of the 32 documents of the test corpus, and from the same acquisition rules defined previously, the following table presents the results for the entire set of concepts present in the linguistic extraction corpus. A set of 1765 concepts of the ontology categorised in topics, attributes (or occurrences) of these topics, associations and roles have been detected in the conceptual trees of the test corpus. Among those concepts, 975 have been correctly instantiated, by the rules, 257 incorrectly instantiated and lastly 533 not instantiated. We are thus obtaining the following rappel of 0,55 and a precision of 0,79.

Table 2. Experimentation results on the 32 documents of the test corpus.

Concept type	Number of concepts in the tree (A)	Number correctly instantiated (B)	Number incorrectly instantiated (C)	Number not instantiated (D)
Topics	585	432	139	14
Attributes	798	329	0	469
Associations	80	69	0	11
Roles	302	145	118	39
Total	**1765**	**975**	**257**	**533**

Table 3. Rappel and precision measures.

Concept type	**Rappel (B/A)**	**Precision (B/B+C)**
Topics	**0,74**	**0,76**
Attributes	**0,41**	**1**
Associations	**0,93**	**1**
Roles	**0,48**	**0,55**
Total	**0,55**	**0,79**

To sum up, even if the precision rate is satisfying for a first experimentation, we notice that an important number of textual units, correctly tagged in the conceptual tree, are not instantiated afterwards, especially the attributes and the associations. Other concepts are not correctly instantiated, mainly the topics. This is due to a redundancy issue coming from conflicting rules. This problem produces other incorrect instantiations of the roles with the non respect of the constraints modelled in the ontology, especially the cardinalities, inducing for a single association many roles of the same type instead of only one.

We are also noticing that it is necessary to introduce more complexity in the context of the tree between the tags generated by the linguistic extraction. For the moment, our acquisition rules are limited to the constaints on the child, parent or sibling nodes. Yet, the ancestor context is particularly important for the creation of topics' attributes. Let's take as an example the tag « /num » : if the direct parent node is « /ARTICLE », the attribute will be instantiated as an article number whereas if this same node is « /APPEAL », the attribute will correspond to an appeal number. The context of the descendant nodes can also bring more exactness wrt the creation of a topic or an association. In Figure 6, the tag « /Names-of-persons » informs that the node deals with the class « Person » in the ontology. Therefore, this class has two sub-classes : « Legal personality » and « Political personality ». An analysis of the descendants of the node « /Names-of-persons », and mainly the presence of one or another nodes « Legal » or « Political », can set the right concept to instantiate.

Figure 6. Contextual analysis example.

```
/Names-of-persons(M. BOUSCHARAIN , president)
    Name(M. BOUSCHARAIN)
    role(president)
       /Role/Legal(president)
```

CONCLUSION AND DISCUSSION

This platform provides an innovative solution for ontology population from linguistic extractions thanks to the definition of acquisition rules. To our knowledge, there is no similar approach in the Semantic Web framework. Of course, other systems (Kyriakov, 2003) are interested with ontology population thanks to linguistic tools but their ontologies are modelled in concordance with the results of their linguistic extractions, at a higher level and without complex semantic relations (n-ary). At the contrary, our approach allows to populate a given ontology from any linguistic tool, once this one extracts the pertinent informationof the domain as a conceptual tree. Taken the issues raised during the first implementation of the system, we are defining the following priorities for our future research works:

- Improvement of the two conceptual tree parsing in order to manage more complexity in the rules thanks to a richer contextualisation.
- Detection of the conflicts caused by recovery problems between rules.
- Checking the respect of the cardinalities, especially for the roles of an association.

The achievement of some of these priorities would rapidly improve the actual system performance, mainly with regards to associations and roles. There is still the problem of coherence and maintenance between the knowledge acquisition rules that might become more and more numerous according to the size of the domain ontology to populate. The manual definition of all the acquisition rules is itself heavy and error-prone. And if the linguistic resources or if the client ontology are modified, then all those rules must be verified and updated by the administrator of these rules.

That's why we propose to develop a formal language to describe the knowledge needed to populate an ontology from a conceptual tree. This language will be inspired from LangText (Crispino, 2003), developed to modelled the linguistic knowledge in the framework of the contextual exploration (Desclès, 1991), (Minel & al., 2001). One of the advantage of this language is the declarative way to formalise the notions of search space, of indicator and of annotation of a textual unit (word, phrase, sentence, paragraph,…). Actually, it is necessary to adapt this language to a conceptual tree parsing and not a textual one. This adapted language will allow a better knowledge maintenance, a greater efficiency in the definition of the concepts to instantiate by the potential conflicts management, and thus a productivity gain for the user. This language is still under development at this time.

Lastly, we would like to emphasize the fact that this system must stay generic enough to being able to define and apply the acquisition rules to any application domain. Moreover the purpose of the knowledge acquisition rules to transform a linguistic tag into an instantiated concept of the ontology.

REFERENCES

[1] Gennari, J., et al. Templates for K-CAP, available at <http://www.kcap05.org>

[2] Panther, J. G., *Digital Communications*, 3rd ed., Addison-Wesley, San Francisco, CA (1999).

[3] Schneider, M., "Guidelines for Bias-Free Writing," PhD Thesis, Indiana University, Bloomington, IN (1995).

AMARDEILH F. & FRANCART T. (2004). A Semantic Web Portal with HLT Capabilities, In *Actes du colloque « Veille Stratégique Scientifique et Technologique »* (VSST2004), Toulouse, (Vol.2), p 481-492.

BONTCHEVA K. & CUNNINGHAM H. (2003). The Semantic Web: A New Opportunity and Challenge for Human Language Technology, in *Proceedings of the Second International Semantic Web Conference*, Workshop on Human Language Technology for the Semantic Web and Web Services, Florida, 20-23 October 2003, p. 89-96.

BOURRIGAULT D., AUSSENAC-GILLES N. et CHARLET J. (2004). Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas, Revue d'Intelligence Artificielle, 18(4), 24 pp.

CRISPINO G. (2003). Une plate-forme informatique de l'Exploration Contextuelle : modélisation, architecture et réalisation (ContextO). Application au filtrage sémantique de textes, Thèse sous la direction de Jean-Pierre Desclés, Université Paris-Sorbonne, Décembre 2003, 241pp.

DESCLÈS J.-P., JOUIS C., OH H-G., MAIRE REPPERT D. (1991). Exploration Contextuelle et sémantique : un système expert qui trouve les valeurs sémantiques des temps de l'indicatif dans un texte, Knowledge modeling and expertise transfer, Amsterdam, p. 371-400.

GRISHAM R. & SUNDHEIM B. (1996). Message understanding conference - 6: A brief history. In *Proceedings of the 16th International Conference on Computational Linguistics*, Copenhagen, p.466-471.

GRIVEL L., GUILLEMIN-LANNE S., LAUTIER C. et al. (2001). La construction de composants de connaissance pour l'extraction et le filtrage de l'information sur les réseaux, In *Filtrage et résumé automatique de l'information sur les réseaux*, 3ème congrès du Chapitre français de l'ISKO International Society for Knowledge Organization, Paris, 5-6 July 2001, 9 pp.

GRUBER T. (1995). A Translation approach to portable ontology specifications. In Knowledge Acquisition, 5(2), p. 199-220.

HANDSCHUH S., STAAB S., CIRAVEGNA F. (2002). S-CREAM – Semi-automatic CREAtion of Metadata, In *Proceedings of the 13th International Conference on Knowledge Engineering and Management* (EKAW 2002), Espagne, 1-4 Octobre 2002, Springer Verlag, p. 358-372.

HENDLER J., HORROCKS I. and al. (2004). OWL web ontology language reference, W3C Recommendation.

KAHAN J., KOIVUNEN M., PRUD'HOMMEAUX E. et al. (2001). Annotea: An Open RDF Infrastructure for Shared Web Annotations, In *Proceedings of the WWW10 International Conference*, Hong Kong, Mai 2001, p. 623-632.

KIRYAKOV A., POPOV B., OGNYANOFF D., MANOV D., KIRILOV A., GORANOV M. (2003). Semantic Annotation, Indexing, and Retrieval, In *Proceedings of the 2nd International Semantic Web Conference* (ISWC2003), Florida, 20-23 Octobre 2003, p.484-499.

LAUBLET P., REYNAUD C. et CHARLET J. (2002). Sur quelques aspects du Web Sémantique, In *Assises du GDR I3*, Eds Cépadues, Nancy, 20 pp.

MINEL J.-L., J-P. DESCLÉS, E. CARTIER, G. CRISPINO, S. BEN HAZEZ, A. JACKIEWICZ. (2001). Résumé automatique par filtrage sémantique d'informations dans des textes. Présentation de la plate-forme FilText, Revue Technique et Science informatiques, 20(3), Hermès, p. 369-395.

ORA L. & SWICK R. (1999). Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation.

PARK J. & HUNTING S. (2003). XTM Topic Maps : Creating and using Topic Maps for the Web, Addison Wesley Eds, Boston, p. 81-101.

VARGAS-VERA M., MOTTA E., DOMINGUE J. (2002). MnM : Ontology Driven Tool for Semantic Markup, In *Proceedings of the Workshop Semantic Authoring, Annotation & Knowledge Markup* (SAAKM 2002), Lyon (France), 22-23 Juillet 2002, p. 43-47.