International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:3, No:4, 2009

Knowledge Acquisition for the Construction of an Evolving Ontology: Application to Augmented Surgery

Nora Taleb, Sellami Mokhtar, and Michel Simonet

Abstract—This work concerns the evolution and the maintenance of an ontological resource in relation with the evolution of the corpus of texts from which it had been built.
The knowledge forming a text corpus, especially in dynamic domains, is in continuous evolution. When a change in the corpus occurs, the domain ontology must evolve accordingly. Most methods manage ontology evolution independently from the corpus from which it is built; in addition, they treat evolution just as a process of knowledge addition, not considering other knowledge changes. We propose a methodology for managing an evolving ontology from a text corpus that evolves over time, while preserving the consistency and the persistence of this ontology.
Our methodology is based on the changes made on the corpus to reflect the evolution of the considered domain - augmented surgery in our case. In this context, the results of text mining techniques, as well as the ARCHONTE method slightly modified, are used to support the evolution process.

Keywords—Corpus, Evolution, Ontology

I. Introduction

IN various domains, and especially those which are specialized, such as in industry and medicine, pieces of knowledge that represent the heritage of the organization (hospital, ...) are numerous, distributed over several sources, and dynamic in continuous evolution.

The first issue is to characterize the transformation of knowledge expressed in many forms (text, experts ...), into a knowledge formalized with the ontology model.

The second issue is to manage evolution in the knowledge domain (specifically the corpus of texts), during the life cycle of the model which represents it (ontology in our case).

Many studies [1, 2, 3, 4, 5] aimed at answering the first question. On the other hand, concerning the second issue, several studies highlight the major importance of ontology evolution, and the lack of approaches to manage it [10, 11, 16].

The main objective of our work is to propose general methodologies, with detailed steps, to manage the ontology evolution based on the knowledge contained in a corpus of texts.
This methodology integrates text mining techniques, and is based on the ARCHONTE method proposed by Bruno Bachimont [18], with a slight modification to maintain the steps of ontological evolution.

Our scope is augmented surgery, and more specifically computer-assisted medical gestures, which aim at assisting the doctor and surgeon in order to perform the least invasive, and most precise diagnostic or therapeutic gestures, with the objective of improving patients care.

This domain is constantly evolving and the modelling of the knowledge and the practices in this domain is inevitably confronted with the paradox that exists between permanence and evolution, between stability and adaptability.
The work takes place in the TIMC-IMAG laboratory[1] , in the context of a work in progress to define Quality in Augmented Surgery [17]. The figure 1 shows the overall structure of an Information System designed to capture the notions of Expected Medical Service (in French SMA: Service Mdical Attendu) and Delivered Medical Service (In French: SMR: Service Mdical Rendu).

The adopted method for our work is an ascending experimental approach, which starts from the concrete encountered problems and moves towards the resolution of underlying scientific questions.
According to this approach, we will first identify the needs of doctors in the domain of augmented surgery, in terms of knowledge representation, then we will retrieve the available ontological structure that we will use as an initial ontology. This ontology evolve following the changes or modifications in the texts corpus related to the ontology.

II. STATE OF THE ART

In this section we present the methods and tools used in the global evolution process.

A. ARCHONTE

Actually, few methodologies propose to guide a knowledge engineer to structure a knowledge domain.

N. Taleb is with the Department of Computing, Annaba univesity, Algeria,23000 e-mail: nora.taleb@imag.fr
M. Sellami is with the Department of Computing, Annaba univesity, Algeria,23000 e-mail: sellami@lri-annaba.org
M. Simonet is with the Laboratoire TIMC, UJF University, Grenoble,38000,France e-mail:michel.simon@imag.fr

[1]University UJF, Medical engineering and complexity techniques, Grenoble

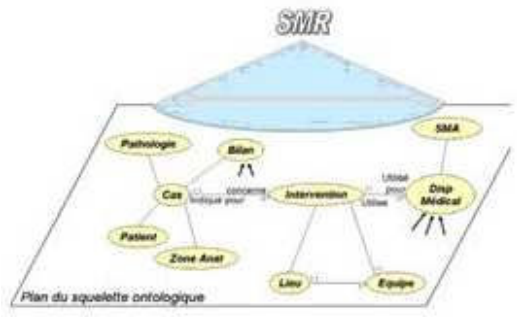

Fig. 1. Overall ontological structure

The ARCHONTE methodology defines precise directives to clarify the concepts using the language; it helps describing the variations in meaning of the considered terms in a context (see figure 2).

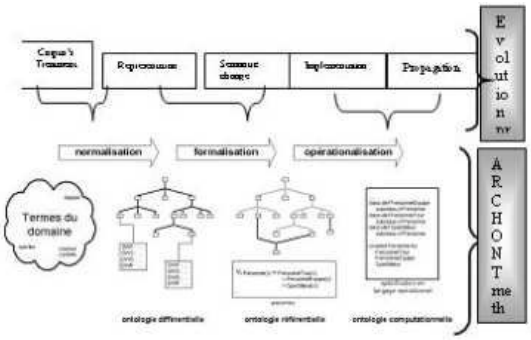

Fig. 2. Combination of evolution process and ARCHONTE Methodology

B. LSTAT

The LSTAT method is a hybrid approach designed in the LRI laboratory[2] . It is used for the automatic construction of an ontology from a text corpus. It combines the principles of linguistic and statistical approaches, in order to avoid proposing to the expert the validation of the terms, which are not identified and not relevant to the domain, by keeping the frequency or the occurrence number.
This approach also aims at minimizing the noise in the list by avoiding proposing terms which belong to certain filters (linguistic, grammatical, punctuation,). For more details see [15].

C. ONTODOM

ONTODOM is a software tool which implements the principles of LSTAT. Its input is a corpus of texts, and its output is a list of concepts and relationships between them. It uses PROTEGE 2000 for a direct representation of the ontology [15].

[2]Laboratoire de Recherche en Informatique- Annaba University-Algeria

D. Text mining techniques

There are essentially three types of text mining techniques:

- Learning lexico-syntactic patterns;
- Techniques for clustering ;
- Statistical techniques and association rules;

1) Learning lexico-syntactic patterns: Lexico-syntactic patterns are based on the study of syntactic patterns between the data associated with two concepts. This is an observation of the relationship realization in the corpus, in order to simplify the vocabulary and syntax. This mapping is a lexico-syntactic pattern. The advantage of this technique is that it is targeted on the lexico-syntactic context; it remains effective in small corpus size. In [19], the experimental evaluation of a large number of patterns was made using the CAMELEON. The results showed that the learning of lexico-syntactic patterns generates a significant number of errors due to the dependence of the corpus.

2) Clustering techniques: There are many classification approaches. The classification approach proposed by [20] consists in classifying documents in collections according to the meaning of each word. It uses a labelled corpus such as WorldNet. For each of the collections the words and their respective frequencies are extracted and compared to other collections. The approach proposed by [21], shows the construction of a domain ontology from text documents by using two algorithms for hierarchical classification. Within this framework the authors propose several algorithms; among them we quote the algorithms for neural networks, decision trees and hierarchical classification algorithms (HCA).

3) Statistical techniques and associations rules: These techniques are based on the calculation of some similarity measures. The association rules describe associations between certain elements. It is a balanced involvement (implication) of the form:
$A \rightarrow B$,
where A = $\{t_1, t_2, ..., t_p\}$ and
B=$\{t_{p+1}, t_{p+2}, .., t_q\}$
The rule $A \rightarrow B$ is interpreted as: all texts containing the terms $\{t_1, t_2, , t_p\}$ also tend to contain the terms $\{t_{p+1}, t_{p+2}, ..., t_q\}$, with some probability given by the confidence of the rule. Several algorithms are available to implement the extracting rules process, e.g., Close and Pascal [35].

The support and confidence are two measures associated with association rules. They are used in order to reduce the number of extracted rules.
The support is given by the number of texts containing the key terms of A and B. While the confidence of a rule is the ratio between the number of texts containing A B and the number of text containing A. when the confidence value is

1, the rule is called correct, otherwise it is approximate. The measures of support and confidence cannot always identify the rules which make sense. Therefore, other measures are also used, such as interest, belief, dependence, novelty and satisfaction [36].
In our case study, the measure of confidence is sufficient to confirm the deletion of a fragment of text.

E. TEXTTOONTO

The methodologies developed in the literature have contributed to the development of several tools. Their goal is to help users to build ontologies. Among the most experienced tools dedicated to the development we quote Kaon [29], Ontoview [30], Ontomanager [31] and TextToOnto [32].

The study of the main tools highlights the lack of features to ensure ontology evolution. These tools do not support the steps of identification and analysis of changes.

TextToOnto is a tool suite developed to support the ontology engineering process by text mining techniques. The usage of the algorithms varies from interactive (the system only makes suggestions) to fully automatic. TextToOnto builds upon KAON. It just supports the addition of texts, but deletion and update are not supported by this tool. The working corpus must be in English or in Spanish, but the French language is not dealt with.

III. ONTOLOGICAL EVOLUTION FROM A CORPUS OF TEXTS

The evolution is a change that takes into account the coherence and the persistence of the ontology. It has not to keep the ontology versions produced overe time [3].
In our study, the evolution is seen with two axes: corpus evolution and ontology evolution.

A. Corpus Evolution

We note that our text corpus is in French. The corpus may be represented by the following vector: $T = (t_1, t_2, ..., t_n)$ where t_i are terms from the corpus, with the following conditions:

- The meaningless terms are removed (le, la, les, du,..).
- All uppercase become lowercase.
- All conjugations are converted to infinitive.
- Synonyms and words with the same root are considered equivalent.

We are in a domain where the knowledge is expressed in natural language. The studied corpus is characterized by its continuous evolution over time. These changes can include: adding a piece of text, deleting a piece of text, change or modification of a text fragment.
Adding a new text According to the studied domain, new knowledge can intervene; in the case of augmented surgery, adding an option in the knee equipment automatically leads to the addition of a text fragment which explains this option in the manual.
Deleting of a text fragment The knowledge domain can evolve because of some irrelevance pieces of knowledge have to be deleted, which leads us to remove the fragment of text representing them.
Corpus update The update can be described as a deletion and an addition of a fragment of text. In our study case it can intervene during the change of a feature an option in the equipment.

B. Ontological evolution

Our methodology do not concern direct changes on the ontology;it deals with ontology evolution when the corpus is updated. Our objective is to present a direct method, which takes into account corpus updates . It ensures the evolution of an ontology by using text processing and text mining tools [7, 8, 12, 13, 14].
The evolution of the ontology consists of five steps [33]:

- Representation of changes.
- Semantic changes.
- Implementation.
- Propagation of changes to other ontologies in relation with the current ontology.
- Validation.

1) Representation of changes: The aim of this step is the edition of elementary or complex changes. An elementary change is not decomposable; examples of elementary changes are the addition, deletion or modification of ontological entities. A complex change consists of several elementary changes; that form together a single logical entity, such as the merging or the separation of ontological entities.

2) Semantic changes: The ontology has to evolve from a consistent state towards another consistent state, i.e., the state where the constraints of the ontological model are respected. In order to resolve the inconsistencies introduced by changes, other additional changes may be necessary.
This step has to deal with the resolution of all the additional changes in a systematic way.

3) Implementation: It consists in implementing the changes once approved by users.

4) Propagation of changes: The aim of this step is to automatically modify the instances in dependent ontologies. The objective is to ensure their consistency with the evolved ontology. In our study we call it internal consistency (IC) and External Consistency (EC). For the formalization of the ontology, we considerd [25, 26].
Considering that the ontology is composed of two semiotics levels:

- The lexical level (L) includes all the terms or the labels indicating the concepts and the relationships.
- The structural level defines the structure (S) of the ontology, it contains the concepts and the semantic relationships built from the conceptual relationship between them.

The structure of the ontology is represented by the tuple: $S := \{C, R, <, X\}$

where:
C, R: are separated sets containing the concepts and the taxonomic relationships .
$<$: CXC is a partial order on C, it defines the concept hierarchy.
X : R $\rightarrow$ CXC is the signature of an associative or taxonomic relationship
The lexicon of the ontology is a tuple
L:$\{L_c, L_r, F, G\}$ whereL_c, l_r are disjoint sets. They contain the labels with the frequencies associated with the concepts and the relationships. F, G: are two references relationships. They provide the access to the terms that are associated with the concepts or the relationships respectively.
We note that a concept can be defined by several terms.
The hierarchy of concepts is defined by a structure:
$S_0 := C, <$
A concept is defined by:
$L_0 := L_c, F.$
The objective is to maintain the consistency of S_0.
Operationally, we use the PROTG tool [24] for the implementation of the ontology. This software integrates the representation language OWL [16], which represents the two following interest:. firstly it implements a Description Logic , and secondly it exploits a tag language that can be used to semantically annotate the text from the ontology for Information Retrieval purposes.

IV. OUR METHODOLOGY FOR EVOLUTION MANAGEMENT

It is essential to have a methodological approach for evolution management. This allows to formally specifyi the changes needed in order to automate them in the application, while ensuring the consistency and the quality of the evolved ontology.

The idea is to categorize the changes in order to formally define their meaning, their scope and their potential involvement. The formal specification of the changes has a more important meaning than a simple formalization of them in the ontology representation language. It prepares the analysis phases and the resolution of the effects of changes. The figure 3 represents the steps used in our methodology.

In our case study, the ontology evolution is the result of the environment evolution of the studied domain (augmented surgery), and more specifically the included knowledge in the corpus which represents the domain.
The proposed methodology is organized according to the following phases

A. Pretreatment step

The choice of the corpus is an essential factor for the ontology evolution process. In our approach, the initial corpus is a collection of texts in French, built from books and users manuals.

For any change in the corpus, the domain expert has to verify its consistency (no redundancy, no contradiction,).

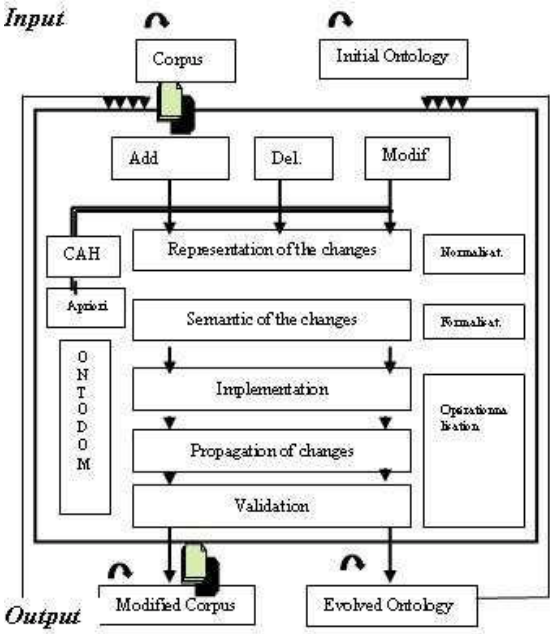

Fig. 3. General scheme for the evolution methodology

B. The evolution Step

It consists in studying the ontology evolution. The MOTO module (ontology processing module) is associated with this step to implement and analyze the ontology changes that are the consequences of the corpus changes. This step can be examined according to the three types of corpus changes detailed in the previous section.

1) Removing of a text fragment: In this case of evolution , we will apply some text mining techniques, such as calculation of confidence and the APRIORI algorithm in order to update the ontology while preserving its coherence.
Let T be the initial corpus: $T = \{t_1, , tn\}$ where t_i: the corpus terms
T_2 is the text to be removed;
We work on the association rule $T \rightarrow T_2$: which means: the text which contains the terms of T, also contains the terms of T_2.
We propose the following APRIORI algorithm to manage the ontological evolution.

2) Addition of a text fragment: In this case, text mining techniques provide an algorithm for hierarchical classification, which is most effective for our case study providing a light modification to fit our needs. The result of the algorithm is the classification of the concepts and the relationships in the ontology.

Before the presentation of the HCA algorithm,we must define the following notions:

- Concept
  a concept Ci is a vector such that

Algorithm 1 Algorithm APRIORI

BEGIN
1: Input
2: $T_c = \{t_1, t_2, , t_k\}$;
{The set of extracted words from The initial corpus with their frequencies}
3: $P_c\{P_1, P_2, , P_p\}$;
{The set of extracted transaction from the initial corpus (a transaction is a Sentence)}
4: $T_s = \{t_1, t_2, , t_{kk}\}$;
{The set of The extracted transactions from the text to be removed (kk¡=k)}
5: $P_s = \{p_1, p_2, , p_{kp}\}$;
{The set of extracted transaction from the texts to be removed (kp¡=p)}
6: Output
{calculate the confidence}
7: $C = \frac{(T_c \cap T_s)}{T_s}$.
if C=1 **then**
the total ontology will be removed and we keep only the initial ontology.
else
for i=1,kk **do**
$C_i = \frac{\{t_i / t_i \in T_c\}}{\{t_i / t_i \in T_s\}}$
if Ci=1 **then**
remove the identifier t_i from the identifiers list
else
$i = i + 1$ {go to the next ti}
end if
end for
13: Repeat until $i = k$
end if
14: Cross the ontology in the search for a concept that has no identifier to remove.
END

$C_i = (T_i, Att_1, Att_2, , Att_j, P_1, , P_k)$
where
T_i: the terms describing the concept;
P_i: the properties (relationships);
Att_i: the attributes.

- Similarity measure

The classification of concepts is based on a similarity measure, which determines if the concepts are independent or equivalent.Its calculation is based on the concept vector (concepts, attributes, and elationships with neighbours).

The calculation of this measure in our case study is comparing way to compare the vectors of concepts (term-term-attribute, attribute).

It is defined as follows:

$Sim(C_i, C_j) = \sum_{i=1} Max(simt(A_{ik}, A_{j1}), , sim(A_{ik}, A_{jm}))$, m=1,,Cj

We note that the vectors of concepts are automatically extracted from the corpus with:

$\sum_{i=1}^{k} Ci = Tn$
The grouping of all the vectors of concepts provides the vector of the added text.

Algorithm 2 Algorithm HCA

BEGIN
1: Input
2: $T = \{t_1, , t_k\}$;{the initial corpus}
3: $Tn = \{t_1, , t_n\}$; {the added text}
4: $O = (C, R, <, K)$; {the global ontology}
5: $C = \{C_1, .., C_n\}$ {all concepts of O with their vectors}
6: $MatriSim[n+m, n+m]$:matrix of similarity with n number of concepts of the new text , m number of existing concepts in the ontology O}
7: S {similarity threshold}
8: Output
9: the classes Symi of synonym concepts ;
for $i = 1, n+m$ **do**
Matrsim[i,i]=;
for $i = 1, n+m$ **do**
for $j = 1, n+m$ **do**
if $matsim[i,j] <>$ **then**
$matrsim[i,j] = X$
end if
end for
end for
end for
END
18: $max = 0$;
19: Repeat;
for $i = 1, n+m$ **do**
for $j = 1, n+m$ **do**
if $matrsim[i,j] > max$ **then**
$max := matrsim[I,j]$;
end if
end for
end for
if $max > seuil$ **then**
$mi := mi - 1 \cup C_i, C_j - C_i, C_j$;
$matrsim[I,j] := ;$
end if
31: Update Matrsim by taking into account the new class;
32: Make the inference by using the new relation;
END

When adding a new text T_2 to the initial corpus, $T = \{t_1, t_2, , t_k\}$, the ONTODOM tool runs in order to automatically extract the first concepts according to the LSTAT approach , and built the vector of terms of T_2, $T2 = \{t_1, , t_n\}$ and the list of vectors of concepts C, at this level, the algorithm of hierarchical classification HCA runs to classify the concepts of the set C in the ontology O, based on a measure of similarity between concepts. The most similar concepts are grouped together in the same class.

The algorithm continues to iterate until it completes the classification of all the concepts of C.

V. Conclusion

In this article, we have provided a state of the art of the tools and methods which are used in our approach, the ARCHONTE method, text mining techniques , the LSTAT approach and its tool ONTODOM.

The initial idea was to use tools which could work on texts.

The first attempt was to examine TEXTTOONTO to integrate it in our prototype of evolution management. The absence of French-language dictionary having an exploitable format by TEXTOONTO, led us to abandon this direction.

The presented methodology is based on principles stated by B. Bachimont and particularly its ARCHONTE method with a light modification. It is presented within the framework of a new approach for the incremental management of ontological evolution from a corpus of texts.

Our contribution is the integration of text mining techniques, by using algorithms easily implementable, in order to limit the intervention of domain experts as maximum as possible, and also to provide to linguists and ontology engineers a detailed processfor ontological evolution.

We are currently working on the integration of the evolution notion in the ONTODOM tool. An experiment of our methodology in the domain of augmented surgery is in progress.

References

[1] H. Assadi, *Construction ďontologies à partir de textes techniques application aux systèmes documentaires* , Ph.D , Paris University, 1998
[2] H.Assadi,D.Bourigault, *Analyses syntaxique et statistique pour la construction ďontologies à partir de textes*, Seyroles, 2000.
[3] B.Audrey, J.Charlet, *Evaluation, évolution et maintenance ďune ontologie en médecine: état des lieux et expérimentation*, 2004.
[4] D.Bourigault, N.Aussenac, J.Charlet, *Construction de ressources terminologiques ou ontologiques à partir de textes*, un cadre unificateur pour trois tudes de cas. Revue dintelligence Artificielle (RIA),87-110, 2004
[5] J.Charlet, *Ontologie pour le web smantique*, action spcifique, 32 CNRS, 2003.
[6] F.Laalam , M.Sellami, *Réalisation et modélisation ďun systme ďaide au diagnostic de pannes*, Journée décole doctorale JED2007, Annaba,Algérie, 2007.
[7] A.Maedche, B.Motic, L.Stojanovic, *Managing Multiple and distributed ontologies in the semantic web*, VLBD journal,2003
[8] A.Maedche, B.Motic , L.Stojanovic, *Ontologies for enterprise knowledge management*, Intelligent information processing, 2003.
[9] F.Natalya, L. Deborah, *Developpement ďune ontologie 101: guide pour la création de votre première ontologie*, article de recherche CA, 2002
[10] M. Fernndez-Lpez, *onto web: a survey on methodologies for developing, maintaining, evaluating and reengineering ontologies*, IST Project IST-2000-29243, 2000
[11] H. Peter, *Ontology management and evolution*; SEKT 2004.
[12] L.Stojanovic, A.Maedche, *User driven ontology evolution management*, 13 th international conference on knowledge management, EKAW 02, Spain, 2002
[13] L.Stojanovic, B.Motic, *Ontology evolution within ontology editors*, international conference on knowledge management, EKAW 02, Spain, 2002
[14] L.Stojanovic, N.Stojanovic, *Usage oriented evolution of ontology based knowledge systems*, international conference, DOA, USA, California, 2002
[15] N.Taleb, M, Sellami, *Approche hybride pour lacquisition des connaissances*, international symposium for programming system, Algiers, 2003
[16] Webont, *Owl web ontology use cases and requirements*, http://www.w3.org/tr/2004/rec-owl-test-20040210
[17] JJ. Banihachemi, A. Moreau-Gaudry, A. Simonet, M. Simonet. *Vers une structuration du domaine des connaissances de la chirurgie augmentéepar une approche ontologique*, Journe francophone sur les ontologies, JFO, Lyon 2008
[18] B. Bachimont , *Engagement sémantique et engagement ontologique: conception et réalisation ďontologies en ingénierie des connaissances*; In "Ingnierie des connaissances Evolutions rcentes et nouveaux dfis", Jean Charlet, Manuel Zacklad, Gilles Kassel, Didier Bourigault; Eyrolles 2000, ISBN 2-212-09110-9, 2000
[19] B.Zghal, M.Aufaure, N.Ben Mustapha *Extraction of ontologies from web pages: conceptual modelling and tourism*, Journal of internet technologies, 2007
[20] E.Agirre., O. Ansa ., E.Hovy ., *enriching very large ontologies using the www*, actes de latelier sur la construction dontologies de la confrence europeenne de lintelligence artificielle, ECAI,2000
[21] L.Khan, F.Luo ., *Ontology construction for information selection*, Proceeding of 14 th IEEE international, 2002.
[22] S.Najla , J.Wassim , *Types de changements et leurs effets sur lévolution de lontologie*, in proceeding JFO, Tunis,2007
[23] A.Maedche et S.Staab, *Mining ontologies from text*,Proceeding of the 12th International Conference on Knowledge Engineering and Knowledge Management, Springer lecture notes in Artificial Intelligence,2000
[24] Protege2000, *téléchargement et guide ďutilisation*, http://protege.stanford.edu/
[25] M.Chagnoux, N.Hernandez, N.Aussenac, *From texts to ontologies: non taxonomical relation extraction*, article de recherche, 2008.
[26] N.Aussenac-Gilles, S.Despres and S.Szulman,*The terminae method and platform for ontology engineering from texts*, IOS press, 2008
[27] P.Cimiano, *Ontology learning and population from text. Algorithms, evaluation and applications*, Springer, Berlin, 2007.
[28] E., Marshman., *Towards strategies for processing relationships between multiples relation participants in knowledge patterns: an analysis in English and French*, In J.benjamins editor, Terminology, volume 13.1, pages 1-34, 2007.
[29] FZI Karlsruhe and AIFB Karlsruhe, *kaon the karlsuhe and semantic web*, Framwork-developers guide for Kaon 1.2.7, 2004
[30] M. Klein , *Versioning of distributed ontologies*, EU/IST, Project Wonder Web, 2002
[31] L.Stojanovic, A.Maedche, *Managing Multiple and distributed Ontologies in the Semantic web*, VLDB journal, Special issue on semantic web, 12, 286-302, 2003
[32] A.Maedche, R. Volz, *the ontology extraction and maintenance framework TextToOnto*, in proceeding ICDM01, workshop on Integrating data mining and knowledge management, 2001.
[33] D.Codruta- Rogozan, *méthodes et outils pour un référencement sémantique évolutif fondé sur une analyse de changements entre versions dontologie*, Projet de recherche Dic 9410, 2005.
[34] G.Galas, *Etudes des principaux algorithmes de data mining*, Article de recherche cole ingnieurs en informatique EPITF, France, 2006
[35] Y.Bastide, R.Taouil, N.pasquier: *un algorithme ďextraction des motifs fréquents*. Techniques et science informatique, 21(1): 65-95, 2002
[36] H.Cherfi, *Etude et réalisation ďun système ďextraction de connaissances à partir de textes*, thèse ďuniversité Henri Poincaré (Nancy 1), 2004