

A Metaontology for Domain Ontology Enriching in an Information Retrieval System

Nesrine Ben Mustapha, Rania Soussi, Hajer Baazaoui, Marie-Aude Afaure

► **To cite this version:**

Nesrine Ben Mustapha, Rania Soussi, Hajer Baazaoui, Marie-Aude Afaure. A Metaontology for Domain Ontology Enriching in an Information Retrieval System. 2èmes journées francophones sur les ontologies, Dec 2008, lyon, France. pp.63-72, 2008. <hal-00831726>

HAL Id: hal-00831726

<https://hal-ecp.archives-ouvertes.fr/hal-00831726>

Submitted on 11 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Metaontology for Domain Ontology Enriching in an Information Retrieval System

Nesrine Ben
Mustapha

Laboratoire RIADI ENSI
Campus Universitaire de
la Manouba 2010
nesrine.benmustapha
@riadi.rnu.tn

Rania Soussi

Laboratoire RIADI ENSI
Campus Universitaire de
la Manouba 2010
soussi_rania@yahoo.fr

Hajer Baazaoui
Zghal

Laboratoire RIADI ENSI
Campus Universitaire de
la Manouba 2010
hajer.baazaouizghal
@riadi.rnu.tn

Marie-Aude Aufaure

Ecole Centrale Paris
Laboratoire MAS Chaire
SAP Business Objects
Grande Voie des Vignes
92 295 Chatenay-
Malabrymarie-
aude.aufaure@inria.fr

ABSTRACT

With the continual increase of the volume of available information on the web, information access and knowledge management become challenging. Thus, adding a semantic dimension to the Web, by the deployment of ontologies, contributes to solve many problems. In the context of the semantic Web, ontologies improve the exploitation of Web resources by adding a consensual field of knowledge. The need for using domain ontology for information retrieval (IR) has been explored by some approaches to better answer users' queries. However, ontology in IR system requires a regular updating, especially the addition of new concepts and relationships. In fact, IR systems are generally based on few number of domain ontology that cannot be extended. This paper proposes an incremental approach for domain ontology learning. This approach is designed to be integrated into IR system based on ontology. It is based on an ontological representation called "Metaontology." The approach presented was tested and evaluated on a tourism ontology built from results of an online IR system based on ontology.

Categories and Subject Descriptors

I.2.6 [Computing Methodologies]: Artificial Intelligence – Learning – *Concept learning, Knowledge acquisition.*

General Terms

Design, Theory.

Keywords

Ontology learning, Metaontology, text mining, information retrieval.

1. INTRODUCTION

Aujourd'hui, le Web est utilisé pour chercher des informations via les moteurs de recherche grâce à l'indexation des pages, des procédures d'extraction et d'analyse d'informations. L'ajout d'une dimension sémantique aux données du Web permet d'améliorer cette recherche d'information. D'où l'apparition de la notion du

Web sémantique [5] qui n'est pas un Web indépendant du Web actuel mais le prolongement du Web que l'on connaît et où on attribue à l'information une signification clairement définie, ce qui permet aux ordinateurs et aux humains de travailler en étroite collaboration [1]. La représentation explicite du sens se fait à travers les ontologies qui sont une spécification explicite, formelle d'une conceptualisation partagée [11]. Les ontologies ont contribué à l'apparition des moteurs de recherche sémantiques. Parmi ceux-ci, nous citons les moteurs de recherche contextuels qui se basent sur un nombre déterminé d'ontologie de domaine. La recherche dans ces moteurs est limitée à un domaine bien déterminé. L'objectif du présent travail est de bâtir un moteur de recherche multi-contextuelle où il serait possible de construire de nouvelles ontologies et de les enrichir. Etant donné que la construction manuelle d'une ontologie est une tâche longue et fastidieuse, plusieurs auteurs ont proposé des approches basées sur l'apprentissage d'ontologies pour améliorer l'automatisation de ce processus. Dans le présent travail, nous nous intéressons principalement à l'apprentissage d'ontologies à partir de textes dans un objectif de l'intégrer dans les systèmes de recherche en ligne à base d'ontologies de domaine. Différentes techniques de fouille de textes sont utilisées. Les connaissances relatives à ces techniques sont capitalisées dans une représentation ontologique dite « Métaontologie ». Cet article est organisé de la façon suivante: la section 2 propose un état de l'art des différentes techniques de fouille de textes qui ont été utilisées pour la construction d'ontologies à partir de textes. La section 3 décrit les limites dégagées lors d'une étude approfondie de ces techniques. Les objectifs et la motivation de ce travail sont présentés dans la section suivante. La section 5 décrit notre proposition qui consiste à définir une Métaontologie et un processus incrémental d'enrichissement d'ontologie à partir des documents extraits du Web. La section 6 présente les résultats expérimentaux. Enfin, nous concluons et donnons les perspectives de ce travail.

2. APPRENTISSAGE D'ONTOLOGIES À PARTIR DE TEXTES

Les approches d'apprentissage d'ontologies s'inscrivent parmi les approches de construction semi-automatique d'ontologies en utilisant des connaissances a priori. Elles consistent à enrichir une petite ontologie dite « minimale » ou « granulaire » avec de nouveaux concepts et de nouvelles relations en utilisant des techniques de fouille de textes. L'apprentissage d'ontologies à partir de textes a été largement employé par la communauté de l'ingénierie de connaissances. Ces approches sont généralement

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

JFO 2008, December 1-2, 2008, Lyon, France.

Copyright 2008 ACM 978-1-60558-373-0/08/0003.\$5.00.

basées sur l'utilisation d'un corpus textuel et l'application des techniques de fouille de textes. Ce corpus doit être représentatif du domaine pour lequel on essaye de bâtir l'ontologie. Nous avons remarqué que ces approches partagent généralement les étapes suivantes :

- La préparation des sources de connaissances qui peuvent être soit un corpus textuel, soit une collection de documents web.
- L'utilisation des connaissances a priori (une ontologie de haut niveau d'abstraction, un thésaurus, un lexique ou bien une ontologie minimale construite manuellement en vue de l'enrichir).
- Le prétraitement de ces sources de connaissances afin de les soumettre à des méthodes d'extraction d'informations.
- L'apprentissage des concepts en se basant sur des techniques linguistiques, statistiques et / ou de clustering.
- L'extraction des relations.
- L'extraction des instances des concepts de l'ontologie
- La validation ou l'intégration de l'ontologie, généralement effectuée par les experts.

Nous nous intéressons dans ce qui suit à présenter les différentes techniques d'apprentissage des concepts et des relations qui ont été utilisées par la plupart des approches d'apprentissage d'ontologies. En effet, à l'aide d'un ensemble de techniques, on tente de projeter dans l'ontologie les connaissances contenues dans les textes en extrayant des concepts et des relations. Nous distinguons principalement trois types de techniques de fouille de textes:

- L'apprentissage des patrons lexico-syntaxiques;
- Les techniques de clustering et/ou de classification ;
- Les techniques statistiques et des règles d'association.

2.1 Apprentissage de patrons lexico-syntaxiques

Les patrons lexico-syntaxiques se basent sur l'étude des régularités syntaxiques entre deux concepts donnés. En effet, il s'agit d'une observation de la réalisation d'une relation dans le corpus¹ afin d'en schématiser le contexte lexical et syntaxique. Cette schématisation constitue un patron lexico-syntaxique et permet d'extraire des couples de mots vérifiant cette relation à partir du corpus. Cette technique a été développée notamment dans [12]. Des patrons lexico-syntaxiques liés à l'hyponymie ont été découverts pour mettre en relation des couples père-fils potentiellement intéressants pour la construction d'ontologies. L'avantage de cette technique est que, étant ciblée sur le contexte lexico-syntaxique, elle reste efficace sur des corpus de petite taille. Toutefois, les régularités linguistiques caractérisent généralement le même type de connaissances, tels que les relations sémantiques. Dans [5], l'évaluation expérimentale d'un grand nombre de patrons a été faite en utilisant l'outil Caméléon. Les résultats ont permis de remarquer que l'efficacité des patrons et leurs significations dépendent du corpus. C'est pour cette raison que l'apprentissage des patrons lexico-syntaxiques engendre un degré important d'erreurs. En effet, les régularités syntaxiques

¹ Par exemple sur la base de l'extraction de passages mettant en jeu deux mots liés par la relation sélectionnée.

relatives aux relations d'hyponymie qui ont été découvertes ne reflètent pas des relations pertinentes dans l'ontologie. Ainsi, une combinaison de cette technique avec les signatures contextuelles ou sémantiques (« Topic Signature») a été proposée par [1] afin d'atténuer les erreurs de la première technique. Cette technique a été utilisée aussi bien pour la découverte des concepts de domaine que des relations (taxonomiques). En effet, nous illustrons un patron lexico-syntaxique identifié par Hearst dans la langue anglaise par l'exemple suivant:

Nom such as Nom1 {Nom2..., (and/ or) Nomi }

2.2 Techniques de clustering et de classification

L'idée de combiner les techniques linguistiques et celles de clustering a été adoptée par l'approche ASIUM [13] en vue d'apprendre des cadres de catégorisation en s'appuyant sur une technique de clustering conceptuel (regroupement de concepts). L'approche de classification proposée par [2] consiste à classer les documents en collections relativement au sens de chaque mot et en utilisant un corpus étiqueté ainsi que Wordnet. Puis, pour chacune des collections formées, les mots et leurs fréquences respectives sont extraits et comparés aux autres collections. Enfin, des signatures sémantiques sont construites en se basant sur une fonction χ^2 en vue de détecter les concepts partageant les mêmes contextes. Enfin, nous citons l'approche proposée dans [14] qui propose la construction d'ontologie du domaine à partir des documents textuels en utilisant deux algorithmes de classification hiérarchique. A partir de ces approches, nous identifions les techniques suivantes :

- La construction de signature contextuelle ;
- Le clustering conceptuel.

2.2.1 Construction de la signature contextuelle

Cette technique est basée sur les principes de la Sémantique Distributive qui admet que "La signification d'un mot est fortement corrélée aux contextes dans lesquels il apparaît". Cette hypothèse peut être généralisée pour couvrir les termes composés au lieu des termes simples. Les contextes peuvent être formalisés sous la forme de vecteurs de mots du contexte, comme dans le cas de signature sémantique de sujet décrit dans [17]. En utilisant les signatures sémantiques du sujet, chaque concept serait représenté en fonction de l'ensemble des mots qui sont en cooccurrence avec lui, et des fréquences de ceux avec lesquels ils apparaissent. Dans ce cadre, plusieurs métriques de similarité, tel que TFIDF [21] ou Chi Square, peuvent être utilisées pour mesurer la distance entre les différents concepts. Un algorithme de classification descendante pour étendre des ontologies existantes avec les nouveaux concepts est décrit dans [1]. En effet, la qualité de la construction des signatures contextuelles (« topics signatures ») décrite dans [2] peut être améliorée en incluant seulement les contextes qui ont en quelque sorte des relations syntaxiques avec les concepts dans l'ontologie. Par exemple, il est possible de considérer seulement la liste des verbes pour lesquels les concepts sont des sujets ou un complément d'objet direct, ou de considérer seulement les adjectifs qui modifient le concept.

De même, des informations contenues dans une ressource linguistique existante comme Wordnet (les synonymes des concepts, les hyperonymes, les antonymes, etc.) sont utilisées pour construire des requêtes qui servent à la récupération des documents pertinents relatifs à un sens d'un concept donné. Les

documents liés au même sens de ce concept sont groupés ensemble pour former des collections dont chacune est relative à un sens d'un concept donné. Puis, les documents dans chaque collection sont traités pour en extraire les mots et leurs fréquences en utilisant une approche statistique. Ces derniers sont comparés avec les données qui couvrent d'autres sens pour le même concept dans les autres collections. Les mots qui ont une fréquence distinctive pour une des collections sont groupés dans une liste. Celle-ci constitue ainsi pour chaque sens de concept, la signature contextuelle ("Topic signature") généralement utilisée dans la construction de résumés de textes [17]. En troisième lieu, pour un mot donné, les concepts associés à son sens de mots sont hiérarchiquement groupés. Pour ce faire, des différentes signatures sont comparées pour découvrir des mots partagés et pour déterminer des chevauchements entre les signatures. Des métriques calculant la distance sémantique peuvent être utilisées à cette fin. Les signatures contextuelles ont été évaluées par leur application dans la tâche de désambiguïsation de sens des mots. Ces premières contiennent des informations considérablement utiles pour cette tâche. Toutefois, l'évaluation de cette méthode en utilisant Wordnet ne suffit pas pour conclure sur son efficacité dans le cas de la construction d'une ontologie spécifique à un domaine.

2.2.2 Clustering conceptuel

Le regroupement conceptuel a recours à un traitement syntaxique des documents à partir desquels on estime construire une ontologie. Des classes sont formées à partir des termes qui apparaissent après le même verbe et la même préposition en appliquant un algorithme de regroupement conceptuel (clustering conceptuel). La difficulté réside dans la labellisation des relations après leur découverte afin de désigner la relation sémantique en question. Pour résoudre ce problème, deux méthodes de clustering "Asium-Best" et "Asium-Level" qui se basent sur l'extraction des filtres syntaxiques, ont été proposées dans l'approche ASIUM [13]. Ces techniques permettent l'apprentissage des relations non taxonomiques entre deux classes de termes. Les relations sont labellisées en fonction du verbe et de la préposition concernée par le filtre syntaxique. Un filtre syntaxique est illustré par cet exemple : <To travel><subject: human><by: vehicule>. En premier lieu, l'analyseur syntaxique fournit automatiquement des expressions nominales associées aux verbes et aux clauses. Par exemple, à partir des filtres syntaxiques suivants, des classes synthétiques sont créées :

- <Voyager> (< sujet : Jean>) (< en : voiture>);
- <Voyager> (< sujet : David>) (< en : train>);
- <Conduire> (< sujet : Helene>) (< objet : voiture>);
- <Conduire> (< sujet : Roland>)(< objet : avion>).

Les classes sont successivement agrégées pour former de nouveaux concepts et une hiérarchie constituant l'ontologie. Les classes formées sont labellisées par un expert pour identifier les concepts qu'elles représentent. Les classes sont composées de groupements de mots suivant le patron: <verbe> <rôle syntaxique |préposition : nom>, comme par exemple « <voyager> <sujet : humain> <par : véhicule>. Les couples <rôle syntaxique : nom> ou <préposition : nom> sont appelés « mots têtes ». La mesure de similarité qui permet d'évaluer la distance entre les classes, et donc de les regrouper, dépend de la proportion d'entêtes communs des expressions nominales relatives aux différentes classes en

prenant en compte leur fréquence d'apparition dans les documents. La méthode a été testée sur un corpus de recettes de cuisine. Lorsque le système est expérimenté à retrouver les couples verbe-argument sur 30 % du corpus, la hiérarchie proposée est valide à 30%.

2.3 Techniques statistiques et les règles d'association

Dans ce cadre, nous citons principalement l'approche implémentée dans l'environnement DOODLE II [25] qui est une extension de DOODLE (A Domain Ontology Rapid Development Environment) [22] et l'approche de [18]. Ces dernières ont recours à des techniques utilisant un dictionnaire de mots vides pour l'extraction des termes pour éliminer les mots ayant un contenu informationnel vide. Plusieurs conclusions contradictoires ont été tirées sur l'efficacité comparative des différentes mesures de sélection. Les résultats présentés dans [14] pour l'extraction des termes du domaine de l'économie à partir de documents du Wall Street Journal ont montré que la fréquence des termes donne de meilleurs résultats que la mesure Tf.Idf. Toutefois, les résultats dépendent toujours de la spécificité du corpus de référence. Plusieurs techniques statistiques sont décrites dans la littérature permettant d'extraire des relations taxonomiques entre termes. Ces premières se basent sur l'analyse des cooccurrences entre termes dans les documents. La cooccurrence correspond à l'apparition simultanée de deux termes dans un texte (document ou fenêtre de N mots). L'ensemble des cooccurrences est représenté par une matrice. Celle-ci est ensuite utilisée pour le regroupement hiérarchique des termes par l'application des méthodes de classification automatique [17] et le regroupement basé sur des mesures de probabilités [24].

Selon l'approche DOODLE II [24], l'apprentissage des relations non taxonomiques est basé sur un algorithme de recherche des règles d'association et la construction de l'espace des mots pour l'extraction des paires de concepts similaires. La construction d'un espace de mot comprend les étapes suivantes:

- L'extraction des 4-gram les plus fréquents en vue de construire une matrice de cooccurrence utile pour représenter le contexte d'un texte;
- La construction de la matrice de collocation est faite dans le but de comparer le contexte de deux 4-grams;
- La construction des vecteurs de contextes, des vecteurs de mots et de vecteurs de représentation des concepts dont l'ensemble constitue l'espace de mots.

Afin de déceler des relations entre termes, les termes apparaissant dans une fenêtre de quatre mots autour des termes centraux sont extraits. Les termes co-occurents fréquemment sont proposés pour être reliés dans l'ontologie, les verbes apparaissant dans le contexte sont proposés pour être les labels de la relation. Ces termes peuvent être déterminés à partir d'une matrice de cooccurrence dans une fenêtre de n mots ou à partir d'un espace de mot construit [21]. L'extraction des relations non taxonomiques entre concepts à partir des règles d'association a été proposée dans [18].

3. SYNTHÈSE ET DISCUSSION

L'état de l'art présenté dans la section précédente nous a permis de dégager les limites suivantes :

• **L'absence d'une évaluation exhaustive des approches et des outils d'ingénierie ontologique:** En effet, chacune des approches est mise en œuvre en appliquant des techniques qui permettent l'enrichissement de l'ontologie avec de nouveaux concepts et de nouvelles relations à partir de textes. Ces techniques sont ensuite implémentées dans un outil. Dans ces travaux, nous n'avons pas trouvé une étude comparative des techniques utilisées pour en déduire les meilleures. Ceci est expliqué par le fait qu'il doit y avoir l'expérimentation de ces approches pour un même corpus relatif à un même domaine et écrit dans une langue déterminée et qu'on doit se référer à des techniques d'évaluation.

• **L'absence d'un cadre méthodologique permettant le guidage des ontologues :** au cours de la construction d'ontologies à partir de textes. En effet, les règles d'extraction n'évoluent pas au cours du processus d'apprentissage d'ontologies. Seules les règles définies dans l'approche proposée par des ontologues sont implémentées dans l'outil qui supporte cette approche. Une fois que les techniques implémentées présentent des erreurs relatives à leur application sur certains corpus, il n'est plus possible de maintenir l'outil sans le réimplémenter pour prendre en compte de ces erreurs. Ceci, est dû au fait que la liaison entre les règles d'extraction d'ontologies à partir de textes et les éléments appris de l'ontologie à construire n'a pas été formalisée de manière explicite. La fouille dans ces liaisons pourrait corriger énormément le processus d'apprentissage.

En effet, ces limites impliquent que l'intervention humaine est de plus en plus requise pour la correction de l'ontologie construite à partir de textes.

L'objectif principal de ce travail est de proposer une solution à ces limites pour faciliter la construction d'ontologie au sein des systèmes de recherche en ligne. Nous détaillons nos motivations et nos objectifs dans la section suivante.

4. MOTIVATIONS ET OBJECTIFS

L'objectif principal de ce travail est de proposer une approche d'apprentissage d'ontologies à partir du Web permettant d'améliorer le processus d'apprentissage tenant compte des limites décrites précédemment. Ainsi, notre proposition devrait :

- Expliciter les règles d'extraction des concepts, relations et axiomes d'ontologies afin de pouvoir maintenir ces règles en fonction des expériences antérieures d'ingénierie ontologique.
- Faciliter l'évaluation de chacune des techniques utilisées en gardant une liaison entre les éléments appris de l'ontologie et les techniques utilisées.
- Faciliter l'apprentissage des axiomes qui sont porteurs de connaissances

Pour cette raison, nous associons l'ensemble des techniques utilisées par ces approches à une boîte à outils utile pour accomplir les phases communes d'un processus d'ingénierie ontologique. Toutefois, bien que chaque technique tente d'améliorer l'apprentissage des éléments ontologiques, le degré d'erreur varie d'une technique à une autre. Ainsi, une capitalisation des résultats disponibles est requise.

Après cet état de l'art des différentes techniques de fouille de textes appliquées pour rendre la construction d'ontologies à partir de textes semi-automatique, nous proposons une métaontologie pour la formalisation des règles d'extraction des concepts et des

relations d'ontologies. Dans ce cadre, nous avons spécifié les connaissances communes entre les ontologues par une grande ontologie de haut niveau d'abstraction que nous avons appelée « Métaontologie ».

Notre principale motivation est de pouvoir intégrer cette approche dans les systèmes de recherche en ligne à base d'ontologies. Ceci, justifiera nos choix lors de la proposition d'une approche incrémentale de construction d'ontologie à partir des pages Web. Celle-ci est fondée sur cette « Métaontologie » qui fera l'objet de la section suivante.

5. UNE METAONTOLOGIE POUR L'ENRICHISSEMENT D'ONTOLOGIE DE DOMAINE

Le terme "Métaontologie" est utilisé dans le domaine de l'ingénierie ontologique comme étant un type d'ontologie mais certains la confondent avec la notion de "l'ontologie générique" ou "noyau ontologique" qui est réutilisable dans différents domaines [26]. Toutefois, une métaontologie est un type d'ontologie qui est conceptuellement à un niveau d'abstraction supérieur par rapport aux autres ontologies et qui a donné lieu à la proposition de nouveaux langages de représentation (OWL, RDF, OIL, etc.). Toutefois, ces métaontologies décrivent les métamodèles d'ontologies et sont dépourvues d'axiomes qui permettent de découvrir ou apprendre de nouveaux concepts. Dans la plupart des travaux, elles ont été utilisées au cours de l'étape de conceptualisation ou lors de la validation et l'évaluation d'ontologies [10]. Notre travail consiste à définir une métaontologie pour l'utiliser au cours de l'enrichissement d'ontologie de domaine à partir des pages Web. Dans notre travail, la métaontologie est assimilée à une ontologie dont le domaine est l'apprentissage d'ontologies. Les concepts et les relations de celle-ci (appelés respectivement métaconcepts et métarelations) servent à former non seulement le métamodèle conceptuel de toute ontologie mais aussi les patrons, les mesures et les éléments textuels qui ont servi pour la découverte des éléments de l'ontologie à construire. Celle-ci comprend également des axiomes ayant pour rôle l'inférence de nouveaux concepts ou de nouvelles relations. Ces axiomes se basent sur les mesures et les patrons spécifiés dans la métaontologie pour inférer de nouvelles instances qui sont des concepts ou de relations de l'ontologie à construire. D'où, seulement une partie de la métaontologie est instancié pour obtenir une ontologie. Les autres informations relatives à la découverte de ses éléments reste disponible dans la métaontologie.

Dans la section suivante, nous détaillerons la spécification de cette métaontologie.

5.1 Spécification de la Métaontologie

Dans le but de faciliter la maintenance des ontologies d'une manière incrémentale, nous proposons de formaliser les règles de découverte de nouveaux concepts et de nouvelles relations par une base de connaissances représentée comme étant une métaontologie. Nous alimentons cette base par les patrons linguistiques, les mesures de similarité entre les concepts calculées selon plusieurs métriques, les termes référents les concepts de domaine, les concepts appris (validés ou rejetés), les relations apprises (validées ou rejetées). Nous gardons le lien entre les concepts appris avec la ou les techniques utilisées et la raison pour laquelle les concepts ou les relations ont été rejetés.

La spécification des connaissances servant à l'apprentissage d'ontologie de domaine à partir du contenu textuel a pour but de formaliser les règles d'apprentissage des constituants d'ontologies de domaine à partir de textes. Dans ce cadre, nous distinguons quatre principaux concepts autour desquels sont construits les autres concepts et les autres relations, à savoir: le concept nommé "concept_de_domaine", le concept nommé "relation_de_domaine", le concept nommé "instance_de_domaine" et le concept nommé "axiome_de_domaine". Nous avons commencé par la construction du modèle conceptuel en spécifiant les métaconcepts, les métarelations, et les axiomes servant à l'extraction des termes qui référencent les concepts de domaine et à la découverte des relations taxonomiques et non taxonomiques à partir du contenu textuel des documents Web.

5.1.1 Connaissances pour l'extraction des concepts à partir de textes

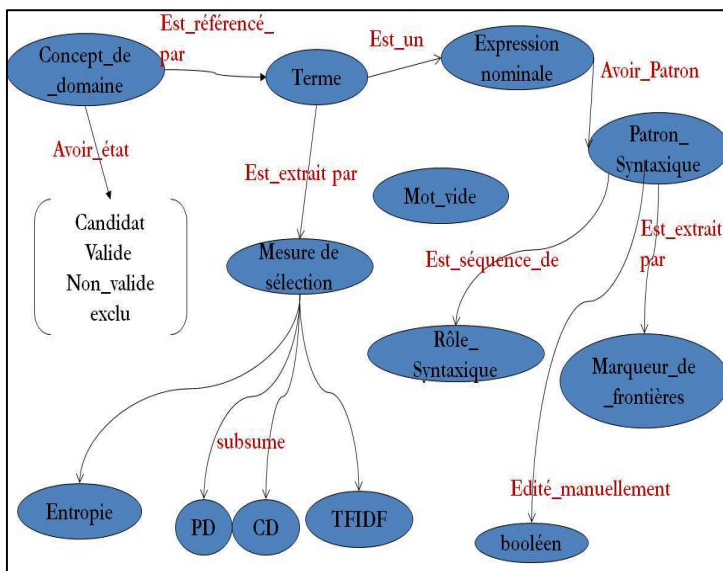


Figure 1. Formalisation des connaissances pour l'extraction des concepts à partir de textes.

Tout concept de l'ontologie de domaine est référencé par un ensemble de termes dont un seul est choisi pour labelliser le concept. Nous avons choisi de combiner toutes les techniques décrites dans la section 2. En effet, les techniques de fouille de textes appliquées pour la découverte de nouveaux concepts et qui ont été spécifiées dans la Métaontologie sont : les méthodes de surface à base des marqueurs de frontières [8] et la sélection des termes candidats en fonction de leurs occurrences dans les documents en appliquant les mesures suivantes à savoir: l'entropie [9], le Consensus dans le domaine (CD) [27], la mesure d'information mutuelle (IM) [16], le TFIDF [19]. Pour chaque technique utilisée, nous avons spécifié les concepts et les relations qui contribuent dans la découverte de nouveaux concepts. Par exemple, nous illustrons la formalisation d'une des techniques statistiques de sélection de termes et les patrons lexicosyntaxiques par la figure 1. Un exemple d'axiome qui permet l'apprentissage d'un concept candidat est illustré ci-dessous :

$$(\forall X,Y \text{ Concept_de_domaine}(X) \wedge \text{avoir_expression}(X,Y) \wedge \text{avoir_patron}(Y, \langle \text{adjectif_Concept} \rangle) \rightarrow \text{Concept_de_domaine}(Y) \wedge \text{hyponyme}(X,Y))$$

5.1.2 Connaissances pour l'extraction des relations à partir de textes

Dans notre métaontologie, nous spécifions les techniques suivantes relative à l'apprentissage des relations taxonomiques et non taxonomiques: les techniques de définition et d'apprentissage des patrons lexico-syntaxiques d'hyponymie [12], le clustering conceptuel et le regroupement hiérarchique en appliquant les mesures issues des travaux sur la recherche d'information [13] à savoir: l'inverse de la distance entre les concepts, la profondeur des concepts dans la hiérarchie, la densité, le contenu en information [18] et la mesure de similarité [19] et l'edgcounting [20]. Nous admettons qu'il est possible de classer les relations taxonomiques suivant les patrons lexico-syntaxiques responsables de leur identification. D'autres classes de relations sont à découvrir par l'apprentissage de nouveaux patrons. L'hyponymie, comme toute relation, est caractérisée par une probabilité d'occurrence qui est calculée sur la base de la fréquence d'apparition du patron dans le corpus et de la fréquence de l'application de ce patron associé à une relation taxonomique bien déterminée existante dans l'ontologie à enrichir. Chaque patron est caractérisé par la fréquence de son existence dans le corpus, indépendamment des concepts de l'ontologie de domaine. Selon notre conception de la métaontologie, la spécification du calcul des probabilités est à la charge des ingénieurs d'ontologies qui décideront du choix du seuil de probabilité à partir duquel, un patron, un concept ou une relation est valide, selon l'application des techniques dans plusieurs corpus textuels et selon le type du corpus utilisé. Cette Métaontologie joue un rôle fondamental dans l'approche de construction incrémentale que nous avons proposée [5]. Nous décrivons brièvement ce processus dans la section suivante.

5.2 UNE APPROCHE INCREMENTALE DE CONSTRUCTION D'ONTOLOGIE DE DOMAINE

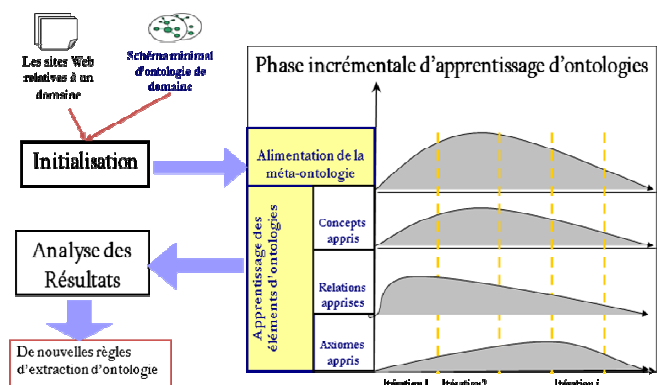


Figure 2. Processus incrémental d'enrichissement de l'ontologie de domaine.

Nous distinguons dans le présent processus trois phases (cf. figure 2) à savoir: une phase d'initialisation, une phase incrémentale d'apprentissage d'ontologie de domaine et finalement, une phase d'analyse des résultats.

La phase d'initialisation revient à la récupération et au prétraitement des sources de données qui sont formées d'une ontologie minimale, une métaontologie, l'ontologie générale Wordnet et un ensemble de documents issus du Web et relatifs au domaine. La deuxième phase est caractérisée par son aspect incrémental et itératif. Chaque itération est formée de deux étapes successives dont la première est une étape d'alimentation de la métaontologie et la deuxième consiste à appliquer les axiomes relatifs à l'apprentissage des éléments de l'ontologie de domaine. Quant à la dernière phase, elle sert à vérifier la cohérence de la métaontologie en faisant l'analyse des résultats d'apprentissage. Ce processus sera intégré dans un système en ligne de recherche d'informations à base d'ontologies pour construire ou enrichir le ou les ontologies à partir des documents textuelles sélectionnés par l'utilisateur et les plus pertinents à sa requête. Ainsi, chaque émission d'une requête de l'utilisateur déclenche une itération du processus incrémental dont les phases feront l'objet des sections suivantes.

5.2.1 Phase d'initialisation

La phase d'initialisation est la première phase du processus incrémental proposé. Elle comprend en premier lieu la conception et la construction d'une ontologie minimale du domaine. Cette ontologie est appelée "minimale" car elle est réduite en termes de nombre de concepts et de relations. Cette approche est destinée à être intégrée dans un système en ligne de recherche à base d'ontologies. Pour ce faire, la phase d'initialisation consiste à collecter les documents Web relatifs à une requête donnée de l'utilisateur en utilisant un système de recherche en ligne SIRO [3] que nous avons conçu et développé pour cette finalité. L'utilisateur peut choisir ou définir le domaine de recherche qui sera celui de l'ontologie à enrichir. L'originalité de ce système est qu'il récupère des documents pertinents et les classe par les sous-thèmes du domaine choisis (ou services de domaine). Cette classification est réalisée à l'aide d'une ontologie appelée « ontologie de service » [5] qui sera détaillée dans la section 5.2.1.2.

5.2.1.1 Classification des pages Web selon les services ou les sous-services de domaine

Nous proposons d'associer à chaque ontologie utilisée dans le système en ligne de recherche une ontologie de service. Celle-ci sert à classer les documents Web par service de domaine. Cette ontologie de services [5] permet de spécifier:

- les services d'un domaine bien déterminé,
- les principales activités fournies ou demandées respectivement par les fournisseurs de services et les demandeurs de services.

Cette ontologie décompose chaque service en activités et chaque activité en tâches. En effet, selon le grand dictionnaire terminologique, une activité est « un ensemble des tâches élémentaires ou des travaux exécutés par un individu ou un groupe et qui conduisent à la réalisation de biens ou de services ». Le modèle conceptuel de l'ontologie est illustré par la figure 3. La relation entre l'ontologie de domaine et l'ontologie de services du domaine est établie par l'association de chacune des tâches au triplet (Concept, relation, Concept). Le but est de déterminer l'ensemble de concepts et de relations qui identifient chaque service. Soient les fonctions "ProjectionService", "ProjectionActivité" et "ProjectionTâche" définies comme suit :

- ProjectionService : Service → Concept_de_Domaine.
- ProjectionActivité : Activité → Concept_de_Domaine.

- ProjectionTâche : Tâche → Concept_de_Domaine.

Nous illustrons ces fonctions par les exemples suivants :

-Exemple 1: ProjectionService("hébergement_hôtel")={"hôtel", "région", "nom", "chambre", "type_chambre"}.

-Exemple 2: ProjectionActivité("vérification_disponibilité ")={"hôtel", "chambre", "type_chambre"}

-Exemple 3: ProjectionTâche("vérification_type_chambre")={"chambre", "type_chambre"}.

A l'aide de ces fonctions, un système de recherche d'information peut déduire les concepts qui définissent un service donné. En effet, les concepts ("hôtel", "chambre", "type_chambre") décrivent le service ("hébergement_hôtel") car

ProjectionService-1({"hôtel", "chambre", "type_chambre"}) = "vérification_type_chambre" alors que

ProjectionService-1("région ")= {"Voyage", "Visite_guidée", "hébergement_hôtel"}.

Ces fonctions sont utiles pour la classification sémantique des documents Web selon les services d'un domaine bien déterminé.

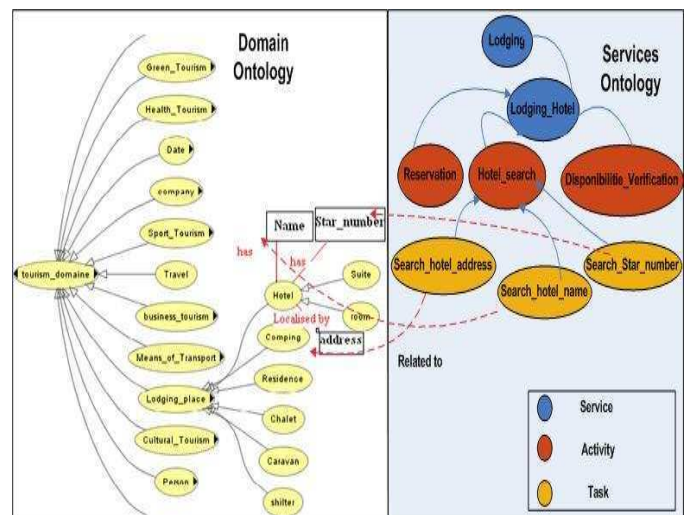


Figure 3. Relation entre l'ontologie de domaine et l'ontologie de service.

Ainsi les documents pertinents sélectionnés par l'utilisateur et classés par service de domaine forment le corpus textuel à partir duquel l'ontologie de domaine serait construite.

5.2.1.2 Prétraitement des documents Web

Cette étape revient à éliminer le balisage et les images des pages Web pour chacun des documents Web sélectionnés, segmenter et étiqueter au moyen d'un étiqueteur syntaxique pour obtenir un corpus de textes où chaque page Web est associée à un fichier textuel segmenté et étiqueté syntaxiquement.

5.2.2 Une itération d'apprentissage d'ontologie de domaine

Etant donné que l'apprentissage de l'ontologie de domaine est confronté à plusieurs problèmes énoncés dans la littérature (problème de redondance, de bruit, de désambiguïsation sémantique et d'incohérence, de choix de termes référant, d'extraction des relations taxonomiques, de construction de corpus d'apprentissage), une seule itération ne suffit pas pour

apprendre tous les éléments de l'ontologie de domaine d'où l'aspect incrémental de notre processus. En outre, un seul corpus ne couvre pas sémantiquement le domaine en question. Ainsi, de nouveaux corpus doivent être préparés pour chaque itération dans le but d'assurer un enrichissement progressif. Ainsi, chaque itération fera l'objet d'un ensemble de documents relatifs à un service ou un sous-service de domaine relativement à une requête émise au système de recherche.

Chaque itération est réalisée par l'accomplissement de deux étapes dont la première concerne l'alimentation d'une instance de la métaontologie apprenante et la deuxième est l'apprentissage des éléments de l'ontologie de domaine sur la base des axiomes définis dans l'instance de la méta-ontologie. Chacune des étapes est décrite dans les sections suivantes.

5.2.2.1 L'étape d'alimentation de la métaontologie

L'étape d'alimentation consiste à appliquer les techniques spécifiées par la Métaontologie pour instancier les métaconcepts et les métarelations. Ces techniques sont appliquées selon le processus décrit dans [5]. En effet, l'application de ces techniques de fouille de textes permet l'instanciation de la métaontologie par les expressions nominales référant les concepts (cf. table1), la signature sémantique et contextuelle de chacun des concepts qui définit le service de domaine auquel l'itération est relative, les mesures de similarité et l'espace de mots, la fréquence des patrons linguistiques dans le corpus, etc.

Le tableau ci-dessous montre la liste des expressions nominales relative au concept « hébergement » (*Lodging*). Cette liste instanciera la métaontologie,

Table 1. . Liste des expressions nominales qui réfèrent le concept « Lodging ».

Les expressions nominales	Les patrons	Les probabilités d'occurrence
dame_paris_lodging	NN_NNS_lodging_VVG	0.95454544
holidays_discount_lodging	NNS_NN_lodging_VVG	0.045454547
paris_holidays_lodging	NP_NNS_lodging_VVG	0.045454547
paris_lodging	NP_lodging_VVG	0.95454544
paris_private_lodging	NNS_adj_lodging_NN	0.90909094
paris_shortterm_lodging	NP_adj_lodging_NN	0.90909094
paris_vacation_lodging	NP_NN_lodging_VVG	0.8636364
rentals_paris_lodging	NNS_NNS_lodging_VVG	1.8636364
short_term_lodging	Adj_NN_lodging_NN	0.90909094
vacation_discount_lodging	NN_NN_lodging_VVG	0.8636364

Nous illustrons l'instanciation de la méta-ontologie par l'exemple de l'expression nominale "vacation_discount_lodging".

Cette dernière instancie les méta-concepts dont les noms sont respectivement: "expression nominale" (*nominal_proposition*) (cf. figure 4).

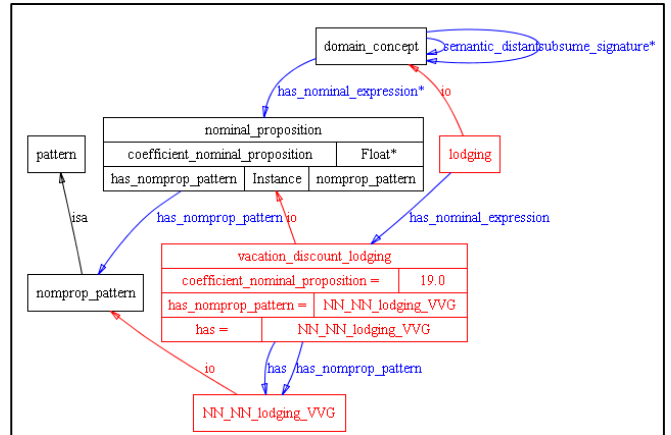


Figure 4. Illustration de l'instanciation du concept « metaconcept » de la métaontologie.

5.2.2.2 L'étape d'apprentissage d'ontologies

L'étape d'apprentissage consiste à découvrir de nouveaux concepts, de nouvelles relations et de nouveaux axiomes de domaine (cf. figure 5).

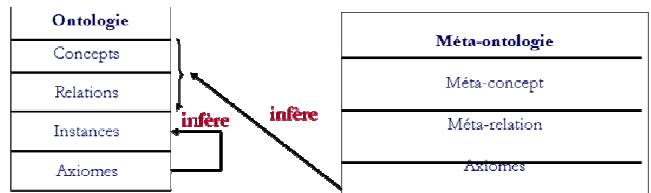


Figure 5. Apprentissage d'ontologies.

Pour ce faire, nous procédons par:

- La mise à jour du corpus textuel sur lequel nous appliquons les axiomes de la métaontologie de domaine en vue d'enrichir l'ontologie minimale de domaine et de vérifier la validité des techniques appliquées dans la première étape. Ainsi, une construction automatique d'un nouveau corpus textuel à partir d'une recherche basée sur les concepts résultant de la projection d'un service dans l'ontologie minimale.

- L'apprentissage de concepts et de relations candidats et la mise à jour de leur état par la valeur « non validé » jusqu'à la confirmation de leur validité par un axiome existante dans la méta-ontologie et ayant un fort coefficient de confiance.

- L'application des axiomes relatifs à l'extraction des concepts et des relations de l'ontologie de domaine et le changement de leur état de validité selon le degré de confiance.

- La génération des concepts et des relations dont l'état est « validé » avec lesquels l'ontologie minimale de domaine est enrichie à la fin de cette étape.

- La suppression des concepts et des relations dont l'état est « exclu » des schémas d'ontologie de domaine.

Au cours de cette étape, les axiomes spécifiés par la métaontologie jouent un rôle fondamental dans l'inférence des concepts et des relations candidats.

Par exemple, nous supposons que :

– Au cours de l'étape de l'alimentation, les expressions nominales relatives au concept « lodging » qui ont été extraites sont : "short term lodging", "lodging 's place".

les axiomes suivants permettent l'apprentissage de deux nouveaux concepts et deux nouvelles relations :

-Si l'expression nominale est conforme au patron « adjectif-concept » alors cette expression est un concept hyponyme au concept. $(\forall X, Y \text{ Concept_de_domaine}(X) \wedge \text{avoir_expression}(X, Y) \wedge \text{avoir_patron}(Y, \langle \text{adjectif_Concept} \rangle) \rightarrow \text{Concept_de_domaine}(Y) \wedge \text{hyponyme}(X, Y))$

-Si l'expression nominale admet le patron suivant « nom_'s_Concept » alors nom est un concept et une nouvelle relation « has_a » lie le concept avec le nouveau concept labellisé par « nom. $(\forall X, Y \text{ Concept_de_domaine}(X) \wedge \text{avoir_expression}(X, Y) \wedge \text{avoir_patron}(Y, \langle \text{nom_'s_Concept} \rangle) \rightarrow \text{Concept_de_domaine}(Y) \wedge \text{relation}(\langle \text{has_a} \rangle) \wedge \text{avoir_domain}(\langle \text{has_a} \rangle, X) \wedge \text{avoir_range}(\langle \text{has_a} \rangle, Y))$.

5.2.3 Phase d'analyse des résultats

La troisième phase du processus de construction d'ontologie de domaine est une phase d'analyse des résultats de l'application de la phase incrémentale d'apprentissage d'ontologie de domaine. En effet, cette phase est menée par l'ontologue. C'est la seule phase où l'intervention humaine est indispensable. En effet, l'ingénieur consulte l'ontologie de domaine générée et détecte les erreurs conceptuelles existantes. Ensuite, à partir de la métaontologie, on recense les techniques mises en jeu dans la découverte des éléments ontologiques erronés dans le but de maintenir les règles spécifiées par la métaontologie. Ainsi, il est possible de mettre au point ces règles en spécifiant l'exclusion de ces éléments erronés en affectant leurs états avec « exclu ». Dans cette optique, nous admettons que la maintenance de la métaontologie permet le réajustement des règles suivant les résultats obtenus en vue d'améliorer la construction de l'ontologie de domaine lors d'une éventuelle exécution de la deuxième phase du processus proposé. En outre la correction de la méta-ontologie engendre la génération d'une ontologie de domaine plus valide et plus riche. En effet, plusieurs nouveaux concepts appris ne sont pas générés pour être dans l'ontologie de domaine étant donné qu'ils nécessitent leur confirmation par l'intermédiaire d'une règle ou de l'intervention humaine. Ainsi, l'ingénieur peut modifier les états de concepts candidats jugés comme étant valides et provoque la génération d'un autre schéma d'ontologie de domaine. Celui-ci devient plus riche ou plus réduit selon la validation de l'ingénieur ontologue.

5.3 OntoCoSemWeb : Expérimentation et application au domaine du tourisme

5.3.1 OntoCosemWeb : outil basé sur la métaontologie pour l'enrichissement incrémental d'ontologie

L'architecture du prototype supportant l'approche proposée comprend les composants suivants (cf. figure 6):

- Un module de prétraitement des sources de données qui permet la préparation du corpus textuel et son étiquetage et l'importation des autres ressources terminologiques et conceptuelles telles que : l'ontologie minimale de domaine, l'ontologie des services de domaine et la ressource linguistique Wordnet".

- Un module d'édition de la méta-ontologie : permet de mettre à jour les concepts et les axiomes d'ontologies en intégrant le Plugin de l'outil Protége-OWL.

- Un module relatif à la génération de l'ontologie de domaine.

- Un module d'alimentation de la méta-ontologie : consiste à projeter les éléments conceptuels dans la méta-ontologie à partir de textes. Ce module implémente l'étape de l'alimentation de la métaontologie du processus proposé dans le cadre de l'approche.

- Un module d'apprentissage d'ontologies de domaine qui met en œuvre un ensemble de techniques d'apprentissage des concepts et des relations. Ce module fonctionne étroitement avec un moteur d'inférence qui permet l'ajout des éléments de l'ontologie au moyen des axiomes spécifiés par la méta-ontologie dans le but d'enrichir une base de connaissances d'extraction d'ontologies à partir de textes extraits du Web.

La métaontologie et l'ontologie de services ont été formalisées en utilisant le langage OWL². Les axiomes de la métaontologie sont formalisés en logique du premier ordre en utilisant le langage SWRL³.

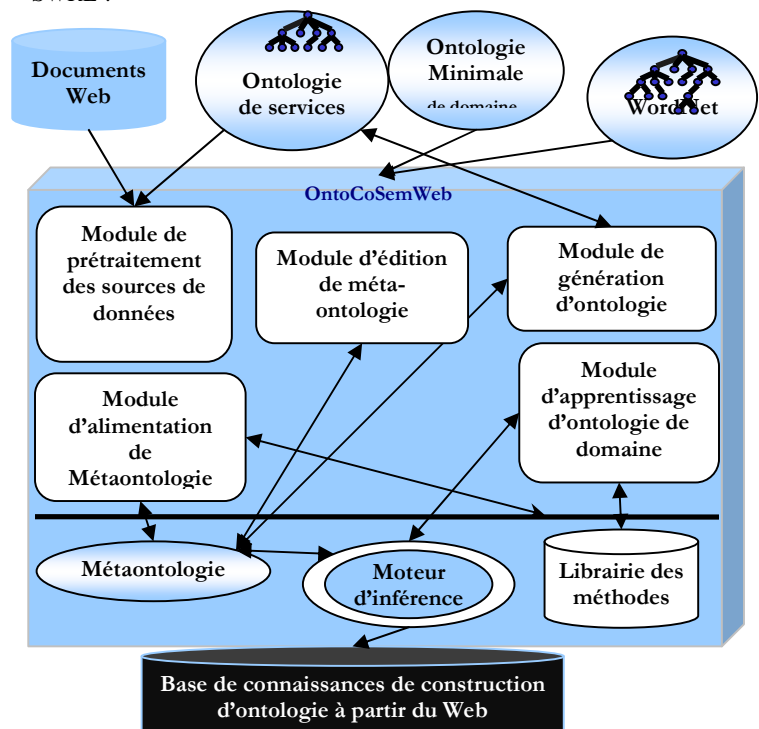


Figure 6. Architecture de OntoCoSemWeb

Les fonctionnalités relatives à chacune des phases identifiées de l'approche proposée sont implémentées en utilisant les outils suivants : (1) Protege2000⁴ pour l'édition de l'ontologie minimale de tourisme, (2) TreeTagger⁵ pour l'étiquetage du corpus textuel, (3) SIRO [3] pour la collecte des sites Web dans le domaine de

² <http://www.w3.org/TR/owl-guide/>

³ <http://www.w3.org/Submission/SWRL/>

⁴ <http://protege.stanford.edu/>

⁵ <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

tourisme. Dans le cadre de l'implémentation du module de construction d'ontologies de domaine, nous avons recensé les bibliothèques, les plateformes et les technologies qui sont jugées utiles. Nous avons utilisé le projet du code source JENA⁶, API DOM (Document Object Model), le plugin SWRL, le moteur d'inférence Bossam, le code source de Protege-OWL. Dans ce qui suit, nous décrivons brièvement SIRO [3], l'outil de récupération des documents classés par thèmes.

5.3.2 SIRO pour la recherche et la classification des documents Web classés par service de domaine

SIRO [3] est constitué de trois principaux modules: un module de traitement de la requête, un module pour la recherche et le traitement des documents et le dernier pour leur classification. Le module de traitement de la requête permet d'enrichir la requête initiale de l'utilisateur en se basant sur les concepts et les relations de l'ontologie de domaine et sur WordNet. La requête enrichie est fournie à un moteur de recherche. Le résultat obtenu est ensuite traité par le module de traitement des documents. Ces documents sont classifiés par services en utilisant le module de classification par service. Ce module aide l'utilisateur à construire une deuxième requête en utilisant les services choisis. Ce système est basé sur :

- L'usage d'une ontologie du domaine combiné avec une ontologie des services ;
- La classification par service des résultats d'une requête utilisée pour améliorer une recherche basée sur les services correspondants; ce qui permet d'assurer un usage plus facile de l'information;
- L'adaptation du modèle vectoriel [23] dans lequel nous substituons les termes par les concepts.

La classification permet d'affecter un document à une des classes ou sous-classes d'un plan de classement ou table de classification, c'est-à-dire dans un des domaines de la connaissance. Un résultat classifié par catégorie permet de faciliter la récupération de l'information et même de détecter d'autres besoins. Lorsque le résultat est traité et affiché à l'utilisateur, il peut faire l'objet d'une classification par service fournie par le module de classification. La requête reformulée par le module de traitement de la requête contient un ensemble de concepts qui appartiennent à l'ontologie du domaine. Ces derniers sont en liaison avec un ensemble de services de l'ontologie de services, plus précisément chaque concept est l'objet d'un ou plusieurs services. Par exemple, le concept « restaurant » est associé au service « Restauration ». Cette liaison est construite manuellement au cours de la construction de l'ontologie de services. En utilisant cette liaison, nous pouvons extraire, à partir des concepts ajoutés aux requêtes, les services qui lui sont relatifs dans l'ontologie des services.

Le modèle vectoriel est de nouveau utilisé, un service est représenté par un vecteur : $Serv_i = (c_1, c_2, \dots, c_N)$ avec N est le nombre de concepts relatifs à un service. Pour chaque document retenu, nous calculons sa similarité avec les services en utilisant la formule du cosinus, puis il est affecté au service avec lequel il est le plus similaire. Ainsi les URL sont affichées par service.

5.4 Observation de l'approche LEO-By-LEMO

Cette approche a été expérimentée par le développement d'un outil OnToCoSemWeb [7]. Dans le but de mener une construction d'ontologies de domaine à partir des documents Web au moyen d'une métaontologie, nous avons appliqué cette approche pour l'enrichissement d'une ontologie de domaine de tourisme qui contient initialement 8 concepts et 6 relations. Les principales phases de l'approche proposée sont respectées. Trois itérations ont permis d'obtenir les résultats suivants au cours de la phase d'analyse des résultats de notre approche incrémentale. Cette phase nous a permis de dégager des remarques importantes sur l'utilisation des techniques linguistiques et d'avoir le graphe illustré dans la figure 7.

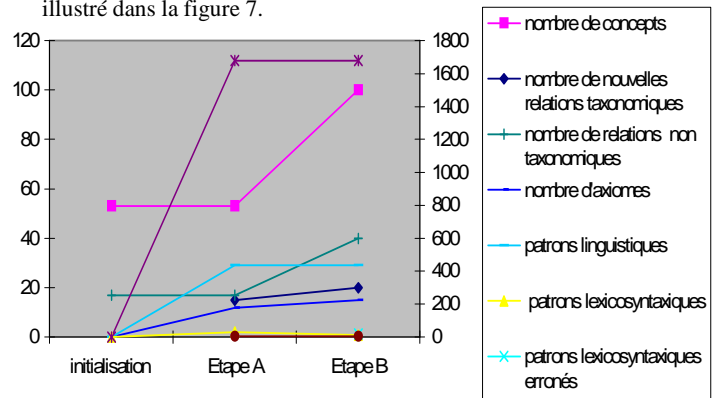


Figure 7. Graphe d'évolution au cours de la troisième itération de l'approche proposée

Nous remarquons que le nombre de concepts et de relations taxonomiques a doublé lors de la deuxième étape après l'application des patrons lexicosyntaxiques. Quant aux relations non taxonomiques, elles ont légèrement évolué depuis l'étape d'alimentation à cause de la difficulté pour trouver des patrons lexicosyntaxiques qui définissent les relations d'hyponymie. Ainsi, l'apprentissage des patrons non taxonomiques se fait essentiellement au cours de la première étape, grâce aux axiomes édités par les linguistes qui s'inspirent des patrons relatifs aux expressions nominales extraites du corpus. Le degré d'erreur des patrons lexicosyntaxiques dans une première itération montre la dépendance des erreurs d'apprentissage aux nouveaux patrons appris à partir du corpus textuel. Ainsi, une deuxième itération utilisant un autre corpus permet la dépréciation des coefficients associés aux patrons qui sont rares afin de ne pas en tenir compte au cours de l'étape d'apprentissage d'ontologie de l'itération suivante.

6. CONCLUSION

Dans le présent article, un état de l'art des techniques de fouille de textes utilisées pour la construction d'ontologies est présenté. Puis, une métaontologie a été proposée dans le cadre d'une nouvelle approche d'enrichissement incrémental d'une ontologie à partir des pages Web. Notre contribution réside dans l'utilisation d'une Métaontologie pour la construction semi-automatique d'ontologies à partir de textes. Ce qui permettrait de faciliter aussi bien la maintenance de l'ontologie construite que les règles d'extraction formalisées par la métaontologie. En outre, cette dernière permet d'assister l'ontologue pour évaluer les techniques utilisées. Une expérimentation de l'utilisation de la Métaontologie

⁶ <http://jena.sourceforge.net/>

proposée dans un outil qui supporte l'approche, « OntoCoSEMWEB » a été menée. Nous avons intégré un système de recherche en ligne SIRO [3] dans cet outil pour automatiser la récupération des pages Web et leur classification sur la base d'une ontologie de service définie [5]. L'approche proposée a l'avantage de limiter l'intervention humaine au cours de la construction d'ontologies. Elle met également à la disposition des linguistes et des ingénieurs d'ontologies une Métaontologie dont l'utilisation contribue à un suivi facile de l'évolution de l'ontologie et à détecter les erreurs d'apprentissage des concepts ou des relations. Dans les travaux futurs, nous allons intégrer l'utilisation d'un système multi-agents pour améliorer l'outil développé.

7. REFERENCES

- [1] Alfonseca E., Manandhar S. « An unsupervised method for general named entity recognition and automated concept discovery ». Actes de la 1er conférence internationale sur General WordNet, , India, 2002.
- [2] Agirre E., Ansa O., Hovy E. et Martinez D. « Enriching very large ontologies using the WWW », actes de l'atelier sur la construction d'ontologies de la conférence européenne de l'intelligence artificielle (ECAI-00), 2000.
- [3] Aufaure M, Soussi R. et Baazaoui Zghal, H. Ben Ghezala H., : SIRO: On- Line Semantic Information Retrieval using Ontologies, The Second International Conference on Digital Information Management (ICDIM'07), october 28-31 lyon,France, 2007.
- [4] Aussenac-Gilles, Jacques M-P. Designing and Evaluating Patterns for Ontology Enrichment from Texts. EKAW 2006: 158-165).
- [5] Baazaoui Zghal, H., Aufaure MA. et Ben Mustapha, N. Extraction of Ontologies from Web Pages: conceptual modeling and tourism, Journal of internet Technologies, 2007.
- [6] Berners-Lee, T, Hendler, J, Lassila, O.: The Semantic Web, Scientific American ; 2001.
- [7] Ben Mustapha, N., Baazaoui Zghal, H. et Aufaure, MA.. A Prototype for knowledge extraction from semantic Web based on ontological components construction, Internet technology, Web information System and Technologies (WEBIST07), 2007.
- [8] Bourigault D. LEXTER, un Logiciel d'EXtraction de TERminologie, Application à l'acquisition des connaissances à partir de textes. Ecole des Hautes Etudes en Sciences sociales, Paris, 1994.
- [9] Brini A., Boughanen M et Dubois D. A Model for Information Retrieval based on Possibilistic Networks. Proceedings of the 12th Symposium on String Processing and Information Retrieval (SPIRE), 2005.
- [10] Gangemi A., C. Catenacci, M. Ciaramita, and J. Lehmann. A theoretical framework for ontology evaluation and validation. In Semantic Web Applications and Perspectives (SWAP) -- 2nd Italian Semantic Web Workshop, Trento, Italy, 2005
- [11] Gruber T.: Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies, special issue on Formal Ontology in Conceptual Analysis and Knowledge Representation. Eds, Guarino, N. & Poli , R., 1993.
- [12] Hearst M.A. Automated Discovery of WordNet Relations. "Wordnet An Electronic Lexical Database". MIT Press, Cambridge, MA, 132-152, 1998.
- [13] Faure D. et Nedellec, C. A corpus-based conceptual clustering method for verb frames and ontology acquisition. LREC workshop on adapting lexical and corpus resources to sublanguages and applications, Granada, Spain, 1998.
- [14] Khan L. et Luo F. Ontology Construction for. Information Selection. Proceedings of 14th IEEE International, 2002.
- [15] Koo S., Lim S.Y. et Lee S.J. Building an Ontology based on Hub Words for Informational Retrieval. Proceedings of the IEEE/WIC International Conference on Web Intelligence, 2003.
- [16] Lebart, L. et Salem A. Analyse statistique des données textuelles Paris: Dunod, 1988.
- [17] Lin C-Y et Hovy E.H The Automated Acquisition of Topic Signatures for Text Summarization. Proceedings of the COLING Conference. Strasbourg, France. Aout, 2000.
- [18] Maedche A. et Staab S. Mining ontologies from text. Proceedings of the 12th International Conference on Knowledge Engineering and Knowledge Management, Springer Lecture Notes in Artificial Intelligence, 2000.
- [19] Resnik P. Using information content to evaluate similarity in taxonomy. Proceedings of the 14th joint conference in Artificial Intelligence, 1995.
- [20] Rada R., Mili H., Bicknell E. et Blettner M. *Development and application of a metric on semantic nets*, IEEE Transaction on Systems, Man and Cybernetics, 19(1), pp 17-30, 1989.
- [21] Robertson S. E. et Sparck Jones K. Relevance weighting of search terms. Journal of the American Society for Information Sciences, pp 129-146, 1976.
- [22] Sekiuchi R., Aoki C., Kurematsu M. et Yamaguchi T. DODDLE: A Domain Ontology Rapid Development Environment, PRICAI98, 1998.
- [23] Salton, G. The SMART Retrieval System - Experiment in Automatic Document Processing. Englewood Cliffs, NJ : Prentice-Hall, 1971.
- [24] Sanderson M. et Croft W.B. Deriving concept hierarchies from text. Proceedings of the 22nd International ACM SIGIR Conference, pp 206-213, 1999.
- [25] Sugiura N., Masaki K., Naoki F., Noriaki I. et Takahira Y. A Domain Ontology Engineering Tool with General Ontologies and Text Corpus. Proceedings of the 2nd Workshop on Evaluation of Ontology based Tools, 2003.
- [26] Uschold M. et Gruninger M. *Ontologies: principles, methods and applications*. The Knowledge Engineering Review, 11(2):93--136, November 1996.
- [27] Velardi P., Fabriani P. et Missikoff M. Using text processing techniques to automatically enrich a domain ontology, Proceedings of the ACM Conference on Formal Ontologies and Information Systems, pp 270-284, 2002

