REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE MOHAMED BOUDIAF - M'SILA
FACULTE DES MATHEMATIQUES ET DE L'INFORMATIQUE
DEPARTEMENT D'INFORMATIQUE

MEMOIRE de fin d'étude

Présenté pour l'obtention du diplôme de MASTER

Domaine : Mathématiques et Informatique

Filière : Informatique

Spécialité : Systèmes d'Informations Avancés

Par: Smaili Asma

SUJET

Annotation automatique Des Pages Web

Soutenu publiquement le : 01 / 06 /2016 devant le jury composé de :

Boudia Malika	**Université de M'sila**	**Président**
Bouzaroura Ahlem	**Université de M'sila**	**Rapporteur**
Boughrara	**Université de M'sila**	**Examinateur**

Promotion : 2015 /2016

Introduction générale

Le web est devenu l'une des plus importantes sources d'information accessible de manière électronique, il constitue une masse d'information gigantesque car Le nombre de ressources sur le web croît de façon exponentielle et le nombre d'usagers augmente chaque jour. Malheureusement, le Web est si énorme et si peu structuré qu'y trouver une information exacte et utile est devenu une tâche très coûteuse en temps. De ce fait, il faut structurer les documents accessibles via Internet et précisément les pages web, ce aménagement est basé un attribut très important "la sémantique".

Cette vision du Web futur dépend de la construction où il est nécessaire d'associer aux ressources du Web des informations exploitables par des agents logiciels afin de favoriser l'exploitation de ces ressources. Les recherches ont mené à la naissance du web sémantique, et toute une série de technologie et de nouveaux concepts.

Pour le Web Sémantique, l'un des aspects les plus importants est de pouvoir manipuler des annotations sémantiques de documents Web, puisque le Web Sémantique permettra aux machines de comprendre la sémantique des documents et des données. Les annotations sémantiques décrivent le contenu des documents, en associant une sémantique à ces descriptions. On peut les considérer comme des Web Sémantique et application à la recherche d'information métadonnées de documents, ressources du Web.

Les ontologies sont l'une des concepts primordiales pour le Web sémantique qui, d'une part, cherche à s'appuyer sur des modélisations de ressources du Web à partir de représentations conceptuelles des domaines concernés et, d'autre part, a pour objectif de permettre à des programmes de faire des inférences dessus. Clairement, la sémantique de l'annotation est fondée sur des vocabulaires dans les ontologies qui sont spécifiées explicitement dans un langage de représentation.

Nous nous posons donc la question : comment faciliter le plus possible l'accès à ces connaissances ?

Dans notre travail, nous avons opté pour l'utilisation des techniques du Web Sémantique et en particulier pour l'utilisation des annotations sémantiques basées sur les ontologies pour faciliter l'accès aux connaissances géographie contenues dans les pages web.

Nous avons décomposé notre mémoire en trois chapitres.

Chapitre 1 : Le Web Sémantique et l'ontologie

Ce chapitre introduit le Web sémantique et spécialement les ontologies, nous commençons par la définition de la notion d'ontologie. Nous présentons ensuite les principaux formalismes

de représentation de connaissances. Nous découvrirons après les méthodologies les plus représentatives de leur construction et quelques domaines de leur utilisation. A la fin, nous présenterons les outils nécessaires de leur développement, à savoir, les langages de représentation, les outils d'édition

Chapitre 2 : L'annotation sémantique

L'annotation sémantique de documents sera introduite, en présentant son apport par rapport à l'annotation classique, à travers la représentation des définitions et des éléments liés à cette nouvelle approche

Chapitre 3 : conception et implémentation

Dans ce chapitre, nous allons présenter, la description de notre approche implémentée. Ainsi que les principaux outils qui seront utilisés dans notre approche.

Enfin, notre mémoire s'achève par une conclusion générale récapitulant le contexte de recherche de notre étude, la démarche suivie, nos contributions et énonce un ensemble de perspectives

Conclusion générale

Dans ce mémoire nous présentons un système d'annotation sémantique automatique pour annoter des pages web. Comme le web est une ressource de données très étendue nous avons choisi d'annoter des pages web pour des sites géopolitiques, de ce fait, notre système pivote au tour d'une ontologie géographique. La feuille de données sémantique est la première étape dans notre système, elle consiste à extraire les mots clés selon leurs poids sémantiques et leurs valeurs statistiques dans la page puis ces mots sont enrichis à travers l'ontologie pour avoir une annotation sémantique.

Or, nous avons rencontré des problèmes qui sont:

• Des difficultés de couvrir tout les aspects géographiques du monde réel et définir précisément les objets géographiques pour l'ontologie géographique.

• Le prétraitement du nettoyer les données, qui est un processus fastidieux et complexe dû principalement à la grande quantité de données dans page web et la structure de la page Web est uniforme.

Nos perspectives de recherche est de compléter cette extension, nous projetons d'appliquer cette méthode sur un plus grand nombre de page Web et d'une complexité plus élevée afin de faire une étude comparative effective pour trouver une nouvelle approche. Et l'intégration des connaissances de l'utilisateur dans le processus d'annotation et l'exploitation de l'annotation dans les systèmes de recherche d'informations.

[1] Jean Charlet, Philippe Laublet , Chantal Reynaud , Web sémantique , Rapport final , CNRS / STIC, décembre 2003

[2] Mohamed Amine Mestiri, Vers une approche web sémantique dans les applications de gestion de conférences, Université Laval, Mémoire de Magister,2007

[3] Philippe Laublet, Chantal Reynaud, Jean Charlet , Sur quelques aspects du Web sémantique , Université de Paris-Sorbonne– CNRS (lalicc) ISHA , Université Paris-Sud – CNRS (L.R.I.) & INRIA (Futurs), Mission de recherche en sciences et technologies de l'information médicale - DPA/DSI/AP-HP,2003

[4] Fabien L. Gandon, Graphes RDF et leur Manipulation pour la Gestion de Connaissances, Université de Nice – Sophia Antipolis, Mémoire d'Habilitation à Diriger les Recherches, novembre 2008

[5] Phuc Hiep LUONG , Gestion de l'évolution d'un Web sémantique d'entreprise , Thèse de Doctorat , Ecole des Mines de Paris , 14 décembre 2007

[6] Tuan Dung CAO , Exploitation du web sémantique pour la veille technologique, Thèse de Doctorat, Université de Nice-Sophia Antipolis, 29 Novembre 2006

[7] Mohamed Khaled KHELIF , Web sémantique et mémoire d'expériences pour l'analyse du transcriptome, Université de Nice-Sophia Antipolis ,Thèse de Doctorat , 4 avril 2006

[8] Samia BOUARROUDJ, Raisonnement sur une ontologie enrichie par des règles SWRL pour la recherche sémantique d'image annotées, Mémoire magister, Universite 20 Aout 1955 Skikda, 2009-2010

[9] DJAMA OUAHIBA , Une approche d'annotation sémantique à partir d'une ontologie multi-points de vue ,Mémoire de magister, Université Mentouri De Constantine, 20/06/2010

[10] MAHIDDINE M ehanna et MISSOUM Mehenna , Conception et réalisation d'une ontologie dans le domaine des hydrocarbures pour la recherche d'information , Mémoire d'ingénieur, Institut National de formation en Informatique (I.N.I) Oued-Smar Alger , Promotion: 2006/2007

[11] Http://dspace.univ-tlemcen.dz/bitstream/112/1062/5/chapitrei.pdf

[12] Soraya Zaidi–Ayad, Une plateforme pour la construction d'ontologie en arabe : Extraction des termes et des relations à partir de textes (Application sur le Saint Coran) ,Universite badji mojhtar annaba ,thèse de doctorat,2012/2013

[13] Fouzia Amourache, construction d'une ontologie pour l'annotation des cv offres d'emploi, Mémoire de magister, Université de Mentouri – Constantine, 01/12/2008

[14] Iana ATANASSOVA, Exploitation informatique des annotations sémantiques automatiques d'Excom pour la recherche d'informations et la navigation, Université Paris-Sorbonne ; thèse de doctorat, 14 janvier 2012

[15] Taibaoui Mohamed, Debbar Djafar, La découverte des concepts sémantiques cachés avec plusieurs niveaux d'abstraction pour la recherche d'images, universite kasdi merbah ouargla ; Mémoire Master ; juin 2013

[16] Charles Abiodun Robert ; l'annotation pour la recherche d'information dans le contexte d'intelligence economique ; Université Nancy 2 ; thèse de doctorat, 16/02/2007

[17] NGUYEN Van Tien , Méthode d'extraction d'informations géographiques à des fins d'enrichissement d'une ontologie de domaine , Ecole Doctorale Des Sciences Exactes Et De Leurs Applications , thèse de doctorat , 15 novembre 2012

[18] Anne-Lyse MINARD , Etat de l'art des ontologies d'objets géographiques, Document réalisé au cours du stage, laboratoire COGIT de l'IGN , Mai – juin 2008

[19] BENYAHIA Kadda, LEHIRECHE Ahmed, LATRECHE Abdelkrim, Annotation Sémantique De Pages Web, http://ceur-ws.org/Vol-547/54.pdf

[20] BOUYACOUB soumia et KAOUADJI Sarra, Enrichissement de la représentation conceptuelle dans la catégorisation du texte en utilisant les mesures de similarité sémantique, Université Abou Bakr Belkaid– Tlemcen, Mémoire Master ,2012-2013

ملخص

عملية إضافة الملاحظات الدلالية إلى صفحات الويب يمكن أن تكون معقدة بالنسبة للإنسان. في هذه المذكرة قدمنا طريقة لإضافة الملاحظات الدلالية آليا لصفحات الويب. يعتمد نظامنا على استخراج الكلمات المفاتيح من صفحة ويب باستخدام الأساليب الإحصائية بالإضافة الدلالات الألفاظ. ثم البحث عن هذه الكلمات في انطولوجيا لاستخراج المفاهيم ذات الصلة وبذلك نظيف ملاحظات دلالية لصفحات الويب من خلال ملف من نوع XML

الكلمات المفاتيح: الويب بالدلالات اللفظية، انطولوجيا، ملاحظات دلالية

Abstract

An annotation task can be complex for a human being. In this paper we focus on automatic semantic annotation of web pages. our system is based on keywords extraction from web page using statistical and semantic methods. Then these keywords are projected on an ontology to extract related concepts to constitute semantic annotations that will be attached to web pages through an XML file.

Keywords: Semantic Web, ontology, semantic annotation

Résume

Une tâche d'annotation peut s'avérer complexe pour un être humain. Dans ce mémoire nous nous intéressons à l'annotation sémantique automatique de pages web. Notre travail consiste à extraire les mots clés à partir du page web en utilisant des méthodes statistiques et sémantiques. Ces mots clés sont par la suite projetés sur une ontologie afin d'extraire les concepts associés pour avoir les annotations sémantiques qui seront attachées aux pages web en biais d'un fichier XML.

Mots clés: web sémantique, ontologie, annotation sémantique