

Apport du Web sémantique dans la réalisation d'un moteur de recherche géo-localisé à usage des entreprises

Frédéric Triou*, Fabien Picarougne*
Henri Briand*

*LINA CNRS FRE 2729 - Équipe COD
École Polytechnique de l'Université de Nantes
rue Christian Pauc, 44306 NANTES Cedex 3, France
{frederic.triou, fabien.picarougne, henri.briand}@univ-nantes.fr
<http://www.sciences.univ-nantes.fr/lina/fr/research/teams/ECD/index.html>

Résumé. La recherche d'une entreprise sur le Web, relative à un savoir-faire particulier, n'est pas une tâche toujours facile à mener. Les outils mis à la disposition de l'internaute ne donnent pas entièrement satisfaction. D'un côté les moteurs de recherche éprouvent des difficultés à faire ressortir clairement le résultat escompté. De l'autre côté, les annuaires spécialisés (type Pages Jaunes) sont tributaires d'une organisation figée, nuisant à leur efficacité.

Face à ce constat, nous nous proposons de créer un nouveau moteur spécialisé dans la recherche d'entreprise, associant Web sémantique et géo-localisation. Cette approche novatrice nécessite l'implémentation d'une ontologie ayant pour objectif la formalisation des connaissances du domaine.

Cette tâche a mis en évidence l'intérêt des structures économiques, maintenues par l'INSEE, et leur utilisation au sein de l'ontologie. Les nomenclatures économiques ont été retenues pour gérer la classification des activités et produits pouvant être dispensés par les entreprises. La structure des unités administratives, telle que gérée au sein du fichier SIRENE, s'est avérée judicieuse pour répondre à la problématique de géo-localisation des entreprises. Une opération de désambiguïsation est réalisée en associant à chaque nœud d'activité les mots clés et synonymes lui correspondant.

Enfin, nous comparons les résultats obtenus par notre moteur à ceux obtenus par le principal moteur de recherche d'activités géo-localisées en France : les Pages jaunes. Que ce soit au niveau de la précision et du rappel, notre moteur obtient des résultats significativement meilleurs.

1 Introduction

Les moteurs de recherche classiques sur le web ont des caractéristiques étonnantes : ils possèdent des milliards de documents dans leur index, ils peuvent traiter des millions de requêtes quotidiennement, ils donnent des réponses très volumineuses quasiment en temps réel et ils nécessitent des ressources informatiques et humaines considérables. On peut dire aujourd'hui