

Focused Web search

Par

Abdallah Benabbes

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Mémoire présenté à la faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en commerce électronique, axe : informatique

Décembre, 2005

© Abdallah Benabbes, 2005

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé

Focused Web search

Par

Abdallah Benabbes

A été évalué par un jury composé des personnes suivantes :

Président rapporteur : Jacques Robert

Direction de recherche : Gilbert Babin

Membre du jury : Paul Mireault

Abstract	2
List of Tables	4
List of Figures	5
Acknowledgement	7
Chapter 1 Introduction	8
Chapter 2 Information Retrieval	11
Chapter 3 Survey of web search	21
Chapter 4 Domain specific web search	29
Chapter 5 Implementation	49
Chapter 6 Testing and Evaluation	79
Chapter 7 Conclusion	102
Bibliography	104
Appendices	108

Résumé

Un nouveau modèle de recherche d'information développé dans ce mémoire traite des problèmes de précision et complétude pour les systèmes de recherche courants. De nos jours, la recherche d'information sur le Web est devenue de plus en plus importante dans toutes les applications informatiques. Il y a beaucoup de moteurs de recherche commerciaux puissants tels que Google, Askjeeve, Yahoo et de MSN etc... Tous ces systèmes de recherche semblent pouvoir accepter des requêtes arbitraires, avec un secteur de recherche large. Cependant, leurs résultats manquent de détails dans la plupart des situations. Ce mémoire se concentre sur une situation différente: des recherches récurrentes dans quelques domaines ciblés. Ceci exige des résultats de recherche plus précis et plus complets.

Pour créer un tel système, nous avons utilisé plusieurs progiciels du domaine public et nous présentons un survol des systèmes disponibles.

Comme expérimentation finale, nous avons utilisé notre application de recherche pour trouver des appels d'offres de 5 villes canadiennes choisies aléatoirement. Les résultats prouvent que nos techniques fonctionnent mieux que les moteurs de recherche conventionnels.

Mot-clé :

domaine spécifique, recherche, recherche documentaire

Abstract

A new information-searching model is developed in this thesis to deal with the precision and completeness problems for current searching systems. Nowadays, web searching has become more and more important in all computer applications. There are many powerful commercial search portals such as Google, Askjeeve, Yahoo and MSN; these searching systems support arbitrary queries, with a wide search area, and their results lack in depth in most situations. This thesis focuses on a different situation: recurrent searches in some specified domain. This requires the search results to be more focused.

To create such a system, we utilized several public domain packages and we present a survey of available systems.

As a final experiment, we tested our search application to find call for tenders (CFT) from 5 randomly selected Canadian cities. Results show that our techniques work much better than conventional search engines.

Keyword:

domain specific, search, information retrieval