

# Recherche d'informations sur le Web basée sur des données d'opinion

Mehdi Adda\*, Rokia Missaoui\*\*, Petko Valtchev\*\*\*

\*Département de mathématiques, d'informatique et de génie, Université du Québec À Rimouski  
mehdi\_adda@uqar.qc.ca,

\*\*Département d'informatique et d'ingénierie, Université du Québec en Outaouais  
rokoa.missaoui@uqo.ca

\*\*\*Département d'informatique, Université du Québec À Montréal  
valtchev.petko@uqam.ca

**Résumé.** Avec la diversification du contenu disponible sur le Web, aussi bien en quantité qu'en qualité, retrouver l'information pertinente est problématique. Les moteurs de recherche d'informations sur le Web ont comme objectif de permettre de retrouver efficacement l'information pertinente. Ces moteurs implémentent différentes stratégies pour déterminer la pertinence des pages Web. Ces dernières sont classées en se basant principalement sur des critères d'évaluation automatisés, ce qui crée un vide entre la pertinence effective d'un contenu et la pertinence calculée. Afin de réduire cet écart, nous présentons un cadre d'application de moteur de recherche où la pertinence d'une page Web est calculée à partir des évaluations des internautes du contenu de la page (texte, image, audio, vidéo,...). Pour illustrer l'exploitation du cadre proposé, nous avons développé le moteur de recherche expérimental *SocialSeeker* (Adda, 2008). Finalement, une étude empirique préliminaire du potentiel du prototype est présentée et discutée

## 1 Introduction

La tâche principale d'un moteur de recherche est d'offrir à ses utilisateurs la possibilité de formuler des requêtes et de leur retourner le contenu qui correspond le mieux à ces requêtes (Adah et al., 1997). Les moteurs de recherche scrutent périodiquement, voire de manière continue, l'Internet pour (ré)indexer des pages Web (Sun et al., 2005). Les stratégies utilisées pour classer ces pages diffèrent d'un moteur à l'autre.

### 1.1 Stratégies de classement des pages Web

Les grands acteurs sur le marché des moteurs de recherche, tels que *Google* (Google, 2008), *Microsoft* (Microsoft, 2008), et *Yahoo* (Yahoo, 2008), gardent secrets les détails des techniques de classement qu'ils utilisent. Cependant, quelques éléments qui jouent un rôle important dans ce classement sont connus (Ding et Chi, 2003).

En premier lieu, les moteurs de recherche prennent en compte aussi bien le titre d'une page