

Développement d'un méta-moteur de recherche sur le Web orienté communauté

Contact : rushed Kanawati
LIPN B309, Tél. : 0149403609
rushed.kanawati@lipn.univ-paris13.fr

I. Contexte

Les moteurs de recherche sur le Web, comme Google, AltaVista, etc. sont aujourd'hui des outils incontournables pour aider les utilisateurs à chercher et localiser l'information sur le Web. La popularité de ces outils est en grande partie due à leur simplicité d'utilisation : les utilisateurs soumettent des requêtes, souvent formées de quelques mots clés et le moteur retourne les documents pertinents pour la requête. Les performances actuelles des moteurs existants en termes de précision et de rappel sont encore assez médiocres. A titre d'exemple la précision de Google ne dépasse pas le 29% et le rappel est de 20% seulement en 2004). Si le faible rappel est souvent justifié, à juste titre, par la grande taille du Web et sa rapidité de croissance, la faible précision est due en premier lieu à la faible spécification des requêtes utilisateurs. Différentes approches sont proposées dans la littérature scientifique afin de répondre à ce problème. Dans ce projet nous nous intéressons à une approche dite « communautaire ». L'idée est de permettre à un groupe d'utilisateurs, qui forment une communauté d'intérêts, de partager leurs expériences de recherche d'information d'une manière implicite et intelligente. En effet, dans une communauté d'utilisateurs (e.g. membres d'un laboratoire de recherche) l'intersection entre les thèmes de recherche sur le Web peut être assez important (certaines études montrent que l'intersection entre les requêtes utilisateurs d'un groupe peut dépasser 40%). Par conséquent, un moteur qui garde trace des requêtes passées soumises par les utilisateurs peut en profiter afin d'améliorer ses performances en filtrant les réponses obtenues de sorte à promouvoir les ressources approuvées par les autres membres de la communauté. L'objet de ce projet est de développer un moteur orienté communauté qui profite des expériences passées du groupe afin de modifier le tri des résultats retournés par les moteurs existants (ainsi le moteur proposé agit comme un méta moteur de recherche).

2. Description du méta-moteur

Le méta-moteur à développer aura une architecture égal à égal (Peer To Peer en anglais). Chaque utilisateur est assisté par un agent logiciel personnel qui garde trace des requêtes envoyées par l'utilisateur, les réponses retournées et les sélections faites par l'utilisateur. Le choix de l'architecture P2P est en partie justifié par la volonté de respecter la vie privée des utilisateurs (i.e les traces des leurs requêtes ne sont pas partagées par d'autres utilisateurs). Un document sélectionné est jugé pertinent pour la requête. En soumettant une requête, l'agent renvoie aussi la requête et les réponses à ses pairs qui vont trier les résultats obtenus en fonction de leurs propres expériences. L'agent émetteur va recevoir les différentes tris proposées par les différentes agents et va opérer une fusion de ces listes triées (la même liste triée différemment) avant de présenter les résultats à l'utilisateur. Une description détaillée de l'architecture du système et de l'algorithme de tri est donnée dans l'article joint à cette proposition.

3. Réalisation

Le projet est à réaliser en Java. Une attention particulière est à faire lors de la conception afin de rendre le prototype facilement extensible pour inclure de nouveaux services ou pour changer la mise en oeuvre des services existants (i.e. modification de l'algorithme de tri). Le prototype devrait servir à une validation grande échelle de l'approche proposée.

