

Examen du cours Moteurs de recherche 2010

M2 Pro, Université Paris Diderot, Paris 7

Michel Habib, Fabien de Montgolfier

March 19, 2011

1 PageRank

Vous avez vu en TP, comment calculer un ordre de type PageRank sur les pages du WEB.

1. Comment cet ordre est actuellement utilisé dans un moteur de recherche ?
2. Si le graphe n'est pas fortement connexe, eut-on calculer PageRank sur chaque composante fortement connexe indépendamment et fusionner les résultats ? (Précisez la méthode de fusion)
3. Est-ce algorithmiquement pertinent (temps de calcul sur le très gros graphe du WEB) ?

2 Détection algorithmique de communautés

La détection de communautés dans une population est un problème à la fois très difficile mais aussi très utilisé. Considérons trois exemples:

1. Un réseau. La donnée disponible est uniquement constituée des liens entre les individus.
2. Chaque individu est décrit textuellement, mais il n'y a pas de liens connus entre les individus.
3. Un réseau. Chaque individu est typé et l'on dispose de la connaissance des liens entre les individus.

Question: Dans chacun des cas, définis précédemment proposez une formalisation (spécification dans un langage mathématique) du problème de la recherche de communautés, ainsi qu'un algorithme ou une heuristique qui corresponde à votre formalisation.

3 Construction d'un nouveau réseau social

Il s'agit de travailler à la spécification d'un nouveau réseau social de rencontre et de partage de connaissance. À l'inscription chaque nouveau membre du réseau peut définir son profil librement à l'aide de mots clés prédéfinis, d'un court texte format libre (une page) et d'une image.

1. Précisez comment vous allez calculer les groupes de communications (ou communautés).
2. L'utilisateur ayant la possibilité de poser des questions aux membres du réseau, à qui transmettre une question donnée? Précisez la méthode que vous préconisez.
Piste: il n'est pas envisagé de transmettre la question à tout le réseau social, sous peine de saturation rapide du réseau et d'un désintérêt rapide (on envisage des millions d'utilisateurs), mais seulement aux membres susceptibles d'être intéressés par cette question.
3. Au regard de vos réponses précédentes, auriez-vous des suggestions à faire en ce qui concerne la définition des profils des membres du réseau social ? Toujours dans la perspective du compromis temps de calcul / pertinence des résultats.
4. Installons un système d'amis choisis (à la Facebook), reprendre les réponses aux questions précédentes.
5. Comment détecter le degré de sincérité d'un profil (tricherie sur l'âge, le sexe, la profession, les diplômes ...) ?

4 Dans la peau des ingénieurs de chez Google

Google vient récemment de déclarer que lors d'une requête les réponses étaient filtrées suivant un profil d'utilisateur.

1. En respectant les lois Françaises, de quoi dispose vraiment un moteur de recherche pour définir un profil d'utilisateur ?
2. Proposer une méthode de calcul de profil et le filtrage des réponses associé.
3. Est-ce vraiment une solution d'avenir pour les moteurs de recherche non spécialisés de filtrer les résultats d'une requête suivant un profil ?

5 Distances

1. Comment définir une distance entre requêtes ?
2. Même question pour une distance entre pages Web.
3. Comment calculer ou approximer ces distances ?