

Notes de cours Moteurs de recherche : Master 2 Pro, Université Paris VII, Denis Diderot

Michel Habib,
email: habib@liafa.jussieu.fr
<http://www.liafa.jussieu.fr/~habib>

8 janvier 2009

1 Objectifs

Comprendre le fonctionnement d'un moteur de recherche, les mécanismes (structures de données, idées, matériels : machines et réseaux) qu'il est nécessaire de mettre en oeuvre pour réaliser un moteur de recherche. Le cours commence par une approche expérimentale et orientée utilisateur et se termine sur quelques problèmes de recherche actuels des concepteurs de moteurs de recherche.

2 Plan détaillé

Introduction : analyse du fonctionnement d'une requête sur un moteur de recherche de type Google, Ask ou Exalead. Le point de vue d'un utilisateur. On constatera la prodigieuse efficacité (i.e. le temps de réponse).

Fonctionnement interne : Structuration des tables d'index, recherche pondérée.

Classement des réponses : Théories liées au calculs de Page Rank et ses variantes.

Indexation : Que faut-il indexer et comment, lorsqu'on a trouvé une nouvelle page ? La composition des lexiques. Fouille de textes.

Fouille sur le Web : Robots fouineurs, recherche perpétuelle, les problèmes du protocole html et des navigateurs.

Recherche d'information : Structuration de l'information, syntaxe et sémantique. Évolution des moteurs de recherche.

Modèles issus de l'IA : Graphe conceptuels, logiques de description.

Recherche de communautés : Analyse de comportements et de traces (logs) sur le NET. Fouille de graphes.

Problèmes de recherche : Evolution des moteurs de recherche, la lutte anti-spam, le google bombing, etc.