

Les modes de recherche sur le Web 2.0

Sommaire

1. Contexte
2. Le collaboratif : les moteurs thématiques
3. La syndication des résultats de recherche
4. Les connaissances : des moteurs *intelligents*

1. Moteurs de recherche



1. Définition

- Un **moteur de recherche** est un logiciel permettant de retrouver des ressources (pages Web, images, vidéo, etc.) associées à des mots.
- Constitué de « robots » (*spiders*, *crawlers* ou agents) qui parcourent les sites à intervalles réguliers et de façon automatique. Chaque page identifiée est alors indexée dans une base de données, accessible ensuite par les internautes à partir de mots-clés.
- L'**indexation** des ressources récupérées consiste à extraire les mots considérés comme significatifs (pratiquement tous) correspondant à chaque ressource. Les mots extraits sont enregistrés dans une base de données organisée comme un gigantesque dictionnaire inverse. La partie *requêtes* du moteur de recherche peut ainsi rapidement retrouver les correspondances. Un algorithme gardé secret est généralement appliqué pour donner un poids variable aux correspondances, afin de pouvoir présenter les résultats des recherches par ordre de pertinence supposé. L'algorithme tient généralement compte du contexte du mot clé (titre, paragraphe, hyperlien...) et de la ressource (ressources liées, popularité du site...)
- On trouve également des **métamoteurs**, c'est-à-dire des sites web où une même recherche est lancée simultanément sur plusieurs moteurs de recherche (les résultats étant ensuite fusionnés pour être présentés à l'internaute)

1. Les moteurs de recherche du Web 2.0

- Moteurs de recherche de blogs
 - Technorati : <http://www.technorati.com/>
 - Blogpulse : <http://www.blogpulse.com/>
 - Etc
- Moteurs de recherche de personnes
 - Myspace : <http://www.myspace.com/>
 - Stalkerati : <http://stalkerati.com/>
 - Etc
- Moteurs de recherche de contenus spécifiques
 - Podcast : <http://www.casteropod.com/>
 - Photos : <http://images.google.fr/>
 - Etc

1. Contexte : le Web 2.0

- Le Web 2.0 est sous-tendu par plusieurs principes qui ont fait évoluer les moteurs de recherche :
 - Collaboratif
 - Syndication
 - Connaissances

2. Collaboratif : les moteurs thématiques

- Une idée ancienne : les moteurs thématiques
 - Les moteurs n'indexent pas tout le Web mais des sites pré sélectionnés
 - <http://droit.francophonie.org>
 - <http://www.ejustice.fr>
 - etc

2. Collaboratif : les moteurs thématiques

- Une nouvelle fonctionnalité : les moteurs thématiques à construire à plusieurs
 - Google co-op <http://www.google.com/coop/>
 - Système qui permet de définir des listes de sites traitant un thème particulier. Les résultats des recherches porteront sur les pages des sites identifiés.
 - Meilleure précision, moins de bruit.

3. La syndication des résultats de recherche

- Les fils RSS
 - Des moteurs de recherche génèrent des flux RSS (MSN search, Yahoo)
 - Des logiciels convertissent les résultats des moteurs en flux RSS (google2RSS etc)
- Les résultats des requêtes sur fil RSS :
 - Possibilité de les intégrer dans des pages
 - Permet de surveiller l'apparition de nouveaux résultats pour une recherche donnée.

4. Les connaissances, des moteurs *intelligents*

- Suggérer des termes proches
 - Certains moteurs de recherche proposent à l'utilisateur des termes proches de ceux de sa requête, ce qui permet de la préciser.
 - Ces termes proches sont identifiés par des méthodes statistiques ou bien en se basant sur des dictionnaires ou ontologies
 - <http://www.exalead.com>
 - <http://www.mooter.com>

4. Les connaissances, des moteurs *intelligents*

- Les connaissances plutôt que les informations
 - Tentatives d'appréhender non plus les informations brutes (présence de tel mot dans telle page d'un site Internet) mais des informations qualifiées ou *connaissances*.
 - Nécessite des pages de sites où les informations sont qualifiées
 - Recherches sur des informations qualifiées

4. Les connaissances, des moteurs *intelligents*

- La folksonomie
 - Les utilisateurs apposent leurs propres tags – ou connaissances - sur les sites ce qui constitue une alternative aux moteurs classiques d'indexation
 - <http://del.icio.us/>

4. Les connaissances, des moteurs *intelligents*

- Semanticmediawiki

http://ontoworld.org/wiki/Semantic_MediaWiki

- Dans Mediawiki :
« Paris est la capitale de la [[France]] »,
- avec semantic mediawiki :
« Paris est la capitale de la [[is capital of::France]] ».
- En typant la relation avec 'is capital of', on ajoute à l'ontologie du wiki, la propriété d'objet isCapitalOf.

5. Bibliographie – liens utiles

- <http://www.wikipedia.org/>
- <http://www.precisement.org/index.htm>
- <http://www.exalead.fr/search>
- Google co-op :
 - Données juridiques publiques
<http://www.google.com/coop/cse?cx=006566037309197248496%3Aix0oomavqxc>
 - Blogs juridiques
<http://www.google.com/coop/cse?cx=004960853941544958525:e5zleck5heq>
 - Droit constitutionnel
http://www.google.com/coop/cse?cx=012977174311700543344%3Axaf1_umv5tg
- http://ontoworld.org/wiki/Semantic_MediaWiki
- <http://www.lespetitescases.net/et-le-wiki-devint-semantique>