

INF3005 – Programmation web avancée

SEO

Jacques Berger

Objectifs

Expliquer le fonctionnement des moteurs de recherche

Présenter ce que vous pouvez faire pour améliorer la visibilité de vos sites

Prérequis

HTML

SEO

Search Engine Optimization

Améliorer l'ordre d'apparition d'un site dans les résultats d'un moteur de recherche

Marché

Analyste SEO

Existe depuis 1996, mais rare à l'époque

Aujourd'hui, de plus en plus fréquent

Presque toutes les grosses boîtes de développement web ont des analystes SEO

Pourquoi?

L'utilisateur d'un moteur de recherche va habituellement cliquer sur un des premiers liens dans les résultats

Génère du trafic, bonne visibilité

Peut être très important pour une entreprise

Fonctionnement

Spider : récupère la page web

Crawler : parcourt les liens de la page et index le contenu de la page

Algorithme

1996-1997

Utilisation des balises meta keywords

1998

Invention de Backrub qui deviendra
Google avec l'algorithme PageRank

Algorithme

PageRank calcule le nombre de liens qui pointent vers un site

Chaque lien possède un poids en fonction de la crédibilité du site qui le contient

Algorithme

Ces algorithmes ne peuvent pas être utilisés seuls car ils sont trop faciles à manipuler

- Link farms

- Link-spamming

- Utilisation abusive des mot-clés

Algorithme

Google utilise un algorithme secret

Considère 250+ indicateurs différents

Fréquence des mot-clés

Balises meta

L'en-tête HTML

PageRank

Liens sortants

Structure du site

etc.

Change régulièrement (500 à 600 fois par année)

Algorithme

Google peut également utiliser votre historique de recherche ou l'historique mondial pour construire ses résultats

Indexation

Les crawlers sont des logiciels d'analyse de texte, ils peuvent facilement indexer le texte mais ils n'arrivent pas à indexer ce type de contenu :

- Les images ou graphiques

- Les applications Flash

- Les PDF (Google le fait)

- Le contenu obtenu par un appel Ajax

- Les documents Word, etc.

Sitemap

Pour être certain que le crawler parcourt votre site au complet, vous pouvez faire une carte du site (un sitemap) et l'envoyer au moteur de recherche

Le sitemap est habituellement un document XML

Sitemap

Le sitemap n'est pas toujours nécessaire

Certains hébergeurs enregistrent votre site dans les moteurs de recherche, et si tout votre contenu est facilement accessible, les crawlers vont le parcourir au complet

robots.txt

Avant d'aller parcourir un site, le spider va vérifier s'il existe un fichier robots.txt à la racine du site web

Ce fichier texte contient des instructions sur ce qui doit être exclut de l'indexation

robots.txt

```
User-agent: *  
Disallow: /notes/  
Disallow: /inf2005/
```

Black Hat vs White Hat

Les SEO qui essaient d'exploiter les failles des moteurs de recherche par des pratiques douteuses sont appelés Black Hat

Ceux qui font leur travail en appliquant les bonnes pratiques de qualité sont appelés White Hat

White Hat

Quelques bonnes astuces pour améliorer votre référencement

Avoir un titre pertinent (title) et une description pertinente (meta description)

Les mot-clés doivent se retrouver dans le contenu

White Hat

Régulièrement mettre à jour le contenu du site

Chaque page à indexer devrait être accessible à partir d'un lien statique

Gros site? Faites un plan du site avec des liens vers les différentes sections

White Hat

Ayez du contenu original et pertinent

Utilisez toujours l'attribut alt pour les images

Chaque élément important pour l'indexation devrait être en texte

White Hat

Facilitez l'interprétation du site par le crawler :
HTML toujours valide!

Ne cachez pas de lien ou de texte

Ne mettez pas de lien vers des sites douteux ou frauduleux

White Hat

Le site doit être responsive

Le site doit utiliser HTTPS

Black Hat

Les techniques Black Hat sont considérées comme étant frauduleuses car elles ont comme objectif de tromper les utilisateurs et les moteurs de recherche

L'utilisation de ces techniques pourrait vous faire bannir des moteurs de recherche

Black Hat

Chaînes de liens non reliés

Cloaking (données cachées)

Pages satellites (traffiquer PageRank)

Documentation

Les moteurs de recherche publient des articles sur les techniques pour être bien référencé par leur logiciel, lisez les attentivement

Utilisez les Google Webmaster Tools

Plus loin...

robots.txt

<http://www.robotstxt.org/>

Google Webmaster Tools

<https://www.google.com/webmasters/tools/home>

Google SEO Starter Guide

<http://static.googleusercontent.com/media/www.google.com/fr/webmasters/docs/search-engine-optimization-starter-guide.pdf>