



Documentation robots.txt

Robots.txt (AUFA)

Référence	Robots.txt
Date de révision du document	08/04/2005
Version du document	1.0
Etat	Terminé
Auteur	Agostinho Marques
Diffusion	Diffusable au client

Objet du document
<p>Ce document contient le descriptif pour la mise en place du fichier robots.txt permettant la configuration de l'indexation des pages Web.</p>

Document établi sous la responsabilité des signataires			
	Nom	Date	Signature
Rédaction	Agostinho Marques	08/04/2005	
Vérification	Dominique Trehen	08/04/2005	
Approbation			

Table des matières

1	Definition	5
2	Mise en place du fichier robots.txt.....	6
2.1	Création du fichier robots.txt	6
2.2	Droit du fichier.....	6
2.3	Complément	6

1 DEFINITION

Les moteurs de recherche, (Google, Yahoo, AltaVista, autres) indexent l'ensemble des sites Web afin de les prendre en compte lors des recherches effectuées par les internautes.

Il est cependant possible de limiter, voire interdire, l'indexation de votre site sur certains critères dont le nom du robot d'indexation propre à chaque moteur de recherche ou les répertoires à indexer.

Le fichier robots.txt contient les informations suivantes :

- User-agent qui est le nom du serveur d'indexation
- Disallow qui est le répertoire à exclure.

Certains sites Web subissent des attaques lors de cette phase d'indexation car leurs pages Web sont créés dynamiquement. C'est notamment le cas des pages Web d'Aleph 500.

Pour plus d'information concernant le fichier robots.txt, vous pouvez consulter le site suivant : http://www.webrankinfo.com/outils/robots/faq_15_0.htm

2 MISE EN PLACE DU FICHER ROBOTS.TXT

L'exemple suivant vous permettra d'éviter l'indexation de vos pages OPAC Aleph 500 tout en étant référencé par l'ensemble des moteurs de recherche.

2.1 Création du fichier robots.txt

Créer le fichier robots.txt dans le sous-répertoire htdocs d'apache.

```
cd $alephe_root/apache/htdocs
vi robots.txt
```

Puis insérer les lignes suivantes.

```
User-agent:*
Disallow: /C/
Disallow: /E/
Disallow: /Q/
Disallow: /S/
```

2.2 Droit du fichier

Taper la commande suivante rendre opérationnel le fichier robots.txt.

```
chmod 644 robots.txt
```

2.3 Complément

Modifier le fichier .index.html pour rester référencé par les moteurs de recherche (en recherche « mot »),

```
vi $alephe_root/apache/htdocs/.index.html
```

Puis ajouter les balises suivantes:

```
<META NAME="DESCRIPTION" CONTENT="Vous pouvez noter ici le petit résumé devant apparaître lors de l'affichage de votre lien dans la page du moteur de recherche">
```

Ainsi que les termes de recherche souhaités avec par exemples :

```
<META NAME="KEYWORDS" CONTENT="Ex Libris, ExLibris, Université xxxx, Bibliothèque Municipale yyyy, Academy, Public Library">
```
