

# robots.txt

Le fichier **robots.txt** est un fichier texte utilisé pour le référencement naturel des sites web, contenant des commandes à destination des robots d'indexation des moteurs de recherche afin de leur préciser les pages qui peuvent ou ne peuvent pas être indexées. Ainsi tout moteur de recherche commence l'exploration d'un site web en cherchant le fichier *robots.txt* à la racine du site.

## Format du fichier robots.txt

Le fichier *robots.txt* (écrit en minuscules et au pluriel) est un fichier ASCII se trouvant à la racine du site et pouvant contenir les commandes suivantes :

- **User-Agent:** permet de préciser le robot concerné par les directives suivantes. La valeur \* signifie "tous les moteurs de recherche".
- **Disallow:** permet d'indiquer les pages à exclure de l'indexation. Chaque page ou chemin à exclure doit être sur une ligne à part et doit commencer par /. La valeur / seule signifie "toutes les pages du site".

Voici des exemples de fichier *robots.txt* :

- **Exclusion de toutes les pages :**

```
User-Agent: *  
Disallow: /
```

- **Exclusion d'aucune page** (équivalent à l'absence de fichier *robots.txt*, toutes les pages sont visitées) :

```
User-Agent: *  
Disallow:
```

- **Autorisation d'un seul robot :**

```
User-Agent: nomDuRobot  
Disallow :  
User-Agent: *  
Disallow: /
```

- **Exclusion d'un robot :**

```
User-Agent: NomDuRobot  
Disallow: /  
User-Agent: *  
Disallow:
```

- **Exclusion d'une page :**

```
User-Agent: *  
Disallow: /repertoire/chemin/page.html
```

- **Exclusion de plusieurs page :**

```
User-Agent: *  
Disallow: /repertoire/chemin/page.html  
Disallow: /repertoire/chemin/page2.html  
Disallow: /repertoire/chemin/page3.html
```

- **Exclusion de toutes les pages d'un répertoire et ses sous-dossiers :**

```
User-Agent: *  
Disallow: /repertoire/
```

## Quelques User-Agents

Voici quelques exemples de *User-Agents* pour les moteurs de recherche les plus populaires :

| Nom du moteur | User-Agent          |
|---------------|---------------------|
| Alta Vista    | Scooter             |
| Excite        | ArchitextSpider     |
| Google        | Googlebot           |
| HotBot        | Slurp               |
| InfoSeek      | InfoSeek Sidewinder |
| Lycos         | T-Rex               |
| Voilà         | Echo                |