

Robots.txt

Fichier servant de directive aux robots d'indexation. Un robot n'a aucune obligation de respecter ce fichier.

Ne rien référencer, pour tout les robots.

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

Ne pas référencer un répertoire ou page, juste pour GoogleBot.

```
User-agent: Googlebot  
Disallow: /repertoire-a/  
Disallow: /page-b.html
```

Attention ! Ce fichier est public, n'importe qui peut voir son contenu. Attention à ce que vous mettez (dossiers sensibles par exemple).

Voici une liste de bots agressifs que l'on peut définir :

```
# http://www.opensiteexplorer.org/dotbot  
User-agent: dotbot  
Disallow: /  
  
# https://ahrefs.com/robot/  
User-agent: AhrefsBot  
Disallow: /  
  
# These fools don't even respect robots.txt  
User-agent: QuerySeekerSpider  
Disallow: /  
  
# http://www.picsearch.com/bot.html  
User-agent: psbot  
Disallow: /  
  
# http://webmeup-crawler.com/  
User-agent: BLEXBot  
Disallow: /  
  
# http://sentibot.eu/  
User-agent: sentibot  
Disallow: /  
  
# http://openlinkprofiler.org/bot  
User-agent: spbot  
Disallow: /
```

```
# http://www.crazywebcrawler.com/  
User-agent: CrazyWebCrawler-Spider  
Disallow: /  
  
# https://ranksonic.com/ranksonic_bot.html  
User-agent: RankSonicBot  
Disallow: /  
  
# http://www.seokicks.de/robot.html  
User-agent: SE0kicks-Robot  
Disallow: /
```

From:
<https://wiki.dureuil.info/> - **GD-WIKI**

Permanent link:
https://wiki.dureuil.info/doku.php/web:robots_txt?rev=1465563209

Last update: **2020/07/24 22:03**

