

Robots.txt

Le fichier robots.txt est un fichier texte situé à la racine d'un site web, qui indique aux robots d'indexation quels sont les éléments du site qu'ils sont autorisés ou non à explorer. Un exemple provenant du site de twitter est accessible à l'URL : <https://google.com/robots.txt>.



```
← → ↻ https://www.google.com/robots.txt
Disallow: /sprint_wml
Disallow: /pqa
Disallow: /palm
Disallow: /gwt/
Disallow: /purchases
Disallow: /local?
Disallow: /local_url
Disallow: /shihui?
Disallow: /shihui/
Disallow: /products?
Disallow: /product_
Disallow: /products_
Disallow: /products;
Disallow: /print
Disallow: /books/
Disallow: /bkshp?*q=*
Disallow: /books?*q=*
Disallow: /books?*output=*
Disallow: /books?*pg=*
Disallow: /books?*jtp=*
Disallow: /books?*jscmd=*
Disallow: /books?*buy=*
Disallow: /books?*zoom=*
Allow: /books?*q=related:*
Allow: /books?*q=editions:*
Allow: /books?*q=subject:*
Allow: /books/about
Allow: /booksrightsholders
Allow: /books?*zoom=1*
Allow: /books?*zoom=5*
Allow: /books/content?*zoom=1*
Allow: /books/content?*zoom=5*
Disallow: /ebooks/
Disallow: /ebooks?*q=*
Disallow: /ebooks?*output=*
Disallow: /ebooks?*pg=*
Disallow: /ebooks?*jscmd=*
Disallow: /ebooks?*buy=*
Disallow: /ebooks?*zoom=*
Allow: /ebooks?*q=related:*
Allow: /ebooks?*q=editions:*
Allow: /ebooks?*q=subject:*
Allow: /ebooks?*zoom=1*
Allow: /ebooks?*zoom=5*
Disallow: /patents?
Disallow: /patents/download/
Disallow: /patents/pdf/
Disallow: /patents/related/
Disallow: /scholar
Disallow: /citations?
Allow: /citations?user=
Disallow: /citations?*start=
```

2.

Les balises Allow serve à indiquer ce qui est autorisé à être indexées.

Les disallow serve à indiquer ce qui n'est pas autorisé à être indexées.

Les User-Agents spécifient les robots d'indexation concernés. Par exemple, "User-Agent: * " s'applique à tous les robots, tandis que des lignes spécifiques peuvent être ajoutées pour des robots particuliers comme par exemple "User-agent: Googlebot ".

User-Agents : Ils spécifient les robots d'indexation concernés. Par exemple, "User-Agent: * " s'applique à tous les robots, tandis que des lignes spécifiques peuvent être ajoutées pour des robots particuliers comme par exemple "User-agent: Twitterbot ".