

<http://d.hatena.ne.jp/e980661979/20100720/1279610838>

<http://www.somethingfine.com/br/searchlab/note02.html>

<http://colo-ri.jp/develop/2012/10/bingbotrobotstxt.html>

--	--
robots.txt	ドメインのルートに仕掛ける。クローラのエージェントを指定したり、挙動に関して指示（お願い）ができる
タグ	コンテンツ内のタグに仕掛ける。細かい設定はできないが、コンテンツレベルで指定（お願い）ができる

robots.txt

ドメインのルートに robots.txt を設置する。

例えば

```
http://hoge hoge.com/robots.txt
```

とか。

すべてのロボットに対し、すべてのコンテンツを不許可とする例

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

特定のロボットに対して、すべてのコンテンツを不許可とする例

```
vi /var/www/html/robots.txt
```

```
User-agent: bingbot  
Disallow: /
```

http://hoge hoge.com/test 以下を侵入禁止にする例

```
User-agent:
```

*

```
Disallow: /test/
```

クローリングの間隔を設定する例

```
User-agent: ロボットの名前  
Crawl-delay: 秒数
```

秒数は、クローラーによって分だったりするので注意。

例えば、bing bot のクローリングの間隔を 60 秒する場合は

```
User-agent: bingbot  
Crawl-delay: 60
```

meta タグ

head 内に書くタグによる指定を行えば、各ページに対する設定を行うことができます。

```
<meta name="robots" content="noindex,nofollow">
```

content 属性の指定は

index

検索データベースへの登録を許可します。

noindex

検索データベースへの登録を禁止します。

follow

このページに含まれるリンクをたぐることを許可します。

nofollow

このページに含まれるリンクをたぐることを禁止します。