

## robots.txt これは何？

プロバイダーからのアクセスログを見ると「Code404-Not Found」ファイルが9.5%もあった。リンク切れがそんなにもあるのかと調べてみた。その結果、小生の頁にはないファイルを探していて、NotFoundになっているのを発見した。その中に、robots.txt と、言うファイル名があった。ネットで調べると、最近、このファイルが使われるようになったとのこと。結論から言うと、ロボット検索を回避するフォルダーやファイルを指定するファイルだ。と、言うことが解った。

<http://machizukan.net/DK/>

## robots.txt 設置場所と書式内容

### robots.txt 設置場所

robots.txt はそのサイトのトップにおく必要があります。

- <http://www.aaa.bbb/robots.txt>
- × <http://www.aaa.bbb/~page/robots.txt>

上記の前者は有効ですが、後者では機能しません。後者は、個人の頁や無料のWeb頁の多くがこれに当てはまります。しかし、設置しても害にはなりません。

### robots.txt 書式内容

書式は、次の内容です。

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

これは、すべて(\*)の検索ロボットに対して、/ではじまるファイル(つまりはすべてのファイル)を検索データベースに登録しないようにします。

実際の書式内容

```
User-Agent: *  
Disallow: /dk/  
Disallow: /cgi-bin/
```

これは、小生の「machizukan.net」に設置したファイルの内容です。

解説:

全ての検索ロボットに、/dk/ と /cgi-bin/ フォルダを検索しないように。と、言う内容です。ここで、/DK/ はあるが /dk/ フォルダはないので、¥dk¥index.html に /DK/ へ強制的に移動するリストを記しておけば、DK へ移動します。  
/cgi-bin/ にも、index.html を設置して、トップ頁へ移動させれば、よいでしょう。この移動の理由は、robots.txt とアドレス欄に打ち込むと、リストが表示され、dk または cgi-bin というフォルダがあることが解ります。もしそこで、フォルダ名を入れれば、その頁が表示され、検索を回避した理由がなくなります。そのため、そのフォルダには、別の頁へ強制的に移動する index.html リストを記するのが鉄則です。  
そして、頁を開くファイルは別の名前になります。例えば、index\_1.html もしくは abc.html 。

<http://machizukan.net/DK/>

## 検索を回避する別の方法

ページのトップの「head」の中に以下の指定を記せば、検索を規制できます。頁ごとに記する必要があります。

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

解説:

noindex - 検索データベースへの記載を禁止。

nofollow - このページに含まれるリンクをたどることを禁止。

<http://machizukan.net/DK/>