

URL	SuzySpider	FurnitureFinder	NosyBot
http://www.marys-antiques.com/private/payroll.xls	×	×	×
http://www.marys-antiques.com/private/suzy-stuff/taxes.txt	✓	×	×
http://www.marys-antiques.com/dynamic/buy-stuff?id=3546	×	×	×
http://www.marys-antiques.com/dynamic/check-inventory?kitchen	×	✓	×

### 9.4.7 HTML的robot-control元标签

robots.txt 文件允许站点管理员将机器人排除在 Web 站点的部分或全部内容之外。robots.txt 文件的一个缺点就是它是 Web 站点管理员，而不是各部分内容的作者所有的。

HTML 页面的作者有一种更直接的方式可以限制机器人访问那些独立的页面。他们可以直接在 HTML 文档中添加 robot-control 标签。遵循 robot-control HTML 标签规则的机器人仍然可以获取文档，但如果其中有机器人排斥标签，它们就会忽略这些文档。比如，因特网搜索引擎机器人就不会在其搜索索引中包含这个目录了。和 robots.txt 标准一样，鼓励但并不强制使用这个标签。

机器人排斥标签是以如下形式，通过 HTML 的 META 标签来实现的：

```
<META NAME="ROBOTS" CONTENT=directive-list>
```

#### 1. 机器人的META指令

机器人 META 指令有几种不同的类型，而且随着时间的推移，以及搜索引擎及机器人对其行为和特性集的扩展，很可能还会添加一些新的指令。最常用的两个机器人 META 指令如下所示。

- NOINDEX

告诉机器人不要对页面的内容进行处理，忽略文档（也就是说，不要在任何索引或数据库中包含此内容）。

```
<META NAME="ROBOTS" CONTENT="NOINDEX">
```