

这种指令的行为很可能是未定义的，肯定会随机器人实现的不同而有所不同。

2. 搜索引擎的META标签

我们刚刚讨论了机器人的 META 标签，可以用来控制 Web 机器人的爬行和索引行为。所有的机器人 META 标签中都包含了 name="robots" 属性。

238

还有很多其他类型的 META 标签可用，包括表 9-5 所示的各种标签。对内容索引型搜索引擎机器人来说，DESCRIPTION 和 KEYWORDS META 标签都非常有用。

表9-5 其他META标签指令

name=	content=	描 述
DESCRIPTION	< 文本 >	允许作者为 Web 页面定义一个短小的文本摘要。很多搜索引擎都会查看 META DESCRIPTION 标签，允许页面作者指定一些短小的摘要来描述其 Web 页面 <meta name="description" content="Welcome to Mary's Antiques Web site">
KEYWORDS	< 逗号列表 >	关联一个由逗号分隔的 Web 页面描述词列表，为关键字搜索提供帮助 <meta name="keywords" content="antiques,mary,furniture,restoration">
REVISIT-AFTER ²⁵	< 天数 >	告诉机器人或搜索引擎应该在指定天数之后重访页面，估计那时候页面可能会发生变化 <meta name="revisit-after" content="10 days">

9.5 机器人的规范

1993 年，Web 机器人社会的先驱 Martijn Koster 为 Web 机器人的编写者们编写了一个指南列表。有些建议已经过时了，但有很多建议仍然非常有用。在 <http://www.robotstxt.org/wc/guidelines.html> 上可以找到 Martijn 的原始论文“Guidelines for Robot Writers”。

表 9-6 是为机器人设计者和操作人员提供的现代更新，这些更新的建议主要还是建立在原始列表的思想和内容之上的。大部分指南都是针对万维网机器人提出的；但它们同样适用于较小规模的爬虫。

注 25：这个指令很可能没有得到广泛的支持。