9.6.5 对结果进行排序，并提供查询结果

一旦搜索引擎通过其索引得到了查询结果，网关应用程序会获取结果，并将其拼成
244 结果页面提供给终端用户。

很多 Web 页面都可能包含任意指定的单词，所以搜索引擎采用了一些很聪明的算法，尝试着对结果进行排名。比如，在图 9-8 中，单词 best 出现在很多文档中；为了将相关度最高的结果提供给用户，搜索引擎要知道应该按照什么顺序来提供结果列表中的文档。这被称为*相关性排名*（relevancy ranking）——这是对一系列搜索结果的评分和排序处理。

为了更好地辅助这一进程，在爬行 Web 的过程中都会进行数据统计。比如，对指向指定页面的链接进行计数有助于判断其流行程度，还可以用此信息来衡量提供结果的顺序。算法、爬行中获取的辅助信息以及搜索引擎所使用的其他技巧都是保守最森严的秘密。

9.6.6 欺诈

在搜索请求得到的前几个结果中没有看到自己想要查找的内容时，用户通常会感到很沮丧，因此，查找站点时搜索结果的顺序是很重要的。在搜索网管们认为能够最
245 好地描述其站点功能的单词时，会有众多因素激励着这些网管，努力使其站点排在靠近结果顶端的位置上，尤其是那些依赖于用户找到它们，并使用其服务的商业站点。

这种对较好排列位置的期待引发了很多对搜索系统的博弈，也在搜索引擎的实现者和那些想方设法要将其站点列在突出位置的人之间引发了持久的拉锯战。很多网管都列出了无数关键字（有些是毫不相关的），使用一些假冒页面，或者采用*欺诈*（spoof）行为——甚至会用网关应用程序来生成一些在某些特定单词上可以更好地欺骗搜索引擎相关性算法的假冒页面。

这么做的结果就是，搜索引擎和机器人实现者们要不断地修改相关性算法，以便更有效地抓住这些欺诈者。

9.7 更多信息

更多有关 Web 客户端的信息，请参见下列资料。

- http://www.robotstxt.org/wc/robots.html
  Web 机器人页面——机器人开发者所需的资源，包括因特网机器人的登记注册。