搜索引擎的工作原理

搜索引擎的工作原理大致可以分为搜集信息、整理信息和接受查询三个部分。

搜集信息：搜索引擎利用称为网络蜘蛛(spider)的自动搜索机器人程序来连接每一个网页上的超链接进行信息搜集。

整理信息：搜索引擎不仅要保存搜集起来的信息，还要将它们按照一定的规则进行编排，以便能够迅速地找到所要的资料。

接受查询：用户向搜索引擎发出查询，搜索引擎接受查询并向用户返回资料。

（1）列举一些同学们知道的搜索引擎及其网址。

（2）搜索引擎通常提供哪两种方式进行信息查询？

（3）利用搜索引擎查询与下列信息有关的站点：

①中国古典四大名著。

②中国的超级计算机发展现状。

四、按关键词搜索信息

利用搜索引擎可以帮助使用者在网络上找到特定的信息，但它们同时也会返回大量无关的信息。因此掌握一些常见的、使用广泛的搜索规则和技巧是很有必要的。

1. 提高搜索准确率的方法

（1）精准搜索。

精准搜索，就是在你要搜索的词上加上双引号（半角），搜索引擎就会完全地匹配你所要的词。例如，在搜索引擎的文字框中输入“电传”，它就会返回网页中有“电传”这个关键字的网址，而不会返回诸如“电话传真”之类的网页。

图11–5　精准搜索

（2）加减运算符搜索。

“+”“—”运算符要求搜索结果中同时包含或不包含特定查询词。例如，在搜索引擎中输入“中央电视台 + 综艺”就表示要查找的内容必须要同时包含“中央、电视台、综艺”这三个关键词（图11–6）；在搜索引擎中输入“电视台— (中央电视台)”，就表示最后的查询结果中一定不包含“中

图11–6　加号运算符搜索

央电视台”（图11-7）。注意：加减号前后须加空格，必要时要加括号。

图11-7　减号运算符搜索

（3）通配符组合搜索。

通配符包括星号（*）和问号（?）。星号（*）表示匹配的数量不受限制。例如，检索式“*百科”，则表示可以检索到百度百科、维基百科、互动百科等词（图11-8）。问号（?）匹配的字符数要受到限制，多用在英文搜索引擎中。例如，当输入“comp?ter”，搜索结果只能找到computer、compater、competer等单词（图11-9）。

图11-8　“*”通配符搜索

图11-9　“?”通配符搜索

（4）布尔检索。

布尔检索是通过标准的布尔逻辑关系来表达关键词与关键词之间逻辑关系的一种查询方法，这种查询方法允许我们输入多个关键词，各个关键词之间的关系可以用逻辑关系词来表示。

① 逻辑“与”。

逻辑“与”一般用“AND”表示，或用空格表示，有的搜索引擎还可以用“&”表示。使用逻辑“与”是为了要求检索结果的页面中同时出现所有输入的检索词，提高查准率。

② 逻辑“或”。

逻辑“或”一般用“OR”表示，或用逗号“，”表示，有的搜索引擎还可以用“|”表示。使用逻辑“或”是为了允许检索结果的 Web 页面中出现输入检索词的任何一个，提高查全率。

③ 逻辑“非”。

逻辑“非”一般用“NOT”表示，有的搜索引擎还可用“！”表示。使用逻辑“非”是为了要求检索结果的 Web 页面中出现 NOT 前的检索词，但排除含有 NOT 后检索词的页面信息，以缩小检索范围。

（5）元词检索。

元词检索即单元词检索。使用元词检索必须保证所用元词记录在检索系

统规范的元词表中。常见的元词有title、image、link、URL等。其格式为：

title：检索词，用于查询网页标题中带有检索词的网页；

image：检索词，用于检索图片；

link：检索词，用于检索链接到某个选定网站的页面；

URL：检索词，用于检索地址中带有某个关键词的网页。

例如在搜索引擎中输入“title：故宫”，就可以查到网页标题中带有故宫的网页，如图11-10所示。

图11-10　元词检索

（6）图片检索。

图片检索除了使用上述的元词image检索外，还可以使用“以图搜图”功能来搜索想要的图片，如百度的以图搜图如图11-11所示。

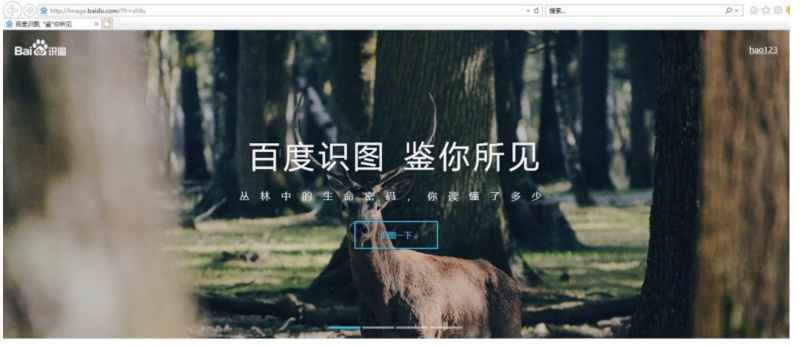

图11-11　以图搜图

2. 搜索结果的准确筛选

在用百度搜索的时候，很多时候搜索出来的结果有成百上千条。如何筛选出和搜索主题相关度最高的结果呢？一般默认地，搜索结果页面排在前面的是相关度较高的，搜索结果以相关度降序排列。在筛选结果时，还可按照时间顺序筛选搜索结果，如图11-12所示。

图11-12　筛选搜索结果

图11-13　搜索结果数量