

# 雅虎、搜狐与 google 搜索引擎的比较

潘 婷

(武汉音乐学院图书馆 武汉 430060)

**摘 要** 本文从信息容量、检索技巧、关键词检索特点、检索繁简等几方面对常用的雅虎、搜狐与 google 搜索引擎进行了介绍与比较,为读者根据需要选择恰当的搜索工具提供一些参考。

**关键词** 搜索引擎 检索策略 信息

自第一个 WWW 搜索引擎——yahoo 出现以来,Internet 上的查询方式焕然一新,而今数百个 WWW 搜索引擎已成为 Internet 的主要查询工具。搜索引擎在帮助用户查询信息方面发挥着重要的作用。下面对比较常用的几种搜索引擎做个介绍与比较。

## 1 雅虎(yahoo!)

Yahoo! 是由大卫·费罗(David Filo)和杨致远(Jerry Yang),美国斯坦福大学电机工程系的博士生于1994年创立的。雅虎在全球共有24个网站,12种语言版本,其中雅虎中国网站(www.yahoo.com.cn)于1999年9月正式开通,它是雅虎在全球的第20个网站。

中文 Yahoo 站点目录分为14个大类,每一个大类下面又分若干子类,搜索十分方便。该站点连接速度快,包含范围广,数据容量大,简便易用,是查询各种信息的好去处。雅虎中国网站(www.yahoo.com.cn)为用户提供了强大的搜索功能,通过其14类简单易用、手工分类的简体中文网站目录及强大的搜索引擎,用户可以轻松搜索到政治、经济、文化、科技、房地产、教育、艺术、娱乐、体育等各方面的信息。

### 1.1 Yahoo! 的使用

在 Yahoo 的主页或任一查询结果返回顶部和底部,你都会看见一个输入框。如果你很清楚你要找的网站(或新闻主题),你可以在输入框内键入你想要找的关键字串(Keyword),然后单击右侧的搜索按钮后,Yahoo 就会从它四个方面的

数据库中找出相匹配的记录,它们是:Yahoo 目录、Yahoo 网点、Yahoo 网上事件和谈话、最新新闻。查询结果返回的是一页与关键词匹配的记录列表,最前面的是 Yahoo 目录链,其后是 Yahoo 网站,网站记录通常由标题(以链接形式出现)和简介组成。如果在 Yahoo 目录和网站中都没有相匹配的内容,Yahoo 则自动利用其内置的查询机制进行整个 WEB 范围的文档查找。

\* 雅虎支持带“+”、“-”等的进阶检索语法。

\* 运用以下几种进阶检索格式,您会获得更精确的检索结果:

- 利用双引号,来查询完全符合关键字串的网站。
- 例如:键入“中文输入”,会找出出包含中文输入的网站,但会忽略包含“中文形声输入”的网站。
- 指定关键词出现的段落。
- 加 t:在关键字前,搜索引擎仅会查询网站名称。
- 加 u:在关键字前,搜索引擎仅会查询网址(URLs)。
- 利用“+”来限定关键字串一定要出现在结果中。

分类类目后面的“@”表示,这个类目会同时出现多个 Yahoo! 中国的不同分类类目如下。

• 范例 1:“时尚”这个类目会同时被放在“艺术”和“社会与文化”的类目下。

• 范例 2:“音乐剧”会被放在“音乐”和“戏剧”的不同类目下。

只要你点击这个含有“@”的类目,就会链接

至 Yahoo! 中国的其它相关类目。

## 1.2 检索结果的排列

Yahoo! 中国搜寻引擎会检索两个部分:

Yahoo! 中国的分类类目和资料库中的网站资讯。Yahoo! 中国搜寻引擎会根据分类类目及网站资讯和查询字串的相关程度而列出相关的 Yahoo! 中国类目和网站。影响相关程度的因素有:和查询字串相同的字串多寡。相同愈多,相关程度愈高。

和查询字串完全符合(Exact Match),相关程度高于部分符合。

和查询字串符合的字串位置。网站名称符合查询字串的相关程度高于网址符合查询字串的网站。

## 2 Google

两位斯坦福大学的博士生 Larry Page 和 Sergey Brin 在 1998 年创立了 Google。2000 年 7 月份,Google 替代 Inktomi 成为 Yahoo 公司的搜索引擎,同年 9 月份,Google 成为中国网易公司的搜索引擎。

Google 检索网页数量达 24 亿,搜索引擎中排名第一;它支持多达 132 种语言,包括简体中文和繁体中文;Google 网站只提供引擎功能,没有花里胡哨的累赘,其速度极快,Google 智能化的“手气不错”功能,提供可能最符合要求的网站;Google 的“网页快照”功能,能从 Google 服务器里直接取出缓存的网页。Google 具有独到的图书搜索功能,Google 具有强大的新闻组搜索功能;Google 具有二进制文件搜索功能(PDF,DOC,SWF 等)。

### 2.1 初阶搜索

最基本的搜索,即查询包含单个关键字的信息。但是,单个关键字搜索得的信息浩如烟海,而且绝大部分并不符合自己的要求,这就需要进一步缩小搜索范围和结果。

### 2.2 基础搜索语法

(1) 搜索结果要求包含两个及两个以上关键字。一般搜索引擎需要在多个关键字之间加上“+”,而 Google 无需用明文的“+”来表示逻辑“与”操作,只要空格就可以了。比如搜得的网页上有“搜索引擎”和“历史”两个关键字。

示例:搜索所有包含关键词“搜索引擎”和“历

史”的中文网页

搜索:“搜索引擎历史”

结果:已搜索有关搜索引擎 历史的中文(简体)网页。共约有 254,000 项查询结果,搜索用时 0.28 秒。

用了两个关键字,查询结果已经少了很多项。但查看一下搜索结果,发现所列的绝大部分结果还是不符合要求,大部分网页涉及的“历史”,并不是我们所需要的“搜索引擎的历史”。怎么办呢?删除与搜索引擎不相关的“历史”。我们发现,这部分无用的资讯,总是和“文化”这个词相关的,另外一些常见词是“中国历史”、“世界历史”、“历史书籍”等。

(2) Google 用减号“-”表示逻辑“非”操作。“A-B”表示搜索包含 A 但没有 B 的网页。

示例:搜索所有包含“搜索引擎”和“历史”但不含“文化”、“中国历史”和“世界历史”的中文网页

搜索:“搜索引擎 历史-文化-中国历史-世界历史”

结果:已搜索有关引擎 历史-文化-中国历史-世界历史的中文(简体)网页。共约有 154,000 项查询结果,搜索时用 0.26 秒。

我们看到,通过去掉不相关信息,搜索结果又减少了将近一半。第一个搜索结果是:

中国电力通信网,里面包括搜索引擎历史/前生今世

[www.dt365.com/list.asp?id](http://www.dt365.com/list.asp?id)

非常符合搜索要求。另外第六项搜索结果:

搜索引擎发展历史=推广中国网

[www.aspl69.com/nous/.htm](http://www.aspl69.com/nous/.htm) 也符合搜索要求。但是,10 个结果只有两个符合要求,未免太少了点。不过,在没有更好的策略之前,不妨先点开一个结果看看。这个名为“搜索引擎发展历史”的网页,我们发现,搜索引擎的历史,是与互联网早期的文件检索工具“Archie”息息相关的。此外,搜索引擎似乎有个核心程序,叫“蜘蛛”,而最早成型的搜索引擎是“Lycos”,使搜索引擎深入人心的是“Yahoo”。了解了这些信息,我们就可以进一步的让搜索结果符合要求了。

注意:这里的“+”和“-”号,是英文字符,而不是中文字符的“+”和“-”。此外,操作符与作用的关键字之间,不能有空格。比如“搜索引擎-文化”,搜索引擎将视为关键字为“搜索引擎”和“文

化”的逻辑“与”操作,中间的“—”被忽略。

(3) 搜索结果至少包含多个关键字中的任意一个。

Google 用大写的“OR”表示逻辑“或”操作。搜索“A OR B”,意思就是说,搜索网页中,要么有 A,要么有 B,要么同时有 A 和 B。在上例中,我们希望搜索结果中最好含有“archie”、“Lycos”、“蜘蛛”等关键字中的一个或者几个,这样可以进一步地精简搜索结果。

示例:搜索如下网页,要求必须含有“搜索引擎”和“历史”,没有“文化”,可以含有以下关键字中人任何一个或者多个:“Archie”、“蜘蛛”、“Lacos”、“Yahoo”。

搜索:“搜索引擎 历史 archie OR 蜘蛛 OR lycos Or Yahoo—文化”

结果:已搜索有关搜索引擎 历史 archie OR 蜘蛛 OR lycos Or Yahoo—文化的中文(简体)网页。共约有 8,960 项查询结果,搜索用时 0.41 秒。我们看到,搜索结果缩小到 8 千多项,前 20 项结果中,大部分都符合搜索要求。

注意:“与”操作必须用大写的“OR”,而不是小写的“or”。在上面的例子中,我介绍了搜索引擎最基本的语法“与”“非”和“或”,这三种搜索语法 Google 分别用“ ”(空格)、“—”、“OR”表示。顺着上例的思路,你也可以了解到如何缩小搜索范围,迅速找到目的资讯的一般方法:目录信息一定含有的关键字(用“ ”连起来),目标信息不能含有的关键字(用“—”去掉),目标信息可能含有的关键字(用“OR”连起来)。

## 2.3 杂项语法

### (1) 通配符问题

很多搜索引擎支持通配符号,如“\*”代表一连串字符,“?”代表单个字符等。Google 对通配符支持有限。它目前只可以用“\*”来替代单个字符,而且包含“\*”必须用“ ”引起来。比如,“ ”以 \* 治国,表示搜索第一个为“以”,末两个为“治国”的四字短语,中间的“\*”可以为任何字符。

### (2) 关键字的字母大小写

Google 对英文字符大小写不敏感,“GOD”“god”搜索的结果是一样的。

### (3) 搜索整个短语或者句子

Google 的关键字可以是单词(中间没有空格),也可以是短语(中间有空格)。但是,用短语做关键字,必须加英文引号,否则空格会被当作“与”

操作符。

## 2.4 进阶搜索

上面已经探讨了 Google 的一些最基础搜索语法。通常而言,这些简单的搜索语法已经能解决绝大部分问题了。不过,如果想更迅速更贴切找到需要的信息,你还需要了解更多的东西。

### (1) 对搜索的网站进行限制

“site”表示搜索结果局限于某个具体网站或者网站频道,如 www.sina.com.cn、“edu.sine.com.cn”,或者是某个域名,如“com.cn”、“com”等等。如果是要排除某网站或者域名范围内的页面,只需用“—网站/域名”。

示例:搜索中文教育科研网站(edu.cn)上关于搜索引擎技巧的页面。

搜索:“搜索引擎 技巧 site:edu.cn”

结果:已搜索有关搜索引擎 技巧 site:edu.cn 的中文(简体)网页。共约有 851 项查询结果,搜索用时 0.17 秒。

### (2) 在某一类文件中查找信息

“filetype:”是 Google 开发的非常强大实用的一个搜索语法。也就是说,Google 不仅能搜索一般的文字页面,还能对某些二进制文档检索。目前,Google 已经能检索微软的 Office 文档如 xls、ppt、doc、.rtf, WordPerfect 文档, Lotus 1—2—3 文档, Adobe 的 pdf 文档, ShockWave 的 swf 文档(Flash)等。其中最实用的文档搜索是 PDF 搜索。PDF 是 ADOBE 公司开发的电子文档格式,现在已经成为互联网的电子化出版标准。目前 Google 检索的 PDF 文档大约有 2500 万左右,大约占有索引的二进制文档数量的 80%。PDF 文档通常是一些图文并茂的综合性文档,提供的资讯一般比较集中全面。

示例:搜索几个资产负债表的 Office 文档

搜索:“资产负债表 filetype:doc OR filetype:xls OR filetype:ppt”

结果:已搜索有关资产负债表 filetype:doc OR filetype:xls OR filetype:ppt 的中文(简体)网页。共约有 2,790 项查询结果,搜索用时 0.35 秒。

使用 Yahoo! 中国搜索时, Yahoo! 本身的数据库以及它的搜索引擎合作伙伴 Google, 组成了所得到的搜索结果。

如果搜索的字词在 Yahoo! 中国的数据库内, 那么搜索结果会在“相关类目”和/或“相关网站”

中。(搜索结果页面上方的工具条上)如果搜索的字词在 Google 的数据库中,那么搜索结果会在“相关网页”中。Yahoo! 中国目录采用专业人工分类,不但可以直接当成目录来浏览,还可以用来搜寻您想要的内容。Google 则是一个全自动搜索引擎,它是利用电脑程序直接在网页抓取相关字。

### 3 搜狐(sohu)

搜狐公司成立于 1996 年 8 月,是由公司创办人张朝阳博士在美国依靠 MIT 媒体实验室主任尼葛洛庞帝先生和美国风险投资专家爱德华·罗伯特先生的风险投资的支持下创办的。搜狐公司于 1998 年推出中国首家大型分类查询搜索引擎,经过数年的发展,每日浏览量超过 800 万。到现在已经发展成为中国影响力最大的分类搜索引擎。累计收录中文网站达 150 多万,每日页面浏览量超过 800 万,每天收到 2000 多个网站登录请求。

#### 3.1 搜狐搜索引擎的特点

搜狐的目录导航式搜索引擎完全是由人工加工而成,相比机器人加工的搜索引擎来讲具有很高的精确性、系统性和科学性。分类专家层层细分类目,组织庞大的树状类目体系。利用目录导航系统可以很方便地查找到一类相关信息。Sohu 的搜索引擎可以查找网站、网页、新闻、网址、软件五类信息。网站和网页这两类信息的区别就象是一本书和书中的每一篇文章一样。Sohu 的网站搜索是以网站作为收录对象,具体的方法就是将每个网站首页的 URL 提供给搜索用户,并且将网站的题名和整个网站的内容简单描述一下,但是并不揭示网站中每个网页的信息。网页搜索就是将每个网页作为收录对象,揭示每个网页的信息,信息的揭示比较具体。新闻搜索可以搜索到搜狐新闻的内容。网址搜索是 3721 提供的网络实名查找。

#### 3.2 检索方法

(1) 关键词查找,用户可以在搜索框中直接输入自己想查找信息的关键词,找到相关信息。这种方法对网站、网页、新闻、网址、软件五类信息都

适用。

用最少的词表达清楚所查信息的主题,比如想查流氓兔动画,只需要输入“流氓兔”就可以了,不需要加上“动画”,因为“流氓兔”就是一个动画作品。少用修饰词,如果检索结果太多,可以用修饰词去掉一些不想要的信息。太长的关键词改用逻辑组合,可以用空格、“+”、“and”等的符号进行逻辑与的组合搜索。比如想查孙燕姿的歌曲,可以输入“孙燕姿”“歌曲”,中间空一格或是加上一个“+”号组合检索就可以了,不能用“孙燕姿的歌曲”来查,那样会把许多相关的信息漏掉。

(2) 目录导航,用户层层点击想查找信息的类目,通过这种方法可以找到相关的一类信息。这种方法只适用于网站搜索。目录导航检索是按照信息所属的类别层层点击查找信息,所以用目录导航检索时首先要考虑清楚想要查找的信息属于哪个类别。比如查找“计算机杀毒软件”,首先浏览搜狐的十八大类,看到“IT”类目,应当是属于这个类目,点击进入下面有“软件”,点击“软件”进入下面有“病毒专区”,再点击“病毒专区”进入下面有“专杀工具下载”,最后点击进入“杀病毒软件”就会找到许多有关病毒软件的网站。目录导航检索的结果如下图所示:

首页>IT>软件>病毒专区>专杀工具下载。

总之,几种搜索引擎互有利弊,各有千秋,可根据自身的需要选取不同的搜索引擎或相互配合使用,达到自己的检索目的。

#### 参考文献

- 1 朱晓云. 搜索引擎检索效率研究. 津图学刊, 2002(3)
- 2 徐家坤. 搜索引擎的实用检索技巧. 科技情报开发与经济, 2003(1)
- 3 卢晓勤. 搜索引擎的选择与检索策略. 情报科学, 2002(4)

潘婷, 武汉音乐学院图书馆。