

第 3 章

商务信息检索与利用

本章学习目的

- 掌握百度搜索引擎的高级使用方法。
- 掌握 360 搜索引擎的高级使用方法。
- 掌握综合搜索引擎的高级使用方法。

实验 3.1 百度搜索引擎的使用

【实验目的】

- 了解百度搜索引擎过程。
- 掌握百度搜索引擎的高级使用方法。

【实验条件】

- 个人计算机一台,基本配置: CPU Core2 以上,内存 2GB 以上,硬盘空间 20GB 以上,100 兆网卡。
- 个人计算机预装 Windows XP 操作系统和浏览器。
- 具有网络连接,可以连接 Internet 网络。

【实验内容和步骤】

打开 IE 浏览器,在地址栏输入“<http://www.baidu.com>”百度的域名,打开百度的首页,如图 3.1 所示。

在百度搜索栏中输入需要搜索的关键词,比如 GDP,查询结果如图 3.2 所示。可以同时输入多个关键词进行检索,“中国人均 GDP”,如图 3.3 所示。搜索引擎希望了解的是 GDP 数字是“中国”的,而且是“人均”的,所以结果网页中一定会有“中国”、“人均”两个关键词。需要运用尝试法,可以先试试你“人均 GDP”或“中国 GDP”试试。实际上,在搜索精确信息的时候,必须考虑的一点就是要找到能够涵盖这一信息的关键词,同时还要考虑可能替换的关键词,这样才能很快搜索到目标。

在查询结果中寻找并单击“中国赶超之后,美国何去何从?”一文,结果如图 3.4 所示。



图 3.1 百度网站的首页



图 3.2 百度搜索“GDP”页面



图 3.3 百度搜索“中国 人均 GDP”页面



图 3.4 打开搜索结果页面

尝试使用“中国 各省 GDP”检索,得到相关检索结果,如图 3.5 所示。

在搜索页面中单击“中国大陆各城市 GDP 排名?”一文,浏览器打开新的窗口,结果如图 3.6 所示。



图 3.5 百度搜索“中国 各省 GDP”页面



图 3.6 “中国大陆各城市 GDP 排名”页面

【相关知识】

1. 搜索技巧

最基本同时也是最有效的,就是选择合适的查询词。选择查询词是一种经验积累,同时也是有规律的。查询词需要表述准确。百度会严格按照您提交的查询词去搜索,因此,查询词表述准确是获得良好搜索结果的必要前提。通常表述不准确的情况是,脑袋里想着一回事,搜索框里输入的是另一回事。例如,要查找 2004 年国内十大新闻,查询词可以是“2004 年国内十大新闻”;但如果把查询词换成“2004 年国内十大事件”,搜索结果就没有能满足需求的了。另一类典型的表述不准确,是查询词中包含错别字。例如,要查找林心如的图片,用“林心如”,当然是没什么问题;但如果写错了字,变成“林心茹”,搜索结果质量就差得远了。不过好在,百度对于用户常见的错别字输入,有纠错提示。您若输入“林心茹”,在搜索结果上方,会提示“您要找的是不是:林心如”。

查询词的主题关联与简练。目前的搜索引擎并不能很好地处理自然语言。因此,在提交搜索请求时,您最好把自己的想法,提炼成简单的,而且与希望找到的信息内容主题关联的查询词。还是用实际例子说明。某三年级小学生,想查一些关于时间的名人名言,他的查询词是“小学三年级关于时间的名人名言”。这个查询词很完整地体现了搜索者的搜索意图,但效果并不好。绝大多数名人名言,并不规定是针对几年级的,因此,“小学三年级”事实上和主题无关,会使得搜索引擎丢掉大量不含“小学三年级”,但非常有价值的信息;“关于”也是一个与名人名言本身没有关系的词,多一个这样的词,又会减少很多有价值信息;“时间的名人名言”,其中的“的”也不是一个必要的词,会对搜索结果产生干扰;“名人名言”,名言通常就是名人留下来的,在名言前加上名人,是一种不必要的重复。因此,最好的查询词,应该是“时间名言”。

根据网页特征选择查询词。很多类型的网页都有某种相似的特征。例如,小说网页,通常都有一个目录页,小说名称一般出现在网页标题中,而页面上通常有“目录”两个字,单击页面上的链接,就进入具体的章节页,章节页的标题是小说章节名称;软件下载页,通常软件名称在网页标题中,网页正文有下载链接,并且会出现“下载”这个词。等等。经常的搜索,并且总结各类网页的特征现象,并应用于查询词的选择中,就会使得搜索变得准确而高效。例如,找明星的个人资料页。一般来说,明星资料页的标题,通常是明星的名字,而在页面上,会有“姓名”、“身高”等词语出现。比如找林青霞的个人资料,就可以用“林青霞姓名 身高”来查询。而由于明星的名字一般在网页标题中出现,因此,更精确的查询方式,可以是“姓名 身高 intitle:林青霞”。intitle,表示后接的词限制在网页标题范围内。这类主题词加上特征词的查询构造方法,适用于搜索具有某种共性的网页。前提是,您必须了解这种共性(或者通过试验性搜索预先发现共性)。

我们在工作和生活中,会遇到各种各样的疑难问题,比如电脑中毒了,被开水烫伤了等等。很多问题其实都可以在网上找到解决办法。因为某类问题发生的几率是稳定的,而网络用户有好几千万,于是几千万人中遇到同样问题的人就会很多,其中一部分人会把问题贴在网络上求助,而另一部分人,可能就会把问题解决办法发布在网络上。有了搜索引擎,我们就可以把这些信息找出来。

找这类信息,核心问题是如何构建查询关键词。一个基本原则是,在构建关键词时,我们尽量不要用自然语言(所谓自然语言,就是我们平时说话的语言和口气),而要从自然语言中提炼关键词。这个提炼过程并不容易,但可以用一种将心比心的方式思考:如果我知道问题的解决办法,我会怎样对此作出回答。也就是说,猜测信息的表达方式,然后根据这种表达方式,取其中的特征关键词,从而达到搜索目的。

例如,上网时经常会遇到陷阱,浏览器默认主页被修改并锁定。这样一个问题的解决办法,应该怎样搜索呢?首先要确定的是,不要用自然语言。比如,有的人可能会这样搜索“我的浏览器主页被修改了,谁能帮帮我呀”。这是典型的自然语言。口语化的搜索词也可以给出适当的答案,但是这样的搜索常常得不到最想要的结果。我们来看这个问题中的核心词汇。对象:浏览器(或者 IE)的主页。事件:被修改(锁定)。“浏览器”、“主页”和“被修改”,在这类信息中出现的概率会最大,IE 可能会出现,至于锁定,用词比较专业化,不见得能出现。于是关键词中,至少应该出现“浏览器”、“主页”和“被修改”,这是问题现象描述。一般情况下,只要对问题作出适当的描述,在网上基本上就可以找到解决对策。

2. 使用多个词语搜索

输入多个词语搜索(不同字词之间用一个空格隔开),可以获得更精确的搜索结果。例如:想了解上海人民公园的相关信息,在搜索框中输入[上海 人民公园]获得的搜索效果会比输入[人民公园]得到的结果更好。

(1) 百度的高级搜索和个性设置。

`http://help.baidu.com/question?prod_en=webmaster&class=123&id=1375`

(2) 提高效率的高级搜索语法。把搜索范围限定在网页标题——intitle 中。网页标题通常是对网页内容提纲挈领式的归纳。把查询内容范围限定在网页标题中,有时能获得良好的效果。使用的方式,是把查询内容中,特别关键的部分,用“intitle:”领起来。

例如,找林青霞的写真,就可以这样查询:“写真 intitle:林青霞”注意,intitle:和后面的关键词之间,不要有空格。

把搜索范围限定在特定站点——site 中。有时候,您如果知道某个站点中有自己需要找的东西,就可以把搜索范围限定在这个站点中,提高查询效率。使用的方式,是在查询内容的后面,加上“site:站点域名”。

例如,天空网下载软件不错,就可以这样查询:“msn site:skycn.com”注意,“site:”后面跟的站点域名,不要带“http://”;另外,site:和站点名之间,不要带空格。

把搜索范围限定在 url 链接——inurl 中。网页 url 中的某些信息,常常有某种有价值的含义。于是,您如果对搜索结果的 url 做某种限定,就可以获得良好的效果。实现的方式,是用“inurl:”,后跟需要在 url 中出现的关键词。

例如,找关于 photoshop 的使用技巧,可以这样查询“photoshop inurl:jiqiao”上面这个查询串中的“photoshop”,是可以出现在网页的任何位置,而“jiqiao”则必须出现在网页 url 中。注意,inurl:语法和后面所跟的关键词,不要有空格。

精确匹配——双引号和书名号如果输入的查询词很长,百度在经过分析后,给出的搜索结果中的查询词,可能是拆分的。如果您对这种情况不满意,可以尝试让百度不拆分查询词。给查询词加上双引号,就可以达到这种效果。

例如,搜索 上海科技大学,如果不加双引号,搜索结果被拆分,效果不是很好,但加上双引号后,“上海科技大学”,获得的结果就全是符合要求的了。书名号是百度独有的一个特殊查询语法。在其他搜索引擎中,书名号会被忽略,而在百度,中文书名号是可被查询的。加上书名号的查询词,有两层特殊功能,一是书名号会出现在搜索结果中;二是被书名号扩起来的内容,不会被拆分。书名号在某些情况下特别有效果,例如,查名字很通俗和常用的那些电影或者小说。比如,查电影“手机”,如果不加书名号,很多情况下出来的是通信工具——手机,而加上书名号后,《手机》结果就都是关于电影方面的了。

要求搜索结果中不含特定查询词如果您发现搜索结果中,有某一类网页是您不希望看见的,而且,这些网页都包含特定的关键词,那么用减号语法,就可以去除所有这些含有特定关键词的网页。

例如,搜神雕侠侣,希望是关于电视剧的,却发现很多关于游戏方面的网页。那么就可以这样查询:“神雕侠侣—游戏”注意,前一个关键词,和减号之间必须有空格,否则,减号会被当成连字符处理,而失去减号语法功能。减号和后一个关键词之间,有无空格均可。

3. 搜索引擎的工作原理

搜索引擎,通常指的是收集了因特网上几千万到几十亿个网页并对网页中的每一个词(即关键词)进行索引,建立索引数据库的全文搜索引擎。当用户查找某个关键词的时候,所有在页面内容中包含了该关键词的网页都将作为搜索结果被搜出来。在经过复杂的算法进行排序后,这些结果将按照与搜索关键词的相关度高低,依次排列。搜索引擎派出一个能够在网上发现新网页并抓文件的程序,这个程序通常称之为蜘蛛(Spider)。搜索引擎从已知的数据库出发,就像正常用户的浏览器一样访问这些网页并抓取文件。搜索引擎通过这些爬虫去爬互联网上的外链,从这个网站爬到另一个网站,去跟踪网页中的链接,访问更多的网页,这个过程就叫爬行。这些新的网址会被存入数据库等待搜索。所以跟踪网页链接是搜索引擎蜘蛛(Spider)发现新网址的最基本的方法,所以反向链接成为搜索引擎优化的最基本因素之一。搜索引擎抓取的页面文件与用户浏览器得到的完全一样,抓取的文件存入数据库。

蜘蛛抓取的页面文件被分解、分析,并以巨大表格的形式存入数据库,这个过程即是索引(index)。在索引数据库中,网页文字内容,关键词出现的位置、字体、颜色、加粗、斜体等相关信息都有相应记录。用户在搜索引擎界面输入关键词,单击“搜索”按钮后,搜索引擎程序即对搜索词进行处理,如中文特有的分词处理,去除停止词,判断是否需要启动整合搜索,判断是否有拼写错误或错别字等情况。搜索词的查询非常快速,对搜索词处理后,搜索引擎程序便开始工作,从索引数据库中找出所有包含搜索词的网页,并且根据排名算法计算出哪些网页应该排在前面,然后按照一定格式返回到“搜索”页面。再好的搜索引擎也无法与人相比,这就是网站要进行搜索引擎优化的原因。

实验 3.2 360 搜索引擎的使用

【实验目的】

- 了解 360 搜索引擎过程。
- 掌握 360 搜索引擎的高级使用方法。

【实验条件】

- 个人计算机一台,基本配置: CPU Core2 以上,内存 2GB 以上,硬盘空间 20GB 以上,100 兆网卡。
- 个人计算机预装 Windows XP 操作系统和浏览器。
- 具有网络连接,可以连接 Internet 网络。

【实验内容和步骤】

打开浏览器输入“http://www. so. com”,打开 360 搜索网页页面,如图 3.7 所示。



图 3.7 360 搜索主页

1. 在指定站点内进行搜索

如果想知道某个站点内是否有您想要的内容,可以将搜索的范围限定在这个站点中。例如,如果想在 360 论坛搜索有关“360 杀毒”的信息,可以在搜索框输入“360 杀毒 site:bbs.360safe.com”,如图 3.8 所示。



图 3.8 指定站点搜索

2. 精确匹配搜索

1) 双引号“”

如果您的查询词很长,360 搜索可能会根据拆分后的查询词给出搜索结果,如果您不想输入的查询词被拆分,可以给您的查询词加上双引号。例如,如果您只想搜索北京大学的相关信息,而不是北京所有大学的信息,您可以在搜索框中输入“北京大学”,如图 3.9 所示。



图 3.9 双引号匹配搜索

2) 书名号《》

在 360 搜索中,中文书名号是可以被查询的,加上书名号的查询词会有两个功能:一是可以在搜索结果中显示带书名号的关键词;二是保证这个词不会被拆分。例如,如果您想搜索书籍“《西游记》”,就可以加上书名号以便精确查找书籍相关的内容,如图 3.10 所示。



图 3.10 书名号匹配搜索

实验 3.3 综合搜索网站的使用

【实验目的】

- 了解综合搜索引擎过程。
- 掌握综合搜索引擎的高级使用方法。

【实验条件】

- 个人计算机一台,基本配置:CPU Core2 以上,内存 2GB 以上,硬盘空间 20GB 以上,100 兆网卡。
- 个人计算机预装 Windows XP 操作系统和浏览器。
- 具有网络连接,可以连接 Internet 网络。

【实验内容和步骤】

用浏览器打开 <http://www.manmanbuy.com/>慢慢买网站,如图 3.11 所示。慢慢买是一个专业的网购比价平台,为用户提供最实用的比价功能及各类省钱方法。

1. 比价搜索

比价是慢慢买的核心功能,让用户轻松对比同一商品在各个商城的不同报价,了解该商品最低价。使用比价方式,打开网站的任何页面,顶部的搜索框内直接输入商品名称或型号进行比价,如图 3.12 所示。

比如在搜索栏中搜索电视“TCL L48A71”就会显示该型号电视在不同商家的报价,在网页最上方给出这个商品报价最便宜的商家,下面列出各个商城这台电视的最新报价。为了保障商品的网购安全和售后保障,网站只收录 B2C 行业商城的报价,如图 3.13 所示。



图 3.11 慢慢买网站首页

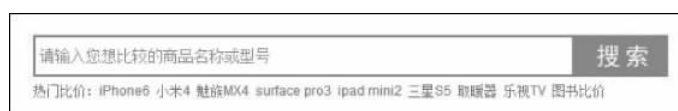


图 3.12 比价搜索(a)



图 3.13 比价搜索(b)

2. 查历史最低价

打开网页的右上部分有该商品的历史价格走势和最低价。通常在网购过程中不良商家的各类活动欺骗手段,比如有的不良商家通常先提高商品的销售价格,然后打着打折促销的名义进行宣传。所以比价网站打开每个商品页面的同时,都会有历史最低价的走势图显示,如图 3.14 所示。



图 3.14 商品价格走势

通过查询最近一周、一个月、三个月这个商品的走势图,可以查看这个商品在近期的平均价格和历史最低价格,为消费者提供购买商品很好的参考。

3. 设置降价提醒

对某个商品设置降价提醒后,当该商品有降价时,第一时间邮件通知。使用方式:打开单个商品页,找到“降价提醒”按钮,单击后在弹出窗口填写接收邮件地址即可,如图 3.15 所示。



图 3.15 设置降价提醒

4. 白菜价

所谓白菜价就是特价商品,比如天猫店铺为了一个款式商品,前期会亏本冲销量,表现形式一般是限时特卖,所以网站每天为用户整理高品质的限时特卖商品。这样对于用户来说便宜一样有好货。输入 <http://baicai.manmanbuy.com/> 进入慢慢买白菜价,如图 3.16 所示。

5. 监控降价信息

清楚了解以上功能后,相信用户在网购过程中再也不会吃亏,感觉买贵了。将慢慢买介绍给其他用户。接下来你可以开始跟大家分享信息了,慢慢买提供了监控各商城降价信息的功能,如果你有时间时刻关注,有可能第一时间发现优质商品的降价信息,bug 价等,并通过省钱控的“我要爆料”进行分享。打开“省钱控”下的“降价监控”就可以打开降价监控页面,如图 3.17 所示。



图 3.16 慢慢买白菜价栏目页面



图 3.17 降价监控页面

【相关知识】

比价(Price Comparison),即价格比较。随着网络购物的发展,各电商平台迅猛扩张,消费者也不再只固定选择一个网购平台,而是综合考虑找到最适合自己的下单方式。在各电商平台,一款相同的产品因为采购渠道、运营成本和促销的不同而产生不同的零售价格,您可以通过比较价格、物流、服务来决定如何购买,从而节约时间和成本。因为有了消费者比价的需求,产生了大量的比价网站。最早的比价网站产生于美国,现在在美国通过比价后再进行购买的网购行为已经接近40%,目前在中国比价逐渐被消费者接受,网购中这一数量也在逐渐上升中。

20世纪90年代末的欧美大陆,互联网繁荣让比价服务有利可图。发展到至今,比价服务已转移至提供比价信息的专业网站和门户网站。20世纪90年代后期,随着人们获取网络资源的渠道越来越便捷,一系列购物门户网站开始兴起。这类网站主要根据商品种类来罗列不同的销售供应商,这些供应商也为此交付一定的费用。这就有些类似于网络在线版的企业黄页。随着网络技术的改进,一种新型的门户网站逐渐诞生,改变了此类网站的盈利模式、特征和提供的功能。这些网站不会聚集销售商提供的数据信息,它们直接从每个商家的网站上搜索和抓取所需要的信息。这也就使得销售商的清单更加全面,同时还可以随时更新数据。在1998年和1999年之间,大量公司研发出类似的价格信息搜索技术,并将之存于企业的数据库中。用户便利用其搜索特定商品,可以从网页上看到针对产品的一系列零售商和价格。广告商并不为排位,而是每次的单击率付钱。成立于1997年的streetprices,就是这一盈利方式的最早实践者,并在1998年新增了商品图像和电子邮件提醒功能。其他价格搜索引擎公司也逐渐演化出价格追踪工具,比如价格下调提醒和价格变化记录的保持。

比价网站产生于美国,不过现如今在中国也有很多比价网站,他们为消费者进行网上购物时找到最便宜、最合理的商品价格提供了极大的便利。主要的比价领域是图书类,即图书比价网(如买书吗?),不过也有提供全方位的比价网站出现(如比价网,安图搜,爱购网,三脉网等)。比价网站一般不收取用户任何费用,而是从消费者在入驻网站的支付活动中赚取利润。商业模式也因公司而异。零售商或者缴纳网站的摊位费,或者为每次用户单击零售网站而付费。甚至用户完成某个特定环节和活动,也可以成为付费的渠道,比如说当用户购买某件商品,或者用电子邮件来注册。比价购物网站从相关的网络营销联盟中获取大量的商品数据传送。也有专门为比价服务从事搜集商品数据的公司。

比价搜索引擎以非人工检索方式、根据您的输入的关键字自动生成到第三方网页的链接,相关内容来自搜索结果所链接到的第三方网站。

“人肉”比价,这个“人肉比价”不是指人的肉的价格,而是指靠网友的力量来比较商品的价格,这是典型的互联网精神。从一些人肉比价网站可以看到,一些商品若积累了一定UGC(用户产生的内容),其价格信息相当可观,作为用户谈价格的情报来源是很有用的,是一种有效的比价形式。

比价网的优缺点:

(1) 比价网站的优点是允许个人同时比较海量的商品,并带有相关的附加服务,省时省力。比价网站一般都可以保证展示商城和产品的合法性与安全性,因为这是比价网站存在的最基本原则与底线。比价网站可以为使用者尽量节省花销,特别是对于产品质量和规格

一致,耐用型的品牌电子数码产品,比价网站提供的最低价格无疑是节省不必要支出的利器。比价网站对用户都是免费的,用户也没有义务去从中购买任何商品。

(2) 比价网站也会造成一些不利的因素。比如会给消费者一种误解,认为比价网站上的最低价就是市场的最低价,实际意义上的市场广阔复杂,不是比价网站可以包括的,并且部分商家并未参与到比价网站的营销方式中,但也有可能提供更低的价格。消费者在消费过程中会逐渐造成对比价网的依赖感。

(3) 比价网站是网络购物体系的健全和网上商城兴起后的产物。随着国内电子商务近些年发展的逐步升温,又由于生活成本不断上升所导致的压力,人们利用比价网站作为消费指南和参考,将不仅仅成为潮流,更将成为一种生活必需。

习题

1. 在百度上搜索“红楼梦书籍”的有关资料。
2. 使用好搜良医搜索“高血压”病的相关注意事项。
3. 使用比价网搜索长虹电视的某一个型号的比价。