

项目一 网络信息资源的获取

【项目描述】

网络信息资源是指以电子资源数据的形式，将文字、图像、声音、动画等多种形式的信息储存在光、磁等非印刷质的介质中，利用计算机通过网络进行发布、传递、储存的各类信息资源的总和。

电子商务的从业者必须了解企业、行业数据动态和信息资源。了解的渠道多种多样，如线上、线下两种渠道。线下渠道主要有市场调研、查询档案文献、座谈等方式，而线上渠道主要有利用网上搜索引擎、利用权威机构的网站、利用网络专业信息资源导航库、利用高校数字图书馆等。

上述方法可以作为寻找网络信息资源的常规方法。当然，在庞杂、无序及动态的信息中寻找自己所需要的信息，除要掌握这些基本方法外，还需要信息收集者掌握一定的技巧，根据不同数据库的特点，从而充分运用各种逻辑检索规则准确表达检索要求，通过运用多种方法和进行多种尝试，最终收集到有用、可靠的信息。

【教学目标】

通过本项目的学习与训练，学生能够完成以下目标：

- (1) 理解网络信息资源对于电子商务的重要性。
- (2) 掌握网络信息资源获取的常用途径或方法。
- (3) 掌握各种获取工具的使用技巧。

任务一 利用搜索引擎获取网络信息资源

【任务目标】

- (1) 理解搜索引擎的概念。
- (2) 了解国内外常用的搜索引擎。
- (3) 掌握主流搜索引擎的使用方法。

一、知识链接

(一) 搜索引擎的概念

搜索引擎 (Search Engine) 是指根据一定的策略、运用特定的计算机程序从互联网上搜集信息, 对信息进行组织和处理后, 为用户提供检索服务, 将用户检索的相关信息展示给用户的系统。

(二) 搜索引擎的分类

搜索引擎按其工作方式主要可分为三种: 全文搜索引擎 (Full Text Search Engine)、目录索引类搜索引擎 (Search Index/Directory Engine) 和元搜索引擎 (Meta Search Engine)。

1. 全文搜索引擎

全文搜索引擎是名副其实的搜索引擎, 国外具有代表性的有 Google、Inktomi、Teoma、WiseNut 等, 国内具有代表性的有百度 (Baidu)。它们都是通过从互联网上提取各个网站的信息 (以网页文字为主) 而建立的数据库, 检索与用户查询条件匹配的相关记录, 按照一定的排列顺序将结果返回给用户, 因此它们是真正的搜索引擎。

从搜索结果来源的角度, 全文搜索引擎又可细分为两种: 一种是拥有自己的检索程序 (Indexer), 俗称“蜘蛛” (Spider) 程序或“机器人” (Robot) 程序, 并自建网页数据库, 搜索结果直接从自身的数据库中调用, 如上面提到的搜索引擎; 另一种则是租用其他引擎的数据库, 并按自定义格式排列搜索结果, 如 Lycos 引擎。

2. 目录索引类搜索引擎

目录索引虽然有搜索功能, 但从严格意义上讲不算是真正的搜索引擎, 仅仅是按目录分类的网站链接列表而已。用户完全可以不用进行关键词 (Keyword) 查询, 仅靠分类目录也可找到所需的信息。目录索引中最具代表性的莫过于大名鼎鼎的雅虎 (Yahoo), 其他的此类引擎还有 Open Directory Project (DMOZ)、LookSmart、About 等。国内的搜狐、新浪、网易搜索也都属于这一类。

3. 元搜索引擎

元搜索引擎是指在接受用户查询请求时, 同时在其他多个引擎上进行搜索, 并将结果返回给用户。具有代表性的元搜索引擎有 InfoSpace、Dogpile、Vivisimo 等 (元搜索引擎列表), 中文元搜索引擎中具代表性的有搜星搜索引擎。在搜索结果排列方面, 有的直接按来源引擎排列搜索结果, 如 Dogpile; 有的则按自定的规则将结果重新排列组合, 如 Vivisimo。

除上述三大类引擎外, 还有以下几种非主流形式:

(1) 集合式搜索引擎。如 HotBot 在 2002 年推出的引擎。该引擎类似 META 搜索引擎, 但主要区别在于不是同时调用多个引擎进行搜索, 而是由用户从提供的 4 个引擎当中选择, 因此将其称为“集合式”搜索引擎更准确些。

(2) 门户搜索引擎。如 AOL Search、MSN Search 等虽然提供搜索服务, 但自身既没有分类目录也没有网页数据库, 其搜索结果完全来自其他引擎。

(3) 免费链接列表 (Free for All Links)。这类网站一般只简单地滚动排列链接条目, 少部

分有简单的分类目录，不过规模比 Yahoo 等目录索引要小得多。

由于上述网站都为用户提供搜索查询服务，为方便起见，我们通常将其统称为搜索引擎。

（三）搜索引擎的市场份额

Statcounter Global Stats（美国国家统计局全球统计数据）显示，2018 年 2 月—2019 年 2 月搜索引擎市场占有率如图 1-1-1 ~ 1-1-4 所示。

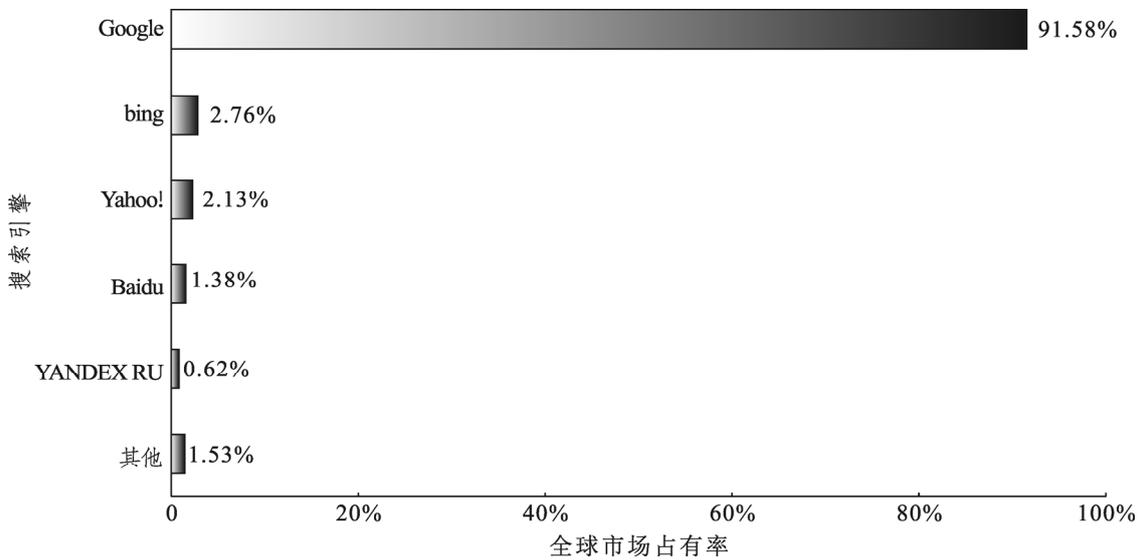


图 1-1-1 2018 年 2 月—2019 年 2 月搜索引擎全球市场占有率排行榜

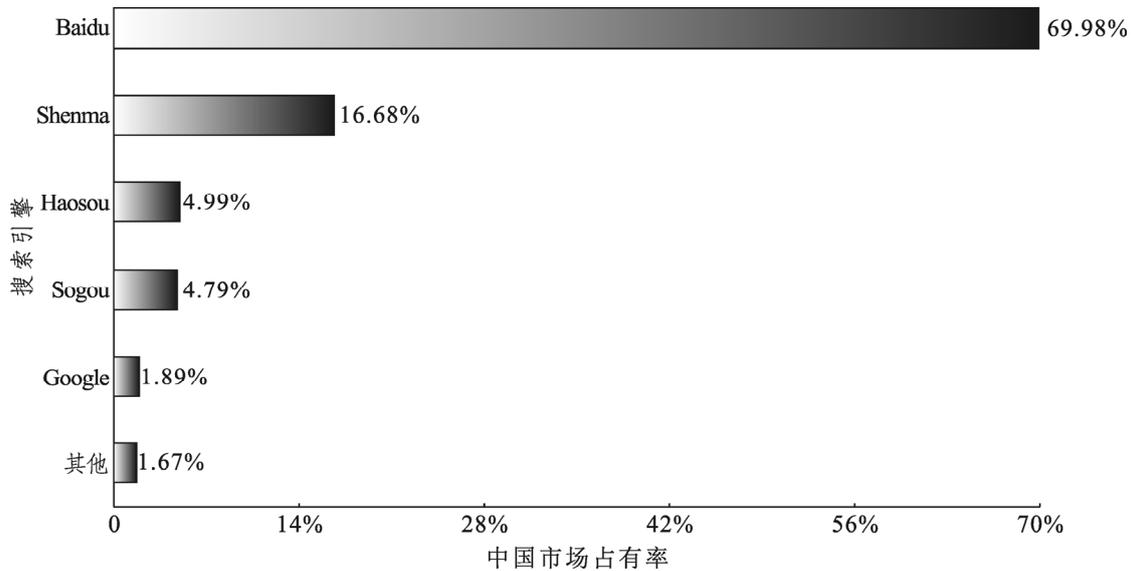


图 1-1-2 2018 年 2 月—2019 年 2 月搜索引擎中国市场占有率排行榜

数据来源：<http://gs.statcounter.com>。

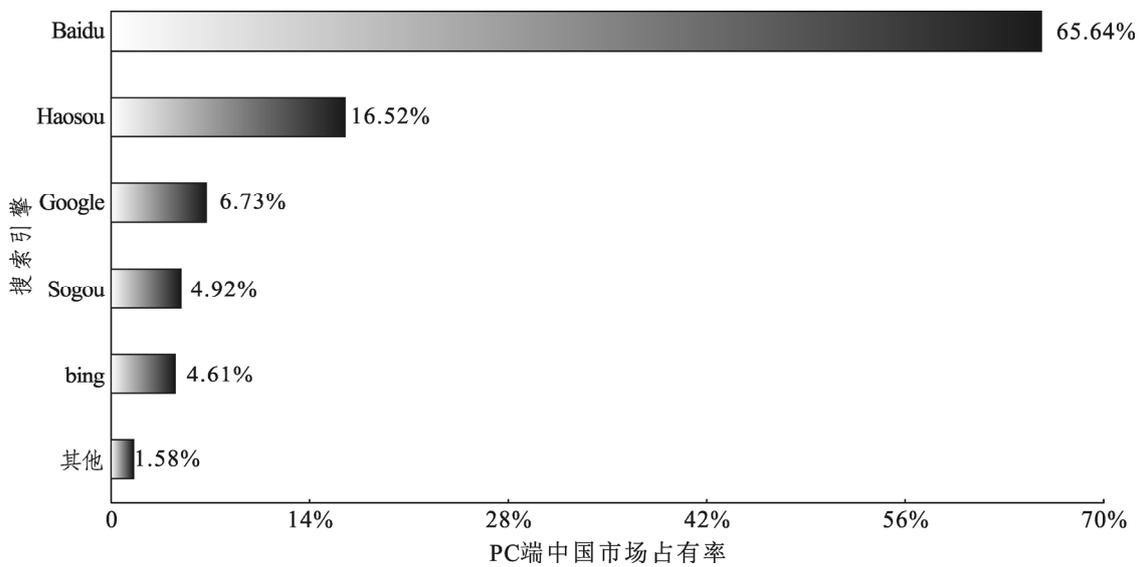


图 1-1-3 2018 年 2 月—2019 年 2 月搜索引擎 PC 端中国市场占有率排行榜

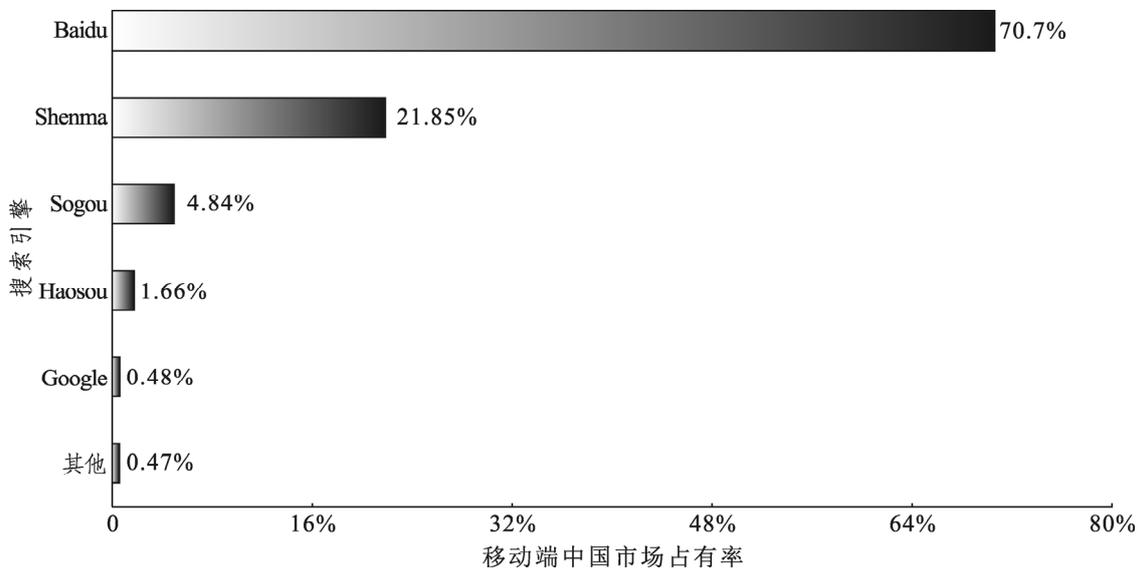


图 1-1-4 2018 年 2 月—2019 年 2 月搜索引擎移动端中国市场占有率排行榜

【拓展知识】

百度搜索是全球最大的中文搜索引擎，2000 年 1 月由李彦宏、徐勇两人创立于北京中关村，致力于向人们提供“简单，可依赖”的信息获取方式。“百度”二字源于中国宋朝词人辛弃疾的《青玉案》中的名句“众里寻他千百度”，象征着百度对中文信息检索技术的执着追求。而“熊掌”图标的灵感来源于“猎人巡迹熊爪”的刺激，与李彦宏的“分析搜索技术”非常相似，从而构成百度的搜索概念，也成为百度的标志形象。

二、实施步骤

通过搜索引擎对网络信息资源进行查找是获取网络信息资源的主要方式。搜索引擎是用来对网络信息资源进行管理和检索的一系列软件，实际上也是一些网页。查找信息资源时，在浏览器的搜索框中输入查找的关键词、短语，或者是其他相关的信息，再通过超级链接，逐一访问相关网站，就可能查找到所需要的信息资源，如 Google、百度、Yahoo 等搜索引擎即以此种方式为用户服务。搜索引擎以搜寻网络信息资源为目标，在一定程度上满足了人们对网络信息资源的查询需求，给用户搜寻信息带来便利，减少了网络浏览的盲目性。为了得到较多的检索结果，在搜索到的网址基础上，可以通过改变 URL 的方法来扩大检索的成果，如改变 HTTP 到 GOPHER 或 FTP，可以获取文本类型的文件资料。

下面，以百度搜索引擎作为实例进行讲解。

(一) 简单搜索

在浏览器中输入百度网址：<https://www.baidu.com>，在搜索框中输入关键词，并点击“百度一下”按钮，就会自动弹出相关的网页和资料（见图 1-1-5）。百度会寻找所有符合查询条件的资料，并将相关的网页或资料展现出来。

关键词，就是输入搜索框中的文字，也就是命令百度寻找的内容，可以是任意中文、英文、数字，或三者的混合体。因此关键词的内容可以是人名、网站、新闻、小说、软件、游戏、星座、工作、购物、论文名称等。可以输入一个关键词，也可以输入两个、三个、四个，甚至可以输入一句话或一段文字。

小技巧：输入关键词后，直接按键盘上的“Enter”回车键，百度会自动找出相关的网页或资料。注意多个关键词之间必须留一个空格，也可在搜索结果中选择需要搜索的类型。



图 1-1-5 简单搜索事例

(二) 高级搜索

1. 减除无关资料

在查询资料的时候，有时需要排除含有某些词语的关键词，以缩小查询范围。百度支持“-”功能，用于有目的地删除某些无关网页，但减号之前必须留一空格，语法是“A-B”。

例如要搜寻关于“武侠小说”，但不含“古龙”的网页，可使用如下方法查询（见图 1-1-6）：武侠小说-古龙。

如关键词和减号之间没有空格，就会被当成连字符处理，从而失去减号语法功能，其对比搜索如图 1-1-7 所示。减号和后一个关键词之间，有无空格均可。



图 1-1-6 减除无关资料搜索事例



图 1-1-7 未加空格搜索事例

2. 结果须同时包含

在关键词的前面使用加号“+”，也就等于要求搜索引擎该关键词必须同时出现在结果中的网页上。

例如搜索“+电脑+传真+电话”，就表示内容中同时包含这三个关键词（见图 1-1-8），其对比搜索如图 1-1-9 所示。



图 1-1-8 结果须同时包含搜索事例



图 1-1-9 未加加号搜索事例

3. 精准搜索

如果输入的搜索词很长，百度在经过分析后，给出的搜索结果中的搜索词可能是拆分的。如果对这种结果不满意，可以尝试让百度不拆分搜索词，给要搜索的关键词加上双引号（半角），可以实现精确查询。

例如搜索“贵阳职业技术学院”，只返回含有“贵阳职业技术学院”这个关键词的网页（见图 1-1-10），而不会出现诸如“贵阳电子职业学校”之类的非精准查询结果的网页（见图 1-1-11）。



图 1-1-10 精准搜索事例



图 1-1-11 非精准搜索事例

4. 并行搜索

使用“A|B”来搜索或者包含关键词A，或者包含关键词B的网页。

例如查询“图片”或“写真”相关资料，无须分两次查询，只要输入“图片|写真”搜索即可（见图 1-1-12）。百度会提供与“|”前后任一关键词相关的网站和资料。



图 1-1-12 并行搜索事例

5. 指定站点搜索

有时候，如果知道某个站点中有自己需要找的材料，就可以把搜索范围限定在这个站点中，提高查询效率。使用方法就是在查询内容的后面加上“site：站点域名”。

例如只在“贵阳职业技术学院”网站中搜索关于“教学”关键字的网页，可以这样搜索教学 site：gyvtc.edu.cn（见图 1-1-13）。



图 1-1-13 指定站点搜索事例

注意，“site:”后面的站点域名，不要带“http://www.”。另外，“site:”和站点名之间，不要加空格，关键字与“site”之间要有空格。

6. 精确匹配

书名号是百度独有的一个特殊查询法。在其他搜索引擎中，书名号会被忽略，而在百度中，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有特殊作用，例如，查名字很通俗和常用的电影或者小说时。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（见图 1-1-14），而加上书名号后，查询“《手机》”结果就都是关于电影“手机”的信息（见图 1-1-15）。



图 1-1-14 手机搜索结果



图 1-1-15 《手机》搜索结果

7. 指定文档类型搜索

平时工作中，需要搜索一些相关文档资料作为参考，这时就可以用到 filetype。例如，搜索“职业教育”关键词的 ppt 文档，可以输入“职业教育 filetype:ppt”（见图 1-1-16）。如果

不限制文档类型，则可以使用 all。



图 1-1-16 指定文档类型搜索事例

【思考与练习】

- (1) 精确搜索“老师学生”网页。
- (2) 搜索“教师”关键词的 Word 文档。

任务二 利用机构网站获取网络信息资源

【任务目标】

- (1) 知晓电子商务行业信息资源常用网站。
- (2) 能够运用网站获取信息资源。

一、知识链接

(一) 网页

网页是构成网站的基本元素，是承载各种网站应用的平台。通俗地说，网站是由网页组成的，如果只有域名和虚拟主机而没有制作任何网页的话，客户仍旧无法访问网站。

(二) 网站

网站（Website）是指在因特网上根据一定的规则，使用 HTML（标准通用标记语言下的

一个应用)等工具制作的用于展示特定内容相关网页的集合。简单地说,网站是一种沟通工具,人们可以通过网站来发布自己想要公开的资讯,或者利用网站来提供相关的网络服务。人们可以通过网页浏览器来访问网站,获取自己需要的资讯或者享受网络服务。

(三) 互联网和电子商务领域获取资源常用网站

如果用户熟悉网络资源的特点和分布状况,了解常用信息资源的发布方式,可以通过国内外重要的科研机构、信息发布机构、学会的网址,及时准确地获得这些权威机构发布的信息。如要了解关于医药研究方面信息,可以到世界卫生组织(<http://www.who.int>)以及其他一些重要的医药学会、协会的网站查找。这些权威机构、学会的网站可以帮助用户及时了解和掌握最新的科研动态。在互联网和电子商务领域,获取资源常用的网站很多,笔者向大家推荐如下几个:

1. 中国互联网络信息中心 (<http://www.cnnic.net.cn>)

中国互联网络信息中心(China Internet Network Information Center, CNNIC)是经国家主管部门批准,于1997年6月3日组建的管理和服务机构,行使国家互联网络信息中心的职责。作为中国信息社会重要的基础设施建设者、运行者和管理者,中国互联网络信息中心(CNNIC)在“国家公益、安全可信、规范高效、服务应用”方针的指导下,负责国家网络基础资源的运行管理和服务,承担国家网络基础资源的技术研发并保障安全,开展互联网发展研究并提供咨询,促进全球互联网开放合作和技术交流,不断追求成为“专业、责任、服务”的世界一流互联网络信息中心。

2. 亿邦动力网 (<http://www.ebrun.com>)

亿邦动力网是国内权威电子商务新闻门户,围绕传统企业做电商、B2B、B2C、外贸电商、移动电商、电商服务、电商政策、电商资本等领域开展信息与研究服务,在电子商务行业拥有巨大的影响力。亿邦动力网的定位是电子商务新闻门户,为中国中小企业提供全面、及时、权威的电子商务新闻资讯。业务范围涵盖传统企业做电商、B2C、B2B、外贸电商、移动电商、国际电商、电商服务、电商资本、电商案例、电商数据、实战社区、电商培训、电商会议、电商行业研究报告等多个领域。

3. 艾瑞咨询 (<http://www.iresearch.com.cn>)

艾瑞咨询集团即iResearch 艾瑞市场咨询,成立于2002年,公司创始人杨伟庆毕业于华东理工大学。杨伟庆1998年开始接触网络广告行业,是中国最早的一批网络营销专业人士。

在创立iResearch公司之前,杨伟庆于1999年年底参与创建商业版本的“网络广告先锋”(WiseCast)网站并负责具体运营。“网络广告先锋”最早由任向晖于1997年5月创立,是中国第一批网络行业从业人员的行业研究资讯网站,主要提供网络广告行业发展资讯。

目前艾瑞市场咨询(iResearch)的主要服务产品有iUserTracker(网民行为连续研究系统)、iAdTracker(网络广告监测分析系统)、iUserSurvey(网络用户调研分析服务)、iDataCenter(网络行业研究数据中心)等。iResearch总部在上海,并于2003年10月在北京成立分部,目前