

3

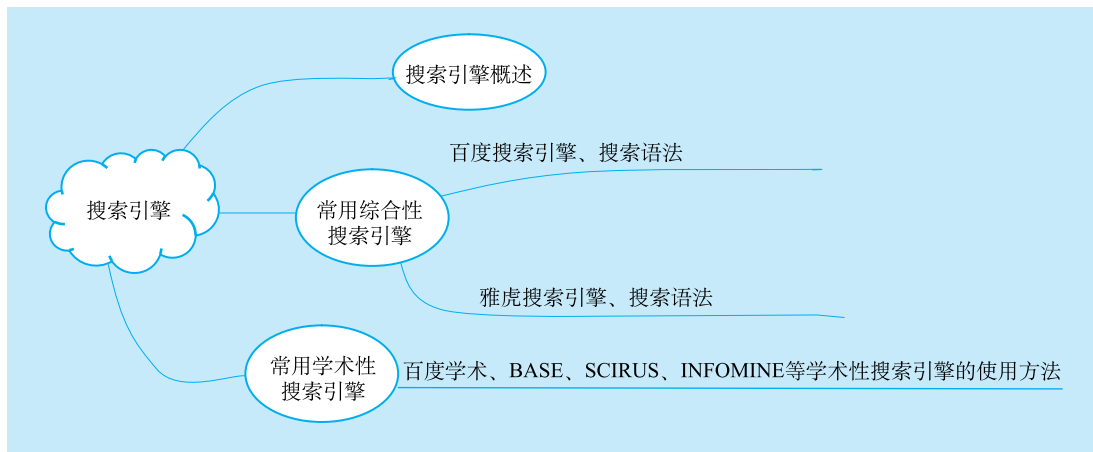
第3章

网络文献信息检索

学习目标

1. 了解搜索引擎的基本概念；
2. 掌握常用综合性搜索引擎的使用方法；
3. 掌握常用学术性搜索引擎的使用方法。

内容框架



3.1

搜索引擎概述

搜索引擎是指互联网上专门提供查询服务的一类网站，这些网站通过网络搜索软件，

又称为网络搜索机器人或网站登录等方式,将互联网上大量网站的页面收集到本地,经过加工处理后建库,从而能够对用户提出的各种查询做出响应,提供用户所需的信息。

搜索引擎与二次文献异曲同工,提供给用户的结果是文献线索(链接及其简介),只不过前者采用了超文本技术,单击链接即可浏览原文。

根据不同的分类标准,可以将搜索引擎分为不同的类型。

3.1.1 根据其搜索的内容范围划分

根据其搜索的内容范围划分,可将搜索引擎分为综合搜索引擎和专业搜索引擎。

▶ 1. 综合搜索引擎

综合搜索引擎是一种查询多学科、多类型信息的搜索引擎,如百度、搜狗等都属于综合搜索引擎。搜狗搜索引擎界面如图 3-1 所示。



图 3-1 搜狗搜索引擎界面

▶ 2. 专业搜索引擎

专业搜索引擎指查询某一专题(如 Newsgroup、机票、购物、学术资源等)或某一领域(如政治、经济、文化、体育等)信息的搜索引擎,也可称为专题搜索引擎,如百度的图片搜索引擎等。百度图片搜索引擎如图 3-2 所示。

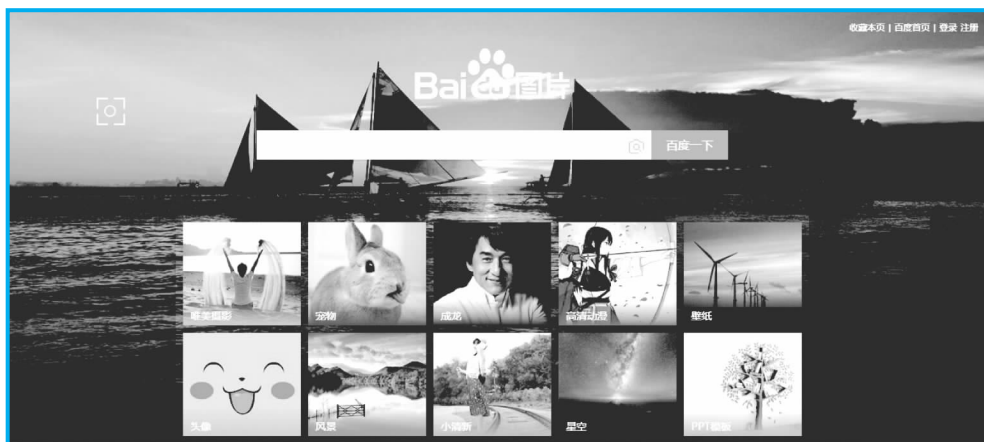


图 3-2 百度图片搜索引擎界面

3.1.2 根据搜索引擎的集成程度划分

根据搜索引擎的集成程度划分,可将搜索引擎分为全文搜索引擎、目录搜索引擎和元搜索引擎。

▶ 1. 全文搜索引擎

全文搜索引擎是名副其实的搜索引擎,国内有著名的百度搜索,国外代表有谷歌。它们从互联网提取各个网站的信息(以网页文字为主),建立起数据库,并能检索与用户查询条件相匹配的记录,按一定的排列顺序返回结果。

根据搜索结果来源的不同,全文搜索引擎可分为两类,一类拥有自己的检索程序(indexer),俗称“蜘蛛”(spider)程序或“机器人”(robot)程序,能自建网页数据库,搜索结果直接从自身的数据库中调用,谷歌和百度就属于此类;另一类则是租用其他搜索引擎的数据库,并按自定的格式排列搜索结果,如 Lycos 搜索引擎,如图 3-3 所示。

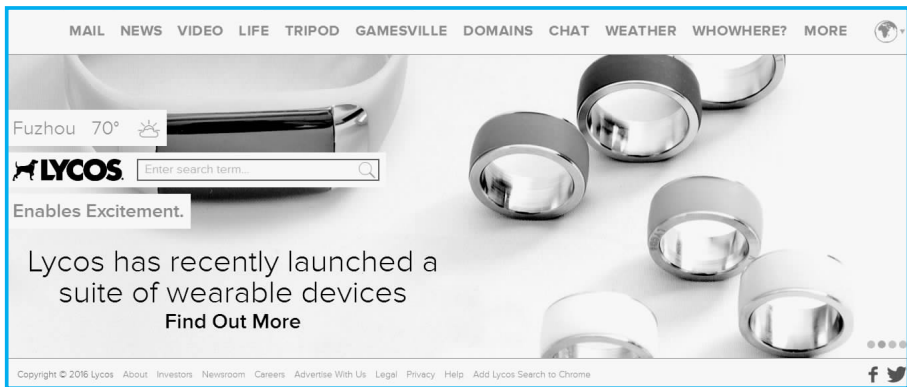


图 3-3 Lycos 搜索引擎界面

▶ 2. 目录搜索引擎

目录索引虽然有搜索功能,但严格意义上不能称为真正的搜索引擎,只是按目录分类的网站链接列表而已。用户完全可以按照分类目录找到所需要的信息,不依靠关键词(keywords)进行查询。目录索引中最具代表性的是雅虎(Yahoo!)、新浪分类目录搜索。雅虎的目录搜索界面如图 3-4 所示。

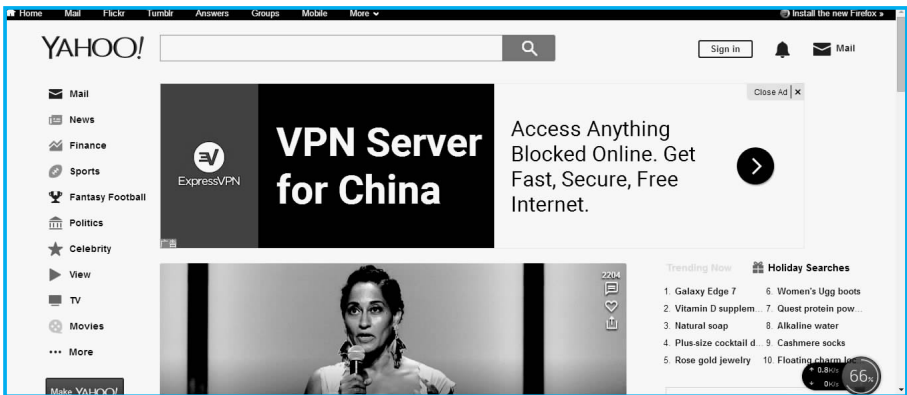


图 3-4 雅虎的目录搜索界面

▶ 3. 元搜索引擎

元搜索引擎(META search engine)接受用户查询请求后,同时在多个搜索引擎上搜索,并将结果返回给用户。著名的元搜索引擎有 InfoSpace、Dogpile、Vivisimo 等,中文元搜索引擎中具代表性的是搜星搜索引擎。在搜索结果排列方面,有的直接按来源排列搜索结果,如 Dogpile;有的则按自定的规则将结果重新排列组合,如 Vivisimo。Dogpile 搜索引擎界面如图 3-5 所示。

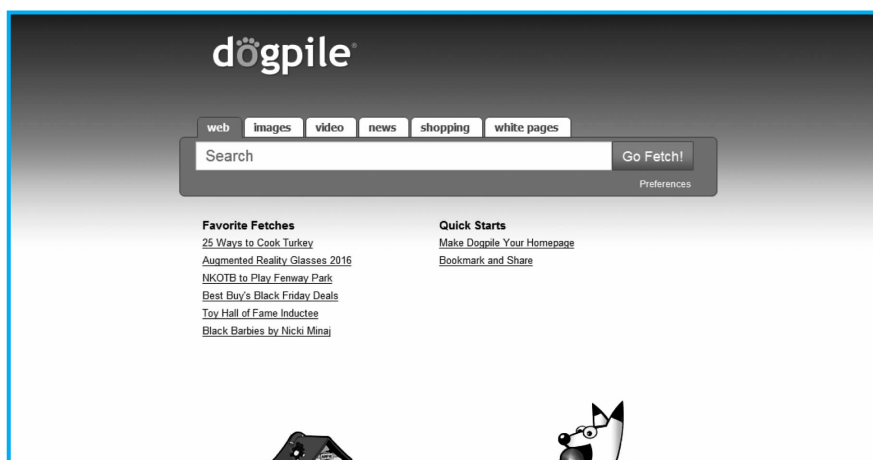


图 3-5 Dogpile 搜索引擎界面

3.2 常用综合性搜索引擎

搜索引擎在我们的日常生活中起着重要的作用,它是一个为大家提供信息“检索”服务的网站,它使用某些程序把互联网上的所有信息归类,以帮助人们在茫茫网海中搜寻到所需要的信息。

常用的综合性搜索引擎是一种可以查询多学科、多类型信息的搜索引擎,如百度、搜狗、Google、Ask、Alta Vista 等。

3.2.1 百度搜索引擎

百度搜索引擎(www.baidu.com)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。百度搜索引擎由四部分组成:蜘蛛程序、监控程序、索引数据库和检索程序。高性能的“网络蜘蛛”程序自动地在互联网中搜索信息,可订制、高扩展性的调度算法使得搜索器能在极短的时间内收集到最大数量的互联网信息。

目前的百度提供基于搜索引擎的各种服务,其中包括以网页、新闻、音乐、图片、视频、地图、文库等内容搜索为主的功能性搜索,以贴吧、知道、百科、空间为主的社区搜索,针对各区域、行业所需的垂直搜索,以及门户频道、IM(百度 Hi)等,全面覆盖了中

文网站世界所有的搜索需求。百度搜索引擎界面如图 3-6 所示。



图 3-6 百度搜索引擎界面

百度搜索引擎支持符号法则、限制检索法则、自然语法规则，以及特殊的查询法则等，主要语法特点如下。

▶ 1. 把搜索范围规定在特定的站点中

有时需要找一些特殊的文档，特别是针对专业性比较强的网站，就可以把搜索范围限定在这个站点中，提高查询效率。在查询内容的后面加上“site: 站点域名”即可使用。

例如，SUV site: www.autohome.com.cn，检索结果如图 3-7 所示。



图 3-7 站点内检索的结果

▶ 2. 把搜索范围限定在标题中

一般情况下，标题是整篇文章的纲要，把查询内容范围限定在网页标题中，可以有利于快速地找到所需要的文章。把查询内容中特别关键的部分，用“intitle:”作为前缀，即可实现检索。

例如，网站优化 intitle: 论坛，检索结果如图 3-8 所示。



图 3-8 标题中检索的结果

► 3. 把搜索范围限定在 URL 链接中

网页 URL 中的某些信息常常有某种有价值的含义，对搜索结果的 URL 做某种限定，就可以获得良好的效果。“inurl:”后跟需要在 URL 中出现的关键词即可实现检索。

例如，想查找关于 Photoshop 的技术类文档，输入 photoshop inurl: jishu 就可以找到相关知识了。其中，“photoshop”可以出现在网页的任何位置，而“jishu”则必须出现在网页 URL 中，检索结果如图 3-9 所示。



图 3-9 在限定的 URL 中检索的结果

► 4. 精确匹配

如果输入的查询词很长，百度在经过分析后，给出的搜索结果可能是拆分的，若不希望拆分的查询词，可以给查询词加上双引号。例如，输入“互联网+”，检索结果如图 3-10 所示。

书名号是百度独有的一个特殊查询语法，加上书名号查询，有两层特殊功能，一是书名号出现在搜索结果中；二是被书名号括起来的内容不会被拆分。

例如，查电影“手机”，如果不加书名号，很多情况下出来的是手机通信工具，而加上书名号后，检索结果就是关于电影方面的了，检索结果如图 3-11 所示。



图 3-10 加双引号的检索结果



图 3-11 加书名号的检索结果

► 5. 要求搜索结果中不含特定查询词

顾名思义，“-”就是你不想要查询的内容。例如，简历模板—毕业生简历，检索结果如图 3-12 所示，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。



图 3-12 搜索结果中不含特定词的检索结果

► 6. 专业文档搜索

百度支持对 Office 文档(包括 Word、Excel、Powerpoint)、Adobe PDF 文档、TRF 文档进行全文搜索。查询方式可以在查询词后面加“filetype:”进行文档类型限定。例如,查找张五常关于交易费用方面的经济学论文,可以输入:交易费用张五常 filetype: doc,检索结果如图 3-13 所示。



图 3-13 在 Word 类型的文档进行检索的结果

3.2.2 雅虎搜索引擎

雅虎(www. yahoo. com)是美国著名的互联网门户网站,也是 20 世纪末互联网奇迹的创造者之一。雅虎是最老的“分类目录”搜索数据库,也是最重要的搜索服务网站之一,在全部互联网搜索应用中所占份额达 36%左右。所收录的网站全部被人工编辑按照类目分类。其数据库中的注册网站无论是在形式上还是内容上质量都非常高。雅虎搜索引擎界面如图 3-14 所示。

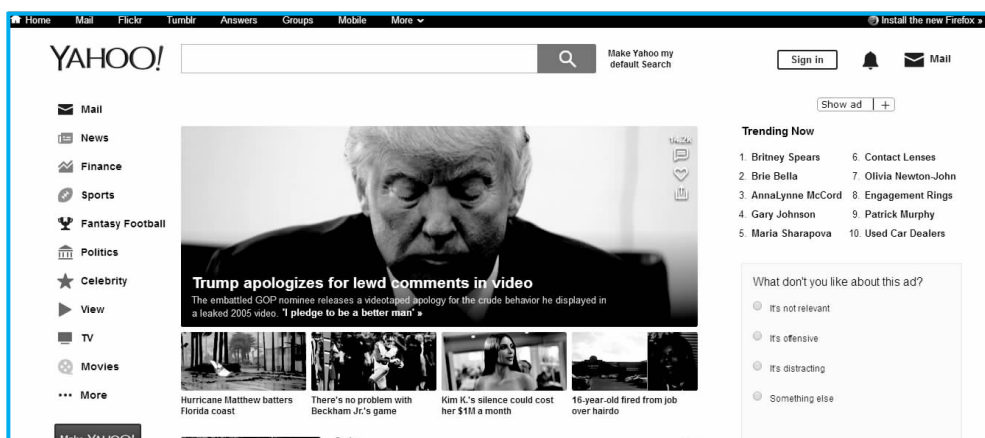


图 3-14 雅虎搜索引擎界面

中文雅虎支持符号法则、限制检索法则、自然语言法则等,主要语法特点如下。

▶ 1. 方括号

运算符“[]”限定关键词在检索结果中出现的顺序。例如，[儿童教育心理健康]，表示检索结果中“儿童教育”在前，“心理健康”在后。

▶ 2. 加号

运算符“+”表示查询结果中一定要出现在“+”后面的字符串。例如，“那英+歌曲”，搜索结果一定是带有那英歌曲的网站。

▶ 3. 感叹号

运算符“!”表示激活雅虎的快捷搜索，可以快速跳转到雅虎网站的其他栏目进行搜索。这是雅虎的一项专有技术。

▶ 4. 用于针对标题进行搜索

title: 加在关键词前面，表示搜索引擎限定在网站的标题中查找。例如，搜索 title: 张学友，搜索引擎会搜索到所有标题中包含“张学友”的网页。

▶ 5. 用于查找所有链接到某个网址的网页

link: 表示搜索引擎限定在网址中查找。例如，link: http://www.yahoo.com 姚明，会搜索到所有链接到 www.yahoo.com 的网页中，包含“姚明”一词的网页。注意搜索时不能缺少“http://”。

▶ 6. 用于限定搜索结果的来源

site: 表示如果知道某个站点中有自己需要找的内容，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，想查询雅虎网站中报导的姚明信息，可以有以下两种方法。

(1) 在搜索框中输入“姚明 site: yahoo.com”(注意中间使用空格隔开)。搜索引擎会搜索到在域名 www.yahoo.com 及其子域名中的所有包含“姚明”一词的网页。

(2) 在搜索框中输入“site: 姚明 yahoo.com.cn”也可以得到相同的搜索结果。注意，site: 后面跟的站点域名不要带“http://”；另外，site: 和站点名之间不要加空格。

3.3 常用学术性搜索引擎

学术性搜索引擎可以满足各领域专业研究用户的学术方面知识的搜索，可以搜索到相关领域的专家评审文献、论文、书籍、预印本、摘要以及技术报告等。

▶ 1. 百度学术搜索引擎(<http://xueshu.baidu.com/>)

百度学术搜索引擎是一个提供海量中英文文献检索的学术资源搜索平台，涵盖了各类学术期刊、会议论文，旨在为国内外学者提供最好的科研体验，界面如图 3-15 所示。

▶ 2. BASE(<http://www.base-search.net/>)

BASE 是德国比勒费尔德(Bielefeld)大学图书馆开发的一个多学科的学术搜索引擎，提供对全球异构学术资源的集成检索服务。它整合了德国比勒费尔德大学图书馆的图书馆目录和大约 160 个开放资源(超过 200 万个文档)的数据，界面如图 3-16 所示。

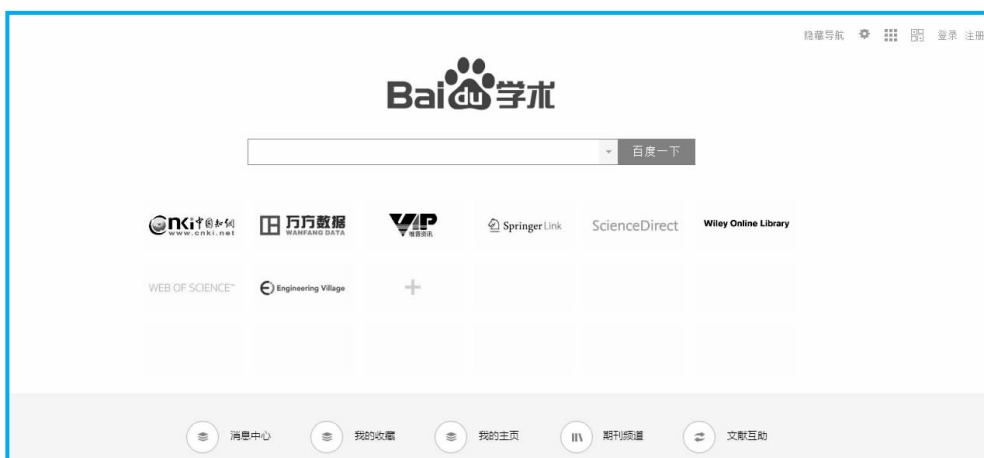


图 3-15 百度学术搜索引擎界面



图 3-16 BASE 学术搜索引擎界面

► 3. SCIRUS(<http://www.scirus.com/>)

SCIRUS 是由爱思唯尔科学公司(Elsevier Science)于 2001 年 4 月推出的迄今为止国际互联网上最全面的科技信息专用搜索引擎。它以自身拥有的资源为主体,对网上具有科学价值的资源进行整合,集聚了带有科学内容的网站及与科学相关的网页上的科学论文、科技报告、会议论文、专业文献、预印本等。其目的是力求在科学领域内做到对信息全面深入的收集,以统一的检索模式面向用户提供检索服务。

SCIRUS 的检索界面友好,简洁方便,如图 3-17 所示。其检索方式可分为基本检索(basic search)和高级检索(advanced search)两种方式。默认方式为基本检索。

(1) SCIRUS 基本检索非常简单,用户仅需输入检索词,按 Enter 键,或单击 Search 按钮即可得到相关资料。在基本检索的操作中,用户可以选择检索结果的来源,如来自期刊或来自网络。用户还可以选择检索结果与输入词组是否精确匹配。

(2) SCIRUS 高级检索支持逻辑检索。“AND”表示检索结果中必须包括所有的检索词、“OR”表示检索结果中至少包含一个检索词、“ANDNOT”表示前面的检索词将一定包含于检索结果中,而后面的检索词则一定不出现在检索结果中。在高级检索的操作中,用户还可以选择检索字段、年限、信息来源等来限定检索结果。

► 4. INFOMINE(<http://infomine.ucr.edu/>)

INFOMINE 是为大学教师、学生和研究人员建立的网络学术资源虚拟图书馆。它建

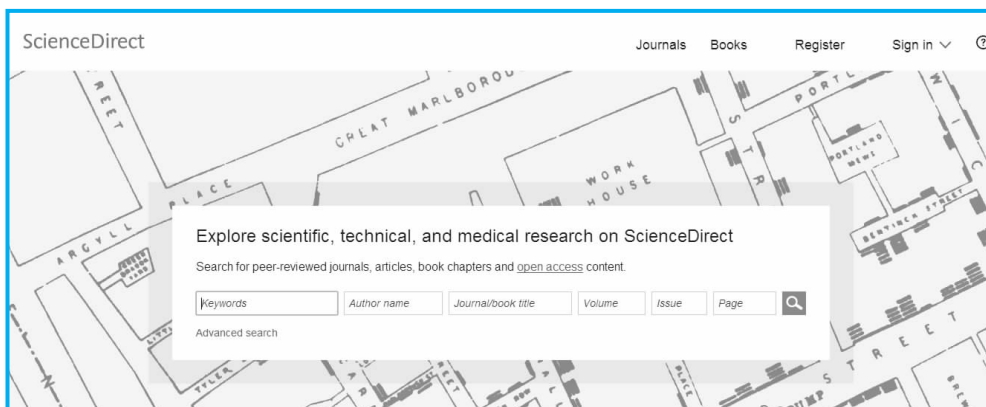


图 3-17 SCIRUS 学术搜索引擎界面

于 1994 年,由加利福尼亚大学、威克福斯特大学、加利福尼亚州立大学、底特律—麦西大学等多家大学或学院的图书馆联合建立,拥有电子期刊、电子图书、公告栏、邮件列表、图书馆在线目录、研究人员人名录,以及其他类型的信息资源 4 万多个。INFOMINE 对所有用户免费开放,但是它提供的资源站点并不都是免费的,能否免费使用取决于用户所在图书馆是否拥有该资源的使用权。

INFOMINE 的检索界面友好,如图 3-18 所示。其检索方法简单易用,检索功能包括基本检索、高级检索和浏览三种方式。

(1) 基本检索:在 INFOMINE 首页的检索框中直接输入检索词(主题词、作者、关键词等),单击 Search 或 Enter 键就可以检索出相关资料。

(2) 高级检索:输入检索词,可使用逻辑检索(AND、OR、AND NOT)或特定符号(*、|、()、“”等)来扩大、缩小检索范围。在高级检索中,通过菜单和下拉菜单中命令的组合使用,可以限定检索范围(如关键词、主题词、资源描述、作者、标题等)、限定检索的数据库范围、限定资源的类型和路径,以及检索结果的显示方式,每页显示的检索结果数和检索结果的排序方式。

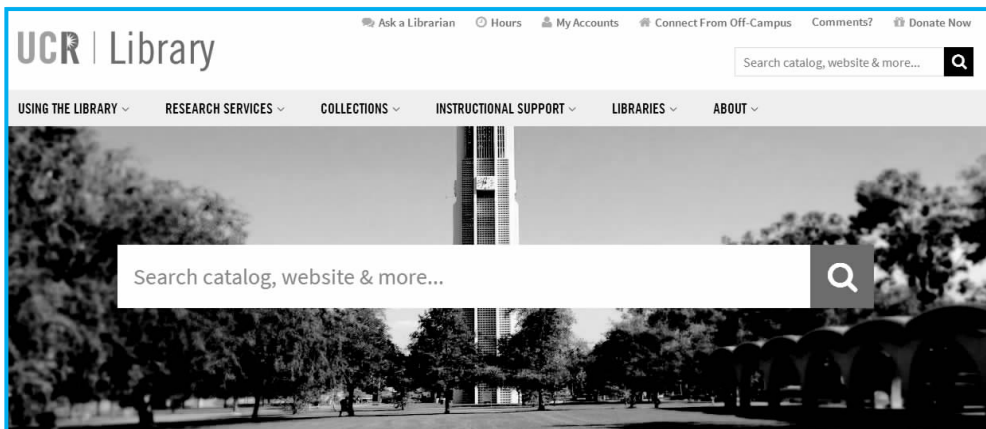


图 3-18 INFOMINE 学术搜索引擎界面

(3) 浏览:对于普通用户而言,浏览是一种有效和常用的检索方式。INFOMINE 在

基本检索、高级检索和每个数据库的页面下，都提供浏览功能，可以从目次表、美国国会主题词表、标题、关键词和作者等途径进行浏览，查找所需的资料。

思考与习题

1. 搜索引擎可以分为哪些类型？
2. 常用搜索引擎有哪些，举例并说明其特点？
3. 百度搜索引擎和谷歌搜索引擎的检索语法各有什么特点？
4. 搜索 2 篇“搜索引擎使用技巧”格式为 doc 的文档，写出相关语法格式，并把搜索结果截图。
5. 用百度和谷歌搜索小说“手机”，关键词分别用普通输入、双引号或书名号输入。写出相关语法格式，把搜索结果截图，并比较哪个检索效果好。
6. 利用百度、雅虎、谷歌三种搜索引擎查找以“4G 手机”为关键词的信息，列出查找结果的前三项标题及 URL，并查看链接是否可用。