

模块二 信息检索技术

——单元2.2 搜索引擎使用



目录

CONTENTS

- 01 | 导入案例
- 02 | 技术分析
- 03 | 知识与技能
- 04 | 案例实现
- 05 | 练习与提高

技术分析

在本任务中，首先介绍搜索引擎的发展、定义和分类，然后比较全面地介绍百度的使用方法和技巧，最后介绍一些国内外常用的搜索引擎。

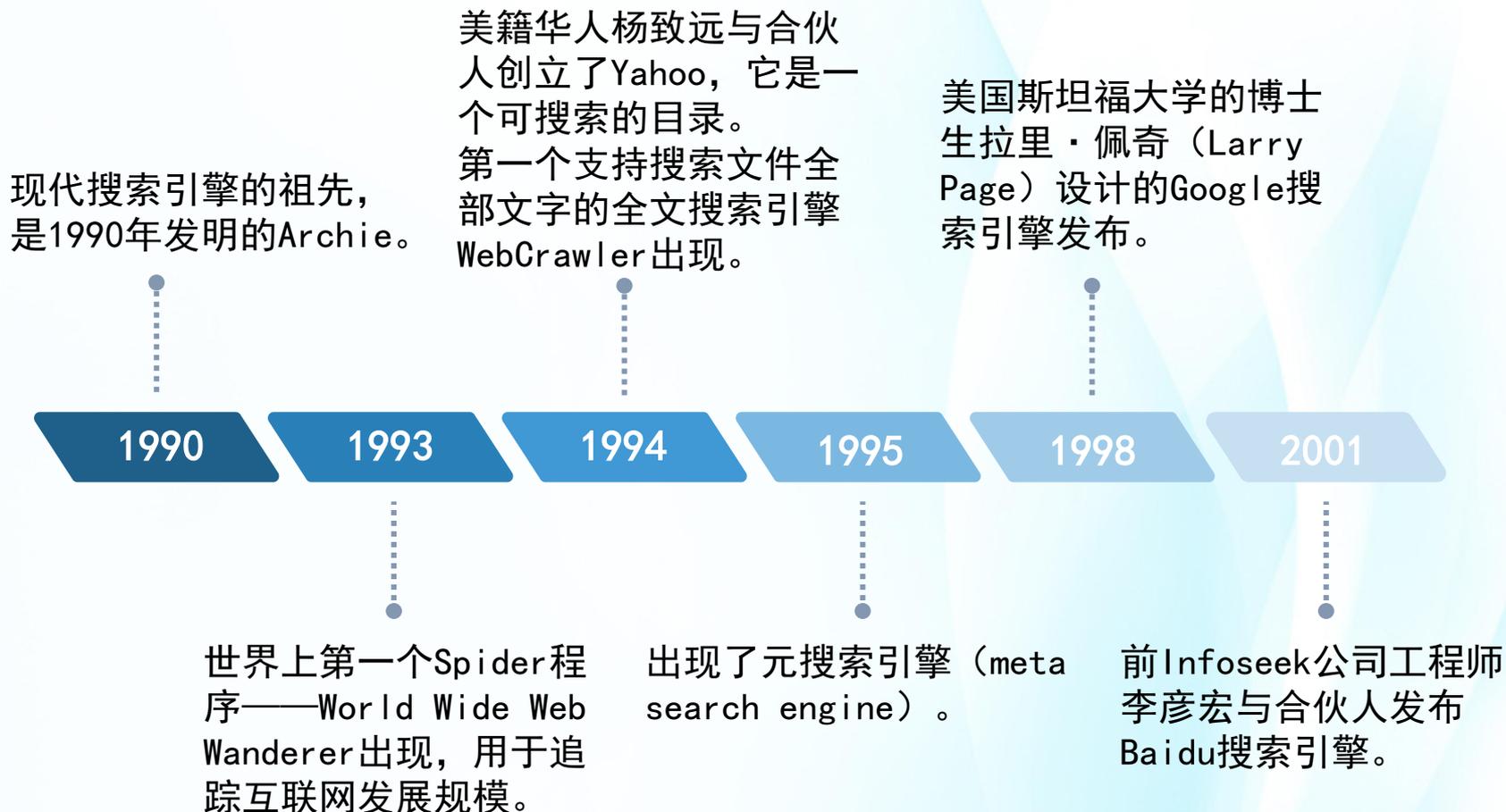
知识与技能



- 一、关于搜索引擎
- 二、百度的搜索方法
- 三、其他常用搜索引擎简介

一、关于搜索引擎

(一) 搜索引擎的发展



一、关于搜索引擎

（二）搜索引擎的定义

所谓搜索引擎，就是根据用户需求与一定算法，运用特定策略从互联网检索出用户要求的信息，在对信息进行组织和处理后，反馈给用户的一门**检索技术**。搜索引擎依托于多种技术，如网络爬虫技术、检索排序技术、大数据处理技术、自然语言处理技术等。

一、关于搜索引擎

(三) 搜索引擎的分类

全文搜索引擎

全文搜索引擎是真正的搜索引擎，国外具有代表性的全文搜索引擎有Google、AltaVista等，国内著名的全文搜索引擎有百度（Baidu）。它们都是通过从互联网上提取的各个网站的信息（以网页文字为主）而建立的数据库中，检索与用户查询条件匹配的相关记录，然后按一定的排列顺序将结果返回给用户。

目录索引类搜索引擎

目录索引虽然有搜索功能，但在严格意义上算不上是真正的搜索引擎，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（keywords）查询，仅靠分类目录也可找到需要的信息。目录索引中最具代表性的是早期的雅虎、早期的搜狐、新浪、网易等。如今，目前搜索引擎都已转向全文搜索引擎。

元搜索引擎

用户只需提交一次搜索请求，由元搜索引擎负责转换处理后提交给多个预先选定的独立搜索引擎，并将从各独立搜索引擎返回的所有查询结果，集中起来处理后再返回给用户。典型的元搜索引擎有360搜索 <https://www.so.com/>。

二、百度的搜索方法

(一) 百度搜索一般方法

在搜索框输入搜索关键词后，会出现“搜索工具”，如图所示。



单击“搜索工具”，展开搜索工具，如图所示。



二、百度的搜索方法

也可以直接进入“高级搜索”选项，方法是：点击百度首页右上角的“设置”，选择“搜索设置”，出现如图所示的百度搜索设置界面。



百度搜索设置界面

二、百度的搜索方法

单击“高级搜索”选项，就进入了是百度高级搜索设置界面，如图所示。



The screenshot shows the Baidu Advanced Search settings window. It has three tabs: '搜索设置' (Search Settings), '高级搜索' (Advanced Search), and '首页设置' (Home Settings). The '高级搜索' tab is active. The interface includes several configuration options:

- 搜索结果:** Three input fields for search results: '包含全部关键词' (Include all keywords), '包含完整关键词' (Include complete keywords), and '包含任意关键词' (Include any keywords).
- 时间:** A label '限定要搜索的网页的时间是' (Limit the time of the web pages to be searched) and a dropdown menu set to '全部时间' (All time).
- 文档格式:** A label '搜索网页格式是' (Search web page format is) and a dropdown menu set to '所有网页和文件' (All web pages and files).
- 关键词位置:** A label '查询关键词位于' (Query keywords are located) and three radio buttons: '网页任何地方' (Anywhere on the page), '仅网页标题中' (Only in the page title), and '仅URL中' (Only in the URL).
- 站内搜索:** A label '限定要搜索指定的网站是' (Limit the search to the specified website) and an input field with the example '例如: baidu.com' (Example: baidu.com).

A blue button labeled '高级搜索' (Advanced Search) is located at the bottom center of the window.

百度高级搜索设置界面

二、百度的搜索方法

(二) 百度搜索高级搜索方法

2. 消除无关性 (-)

相当于逻辑“非”的操作，用于排除无关信息，有利于缩小查询范围。百度支持“-”功能，用于有目的地删除某些无关网页，语法是“A -B”。如：要搜寻关于“花城广场”，但不含“广州”的资料，可使用：花城广场 -广州。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。

二、百度的搜索方法

(二) 百度搜索高级搜索方法

3. 并行搜索 (|)

逻辑“或”的操作，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。使用同义词作关键词并在各关键词中使用“|”运算符可提高检索的全面性。如：“计算机|电脑”搜索即可。

二、百度的搜索方法

(二) 百度搜索高级搜索方法

4. 把搜索范围限定在网页标题中 (intitle:)

网页标题通常是对网页内容的归纳。把查询内容范围限定在网页标题中，就会得到和输入的关键字匹配度更高的检索结果。方法是：在搜索关键字前加“intitle:”。如：intitle:电子商务。注意，intitle:和后面的关键词之间，不要有空格。

二、百度的搜索方法

(二) 百度搜索高级搜索方法

5. 把搜索范围限定在特定站点中 (site:)

有时候，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，能提高查询效率。使用的方式，是在查询内容的后面，加上“site:站点域名”。如site:sina.com.cn。**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。

二、百度的搜索方法

(二) 百度搜索高级搜索方法

6. 把搜索范围限定在url链接中 (inurl:)

网页url中的某些信息，常常有某种有价值的含义。于是，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:”，前面或后面写上需要在url中出现的关键词。如：电子商务 inurl:sina.com.cn可以查找关于新浪网中的电子商务网页。上面这个查询串中的“电子商务”，是可以出现在网页的任何位置，而“sina.com.cn”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。

二、百度的搜索方法

(二) 百度搜索高级搜索方法

7. 特定格式的文档检索 (filetype:)

百度以“filetype:”来对搜索对象做限制，冒号后是文档格式，如pdf、doc、xls等。通过添加“filetype:”可以更方便有效的找到特定的信息，尤其是学术领域的一些信息。如：电子商务
filetype:PDF。

二、百度的搜索方法

（二）百度搜索高级搜索方法

8. 精确匹配/电影或小说（《》）

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：

- 书名号会出现在搜索结果中；
- 被书名号扩起来的内容，不会被拆分。

请对比 “《电子商务》” 与 “电子商务” 的区别。

二、百度的搜索方法

(二) 百度搜索高级搜索方法

9. 百度快照

百度快照功能在百度的服务器上保存了绝大多数网站的大部分页面，当不能链接所需网站时，百度暂存的网页可用来救急。而且通过百度快照寻找资料要比常规链接的速度快得多。因为百度快照的服务稳定，下载速度极快，不会再受死链接或网络堵塞的影响。在快照中，用户搜索使用的关键词均已用不同颜色在网页中标明，一目了然。点击快照中的关键词，还可以直接跳到它在文中首次出现的位置，使浏览网页更方便。

三、其他常用搜索引擎简介

（一）搜狗搜索

搜狗搜索引擎是搜狐公司强力打造的第三代互动式搜索引擎，“搜狗”搜索引擎可以使网站用户不离开网站就可以进行搜索，用户能借助智能的“搜狗”搜索引擎找到他们真正需要的信息。既方便用户使用，提升用户体验，又提高网站的用户黏性。

三、其他常用搜索引擎简介

（二）神马搜索

神马搜索是UC（优视）和阿里合作推出的移动搜索引擎。神马搜索的创新方向如下：



传统PC搜索的方式是，用户输入一个文本关键词，最后得到文本搜索结果。神马搜索关注输入的移动特性，比如语音输入、拍照输入、点击输入的方式。



传统搜索结果追求的是“全”，在此基础上用文本链的方式，也就是搜索的目录列表这种方式呈现搜索结果。神马搜索关注搜索结果的“准”，也就是搜索结果的高质量。



传统搜索的相关性是基于文本关键词的，神马搜索关注从“一致性搜索”逐渐向“个性化搜索”的过渡，会按照用户的特点展现不同的搜索结果。

三、其他常用搜索引擎简介

（三）360搜索

360搜索，属于元搜索引擎，是搜索引擎的一种，是通过一个统一的用户界面帮助用户在多个搜索引擎中选择和利用合适的（甚至是同时利用若干个）搜索引擎来实现检索操作，是对分布于网络的多种检索工具的全局控制机制。而360搜索+，属于全文搜索引擎，是奇虎360公司开发的基于机器学习技术的第三代搜索引擎，具备“自学习、自进化”能力，可以发现用户最需要的搜索结果。

三、其他常用搜索引擎简介

（四）必应搜索

微软必应（英文名：Microsoft Bing），原名必应（Bing），是微软公司于2009年推出，用以取代Live Search的全新搜索引擎服务。为符合中国用户使用习惯，Bing中文品牌名为“必应”。在Windows Phone系统中，微软也深度整合了必应搜索，通过触摸搜索键引出。必应搜索改变了传统搜索引擎首页单调的风格，通过将来自世界各地的高质量图片设置为首页背景，并加上与图片紧密相关的热点搜索提示，使用户在访问必应搜索的同时获得愉悦体验和丰富资讯。

案例实现

我们可以通过百度识图功能来搜索这幅画的信息。

方法一：如果使用电脑来进行搜索，可以打开百度搜索引擎以后，单击“百度一下”左侧的“按图片搜索”按钮，进入百度识图状态，点击“选择文件”按钮，选择要识图的图片，即可得到这幅画的作者介绍，据此结果进一步搜索“富春山居图”，即可得到这幅画的详细信息。

案例实现

我们可以通过百度识图功能来搜索这幅画的信息。

方法一：使用电脑来进行搜索。

1 打开百度搜索引擎

2 单击“百度一下”左侧的“按图片搜索”按钮。

3 进入百度识图状态。



“百度识图”按钮



百度识图状态

案例实现

我们可以通过百度识图功能来搜索这幅画的信息。

方法一：使用电脑来进行搜索。

4

点击“选择文件”按钮，选择要识图的图片，即可得到这幅画的作者介绍，据此结果进一步搜索“富春山居图”，即可得到这幅画的详细信息。



百度识图后的结果

案例实现

我们可以通过百度识图功能来搜索这幅画的信息。

方法二：如果用手机，则可打开百度APP，点击“百度识图”按钮，进入相机拍摄状态，拍摄图片，即可得到这幅画的相关信息。



练习与提高

1. 什么是搜索引擎?
2. 利用百度搜索引擎 (<http://www.baidu.com/>) 完成以下检索。

(1) 检索出你所学专业的某所学校的人才培养方案, 了解本专业将学习的专业课程, 并将相关信息填入下面的空行中。

专业名称: _____。

检索词 (检索表达式): _____。

人才培养方案的链接地址: _____。

本专业的主要专业课程: _____。

练习与提高

2. 利用百度搜索引擎 (<http://www.baidu.com/>) 完成以下检索。

(2) 选择目前正在学习的一门专业课程, 检索该课程的教学大纲 (doc或docx格式) 及教学课件 (ppt或pptx格式), 将相关信息填入下面的空行中。

课程名称: _____。

教学大纲检索: 检索词 (检索表达式) _____。

教学大纲下载链接地址: _____。

教学课件检索: 检索词 (检索表达式) _____。

教学课件下载链接地址: _____。

3. 在百度的高级搜索框里可以使用哪些搜索运算符?