

第六章

开放资源的获取

《文献检索与科技论文写作》

目录

CONTENTS

1

典型搜索引擎及工作原理

2

开放资源检索



典型搜索引擎及 工作原理

6.1 典型搜索引擎及工作原理

自1990年第一个搜索引擎Archie推出以来，各种搜索引擎层出不穷。现如今，搜索引擎的定义已经不仅仅是局限于狭义上的搜索引擎本身了，除了百度、谷歌这样的搜索引擎，很多手机应用软件里的搜索功能，都属于广义上的搜索引擎。搜索引擎可以分为五代。

第一代：分类目录时代。导航网站是搜索引擎第一代的代表。主要是一些分类网址，用户可以从这个分类目录里找到自己想浏览的网页。

第二代：文本检索时代。搜索引擎查询信息的方法是通过用户所输入的查询信息提交给服务器，服务器通过查阅，返回给用户一些相关程度高的信息。这代的搜索引擎的信息检索模型主要包括例如布尔模型、概率模型或者向量空间模型。通过这些模型来计算用户输入的查询信息是否与网页内容相关程度高低，将相关度高的则返回给用户。采取这种模式的搜索引擎主要是一些早期的搜索引擎，例如像Alta Vista、Excite等等。

第三代：整合分析时代。主要采用外部链接形式，通过每个网站的推荐链接数量来判断一个网站的流行性和重要性。然后搜索引擎再结合网页内容的重要性来和相似程度来改善用户搜索的信息质量。谷歌首先使用这种模式并且大获成功，这一成就在当时引起了学术界和其他商业搜索引擎的极度关注。

6.1 典型搜索引擎及工作原理

第四代：用户中心时代。主要是以用户为中心。当客户输入查询的请求时候，同一个查询的请求关键词在用户的背后可能是不同查询要求。例如用户输入的是“苹果”，那么作为一个想要购买iPhone的用户和一个果农来说，那么要求就是大大地不一样。甚至是同一个用户，所查询的关键词一样，也会因为所在的时间和所在的场合不同而返回的结果不同的所有主流搜索引擎，都在致力于解决同一个问题：怎样才能从用户所输入的一个简短的关键词来判断用户的真正查询请求。

移动设备的使用者，即使不向移动搜索互联网授权任何特征信息，移动搜索互联网仍然可以通过移动设备使用者搜索时的大量特征，比如上网的时间习惯、操作习惯、内容归类去逐渐勾勒出这人使用者的特征信息，这种“推测式”算法的可能性也是由于移动设备具有唯一性、随身性而产生的。这些使用者的蛛丝马迹，都是未来移动互联网搜索引擎进行“商业数据挖掘”的巨大宝藏。移动互联网搜索的必然使命，就是：提供精准到个人的搜索。

第五代：生活生态圈。第五代搜索引擎是基于物联网的搜索，物联网搜索拥有更广阔的搜索空间，能预测到物联网一个最典型的应用就是：寻物或物寻人。比如远程看管小孩、老人，或搜索走失小孩。同时，还有可能东西找你，比如AI（人工智能）泊车后超过某个时间点，车主动呼叫你；饭煮好后电饭煲主动呼叫你。

6.1 典型搜索引擎及工作原理

目前，在世界范围内较流行的搜索引擎有：谷歌（Google）、必应（Bing）、百度（Baidu）、雅虎（Yahoo）、Yandex、Ask、Duckduckgo、Naver、AOL、Seznam、有道、搜搜、搜狗、即刻、神马、360等。谷歌在全世界的搜索引擎份额超过了70%，百度和必应搜索引擎的全世界份额接近10%。必应是微软旗下的搜索引擎，它在多个国家都占据了搜索引擎排名第二的份额，虽然很多人往往会忽视它，但是每个月也有13亿人通过必应进行搜索。百度在我国占据了搜索份额的70%以上。

搜索引擎的基本工作原理包括如下三个过程：首先在互联网中发现、搜集网页信息；同时对信息进行提取和组织建立索引库；再由检索器根据用户输入的查询关键字，在索引库中快速检出文档，进行文档与查询的相关度评价，对将要输出的结果进行排序，并将查询结果返回给用户。

6.1 典型搜索引擎及工作原理

① 搜集信息：搜索引擎的信息搜集基本都是自动的。搜索引擎利用称为网络蜘蛛的自动搜索程序来连上每一个网页上的超链接。

② 整理信息：搜索引擎整理信息的过程称为“创建索引”。搜索引擎不仅要保存搜集起来的信息，还要将它们按照一定的规则进行编排。

③ 接受查询：用户向搜索引擎发出查询，搜索引擎接受查询并向用户返回资料。搜索引擎每时每刻都要接到来自大量用户的几乎是同时发出的查询，它按照每个用户的要求检查自己的索引，在极短时间内找到用户需要的资料，并返回给用户。



开放资源检索

6.2 开放资源检索

6.2.1 概述

Open Access (简称OA) 通常被译作开放获取、开放存取、开放共享、开放访问、开放阅览等。OA兴起于20世纪90年代, 是国际科技界、学术界、出版界、信息传播界为推动科研成果利用网络自由传播而发起的运动。目的是促进科学信息的广泛传播, 学术信息的交流与出版, 提升科学研究的共利用程度, 保障科学信息的长期保存。开放获取不是一个技术问题, 而是一个观念和文化问题。

“开放获取”是在基于订阅的传统出版模式以外的另一种选择。通过数字技术和网络化通信, 任何人都可以及时、免费、不受任何限制地通过网络获取各类文献, 包括经过同行评议过的期刊文章、参考文献、技术报告、学位论文等全文信息, 用于科研教育及其他活动。唯一的限制: 要求保证作者拥有保护作品完整性的权利, 同时在使用作者作品时注意相应的引用信息。这是一种新的学术信息交流的方法, 作者提交作品不期望得到直接的金钱回报, 而是提供这些作品使公众可以在公共网络上利用。

开放获取的核心特征是在尊重作者权益的前提下, 利用互联网为用户免费提供学术信息和研究成果的全文服务。开放获取是基于互联网的学术传播机制。

6.2 开放资源检索

开放获取有两种具体形式：

(1) 依托互联网，采用发表付费，阅读免费的形式。发表付费：对于作者，文章以开放获取的形式发表，作者或者作者所在机构支付费用，来获得更多阅读和引用的模式，显著提高引用率。阅读免费：任何人都可以免费获取开放出版的文献全文。

(2) 作者自建文档向外界免费开放。作者将自己的研究成果以电子全文形式存放在一个中心服务器或WEB网页上供同行免费利用的一种方式。开放获取的实现通道主要包括：开放获取期刊、开放仓储（机构知识库、学科知识库）、电子预印本。

1 开放获取期刊（OA Journals）是一种论文经过同行评审的、网络化的免费期刊，旨在使所有用户都可以通过因特网无限制地访问期刊论文全文。

2 开放仓储（Open Repositories and Archives）是研究机构或作者本人将已发表或未公开发表的文献存储在学科知识库、机构知识库或个人网站上供免费获取。

6.2 开放资源检索

3 知识库包括学科知识库和机构知识库。学科知识库：对某机构拥有的某个学科领域的各种类型的资源进行收集、整理、描述、组织、索引，实现长期保存和广泛传播，实现资源的共享和利用的知识库。

4 电子预印本（e-print）指科研工作者的研究成果还未在正式出版物上发表，而出于和同行交流目的自愿先在学术会议上或通过互联网发布的科研论文、科技报告等文献。

开放获取的主要影响：对于作者来说，扩大了学术成果传播的范围（打破了地域、学科、行业的限制，读者群得到拓展），缩短了传播的时滞，提高了成果的影响力和显示度，但需要支付高昂的OA出版费用；对于读者来说，消除了价格壁垒，提高了读者获取信息的便利性和获取信息的范围；对于研究机构来说，学术成果向更广的范围的传播，彰显了机构的科研实力，提高了机构的知名度和显示度，有助于机构树立行业威望，提升行业地位；对于资助机构来说，科研成果得到了更广泛的传播，使得投资产生了更大的社会效益；对于出版商来说，扩大了期刊文章的使用率和影响，直接的效果是提高了文献的引文量，但也影响了发行量，降低了利润，分流了优质稿源。

开放获取存在的主要问题：经济问题、质量控制问题、知识产权保护问题。目前尚未建立起合理有效的经费支持模式，当前的经费支持模式主要有：作者付费、机构成员制度（会员制）、各类资助（基金支持）。

6.2 开放资源检索

6.2.2 开放获取资源介绍及链接

6.2.2.1 开放获取期刊

1. cnpLINKer (中图链接服务)

由中国图书进出口(集团)总公司开发,提供国外期刊网络检索的系统,于2002年底开通运行。目前系统共收录了国外1 000多家出版社的46 693种期刊的目次和文摘数据,并保持时时更新。其中包括29 309种Open Access Journals供用户免费下载全文。cnpLINKer在线数据库检索系统是中图公司为了读者用户方便快捷地查阅国外各类期刊文献而组织开发的综合网络平台。

内容涵盖哲学、社会科学、政论时事、军事、经济、语言文字学、文学、艺术、历史、自然科学、地球科学、生物科学、医药卫生、农业科学、工程技术、运输工程、环境科学与技术17个学科领域。网址: <http://cnplinker.cnpeak.com>。

6.2 开放资源检索

2. Socolar

Socolar由中国教育图书进出口公司开发，为用户提供重要的OA资源的一站式服务，实现以下功能：全面系统收录重要的OA资源，包括重要的OA期刊和OA仓储，为用户提供多层次的浏览、检索和全文链接服务。目前已收录期刊超过8 000种、仓储超过1000个，揭示的文章数量近1600万篇，内容涵括所有的学科领域，年增加数据量超过200万篇文章。该网站也为学者提供学术文章和预印本的OA出版和仓储服务。网址：<http://www.socolar.com>，主页如下图所示。



Socolar主页

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/507020033046006153>