

第5章 中文文献数据库及使用方法

学习目标

1. 熟悉常见文献数据库的分类方式,能够根据实际需要进行有针对性的文献数据库选择。
2. 了解常见文献数据库资源的获取方式,掌握常见的文献检索技术。
3. 了解国内常用的文献数据库的特点,熟练掌握基于主题、关键词、作者、文献类别等不同检索标准的文献检索方法。
4. 尝试使用高级检索方法进行精准文献检索。

在网络信息资源管理活动中,数据库技术具有非常广泛的应用领域,它是实现资源共享、节省开支、提高系统反应能力、提高系统工作质量和服务水平的重要手段和技术保证。文献数据库是专门用来对文献进行组织管理的数据库。

不同的数据库在服务范围、内容涵盖及检索方式上各有差别,只有了解并熟练掌握各类文献数据库的检索特性和功能,才能更好地获取所需信息。在我国,CNKI 数据库、万方数据资源系统、维普期刊全文数据库、人大复印报刊等全文数据库,中国高等教育文献保障系统、读秀学术搜索等大型综合性检索系统是获取文献信息的主要渠道。

图书馆电子数据库资源以其内容丰富和方便使用的优点已成为各级图书馆信息资源建设的重点,也成为广大读者查找和利用文献信息的主要阵地。各个图书馆在数据库信息资源的发布方式上可能有所区别,例如有的是通过购买和安装各种数据库数据的形式,有的是通过制作镜像站点的形式。不管是何种形式,其对资源的使用方式基本是一样的。

5.1 文献数据库综述

5.1.1 常见文献数据库的分类

从不同的角度看待文献数据库,可以将文献数据库分为不同的类型,常见的划分方法有以下 3 种。

1. 按语种划分

按照语种进行划分,文献数据库分为中文数据库和外文数据库。

2. 按文献类型划分

按照文献类型划分,文献数据库分为电子图书数据库、电子期刊数据库、学位论文数据库、会议论文数据库、特种文献数据库、多媒体资源数据库等,不同类型的数据库存储的数据类别也不同。

3. 按照信息类型划分

按照数据库存储的信息类型划分,文献数据库分为文献线索型数据库、全文型数据库、术语型数据库、事实型数据库、数值型数据库、多媒体数据库等。其中,文献线索型数据库又分为书目型数据库、索引型数据库、文摘型数据库。

5.1.2 数据库资源的常用检索方法

1. 浏览法

浏览法即通过数据库提供的浏览功能查找目标文献。浏览功能一般提供两种浏览途径,分别是数据库平台子库(资源导航)的浏览和基于学科或出版物名称的浏览。

数据库平台子库浏览主要是通过浏览综合数据库中每个子库熟悉数据库平台并定位目标内容,该方法了解一个数据库概况的主要窗口;基于学科或出版物名称的浏览一般应用在外文全文数据库中。

2. 检索法

检索法是从数据库中获取资源的最重要的方法。常见的数据库会提供简单检索、高级检索和专业检索 3 种检索方式。

1) 简单检索

简单检索(Basic Search, Easy Search, Quick Search)即一框式检索,是人们普遍使用的文献检索方法,该方法一般适用于精确文献的查找,不适合查找批量文献。

2) 高级检索

高级检索(Advanced Search)又称菜单式检索,是一种通过增加限定条件以缩小检索文献范围的检索方法。

3) 专业检索

专业检索(Expert Search)又称检索式检索,该方法一般被专业检索人员用来进行精准的文献检索。

需要说明的是,个别外文文摘数据库只提供高级检索和专业检索(如 EI、WOS),也有一些外文全文数据库只提供简单检索和高级检索,不提供专业检索(如 SDOL、Springer 等),而且高级检索与搜索引擎的检索类似,字段更加模糊。提供哪些类型的检索一般需要

根据数据库拥有的资源特点及数据库的功能定位决定,通常文摘数据库的检索方式更加强调专业,而全文数据库的检索方式相对模糊。

除上述一般检索方式之外,某些数据库还提供了特殊检索方式,这些检索方式通常分为叙词检索及其他特殊检索。

1) 叙词检索

叙词检索是一种基于受控词检索的方法,如 EI、CSA 数据库的 Thesaurus search 和 EBSCO 数据库的 Subjects 等都采用了叙词检索方法。

2) 其他特殊检索

特殊检索方式是除了上述检索方式之外的检索方法,如 WOS 平台的被引参考文献检索、CA 数据库的化学分子检索、中国知网的作者发文检索和句子检索等都是特殊检索方式。

当用户进行检索时,通常需要基于字段实现,字段分为常见字段和特殊字段。

1) 常见字段

常见字段为主题、作者、关键词等字段,详细介绍如表 5-1 所示。

表 5-1 常见检索字段

中 文 字 段	英 文 名 称
题名	Title
主题(题名-关键词-摘要、复合字段)	Subject-Title-Abstract Topic
文献出版物(文献来源、期刊名称)	Source Title Publication Name
作者	Author
机构(作者单位)	Author Affiliation

需要说明的是,主题字段又称“题名-关键词-摘要字段”或“复合字段”,基于主题搜索时,数据库平台一般会在题目、关键词和摘要字段中进行匹配。

2) 受控字段

如 EI 数据库中的 EI main heading、Controlled Term、Uncontrolled Term 等都属于受控词字段。

3) 与引文索引相关的字段

常见的与引文索引相关的字段有被引频次(中国知网)、Cited Author 和 Cited Title (WOS 核心合集)等。

3. 个性化分析与设置方法

当前,很多数据库都提供了个性化分析功能,通常分析都以可视化的形式展示给用户,而且用户还可以导出分析结果。对检索结果进行分析是用户经常使用的个性化分析对象,用户可以对某个主题内的年代、作者、机构、关键词、出版物等各种字段进行分析。

此外,很多平台也提供了个性化设置功能,如提供邮件通告、被引追踪、RSS 订阅、期刊收藏等个性化功能,要使用这些功能,一般都需要提前进行平台用户注册。

5.2 CNKI 中国知网

5.2.1 简介

1. CNKI 发展简介

国家知识基础设施(National Knowledge Infrastructure, NKI)的概念是由世界银行在《1998 年度世界发展报告》中提出的。NKI 是一个国家生产、传播、扩散与应用知识的互动式知识网络,它基于现代化基础设施和数字化技术产生,以互动的方式将从事知识生产、知识应用的机构和个人关联到一起,形成一个知识网络。其中,知识库是其核心和关键。在知识库中,不但存放着经过重新加工整理的专业知识,还同时建立了各种知识之间的桥梁和网状关联结构,人们借此能够以最方便的方式获取和交流知识,从而进行高效率的知识生产与应用。

1999 年 3 月,王明亮提出建设中国知识基础设施工程(China National Knowledge Infrastructure, CNKI)的提议,该提议旨在全面打通知识生产、传播、扩散与利用等各环节信息通道,总目标是打造支持全国各行业知识创新、学习和应用的交流合作平台。

从建立至今, CNKI 经历了两个主要发展阶段,分别是 CNKI 1.0 阶段和 CNKI 2.0 阶段。CNKI 1.0 是在建成《中国知识资源总库》基础工程后,从文献信息服务转向知识服务的重要转型。CNKI 1.0 的目标是面向特定行业领域知识需求进行系统化和定制化知识组织,构建基于内容内关联的“知网节”,并进行基于知识发现的知识元之间的关联关系挖掘,该研究代表了中国知网服务知识创新与知识学习、支持科学决策的产业战略发展方向。在 CNKI 1.0 基本建成以后,中国知网充分总结行业知识服务的经验教训,以全面应用大数据与人工智能技术打造知识创新服务业为新起点, CNKI 工程跨入了 2.0 时代。CNKI 2.0 的目标是将 CNKI 1.0 基于公共知识整合提供的知识服务深化到各行业机构,知识创新的过程与结果相结合,通过更为精准、系统、完备的显性管理以及嵌入工作与学习具体过程的隐性知识管理,提供面向问题的知识服务和激发群体智慧的协同研究平台,其重要标志是建成“世界知识大数据(WKBD)”、建成各单位充分利用“世界知识大数据”进行内外脑协同创新、协同学习的知识基础设施(NKI)、启动“百行知识创新服务工程”、全方位服务中国世界一流科技期刊建设及共建“双一流数字图书馆”。

2. CNKI 资源简介

CNKI 亦解读为中国知网,是具有知识的整合、集散、出版、传播功能的知识门户网站。CNKI 工程是以实现全社会知识资源传播共享与增值利用为目标的信息化建设项目,由清华大学和清华同方发起。CNKI 经过多年的能力,采用了自主开发并具有国际领先水平的数字图书馆技术,建成了世界上全文信息量规模最大的 CNKI 数字图书馆,并建设了《中国知识资源总库》及 CNKI 网络资源共享平台,通过产业化运作,为全社会知识资源高效共享提供最丰富的知识信息资源和最有效的知识传播与数字化学习平台,成为了全球最大的知识门户网站。

《中国知识资源总库》是 CNKI 的核心资源,其囊括了自然科学、人文社会科学及工程技术等各领域的知识,收录了包括期刊、报纸、优秀硕博学位论文、全国重要会议论文、中小学多媒体教辅等上千种数据库,其所含的重点数据库主要如下。

(1) 学术期刊库。实现中文和外文期刊的整合检索。其中,中文学术期刊有 8560 余种,含北大核心期刊 1970 余种,网络首发期刊 2240 余种,最早可回溯至 1915 年,共计 5910 余万篇全文文献;外文学术期刊包括来自 80 个国家及地区 900 余家出版社的期刊 7.5 万余种,覆盖 JCR 期刊的 96%,Scopus 期刊的 90%,最早可回溯至 19 世纪,共计 1.1 余亿篇外文题录,可链接全文。

(2) 学位论文库。包括中国博士学位论文全文数据库(CDFD)和中国优秀硕士学位论文全文数据库(CMFD),是目前国内资源完备、质量上乘、连续动态更新的中国博硕士学位论文全文数据库。该库出版了 510 余家博士培养单位的博士学位论文 40 余万篇,780 余家硕士培养单位的硕士学位论文 480 余万篇,最早可回溯至 1984 年,覆盖基础科学、工程技术、农业、医学、哲学、人文、社会科学等各个领域。

(3) 会议论文库。重点收录 1999 年以来中国科协系统及国家二级以上的学会、协会、高校科研院所、政府机关举办的重要会议以及在国内召开的国际会议上发表的文献,部分重点会议文献可回溯至 1953 年,目前已收录国内会议、国际会议论文集 4 万本,累计文献总量达 350 余万篇。

(4) 中国重要报纸全文数据库。中国重要报纸全文数据库是以学术性和资料性报纸文献为出版内容的连续动态更新的报纸全文数据库。报纸库收录并持续更新 2000 年以来出版的各级重要党报、行业报及综合类报纸 500 余种。

(5) 中国图书全文数据库。中国图书全文数据库又称心可书馆,以中国知网海内外 2 亿专业读者为服务对象,集图书检索、专业化推荐、在线研学、在线订阅功能于一体。通过参考文献、引证文献等关联关系,实现图书内容与其他各类文献的深度关联融合。目前已收录精品专业类图书 14956 本,覆盖人文社科、自然科学、工程技术等各领域,并实时更新。

(6) 学术辑刊库。学术辑刊库收录自 1979 年至今国内出版的重要学术辑刊共计 1060 余种、30 余万篇。辑刊的编辑单位多为高等院校和科研院所,其内容覆盖自然科学、工程技术、农业、哲学、医学、人文社会科学等各个领域。

CNKI 平台提供了单库检索和跨库检索两个功能。此外,CNKI 还提供知识问答、百科、词典、概念、手册等检索功能,还可以借助中文引文数据库实现引文检索。中国知网收录了学术期刊、学位论文、会议、图书、报纸、年鉴、专利、标准、成果、学术辑刊、法律法规、政府文件、企业标准、科技报告、政府采购等多种文献资料。

3. 两类基本检索方式

CNKI 数据库跨库检索平台可以针对学术期刊、学位论文、会议论文、年鉴、专利、标准、成果、报纸、图书、学术辑刊进行同时检索,如图 5-1 至图 5-3 所示。

跨库检索是指将检索范围限定在多个数据库中,如同时在学术期刊数据库、学位论文数据库中检索。CNKI 默认为跨库检索,其中包括学术期刊、学位论文、会议、报纸、专利、成果、图书和学术辑刊,用户可以根据自己的实际需要添加来源类型,实现跨库检索范围的设置。跨库检索包括初级检索、高级检索和专业检索。跨库检索实现了用户在一个界面下检



图 5-1 中国知网首页



图 5-2 检索文献来源设置



图 5-3 检索结果示意

索多个数据库的功能,省去了多个库逐一登录、逐一检索的繁杂,检索过程简单、快捷,检索界面格式统一,减轻了用户的使用负担。

单库检索即针对一个文献来源库进行检索。例如要找期刊文章就去学术期刊数据库,找学位论文就去学位论文数据库检索。要实现单库检索,用户需要将 CNKI 默认勾选的多个数据库取消,仅保留要检索的数据库即可。

5.2.2 检索方法与技巧

1. 传统检索方法

CNKI 传统检索方法是通过在 CNKI 主页的搜索框中输入关键词获得查询结果的方法,图 5-4 和图 5-5 是以“人工智能”为主题词进行文献资料搜索的示例。



图 5-4 在 CNKI 中国知网检索主页输入检索词



图 5-5 搜索结果示意

由图 5-5 可以看出,以“人工智能”为主题词进行搜索获得了数十万条结果,其来源包括学术期刊、学位论文、会议论文、报纸、图书、标准、成果、学术辑刊、特色期刊及视频。如此多的检索结果是否都需要查看?需要查看哪些?读者需要进一步简化检索结果,此时可以基于检索需求,根据第 1 章介绍的文献类型的特点及应用场合,通过减少文献来源库种类的方法实现,如设置检索结果来源为学术期刊或学位论文等。

2. 一框式检索

一框式检索可以在 CNKI 主页检索字段处选择特定的字段类型提高检索结果的准确性。

CNKI 提供包括主题、篇关摘、关键词、篇名、全文、作者、第一作者、通讯作者、作者单位、基金、摘要、小标题、参考文献、分类号、文献来源、DOI 在内的 16 种字段类型,下面介绍每个字段类型的检索范围,以方便读者根据自己的实际需要选择相应的字段聚焦检索范围。



图 5-6 CNKI 检索字段示意

1) 主题检索

主题检索是指在中国知网标引出来的主题字段中进行检索,该字段内容包含一篇文章的所有主题特征,同时在检索过程中嵌入了专业词典、主题词表、中英对照词典、停用词表等工具,并采用关键词截断算法,对低相关或微相关文献进行截断。

2) 篇关摘检索

篇关摘检索是指在篇名、关键词、摘要范围内进行检索,具体参见下文关于篇名检索、关键词检索、摘要检索的介绍。

3) 关键词检索

关键词检索的范围包括文献原文给出的中英文关键词和对文献进行分析计算后机器标引出的关键词。机器标引的关键词基于对全文内容的分析,并参考了专业词典,可以解决文献作者给出的关键词不够全面准确的问题。

4) 篇名检索

期刊、会议、学位论文、辑刊的篇名为文章的中英文标题;报纸文献的篇名包括引题、正标题、副标题;年鉴的篇名为条目题名;专利的篇名为专利名称;标准的篇名为中英文标准名

称;成果的篇名为成果名称;古籍的篇名为卷名。

5) 全文检索

全文检索是指在文献的全部文字范围内进行检索,包括文献篇名、关键词、摘要、正文、参考文献等。

6) 作者检索

期刊、报纸、会议、学位论文、年鉴、辑刊的作者为文章的中英文作者;专利的作者为发明人;标准的作者为起草人或主要起草人;成果的作者为成果完成人;古籍的作者为整书著者。

7) 第一作者检索

只有一位作者时,该作者即为第一作者;有多位作者时,将排在第一个的作者认定为文献的第一作者,也就是责任人。

8) 通讯作者检索

目前,期刊文献对原文的通讯作者进行了标引,可以按通讯作者查找期刊文献。通讯作者指课题的总负责人,也是文章和研究材料的联系人。

9) 作者单位检索

期刊、报纸、会议、辑刊的作者单位为原文给出的作者所在机构的名称;学位论文的作者单位包括作者的学位授予单位及原文给出的作者任职单位;年鉴的作者单位包括条目作者单位和主编单位;专利的作者单位为专利申请机构;标准的作者单位为标准发布单位;成果的作者单位为成果第一完成单位。

10) 基金检索

根据基金名称可检索受到此基金资助的文献。支持基金检索的资源类型包括期刊、会议、学位论文、辑刊。

11) 摘要检索

期刊、会议、学位论文、专利、辑刊的摘要为原文的中英文摘要,原文未明确给出摘要的,提取正文内容的一部分作为摘要;标准的摘要为标准范围;成果的摘要为成果简介。

12) 小标题检索

期刊、报纸、会议的小标题为原文的各级标题名称;学位论文的小标题为原文的中英文目录;中文图书的小标题为原书的目录。

13) 参考文献检索

检索参考文献中包含检索词的文献,支持参考文献检索的资源类型包括期刊、会议、学位论文、年鉴、辑刊。

14) 分类号检索

通过分类号检索可以查找到同一类别的所有文献。期刊、报纸、会议、学位论文、年鉴、标准、成果、辑刊的分类号指中图分类号;专利的分类号指专利分类号。

15) 文献来源检索

文献来源指文献出处,期刊、辑刊、报纸、会议、年鉴的文献来源为文献所在的刊物;学位论文的文献来源为相应的学位授予单位;专利的文献来源为专利权利人/申请人;标准的文献来源为发布单位;成果的文献来源为成果评价单位。

16) DOI 检索

输入 DOI 号可检索期刊、学位论文、会议、报纸、年鉴、图书。国内的期刊、学位论文、会

议、报纸、年鉴只支持在知网注册 DOI 的文献。

3. 知识元检索

知识元是显性知识的最小可控单位,是无可再分割的具有完备知识表达的知识单位,即能够表达一个完整的事实、原理、方法、技巧等。知识元检索即对一个个完整的知识元进行检索,可以将知识元看作一篇文章。

单击中国知网首页“文献检索”下方的“知识元检索”按钮,就可以进入知识元检索。知识元检索可检索的资源类型包括知识问答、百科、辞典、手册、工具书、图片、统计数据、指数、方法、概念。如用户要查阅“人工智能”的概念定义,可以在知识元检索的检索词输入框中输入“人工智能”,选择检索类型为“概念”,输入完成后按 Enter 键或单击输入框右侧的“搜索”图标 ,就可以看到检索结果,如图 5-7 至图 5-9 所示。



图 5-7 知识元检索入口



图 5-8 知识元检索示例(1)