



大千世界,如何选择信息源

——使用航海指南针



信息检索无时不有、无处不在

可以毫不夸张地说,在学习、生活和工作中信息就像空气一样,是无处不在、无时不缺的必需品。小到考虑今天出门是否带伞,大到完成科研攻关项目,都需要相关信息的支持。只是有时信息随手可得,有时信息的查找与整个实际问题的解决紧密联系在一起,使得我们没有关注到查找信息这一过程。

信息源种类繁多、特点各异

现实生活中,报纸、期刊、杂志、网络等都时刻在传递着各式各样的信息。这些发出信息的实体可统称为信息源。丰富的信息源使我们生活在一片信息的海洋中,有人能在海洋中畅游,有人却会迷失方向,区别就在于是否会使用航海指南针,即是否了解并能有效地选择信息源。

因此,以下将主要介绍如何明确信息需求,然后根据信息需求以及不同信息源的特点选出最有效的信息源,以快速、正确地找到解决信息问题的渠道。

3.1 界定问题

解决问题的前提是对问题本身有明确的认识,信息问题也不例外。然而在实际的检索中这一点却常常被忽略。人们往往有这样的体验:在对自己要查找的对象所知甚少的情況下匆匆上马,就犹如在错误方向上奔跑,速度越快,离目的地越远,这时再重新定位问题已浪费了大量时间和精力。所以界定问题是信息检索的第一步,其内涵主要包括分析研究问题、建立背景知识、拟定主题概念。

1. 分析研究问题



所谓“工夫在诗外”，信息的获取与利用也是这样。这里的“诗外”就是自身对研究问题的了解程度。要查找信息资料，首先必须对问题进行分析研究，以确定该问题需要查找的文献类型、所属领域、地域范围和哪一时期的文献资料。还必须明确研究目标，查找的目的是为了成果查新或鉴定还是为了撰写论文，或是为了解决某个实际问题，是属于一般性的研究问题还是属于具有创新性的基础研究，以确定检索的深度和广度。对研究问题的了解越多，所走的弯路就越少。实际上当你分析完问题后，就已经站在了较高的起点上。

对问题的分析可以从研究目标、时间、地域及事件等方面进行，如表 3-1 所示。

表 3-1 研究问题分析——以“车用替代燃料的应用情况研究”为例

研究问题	车用替代燃料的应用情况研究
研究目标	形成调研报告供公司参考，以进一步考察生产车用替代燃料的可行性
时间范围	近 10 年
地域范围	国内、国外
背景/事件	车用替代燃料的蓬勃兴起和广泛使用

经过初步分析之后，再进一步思考以下几个问题，可以更快、更准确地把握研究问题：

- (1) 研究问题本身包含的主题概念有哪些？
- (2) 研究问题中可能包括哪些子问题？涉及哪些学科领域？
- (3) 子问题是否过多？是否需要缩小研究问题的范围？
- (4) 研究问题是否过于狭窄，意义不大？是否需要扩大研究问题的范围？
- (5) 是需要原始资料还是经过专家学者整理过的资料？是需要一般性的信息还是学术性信息？
- (6) 对信息的新颖性要求有多高？
- (7) 地域范围的选择是越宽泛越好还是越典型越好？

2. 建立背景知识



分析研究问题过程中的许多问题往往不能一步回答得比较准确。这主要是由于该问题与自身现有的知识脱节较大，因而无法准确地界定。这时就需要通过初步检索，建立背景知识。阅读与研究主题相关的概览性资料，是建立背景知识的最佳渠道。概览性资料通常出现在前面说过的三次文献中，如百科全书、字词典及年鉴等参考资料中。通过阅读这些资料，可以加深对研究问题的理解，明确检索目标。同时对问题涉及的名词术语进行记录，了解清楚该研究对象的学名、俗名、商品名、缩写、同义词等，特别是在用非母语检索工具之前，这些名词术语花样繁多的英文表达更是不能疏漏。这将给后续拟定主题概念工作带来极大

便利。

建立背景知识的过程往往也是辨别待解决信息问题研究的必要性及可行性的重要时机。通过对问题的初步了解,你可能发现该问题的提法本身就存在问题,需要重新斟酌;也有可能该问题包含很多子问题或涉及过多自身并不熟悉的学科领域,解决起来比较困难,这时就应该及时调整研究问题,以免解决时间拖延过长,影响后续工作的进行。

以“车用替代燃料的应用情况研究”为例,Amy 对于“车用替代燃料”的具体含义并不清楚,需要查找资料进一步理解。因此,她登录百度搜索引擎,使用“车用替代燃料”为检索词,在维普咨询网查得:车用清洁替代燃料包括天然气、液化石油气、氢能源、二甲醚、醇类燃料及生物燃料。通过进一步的文献研究得知“车用替代燃料”包括的能源类型十分丰富,如表 3-2 所示。

表 3-2 车用替代燃料包含的主题概念

信息问题	车用替代燃料
别名	车用代用燃料
包含的主题概念	天然气、生物柴油、生物燃料、煤基燃料、醇类燃料(包括乙醇燃料、甲醇燃料)、液化石油气(LPG)、CNG、二甲醚(DME)、合成柴油、生物燃料、气体燃料、燃料电池、清洁燃料、氢能、电能、纯电动、混合动力

这时 Amy 意识到如果继续上述信息问题的研究,将需要对列表中的概念进行逐一检索,分别了解其应用情况,工作量是很大的,在规定的时间内很难面面俱到地将这几个方面调研清楚。于是她决定缩小研究问题的范围,结合原有的知识结构她选取了其中一点——“乙醇燃料”为研究对象。

为进一步明确乙醇燃料的含义,她登录百度百科,输入“乙醇燃料”,查得:车用乙醇燃料也称为乙醇汽油,是指在不含 MTBE 含氧添加剂的专用汽油组分油(由炼油厂或石油化工厂生产的用于调和车用乙醇汽油的调和油)中,按体积比加入一定比例(我国目前暂定为 10%)的变性燃料乙醇,由车用乙醇汽油定点调配中心按 GB 18351—2004 的质量要求,通过特定工艺混配而成的新一代清洁环保型车用燃料。理解了这一核心概念之后,Amy 又查找了其他相关信息资源进一步理解这一研究问题的具体含义,确定了研究时间范围、地域范围等具体检索要求。

3. 拟定主题概念

同样的研究对象/方法,国内外不同的研究者在发表文章时对其称谓可能不同,再加上对领域了解不够透彻,仅用自己脑子里固有的主题词去查,往往很多文献都找不着,影响检索的效果。这时拟定主题概念的重要性就体现出来了。通过对背景知识的了解,确定了研究问题的范围、时间、地域等之后,能够帮助我们找出信息问题的主题概念,并用尽可能多的词加以表述。对于要查找外文资料的问题,则还需要将其翻译成所需语言。将主题概念及

相关词一一列出，会极大地提高后续查询操作的查全率，尽可能多地将相关资料一网打尽。

如何实际操作呢？你可以结合背景知识，打开思路，尽可能详细地用一小段话将研究问题阐述清楚，然后从中找出有明确含义的专指名词，结合一定的查询操作，逐个考虑该词能否作为主题词使用。对于可用的词语，寻找它的上位词（概念范围大于它的词语）、下位词（概念范围小于它的词语）、同义词及相关词（见表 3-3）。在此过程中还可以利用相关参考书（如“中国分类主题词表”等）帮助查找。

表 3-3 主题概念分析——以“车用乙醇燃料的应用情况研究”为例

研究问题	车用乙醇燃料的应用情况研究	
主题概念	车用乙醇燃料(methanol gasoline)	
	上位词	发动机燃料、发动机代用燃料、醇类燃料、醇燃料、替代能源 (vehicle alternative fuel, substitutive fuel for automobiles)
	同义词	酒精汽油、乙醇汽油(gasanol)
	相关词	环境保护(environment protection)



Tip 小提示

是否感觉通过上面的方法拟定主题词虽然能“一网打尽”，但有些费时费力，想要找到更快捷的方法呢？这里就告诉你一个小技巧。

我们知道，描述同类问题的文献在关键词的选取上会有很大相似性。依据这一点，可以就这些同类问题文献的作者给出的关键词对主题概念进行扩展。先粗泛地选取一两个词在数据库中检索，通过对检索结果的阅读，找出那些符合信息需要的文献，然后将这些文献所给出的关键词加入到本问题的主题词集中，继续检索，继续选择相关文献并提取关键词……如此循环几次，就方便地完成了主题概念的拟定。

得出了主题概念之后，就可以将其应用在各个数据检索系统中进行检索了。是选取其中的上位词检索以扩大检索范围、获得更多相关文献，还是只选取同义词以提高检索的查准率？这取决于采取的检索策略。

需要特别指出的是，在电子检索越来越方便的今天，可以使用很多宽泛的“自由词”作为主题概念进行检索。它的选取可以非常宽泛，只要在文章中出现过，就能被检索到。但由于是自由词，作者的使用非常随意，这就需要我们z把可能出现的所有自由词尽可能考虑完全，才不至于漏检。在检索的初期建议尽量使用。然而，当对问题有了一定了解后，应选取一些规范的、被数据库生产商用来做标引的“主题词”进行检索。例如，对于化学物质确定的课题，采用 CSA(剑桥科学文献)登记号检索是一种既准又全又快的方式。此外，Ei 中的 Ei Controlled Term 和 Ei Main Heading 途径均是经标引人员规范处理过的检索途径。

3.2 选择信息源

界定问题之后,就需要考虑从何处去查找信息这一问题。信息技术的广泛使用大大扩展了获取信息的渠道。一方面,我们需要了解现实生活中有哪些可获得的信息源;另一方面,我们更要明确不同信息源所提供的信息在时效性、深度、广度等方面都有很大区别。如何根据其不同特点,选出符合界定问题的信息源,是我们必须掌握的一项技能。

首先,从大类上可以将信息源分为以下几类:

- (1) 纸质期刊;
- (2) 各类文献数据库;
- (3) 网页;
- (4) 图书;
- (5) 杂志;
- (6) 报纸。



上述几个信息源看起来都很熟悉是吗?可是你了解它们各自的特点,知道解决什么问题应该查找哪个或哪些信息源吗?我们不妨设置几个场景,由你来考虑应该查找哪种信息源。

场景 1: 查阅最近一个场次的美国职业篮球联赛(NBA)比分。

场景 2: 要参加某公司的面试,想了解该公司的基本情况。

场景 3: 想查找关于一部近期上映电影的影评,以决定周末是否去电影院看看。

场景 4: 写一篇小论文,需要了解计算机的发展历史。

场景 5: 写一篇有关商业智能系统研究现状的报告。

是否没有想象中的那么容易区分?但了解了下文的信息交流模式及各信息源的特点之后,你就能根据问题的特点快速准确地选择信息源了。

1. 信息交流模式

首先,我们必须意识到信息有学术信息与一般性信息的区别。你可能从公司网站介绍和某篇研究同类产品属性的报告中了解到同一产品,但由于生产信息的目的及来源不同,检索的渠道及获得的信息量都有很大区别。比如说,待解决的信息问题有一定深度要求时,应尽可能选择期刊、图书、各类文献数据库,减少对非专业网站的检索,这会过滤掉大量看似相关但无助于问题解决的信息。分别学习两种不同的信息交流模式,有助于理解不同信息源具有不同特点的原因,进一步区分信息源。

学术信息交流是一个不断循环、螺旋前进的过程,如图 3-1 所示。研究人员或者申请发明保护,产生专利文献;或者在研究过程中将研究成果写成会议论文、研究报告与同行交流。会议论文、研究报告等信息源的时效性较强,信息量也较大,但是由于它们通常是非正式出版物,因此很多时候不易获得。得出较好的研究成果后,研究人员往往会写成论文在该领域的学报等期刊上发表,产生期刊论文。期刊论文是学术研究的重要产物,一般都有较高的理论或实践意义,是学术信息交流中非常重要的一环。经过大致半年的时间,多数期刊论文、会议论文等会被各类文献数据库收录,读者可以通过网络方便地检索、浏览这些文献资源。有的研究人员进一步将结论性研究成果写成图书,从而使他人获得对这一领域全面完整的认识。为保存、组织和利用上述文献资源,图书馆通过购买图书和文献数据库等,建立 OPAC 系统(简单说来是一种在线书刊索引),形成文献信息门户,方便读者的文献检索和获取。

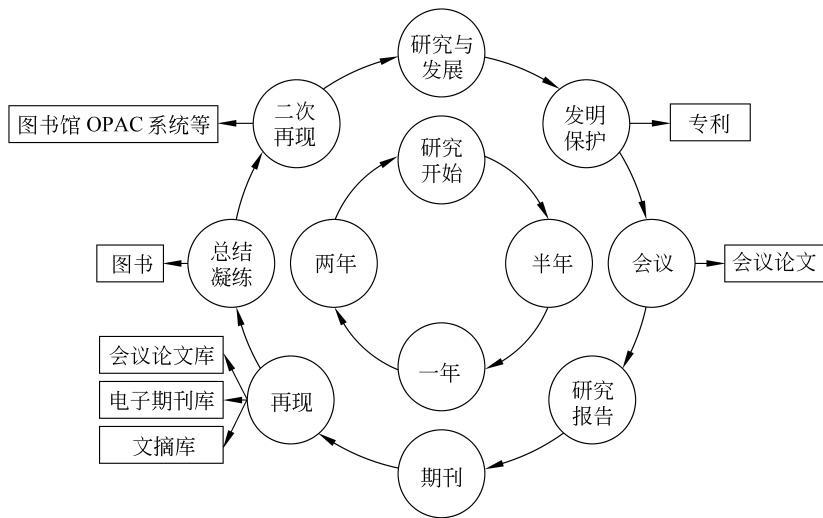


图 3-1 学术信息交流模式

一般性信息交流则是线性向前的,如图 3-2 所示。事件发生(如一场球赛、一次暴雪等)之后,网络、报纸等首先进行新闻报道,概要地描述事件本身。进而通过对相关事件的挖掘,分析事件的成因影响等,刊登在各类杂志上。经过一段时间的积累,著者对特定主题进行深入剖析,写成书籍,有的还可能形成电子书在网上流传。

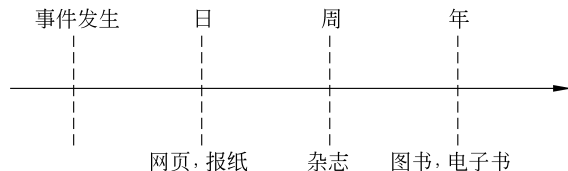


图 3-2 一般性信息交流模式

2. 不同信息源的特点比较

不同信息源在信息交流过程中所起的作用不同,也就形成了各自的特点。表 3-4 中对不同信息源及其特点进行了简单总结,以供在选择信息源的过程中参考。

表 3-4 不同信息源比较

信息源	信息内容	检索渠道	特点
网页、报纸	- 一般性信息,适合于收集关于 3W 问题(who, what, where,即某人或某组织在什么地方做了什么事情) - 无法深入了解原因(why)等较深入的问题	可利用网络上的搜索引擎	- 提供事件发生经过的即时报道 - 不同来源的不同信息内容重复性高 - 网页动态变化,不能长期保存,引用困难
杂志	仍较强调报道性质,适合收集 4W 问题(即 who,where, what, why)	索引方式较少,多依赖即时翻阅	- 与报纸比较,杂志对事件的报道较为详细,提供分析资料、统计等 - 通常不提供信息来源及参考文献
纸质期刊	是研究结果,详细且具有理论架构地进行探讨	各大图书馆的 OPAC 系统	- 研究对象及视角新颖,可把握研究前沿 - 对特定主题进行深入探讨 - 提供参考文献等相关注释,以此为入口可深入查找其他相关信息 - 提供客观的统计及图表
各类文献数据库	是研究结果,详细且具有理论架构地进行探讨	数据库商提供的检索平台及许多图书馆提供的统一检索系统	- 由于使用电子介质,不受地域限制,检索、下载和使用方便 - 与最新的纸质期刊相比,存在一定的时间滞后
图书	- 提供该主题的深入性分析资料 - 系统地编辑相关的学术性文章	各大图书馆的 OPAC 系统	- 提供特定主题的深入剖析,结论一般比较成熟,论述全面 - 由于生产周期相对较长,信息具有一定的时滞 - 提供参考文献等相关注释,以此为入口可深入查找其他相关信息

3. 信息源的选择策略

选择信息源这一步骤往往会被许多检索用户忽略。人们总是不自觉地根据“最省力原则”,利用最易获得的信息源进行检索。网络免费、便捷的特性使得它成为人们的首选信息源。因此大多数人在产生信息需要后的第一反应都是:上网——登录搜索引擎——浏览各类网页,希望从中能找到所需要的信息。这一步骤看似快捷,但由于网络资源本身的无序和缺乏深度,再加上搜索引擎机械利用关键词匹配相关性判断存在误差,往往检索耗费的时间

不短,查得的有用信息却寥寥无几,很容易挫伤人们查找信息解决问题的积极性。

难道说通过检索网络信息解决不了问题吗?当然不是。只是我们需要根据信息问题,结合不同信息源的特点,有选择地使用。

对于查找 NBA 比赛比分这类一般消息,选择浏览网页、报纸无疑是既准确又快捷的方式。而当需要了解该事件的更全面、更详细的介绍时,杂志可能是更好的选择。杂志的编者会结合统计数据,通过逻辑推理等方式对事件的前因后果进行一定深度的剖析,满足你的查找要求。像上文所提到的查看一部近期电影的影评就应选择杂志作为首选查找对象。

对于要求一定深度的研究性信息问题,如果仍主要通过搜索引擎检索网络信息源,将会产生以下问题:①返回的网页很多,但很多并不相关,需要从中筛选有价值的信息,花费时间较长。②事实性数据大量重复,有深度的信息却很少。③信息发布人各异,缺乏有效评定机制,导致出现许多失真甚至互相矛盾的信息,让人无从把握真实情况。这时如果能够了解一些相关专业机构发布信息的网址,如国家专利局、标准化委员会提供的专利标准及各种开放获取资源等,则会极大地提高检索效率。

网络上免费、有价值的学术信息资源是非常有限的。在解决研究性信息问题时,会议论文、研究报告、专利、期刊、各类文献数据库、图书等就成为重要的信息源。具体说来,会议论文和研究报告虽然时效性相对较强,但不易获得,需要关注相关研究机构及研究人员,积累获取渠道。文献数据库中收录了多种连续性的期刊以及部分会议论文、研究报告,且经过专业人员的组织整理,文献内容深刻全面,检索方便,获取容易,应作为检索的重要信息源。目前,许多图书馆等信息服务单位,都建立了统一的检索系统,用户可以像使用 Google 一样,在一个检索平台上方便地查找并下载众多文献数据库的数据。纸质期刊和图书作为重要的学术研究产出形式,是重要的学术信息源。其中,期刊以其研究对象及视角的新颖、详细且具有理论架构的探讨方式,能够帮助我们把握研究前沿。因此,在做开题报告、课题研究时尤为有用。通过阅读图书能把握基本原理,理清学科的发展脉络,这对于在研究初期夯实基础,在研究过程中查阅相关公式、方法都非常必要。因此,写论文、写报告、解决有一定难度的问题时,图书馆所拥有的这些信息资源应该成为首选。

3.3 信息获取的思路

面对不同的信息问题和特点各异的信息源,许多人经常犯“东一榔头,西一棒槌”的毛病,零零散散地查了许多东西后却总是感觉检索效果不佳。这在很大程度上是由于没有形成全局的信息获取思路而造成的。因此在深入学习每种检索工具的具体使用方法前,对信息获取的思路有一个完整的认识是非常必要的。

首先,我们需要明确,必须根据信息问题的不同特点来选择相应的信息源进行查询。如需要一般性、相对粗浅的信息,阅读网页是最便捷的选择;在面临研究性信息问题时,可检索

学术数据库(检索文摘库或全文数据库)获得更全面和系统的研究结果;对于现有数据库中无法获取的全文,可咨询图书馆或各地文献情报机构获取文献原文传递服务或上门借阅印刷型资源,如图 3-3 所示。

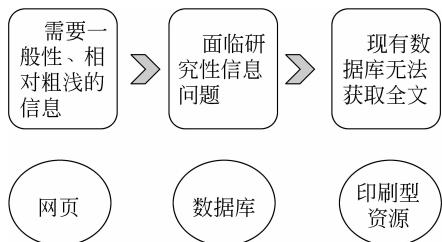


图 3-3 信息获取的思路

其次,我们更需要了解,尽管获取信息的渠道各有不同,检索方式也各异,但有一点是共通的,即无论选择了何种信息获取的渠道,都应首先获取题录信息,在对题录信息进行阅读、获得对该问题现有研究思路及结果的全面把握之后,有方向

性地选取其中最有价值的题录,依据一定的方式去获取全文。把握住这一点有助于我们撇开繁杂的检索工具使用等细节问题,而抓住问题解决的核心,既全又准又迅速地完成信息的获取。

1. 题录信息的获取方法



工程硕士遇到的问题绝大多数属于工程实际问题,因此有条件检索专业数据库的话,应尽可能全面地使用各个专业数据库进行检索,以避免漏查,这样虽然可能查文献的时间多了,但通过阅读文献信息,能接触到最新的研究成果,从而理清解决问题的思路。这比盲目做试验花的时间还是要少的。

如果需要查找英文文献,建议优先在 Ei 和 SCI 这两个综合性的文摘数据库检索。Ei 偏重工程,不报道纯理论,会议论文收录的多。SCI 则收录涵盖所有自然学科、侧重基础理论方面研究的文章,它的文献品位高、可信度大,并可以对文章的水平、著者的学术水平做出评价,寻找热点课题等。如果需要查找中文文献则更为方便,中国知网、维普、万方都提供文摘检索,收录量都比较大。为稳妥起见,有条件的也要每个库都检索一遍。

对于没有条件检索以上文摘库的读者,一个比较好的办法是在 Google Scholar 里查,有线索了再去查全文。因为 Google Scholar 提供的信息比 Google 搜索的学术性强,而且现在与其有协议的出版社也越来越多,数据库是相当全的;也可根据自己的课题学科选择相应的几个全文数据库进行检索。注意:国外的全文数据库特别是期刊数据库往往是出版商自己建的,不包括其他出版商的期刊,因此最好对几个全文数据库都进行检索。

在对文献文摘的检索过程中最重要的是选定检索词以及明确检索词之间的关系,即制定检索策略。好的检索策略往往不能一蹴而就,而是根据检索结果的数量多寡及对结果相关性的判断不断调整得到的。“变”是检索策略的永恒主题。需要经过“检索→阅读→策略调整→再检索……”的过程,不断调整,不断完善,如图 3-4 所示。

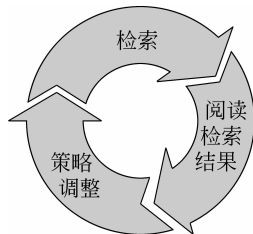


图 3-4 检索的过程

如何将拟定好的主题概念作为检索词, 获得好的检索效果? 首先应考虑选用其中比较规范、概念基本对等的词语进行检索。如果发现检索结果太多, 则应提高检索词的专指度, 选用下位词或专指性较强的自由词检索。并通过阅读比较相关的检索结果, 找出新的检索词作为概念限定。如果发现检索结果太少, 则需要尽可能多地选用同义词、同义的相关词、缩写和全称进行扩检; 或降低检索词的专指度, 选用上位词或相关词进行检索。

合理利用布尔逻辑算符 OR, AND 和 NOT。正确使用这些算符, 才能使拟定的检索词发挥作用。对于过多的检索结果, 可以用 AND 将增加的概念加入检索词, 用 NOT 排除无关的术语; 如果相关文献过少, 则应更多地使用 OR 算符把增加的同义词或相关词连接起来, 而少用 AND。



Tip 小提示

在文章的发表过程中, 可能存在着“二八定理”, 即 20% 的作者或课题组发表了该领域 80% 的文章。对该“领军人物”或课题组进行全方位检索, 收获会很大。因此在实际操作过程中除用关键词途径检索外, 也可根据自己已找到或导师推荐阅读的某相关文章的作者(研究机构)进行搜索, 通常同一作者(研究机构)研究的方向都围绕一个中心, 先看他所著的一篇文章, 然后再搜其他的, 得到的很可能就是所研究问题的延续或另外一种解决办法。

此外, 阅读相关文献的参考文献也是扩展检索结果的一个好方法。参考文献往往包含的是作者在研究该问题时得到过启发的研究结果, 由于研究的问题一致, 你在阅读这些研究结果时往往也能收获颇丰。

2. 全文文献的获取方法



检索的目的就是为了获取具有参考价值的全文文献。一些本可以获取的全文文献, 由于不熟悉其获取的方法和技巧而被迫放弃, 这就有些功亏一篑了。

全文文献的获取可分为直接获取和通过图书馆等文献服务机构获取两种方式。可遵循“先电子后印刷”、“先近后远”的原则。

直接获取比较简单, 在全文数据库中(如中国期刊全文数据库、维普期刊数据库、万方数据资源、Elsevier 电子期刊全文库、PQDD 博硕士论文数据库等), 均可以直接获取电子版的全文献。

对于在现有的全文数据库中未收录的文献, 或在迫于条件限制无法检索这些数据库的情况下, 可以登录图书馆等文献服务机构网站检索, 看是否收录了该文献的纸制版本。如果有, 可办理相关借阅手续借出或阅览。还可以利用图书馆等文献服务机构的原文传递、代查代检等服务, 来获取全文文献。