

Windows 篇

- 第 1 章 文献信息检索
- 第 2 章 生物信息数据资源
- 第 3 章 序列比对
- 第 4 章 核酸序列分析
- 第 5 章 蛋白质序列分析
- 第 6 章 基因表达分析
- 第 7 章 进化分析
- 第 8 章 非编码 miRNA 分析

第 1 章 文献信息检索

文献搜索、信息收集是进行研究课题设计前的必要环节，只有了解其他研究人员已经做过的研究内容及取得的研究成果，才有可能针对性地提出新科学问题，并进行实验方案的设计，最终通过科学实验来进行验证，获得新的研究成果，推动科学的不断发展，避免闭门造车，或者重复他人已经做过的研究。

1.1 文献资源的分类

走进一座图书馆，面对庞大的书籍库存，如何快速找到所需的书籍资料呢？答案必定是进行图书检索。而可检索的前提是图书经过分类整理。对于所有的文献资源，分类是其管理与应用的前提。

中国知网(www.cnki.net)收录了不同辞典记录的“文献”一词的定义，归纳起来可定义为：文献是有历史价值或参考价值的图书资料或者记录知识的一切载体。文献的记录形式有语言、文字、图像、音频、视频等，文献载体可分为纸版、缩微版、声像版、电子版 4 类。通常科研所说的文献资源主要指图书、期刊等有一定科研价值信息的纸版或电子版的文献。

1.1.1 根据出版形式进行分类

文献资源可以根据出版形式的不同分为图书、期刊、会议资料、研究报告、专利说明书、学位论文、政府出版物、标准、新闻报纸等，目前这些不同形式的出版物会同时有纸版和电子版两种不同的保存形式，其内容与时效性也各不相同(见表 1-1)。

表 1-1 文献出版物类型及其特点

类型	特点
图书	对现有知识的总结，思想内容较成熟，知识成体系，更新较慢，适合进行系统学习、全面了解某方面的知识
期刊	定期出版，如周刊、月刊、季刊等，篇幅有限，所录文献多揭示少数几个知识点，知识较新，是科研参考最多的文献类型
会议资料	因学术会议召开所形成的资料，其中尤以会议报告、论文集最有价值
研究报告	通过调研或者实验等过程收集数据并分析形成结论报告出来，结果具有参考价值
专利说明书	可授予专利权的发明及实用新型专利，强调新颖性、创造性、实用性，以保护新的发现并促使其转化为生产力。一般新的研究成果在公开发表之前可申请专利，公开发表后的新成果即使作者本人也无法申请专利(专利法列出的例外情况除外)
学位论文	分学士论文、硕士论文、博士论文 3 级，是学生对所做研究结果的系统总结，有较详细的研究方法、所用仪器、试剂耗材及操作步骤等可供借鉴，尤其适合科研新人参考，并且基本级别越高，参考价值越大
政府出版物	官方出版物，权威性较大，可信度较高
标准	为了统一规范而引入，各行各业的生产实践活动必须遵守相应标准
新闻报纸	报道最近发生的事，时效性最强，但科学性、严谨性方面会比期刊论文等稍差

1.1.2 综合分类法

根据文献信息的内容、形式、体裁、读者用途等不同,不同国家有不同的分类方法,目前国际上比较著名的是杜威十进分类法(Dewey Decimal Classification, DC/DDC),国内影响较大的是中国图书馆图书分类法和中国科学院图书分类法。中国图书馆图书分类法(简称中图法),由中国图书馆、中国科学技术情报所等36个单位于1971年共同编制完成,并于1975年正式出版,目前已修订至第五版,于2010年出版。中图法将图书分为5大部类、22类,其中22类以英文字母表示,各类下面用阿拉伯数字进行进一步分类。我们走进图书馆会发现,阅览室、书库排布基本是按照5大部类进行分类的,如“哲社书库”、“自然科学书库”等;而在书架旁边随意取一本书,其编号基本都是按照22类进行分类编号的。中图法的5大部类及22类关系如表1-2所示。

表 1-2 中图法分类主要类目名称及关系表

5 大部类	22 类字母符号	22 类对应类目名称
马列主义、毛泽东思想	A	马克思主义、列宁主义、毛泽东思想、邓小平理论
哲学	B	哲学、宗教
社会科学	C	社会科学总论
	D	政治、法律
	E	军事
	F	经济
	G	文化、科学、教育、体育
	H	语言、文字
	I	文学
	J	艺术
	K	历史、地理
	自然科学	N
O		数理科学和化学
P		天文学、地球科学
Q		生物科学
R		医药、卫生
S		农业科学
T		工业技术
U		交通运输
V		航空、航天
X		环境科学、安全科学
综合性图书	Z	综合性图书

示例 1-1 中图分类号的使用及其意义

吴祖建等编著的《生物信息学分析实践》,科学出版社2010年出版,ISBN号(International

Standard Book Number, 国际标准书号)为 978-7-03-027831-9, 其中图分类号是 Q811.4, 其所涉及的中图分类号如表 1-3 所示。

表 1-3 中图分类号及其意义示例说明

5 大部类	字母编号	对应名称
自然科学	Q	生物科学
	Q81	生物工程学(生物技术)
	Q811	仿生学
	Q811.4	生物信息论

ISBN 号由 13 位数字组成(2007 年前是 10 位数字), 分 5 段, 由短线隔开, 示例 1-1 的 ISBN 号意义是: 978 代表图书, 7 代表汉语, 03 是科学出版社编号, 027831 是科学出版社对该书的编号, 9 是该书的校验码。

1.1.3 标识码及编号

除了文献的中图分类号, 为了提高文献检索结果的适用性, 中国学术期刊(光盘版)编辑委员会制定了《中国学术期刊(光盘版)检索与评价数据规范》, 并得到广泛的认可, 该规范要求对文献标识一个文献标志码, 并设置了以下 5 种标志码:

- A——理论与应用研究学术论文(包括综述报告)
- B——实用性技术成果报告(科技)、理论学习与社会实践总结(社科)
- C——业务指导与技术管理性文章(包括领导讲话、特约评论等)
- D——一般动态性信息(通讯、报道、会议活动、专访等)
- E——文件、资料(包括历史资料、统计资料、机构、人物、书刊、知识介绍等)

不属于上述各类的文章及文摘、零讯、补白、广告、启事等不加文献标识码。

文章编号是为了给文章在全世界范围内取的唯一标识码, 其结构为:

XXXX—XXXX(YYYY)NN—PPPP—CC

其中各字段的意义如表 1-4 所示。

表 1-4 文章编号各字段意义表

字段	意义	位数
XXXX—XXXX	为文章所在期刊的国际标准刊号(ISSN, 参见 GB9999)	8 位
(YYYY)	为文章所在期刊的出版年份, 并置于小括号内	4 位
NN	为文章所在期刊的期次(不显示卷数)	2 位, 不足补 0
PPPP	为文章首页所在期刊页码	4 位, 不足补 0
CC	为文章页数	2 位, 不足补 0

大多数外文文献都有 DOI 号(Digital Object Identifier, 数字对象标识符), 其具体编号规则参见 http://www.doi.org/doi_handbook/2_Numbering.html。文献的 DOI 号可以在 <http://dx.doi.org/>或者 <http://doi.org/> 网站进行注册、解析、管理等, 并可直接链接至该文献的存放地址; 如果文献的存放地址有变动, DOI 号码不变, 其链接地址会改为文献的新地址。DOI 编号系统已经被 Elsevier、Blackwell、Springer、John Wiley 等国外大型出版社采用, 国内的中

国科技信息研究所和万方数据公司联合申请成立了首个 DOI 注册机构,进行中文文献的注册、解析、管理等服务,其网址为 <http://www.chinadoi.cn>。中文文献的 DOI 注册申请工作正在逐步发展过程中,并可能逐步取代我国给文章所做的“文章编号”。

1.2 文献的格式

专业期刊文献是进行科学研究参考最多的文献类型,各期刊均对其收录的论文格式有统一的规定,从而便于我们迅速在文献中查找相关内容。完整的期刊文献格式规范请参见《中国学术期刊(光盘版)检索与评价数据规范》,这里重点介绍期刊文献如下几部分的内容及其作用。

标题: 是全文最核心的部分,介绍了全文最主要的内容或者观点,是画龙点睛之处,一般作者都会仔细斟酌取一个既有吸引力又对全文有概括性或代表性的好标题。

作者: 对文章撰写或者实验有实质帮助的人,按照对文章内容贡献的大小进行前后排名,第一作者是对文章贡献最大的人,是主要的实验操作者或者主要论文撰写者;通讯作者指导实验的安排及文章写作,对文章内容的科学性、真实性负责,一般是实验室负责人或 PI(principal investigator),也可以联系他获得全文或者文中所使用的材料或者索取有关文章内容进一步的信息。通讯作者和第一作者也可以是同一人。仅为实验提供材料的人(如提供载体、菌株、样本等)可放在致谢部分进行感谢。

摘要: 概括文章的主要成果、主要方法等。绝大多数的文章标题、作者、摘要这些内容信息均可免费获得。对于中文文献来说,其中英文标题及摘要内容需要对应,中文文献在外文数据库收录时一般收录其英文标题及摘要信息,便于实验成果信息的国际学术共享。

关键词: 与文献内容密切相关的 3~5 个词汇,一般会在标题、摘要中出现,或者在全文中多次出现。中英文形式的关键词需一一对应,文献检索时所用的检索词大多来自关键词。

正文: 提供文章的全部内容,并可进一步分为若干组成部分,根据文章属于 Letter(给编辑的信)、Research paper(研究性论文)、Reviewe paper(综述性论文)等的不同,其正文内容的组成部分差异很大。研究性论文大致会被细分为几个部分: Introduction(前言)、Material and methods(材料与方法)、Results(结果)、Discussion(讨论)、Conclusion(结论)。这几部分内容在文中的顺序可能会有差异,有些项目会合并在一起,但整体来说这些内容在文中均会有所体现。这种划分方法有利于研究者快速从文献中获得所需信息。

致谢: 对提供文章所用部分材料,或者为文章的撰写提供意见或建议的非作者人员可在这部分进行致谢;有杂志将项目支持基金放在这一部分,也有杂志将项目支持基金单独列出。

参考文献: 对于文章引用他人的结论、数据、图片、算法、软件等需要列出所参考的文献以示尊重,参考文献的格式必须根据各个杂志的要求进行统一整理。

示例 1-2 中文文献格式

请参考图 1-1, 示例如下。

彭玉林等, Y 两优 2 号在安徽舒城低海拔地区“白亩方”单产突破 12.5t/hm²栽培技术, 杂交水稻, 2013,28(6): 50-52.

中图分类号: S511.048; S318。如果全文内容涵盖多个中图分类号范围,会产生多个分类号,这些分类号写在一起,用分号隔开。

文献标识码: B, 结合全面的说明及文章的内容, 说明该文章是实用技术成果报告。

文章编号: 1005-3956(2013)06-0050-03, 其中“1005-3956”是《杂交水稻》的国际标准刊号(ISSN);“(2013)”指的是2013年出版;“06”代表该杂志的第6期(注意未显示卷数28);“0050”代表首页是第50页;03代表该论文共3页,即50-52页。

• 50 •

杂交水稻(HYBRID RICE), 2013, 28(6): 50-52

Y 两优 2 号在安徽舒城低海拔地区“百亩方” 单产突破 12.5 t/hm² 栽培技术

Cultural Techniques of Y Liangyou 2 Yielding 12.5 t/hm² on a Scale of 6.67 hm² at a
Low-elevation County of Shucheng, Anhui

彭玉林¹, 李 鸿², 何森林², 姜国泉³, 吴朝晖¹, 闻尉宏⁴, 袁隆平^{1,*}

(1. 湖南杂交水稻研究中心, 湖南 长沙 410125; 2. 舒城县农业委员会, 安徽 舒城 231300;
3. 六安亿牛生物科技有限公司, 安徽 六安 231300; 4. 湖南丰惠肥业有限公司, 湖南 长沙 410125)

摘 要: 2012年Y两优2号在安徽舒城低海拔地区进行“百亩方”(7.4 hm²)高产攻关,取得了12.5 t/hm²的高产。总结了其高产栽培技术。

关键词: 两系杂交稻; Y两优2号; 高产; 栽培技术

中图分类号: S511.048; S318 文献标识码: B 文章编号: 1005-3956(2013)06-0050-03

图 1-1 中文文献格式示例图

示例 1-3 英文文献格式

请参考图 1-2, 示例如下。

Zhang L et al. GraP: platform for functional genomics analysis of *Gossypium raimondii*. Database, 2015: bav047.

Original article

GraP: platform for functional genomics analysis of *Gossypium raimondii*

Liwei Zhang^{1,†}, Jinyan Guo^{1,2,†}, Qi You¹, Xin Yi¹, Yi Ling¹, Wenying Xu¹,
Jinping Hua^{2,*} and Zhen Su^{1,*}

¹State Key Laboratory of Plant Physiology and Biochemistry, College of Biological Sciences, China Agricultural University, Beijing 100193, China and ²College of Agriculture and Biotechnology, China Agricultural University, Beijing 100193, China

Correspondence may also be addressed to Jinping Hua. Tel: +86-10-62734748; Fax: +86-10-62734748; Email: jinping_hua@cau.edu.cn

*Corresponding author: Zhen Su Tel: +86-10-62731380; Fax: +86-10-62731380; Email: zhensu@cau.edu.cn

[†]These authors contributed equally to this work

Citation details: Zhang,L., Guo,J., You,Q., et al. GraP: platform for functional genomics analysis of *Gossypium raimondii*. Database (2015) Vol. 2015: article ID bav047; doi:10.1093/database/bav047

图 1-2 外文文献示例格式

Original article: 指原创性的研究论文, 对应其他形式的论文包括 Letter(给编辑的信)、Review(综述文章)、Clinical Research(临床研究)。

Article ID: 杂志对文章的编号(bav047)。

Doi: doi 号码 10.1093/database/bav047, 由前缀和后缀两部分组成, 以“/”隔开, 可用于查询追踪该文献, 由文献所在杂志申请获得。“10”所有的 DOI 号码均以 10 开头; “1093”是 DOI 登记机构代码唯一编号, 这两部分是 DOI 的前缀。后缀部分“database/bav047”是登记机构给予文献的唯一编号。

1.3 文献检索

1.3.1 文献检索词的来源

文献检索词, 就是文献检索时所选定的关键词。检索词可来源于全文的任何部分, 实际检索时根据检索目的不同来确定, 并且在搜索过程中可以进行适当修正以获得最佳结果。文献检索词必须出现在全文中, 其出现的位置不同会导致搜索文献结果与预期所需要的文献结果的相关性有很大差异, 以检索结果与预期结果的相关性来排列检索词在全文出现的位置, 大致有这样的顺序: 标题(Title)>关键词(Keywords)>摘要(Abstract)>全文(Fulltext 或者 Allfield)。但对于以特定目的的检索, 这种顺序就不一定正确了, 如支持基金检索, 检索词出现在全文中; 作者发文检索, 检索词出现在作者栏等。

各数据库均规定了特定的检索范围及其标准缩写形式, 如中国知网(National Knowledge Infrastructure, CNKI)及 ScienceDirect 数据库的常用检索范围见表 1-5; 美国国立生物信息中心文献数据库 PubMed 的常用检索范围见表 1-6。

表 1-5 CNKI 常用检索范围及与 ScienceDirect 数据库相关检索范围对应表

英文缩写	表示方法	意义	对应的 ScienceDirect 检索范围
SU	主题	题名+摘要+关键词	Abstract, Title, Keywords
TI	题名	文献标题	Title
KY	关键词	文献关键词	Keywords
AB	摘要	文献摘要	Abstract
FT	全文	文献全文	Full Text
AU	作者	文献所有作者	Authors
FI	第一责任人	一般是文献主要通讯作者	Specific Author
AF	机构	文献作者所属单位	Affiliation
JN	中文刊名&英文刊名	文献所在杂志名称	Source Title
RF	引文	文献参考文献	References
YE	年	文献发表年份	—
FU	基金	文献支持项目基金	—
CLC	中图分类号	文献的中图分类号	—
SN	ISSN	杂志的国际标准刊号	ISSN
CN	统一刊号	中国标准统一报刊号	—
IB	ISBN	国际标准书号	ISBN
CF	被引频次	文献被应用的次数	—

表 1-6 PubMed 常用检索范围表

检索范围	缩写	中文翻译	意义
Affiliation	AD	作者单位	所有署名作者单位, 有多个可全部列出
All Fields	ALL	所有范围	将检索词自动智能分类检索, 若检索词有引号或者通配词则不进行自动智能分类
Author	AU	作者	作者名检索, 可智能匹配改写检索词
Author-Corporate	CN	署名作者	文章所有作者
Author-First	1AU	第一作者	排名第一位的作者
Author-Full	FAU	作者全名	利用作者全名进行检索
Author-Identifier	AUID	作者编号	部分出版社给作者唯一身份编号
Author-Last	无	最后作者	排名末尾作者, 一般为通讯作者
Book	无	书籍	参考书籍或其章节
Date-Completion	DCOM	完成时间	NLM 编辑完成时间
Date-Create	CRDT	创建时间	文献第一次创建时间
Date-Entrez	EDAT	Entrez 时间	文献放入 PubMed 数据库时间
Date-MeSH	MHDA	MeSH 时间	文献建立 MeSH 检索词的时间
EC/RN Number	RN	EC/RN 号	FDA、EC、CAS 等给化合物或酶编号
Editor	ED	编辑	书或其章节的编辑
Filter	无	过滤	设置过滤参数缩小检索范围
Grant Number	GN	支持基金	项目支持基金或机构名称或编号
ISBN	无	ISBN 号	国际标准书号
Investigator	IR	PI 名	作者所在实验室负责人(PI)
Investigator-Full	FIR	PI 全名	作者所在实验室负责人全名
Issue	IP	期数	文章所在杂志当期期数
Journal	TA	杂志名	文章所在杂志名
Language	LA	语言	文章所使用语言
Location ID	LID	文章编号	出版社给文章的编号或者 doi 号
MeSH Major Topic	MAJR	MeSH 主题	文章所对应的 MeSH 主题
MeSH Subheading	SH	MeSh 下级词	文章所对应 MeSH 下级主题词, 缩小范围
MeSH Terms	MH	MeSH 关键词	NLM 总结的文章关键词
Orther Term	OT	非 MeSH 词	非 MeSH 关键词
Pagination	PG	首页码	文章首页页码
Pharmacological Action	PA	药效	药物药效
Publication Type	PT	文章类型	文章类型如 Review、Letter 等
Publisher	PUBN	出版社	书籍出版社
Secondary Source ID	SI	其他编号	指的查询内容在其他数据库中的编号
Text Word	TW	全文词汇	单词或数字出现在标题、摘要、MeSH、发表时间、作者等处
Title	TI	标题	标题
Title/Abstract	TIAB	标题/摘要	标题或摘要
Transliterated Title	TT	翻译标题	从非英文标题翻译成英文标题
Volume	VI	卷数	文献所在杂志当期卷数

1.3.2 搜索数据库选择

根据查询的目的、数据库收录的文献类型、个人或机构所拥有的数据库查询权限等实际

情况进行文献数据库的选择(见表 1-7)。各数据库还会对其所收录的数据库细分成若干亚库,比如中国知网包含期刊、学位论文、报纸等亚库。

表 1-7 常见中外文献数据库

名称	网站	收录文献类型
中国知网	http://www.cnki.net	期刊、报纸、专利等
维普资讯	http://www.cqvip.com	期刊为主
万方数据	http://www.wanfangdata.com.cn	期刊、报纸、专利等
Elsevier	http://www.sciencedirect.com	期刊、书籍
Springer	http://link.springer.com	期刊、书籍
ISI web of science	http://wokinfo.com	期刊
Wiley InterScience	http://onlinelibrary.wiley.com/	期刊、书籍等
ProQuest	http://pqdt.calis.edu.cn	外文学位论文
学位论文	中国知网或者万方数据等	中文学位论文
Pubmed	http://www.ncbi.nlm.nih.gov/pubmed	期刊
Highwire	http://highwire.stanford.edu/cgi/search	期刊
中国知识产权局	http://www.sipo.gov.cn	中国专利
美国专利局	http://www.uspto.gov	美国专利
Patentlens	http://www.lens.org/lens	国际专利
Priorsmart	http://www.priorsmart.com/	国际专利
专利下载	http://www.drugfuture.com/patent/	多国专利全文 pdf 下载

各科研院所都会购买一定量的文献数据库查询权限,另有一些免费数据库提供一定量的全文文献,就生物学文献来说,通过以上数据库的查询,基本上 80%以上的中外文文献均可检索到。另外,各数据库所收录的文献可能存在重复,查询时需注意尽量避免重复查询。

1.3.3 检索式构建

1. 逻辑词的应用

“并且”或者“AND”:取各检索词的交集,会减少检索文献范围,同时提高检索的精度,如“A 并且 B”,检索结果中必须同时出现检索词 A 和检索词 B。

“或者”或者“OR”:取各检索词的并集,会扩大检索文献范围,同时降低检索的精度,如“A 或者 B”,则检索结果中出现检索词 A 或者检索词 B 即可。

“并不”或者“AND NOT”:在该逻辑词前面出现的检索词的结果中去掉其包含后面检索词的结果而最终获得的结果,如“A 并不 B”,则检索结果中出现 A 同时不能出现 B。

2. 截断词的应用

以英文为关键词进行搜索时,由于有时态、单复数、名动词等差异,需要进行截断搜索。截断词的使用必须配合使用相关符号,代表检索词的全面或者后面有若干个字母存在也算作匹配。这种匹配分有限匹配和无限匹配,其中有限匹配表示检索词前面或者后面出现的字母个数固定,一般以一个为多见;无限匹配表示检索词前面或者后面出现 0 至多个均算作匹配。这样的符号在不同数据库中的使用不一样,可以查看相关数据库的说明。常见的符号有“?”“*”“\$”等,一般来说,“?”代表有限匹配一个字母,“*”代表无限匹配 0 至多个字母,比如 miR*,可能代表 miRNA、miR-1 等多个词。

3. 单引号或者双引号的使用

单引号或双引号代表进行精确搜索，比如“转基因橡胶”用双引号，如果让数据库自动分段可能分成“转+基因+橡胶”3段或者“转基因+橡胶”两段进行检索；如果‘转基因橡胶’用单引号，则强制要求将其作为一个词进行搜索，这样会提高检索的精度(见表 1-8)。

表 1-8 常用检索逻辑词、符号及其意义

名称或符号	意义	用法
AND	同时匹配	如 A AND B: A 与 B 同时出现
OR	择一匹配	如 A OR B: A 与 B 任一个出现即可
NOT	否定后位条件	如 A NOT B: A 出现, 同时 B 不能出现
()	圆括号, 改变条件运算顺序	如 A AND B OR C: A 与 B 均出现, 或者 C 出现; A AND (B OR C): A 出现, 同时 B 与 C 任一出现
[]	方括号, 改变条件运算顺序	在复杂的运算式使用, 里面包含方括号
?	代替任意一个字母	A?B:代表 A 与 B 之间有一个字母
*	代替任意多字符	A*: 代表以 A 开头的所有单词
“ ”	双引号	代表精确搜索
‘ ’	单引号	强制作为一个词进行搜索

4. 检索式的构建

一般数据库都提供检索式构建的模块，检索者仅需输入检索词并进行选择即可自动建立检索式。比如在中国知网的检索式：

TI=(橡胶树*'miRNA')-'抗逆'

表示在标题(TI)中同时出现“橡胶树”及“miRNA”两个搜索词，同时不出现“抗逆”这个搜索词的结果。这里的“*”代表“并且包含”的意思，注意与英文搜索时的无限匹配进行区分；“-”代表“并且不”的意思；类似的还有“+”代表“或者包含”的意思。

如果将上面的搜索式修改为：

TI='橡胶树'*(('miRNA'-'抗逆'))

表示在标题(TI)中寻找这样的结果，出现“miRNA”而不出现“抗逆”的结果，并与“橡胶树”这个搜索词进行交集。

TI'橡胶树'*(('miRNA'+ '抗逆'))

表示在标题(TI)中寻找这样的结果，出现“miRNA”或者出现“抗逆”的结果，并与“橡胶树”这个搜索词进行交集。

5. 检索结果的保存

将检索得到的结果保存或输出，如果数据库支持 EndNote、Reference 等文献管理软件或者其他的文献管理软件，可以按照这些软件的相应格式将检索结果批量保存至本地计算机；如对于 Pubmed 数据库，如果在该数据库中有注册账号，则可直接通过该账号保存本次检索结果，下次可通过注册账号直接调取，也可发送至电子邮箱。如果需要下载全文，则需要一篇一篇地单独下载，一般的文献数据库均不支持批量全文下载。

1.3.4 检索结果的处理

1. 检索结果较少时的处理

- ① 扩大检索词检索范围，比如检索词范围从“标题”或者“摘要”扩大到“全文”；
- ② 检查其他同义的检索词是否存在，比如 miRNA 与 microRNA，可将同义的检索词同时进行检索；

③ 查看是否有其他限制条件存在，部分数据库会将上一次搜索的条件保存为下一次搜索的默认条件，比如文献发表时间限制、语种限制等，如果是这种情况，可去掉相应限制条件；

- ④ 对检索词使用模糊搜索，这种处理方法尽量少用，避免检索结果中出现无关的文献。

2. 检索结果较多时的处理

- ① 缩小检索词范围，比如从“全文”缩小至“题目”或者“摘要”；
- ② 增加检索的限制条件，如文献发表时间、语种范围、文献类型等；
- ③ 增加检索词，或者用新的检索词在结果中进行二次检索等。

1.3.5 CNKI 数据库查询举例

1. CNKI 主页面介绍

CNKI 主页如图 1-3 所示，主页包含客户身份，如海南大学的客户身份就是海南大学；检索入口，这里可以进行不同目的的检索，如期刊检索、硕博学位论文检索、夸库检索，或者可以链接至新页面进行出版物检索、高级检索；不同分类的子库，单击这些子库可以链接至新的页面进行相应的子库检索。

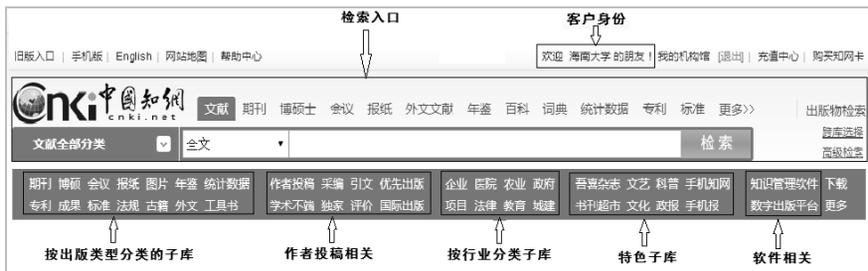


图 1-3 CNKI 主页检索相关项目示意图

2. 使用默认参数进行检索

在首页进行检索时，可以在检索框输入检索词，如以“橡胶树”为搜索词，单击“检索”按钮，即可开始检索。此时未经修改的默认设置如下。

检索文献的分类：指文献的中国图书馆文献分类法进行的分类，默认选择所有分类。

检索词的位置：全文。

数据库选择：默认选择夸库检索，默认包含的子库包括期刊、特色期刊、博士、硕士、国内会议、国际会议、报纸、学术增刊、商业评论 9 个子库；其他非默认包含的子库包括年鉴、专利、标准、成果 4 个子库(见图 1-4)。



图 1-4 夸库检索选项内容

点击搜索后即可获得搜索结果，以“橡胶树”为搜索词，使用默认检索条件得到 8936 条检索结果，该检索结果会随着时间的增长、数据库文献的增加而发生变化。

3. 检索结果页面介绍

检索结果页面大致分为如下 5 个板块(如图 1-5 黑框所示)。

- “1”：分组浏览搜索结果，可以按照来源数据库、学科、发表年度、研究层次、作者、机构、基金进行分组。默认显示的是按照发表年份显示搜索结果，点击相应年份可以呈现相应年份的结果，如点击“2015”会呈现 2015 年发表的 182 条文献搜索结果。
- “2”：显示具体搜索结果。其“排序”可以按照主题、发表时间、被引次数、下载次数进行排序，默认按照主题排序。搜索结果默认是按照列表方式显示，通过点击“切换到摘要”可改为摘要方式显示。“每页记录数”有 10 条、20 条、50 条 3 项可选，默认每页显示 20 条。并且显示了总的结果数“8964 条”，本页在可显示总页数中的位置为“1/300”，如果超过 300 页则显示前 300 页，本页为第 1 页，点击“下一页”可切换至当前页的下一页。点击所列出的搜索结果前面的选择框可选中该条记录，再次点击可取消选择；点击标题栏的选择框可一次选中当前页面的所有记录，再次点击可取消当前页面所有记录；也可点击“清除”取消全部选择项；对所选择的文献可点击“导出/参考文献”进行保存，也可以点击“分析/阅读”进行分析或查看。
- “3”：文献来源刊物，按照搜索的结果倒序排列。
- “4”：对检索词进行智能匹配，相应其他关键词的搜索出现的文献条数。
- “5”：记录检索历史，便于在不同检索历史结果间进行切换。

检索: 全文:橡胶树 x

分组浏览: 来源数据库 学科 发表年度 研究层次 作者 机构 基金

2015(182)	2014(695)	2013(707)	2012(706)	2011(580)	2010(619)	2009(575)	2008(476)
2007(373)	2006(388)	2005(238)	2004(220)	2003(193)	2002(165)	2001(146)	>>

排序: 主题排序 发表时间 被引 下载

切换到摘要 每页记录数: 10 20 50

(0) 清除 导出/参考文献 分析/阅读

找到 8,936 条结果 浏览 1/300 下一页

<input type="checkbox"/>	题名	作者	来源	发表时间	数据库	被引	下载	预览	分享
<input type="checkbox"/>	1 橡胶树EST-SSR标记的开发与应用	安泽伟, 赵彦宏, 程汉, 李维国, 黄华孙	遗传	2009-03-15	期刊	49	543		
<input type="checkbox"/>	2 橡胶树胶乳均一化全长cDNA文库的构建与EST测序分析	曾日中, 段翠芳, 慕智敏, 代龙军, 黎瑜	中国农业科学	2011-02-16	期刊	10	301		
<input type="checkbox"/>	3 乙烯利刺激橡胶树增产机制研究进展	庄海燕, 安峰, 张颖新, 白登忠	林业科学	2010-04-15	期刊	12	337		

文献来源

- > 热带农业科学 (659)
- > 热带农业科技 (594)
- > 热带作物学报 (576)
- > 世界热带农业信息 (486)
- > 中国农垦 (203)

关键词

- > 橡胶树 (844)
- > 巴西橡胶树 (461)
- > 天然橡胶 (231)
- > 橡胶 (134)
- > 基因克隆 (107)

检索历史:

- > 橡胶树 (检索痕迹 清空)

图 1-5 检索结果页面图

4. 搜索结果分析

文献选好后, 点击“分析/阅读”可进行文献的分析或者文献的阅读。对所选文献进行再次选择, 点击“分析”按钮进入“文献分析中心”页面; 点击“阅读”按钮进入“文献在线阅读”页面。在文献的分析方面, 可进行如下分析。

- 文献互引图: 所选文献的相互引用关系分析。
- 参考文献: 所选文献的参考文献及其出现频次分析, 并依据被引频次列出前 50 项。
- 文献共引分析: 依据引用相同的文献将所选文献分组分析。
- 引证文献: 所选文献的引证文献及其出现频次, 列出前 50 项。
- 文献共被引分析: 所选文献中有相同引证文献的分组分析。
- 关键词分析: 所选文献中的高频关键词, 列出前 10 个。
- 读者推荐分析: 所选文献相关的读者推荐文献分析。
- H 指数分析: 所选文献中至少有 H 篇文献被引频次不少于 H 次分析。
- 文献分布: 所选文献分组情况, 如按照年份分组、按照机构分组等(见图 1-6)。

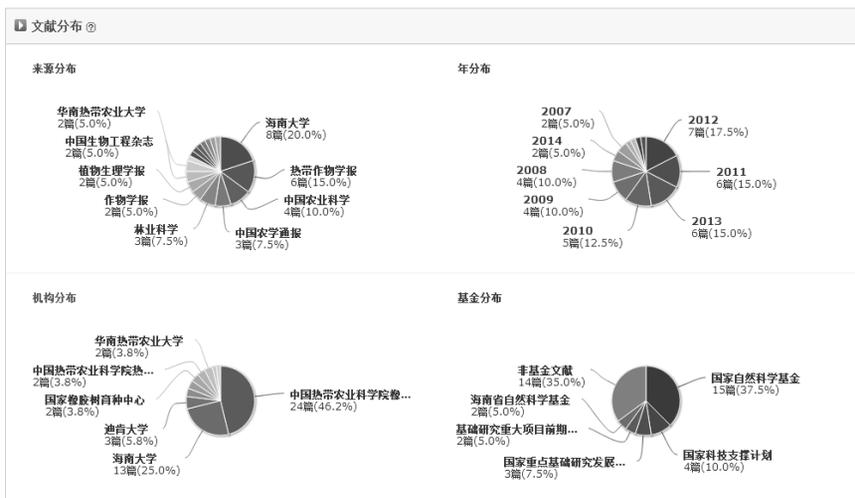


图 1-6 文献分布分析实例

在文献阅读页面, 所选择的论文生成论文集的形式(见图 1-7), 阅读窗口显示两栏, 左边栏显示所选的所有文献, 点选其中任一篇, 则在右边栏会详细显示文章内容。左边栏可以收起隐藏, 右边栏可以进行全屏显示。

5. 文献搜索结果保存

选中相关文献后, 点击“导出/参考文献”按钮, 进入“文献管理中心_导出”页面, 再次选择文献, 可选择相关按钮进行文献的“导出/参考文献”、“定制”、“生成检索报告”等操作。点击“导出/参考文献”按钮进入“文献管理中心-文献输出”页面, 该页面大致分为 3 个板块(见图 1-8, 用五角星标出), 各板块所包含的内容意义如下。

- “1”: 文献显示格式, 常见的如“CAJ-CD 格式引文”格式, 或者使用文献管理软件所对应的格式, 如 CNKI E-Learning、EndNote 等。

- “2”：文献保存形式，可以“复制到剪贴板”再粘贴至其他文件；直接“打印”出来；或者通过“txt”、“xls”、“doc”等文件格式导出保存。这样保存的是文献摘要等信息，文献全文部分需要分别下载保存。
- “3”：针对“1”所选的文献显示格式进行针对性的显示，如图 1-8 显示的是“CAJ-CD 格式引文”所对应的文献格式，保存时的内容就是本处显示的内容。



图 1-7 文献阅读实例

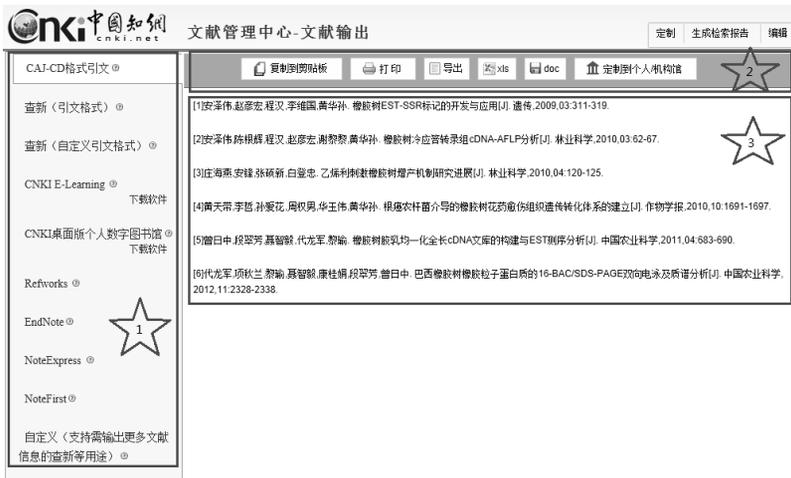


图 1-8 文献保存页面

6. 其他检索方式

点击“高级检索”，进入高级检索页面，该页面包含其他检索页面链接：专业检索、作者发文检索、科研基金检索、句子检索、文献来源检索(见图 1-9)。各检索页面的主要功能如下。

- 高级检索：有更多的“内容检索条件”及“检索控制条件”的选择，可以辅助建立复杂的检索式进行更精确的检索。
- 专业检索：进行文献的专业检索，可以按照数据库提供的检索规则进行详细的检索式构建，之后进行更精确的检索。

- 作者发文检索：检索数据库中某作者发表的全部文献。
- 科研基金检索：检索某项目基金支持所发表的文献。
- 句子检索：可以在全文中的同一个句子或者同一段落进行搜索词的检索。
- 文献来源检索：可以检索某刊物所发表的文章。

高级检索页面其他栏目的内容意义如下。

- 文献分类目录栏：是大致按照文献内容进行的分类，共分 10 项(见图 1-9)。与生物学相关的分类主要是“基础科学、农业科技、医药卫生科技”3 项。
- 文献来源栏：显示文献来源杂志及其文献数量。

该页面的其他栏目与之前介绍的一致，这里不再重复。



图 1-9 高级检索页面图

1.3.6 Elsevier 数据库检索举例

1. 简易检索

Elsevier 主页可进行快速的简易检索，也可点击“Advanced search”按钮链接至高级检索页面进行检索(见图 1-10)。简易检索可在“Search all fields(全文)”、“Author name(作者)”、“Journal or book title(期刊或者书籍)”、“Volume(卷)”、“Issue(期)”、“Page(页码)”6 个检索框进行查询。检索词输入检索框则在对应范围进行查询。



图 1-10 Elsevier 简易检索栏

2. 检索结果页面介绍

在简易检索页面“Search all fields”检索框输入“hevea brasiliensis”，则表示在全文搜索检索词“hevea brasiliensis”，检索结果页面见图 1-11。该结果页面主要分为 3 栏：左边栏(Refine filters)、右上栏(Search result)、右下栏(具体的检索结果)。

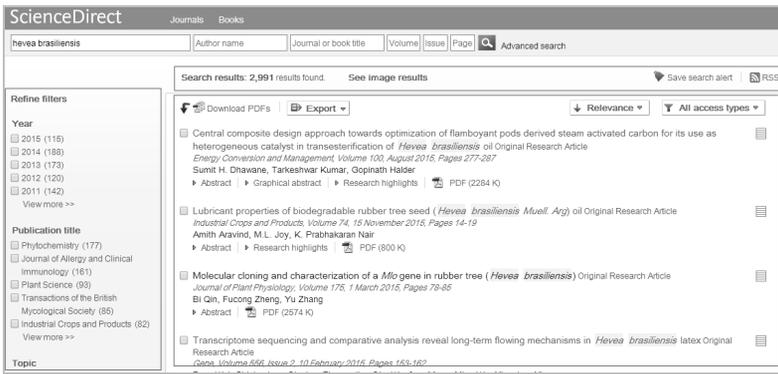


图 1-11 检索结果页面

左边栏可对检索结果进行进一步限制，限制因素(Refine filters)包括：

- **Year(发表年份)**：按照文献发表时间顺序倒序排列，默认显示最近 5 年，可通过点击“view more”查看近 20 年的文献数据。
- **Publication title(文献标题)**：检索文献所在的杂志或书籍标题，按照各发表物所检索到的文献数量倒序排列，默认显示文献数最多的 5 份发表物，可通过点击“view more”查看前 20 份发表物及其所发表的相应文献数。
- **Topic(主题)**：对检索文献对应的主题进行分析，按照出现次数进行排序，频次出现最高的前 20 个作为 Topic(主题词)，显示各主题词所对应的文献数，倒序排列，默认显示前 5 个主题词。
- **Content type(内容形式)**：有 Journal(包括杂志及连续出版书)、Book(包括书籍、连续出版书、手册及 Reference Work)、Reference Work (ScienceDirect 数据库发表的其他形式的文献)。

右上栏是检索结果栏，显示共有 2991 条检索结果，点击“See image results”按钮可查看图形形式的结果；点击“Save search alert”按钮可在网页保存搜索结果(需注册个人账号)；点击“RSS”按钮可保存检索式，并可定期发送新的检索结果(需下载安装 RSS 阅读软件)。

右下栏显示具体搜索结果，相关链接如下。

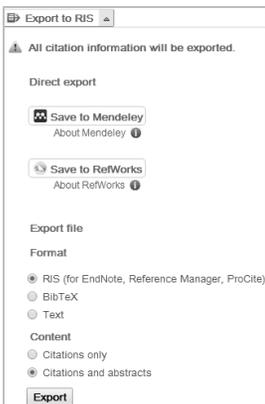


图 1-12 export 文献保存内容

- **download PDFs 按钮**：可批量下载所勾选文献的全文。
- **Export 下拉菜单**：将检索结果(摘要等信息)导出保存，如未勾选任何条目则默认保存全部搜索结果；可直接保存至网络数据库 Mendeley 或者 RefWorks；也可以文件的形式保存。其中的文献格式可以是 RIS (Endnote 等文献管理软件支持格式，默认参数)、BibTeX (Bibshare 文献管理软件识别格式)、Text (纯文本格式)；内容可以按“Citations Only”(仅包含引文信息)、或者“Citations and Abstracts”(引文加摘要，默认参数)的形式保存。选择完毕后，点击下拉框下面的“Export”按钮即可输出保存(见图 1-12)。
- **Relevance 下拉菜单**：是文献排列方式，可按照“Relevance(相关性)”排序或者按照“Data(发表日期)”进行排序。

- All access types 下拉菜单：文献获得方式，有“**All access types** (所有可能形式，默认参数)”、“**Open Access articles** (免费自由使用文献)”、“**Open Archive articles** (定期免费使用文献)”。

在具体文献条目的选择框，点击选中该条目，再次点击可取消选择；点击文章标题可链接查看网页格式的全文；点击“**Abstract**”按钮在当前页打开文献摘要，再次点击则关闭摘要；点击“**PDF**”按钮下载单篇文献全文。

3. 高级检索

点击“**Advanced search**”按钮进入高级检索页面(见图 1-13)，该页面可进行检索式的辅助构建；点击“**Expert search**”按钮可进行专业搜索，自建搜索式。

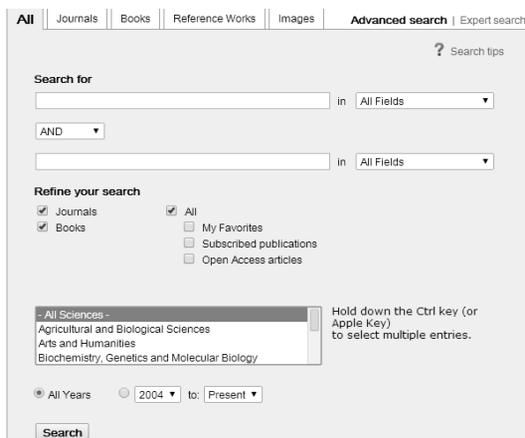


图 1-13 高级检索页面

在高级检索页面还可以选择“**All**”、“**Journals**”、“**Books**”、“**Reference Works**”、“**Images**”子项进行针对性的检索，下面以“**All**”子项为例介绍高级检索，其他子项的检索过程类似。

在高级检索页面有两个文本输入框可输入检索词，并选择对应的检索词范围，默认是“**All Fields**” (全文)；两个输入框的关键词之间可选择逻辑词“**AND**”、“**OR**”、“**AND NOT**”，默认选择是“**AND**”。检索词输入完毕后可进一步选择限制条件，出版物的限定有“**Books** (书)”和(或)“**Journals** (杂志)”；出版物范围的限定有“**All**”、“**My Favorites** (个人喜好清单, 需创建)”、“**Subscribed publication** (注册出版物)”、“**Open Access articles** (开放存取文献)”；出版物主题范围的选定，可多选，默认选择“**All Sciences** (所有学科领域)”；出版时间限制，默认是“**All Years**”，最早文献发表时间可追溯至 1823 年，因此时间选择范围从 1823 年至“**Present**” (现在)。将检索词及各限制范围设置好后即可进行文献检索，检索结果页面与简易检索结果页面类似，这里不再重复。

1.4 文献信息的价值判断及阅读

1.4.1 文献的价值判断

在浩如烟海的文献中，如何判断文献的价值呢？当然对于不同的人来说，同一篇文章的价值是不一样的，但整体来说可从以下几方面对文献价值进行初步判断。

- 杂志类型：杂志的影响力直接决定了文章的价值，比如 *Nature*、*Science*、*Cell* 等杂志发表的文献，创新性强、可信度高，一般来说价值也高。杂志影响力的标志在国际上是以是否被著名的三大检索系统收录来判断的；在国内则是以是否为核心期刊来判断的。国际三大检索工具是：科学引文索引(Science Citation Index, SCI)、工程索引(Engineering Index, EI)、科技会议录索引(Index to Science & Technical Proceedings, ISTP)。中文核心期刊主要有 3 种：中文核心期刊、中文社会科学引文索引、中国科学引文数据库来源期刊(中国科学院文献情报中心发布)。
- 通讯作者：通讯作者对文章的科学价值及可信度负责，因此通讯作者的影响力也从侧面反映了文章的价值。
- 发表时间：最新发表的文章，参考的文献会更丰富，结论更接近真理，价值越高，一般来说尽可能参阅近 5 年发表的文章。
- 相关度：对于特定的研究人员，其搜索文献时搜索词出现在标题、关键词或者摘要部分，所获得的文献与其研究领域更接近，对其来说更有参考价值。

三大检索系统简介如下。

- SCI：收录自 1900 年至现在，涵盖 150 个学科领域的期刊约 8500 多种，由美国科学情报所(Institute for Scientific Information, ISI)创刊于 1964 年，目前属于世界最大信息提供商汤森路透集团(Thomson Reuters)。引入“影响因子(Impact Factor)”的概念来评价刊物的影响力，影响因子越高，说明该刊物的文献被引用的次数越多，表明刊物的影响力越大。目前 SCI 影响因子已经被国内广泛应用于评价个人能力及科研机构的社会影响力，研究生需要发表一定影响因子的文章才能拿到学位，科研人员需要一定影响因子的文章才能谋得一定的科研职位。
- EI：收录工程技术领域的超过 5000 种工程期刊、会议文集和技术报告，由美国工程情报公司(Engineering Information Co.)创刊于 1884 年，是工程技术领域最权威的文獻索引工具，有月刊和年刊，于 1967 年开发电子版，并命名为“Compendex”，目前属于 Elsevier 公司。
- ISTP：目前改名为 CPCI(Conference Proceedings Citation Index)，分为自然科学版(Conference Proceedings Citation Index-Science, CPCI-S, 原 ISTP)和社会科学版(Conference Proceedings Citation Index-Social Science & Humanities, CPCI-SSH)，涵盖自 1990 年起的超过 256 个领域的 160 000 个会议记录，每周更新一次，放在 Web of Science 检索平台里面的 ISI Proceedings。

国内主要核心期刊的评价体系如下。

- 中文核心期刊：收录于《中文核心期刊目录总览》，由北京大学图书馆等 27 家单位科研人员完成的科研成果，目前已经是第六版，2011 年发布。
- 中文社会科学引文索引(CSSCI)：南京大学中国社会科学研究评价中心发布的引文数据库，2014—2015 来源期刊已经发布。
- 中国科学引文数据库(CSCD)：中国科学院文献情报中心发布，分核心库与扩展库两部分，每两年遴选一次，2015—2016 年所收录的期刊已发布。