

# 基于博士学位论文引文分析的文献需求研究

刘远翔, 华中农业大学图书馆

图书馆

## 第一章 研究背景、目的及意义

### 1.1 研究背景

#### 1.1.1 引文及引文分析

引文是一篇完整论文的重要组成部分,是读者文献需求的集中反映。因此,通过对引文的研究,能够准确地了解学科文献需求和利用的规律和特点,能够较全面地反映该重点学科领域文献的保障度和支持力,从而为重点学科文献资源的合理布局和建立重点学科文献资源保障体系提供科学依据

将引文分析作为期刊价值的客观测量指标是在1927年由P. L. K 格罗斯(P. L. K. Gross)首先提出的,他们统计了化学专业的某些期刊论文的参考文献并进行了分析,得出了化学教育方面的核心期刊。从此以后,陆续有研究者采用Gross方法对科技领域期刊作了评价研究。图书情报领域逐渐接受这一方法,这也成为图书馆文献资源建设工作的依据。

#### 1.1.2 学科建设与引文分析

学科建设是我国高校内涵建设的主要内容,加强学科建设是高等院校增强核心竞争力,凸现办学特色,提高教研、科研水平和社会服务能力的一项战略性、全局性和基础性工作。大学图书馆作为学校的文献信息中心,对学科建设起着不可替代的作用。

农业资源与环境学科隶属资源与环境学院农学门类,为湖北省一级重点学科。该学科特色鲜明,门类齐全,设有土壤学、植物营养学、资源环境信息工程等3个博士点。其中土壤学、植物营养学等2个学科为湖北省特色学科,资源环境信息工程为国家“211工程”三期重点建设学科点。

### 1.2 研究目的

这将对指导图书馆的文献资源建设具有重要的意义,进而为学校该学科建设和发展提供了重要的支撑力。作为大学图书馆,尤其是研究型大学图书馆,研究生是其服务的重要组成部分,因此,通过对博士生学位论文引文的统计分析,可以从一个侧面反映我校该学科论文学术水平,研究生对文献信息利用的能力和科研创新水平。同时我们也可以了解博士生利用

文献的规律，提高图书馆对相关专业的文献收藏水平，为图书馆的服务提供质化、量化实证依据。

## 第一章 研究内容、思路及创新点

### 2.1 研究内容

本文选择农业资源与环境学科方向的博士生学位论文，对其所附引文进行统计分析。博士学位论文具有较高的科研水平，在一定程度上代表着该学科的前沿和创新。所以选择博士学位论文的引文作为本课题的研究内容。

在众多的文献需求调查方法中，引文分析方法是一种科学、直观而可靠的定量分析方法。因为文献的被引是读者阅读了该文献并从中获得帮助的重要标志，而且引文分析的研究方法简单，数据的获取更迅速、经济而准确。引文分析的数据可取自期刊论文、学位论文、专著、书目和索引等文献中的引文。作为高校文献需求的调查，该校教师的出版物和学生的学位论文的引文分析最能反映本地文献的需求特点。教师是高校最重要的读者群，其出版物引文来源也反映较高层面的文献需求范畴，但教师的出版物比较分散，收集的难度大，发表在期刊论文中的引文数量少。而本校博士生是在他们的导师指导下完成学位论文，博士学位论文的学术水平较高，博士论文的引文既可反映学生的文献需求，也间接代表了指导老师的文献需求状况。

本研究所需要获取的信息包括引文量、各种不同的引文类型占比、不同语种引文的比例、引文出版年代的分布以及中外期刊引用量等。

引文量即引文数量，是指一篇论文中所引用参考文献的数目，反映引文的多寡。就是博士生撰写论文时参考并接受了多少别人或前人的成果，引文量是引文分析中最基本的测度。引文量的多少，在一定程度上反映了博士生对相关学科文献信息的了解和吸收能力。

引文的类型分布是分析引文文献类型，可以掌握读者的阅读动态和文献来源的构成比例，有利于确定各类文献载体的情报价值。

引文语种是测试博士生利用国内外文献的指标。通过对引文语种的分析，可以了解博士生吸收信息的能力以及外语水平，同时对图书馆外文文献的馆藏布局具有重要的参考价值；

对引文的出版年代进行分析，可以了解文献最佳利用时间，揭示专业文献的老化规律和用户利用文献的高峰期；

基础文献也称为关键论文是社会影响较大的论文，其影响大小可由所发论文的被引频次来确定。通过引文类型的统计结果看中英文期刊在博士生引文类型中所占比重是否较大，对学术论文的引用频次和馆藏进行分析，可以了解我校图书馆馆藏期刊对教学和科研的保障程度。

## 2.2 研究思路

引文分析已成为重要信息计量的方法，其应用可大分为下列七类：(1)文献研究。(2)文献类型之研究。(3)使用者研究。(4)科学家写作历史研究，依据科学发展过程中论文撰写模式，分析科学家写作历史。(5)评鉴文献书目计量学。(6)信息检索。(7)图书馆馆藏发展。

本项目以我校资源与环境学院农业资源与环境学科近3年博士论文的引文为研究对象，分别从引文时间、引文语种、引文半衰期等方面进行统计分析。旨在指导我校图书馆馆藏文献建设，同时此研究方法可供其他图书馆参考借鉴。

## 2.3 创新点

本次项目研究拟打破通过抽样调查馆藏现刊的阅览情况来研究期刊利用率的研究模式，而是从研究对象的样本和网络数据库入手，全面统计农业资源与环境学科方向的博士学位论文，从其引用文献的类型入手来得到相关数据，并与我校图书馆所订的中外文现刊进行比较、分析，得到期刊利用方面的一些参考数据，以期找到某些规律并为指导将来期刊收藏模式提供依据，为学科建设发展提供多样化的文献信息资源保障。另外，本文计算得出今年来该学科博士论文引文文献半衰期，于是该学科文献更新要求得以量化，为后续文献订阅服务提供依据。

## 第二章 研究过程、方法及步骤

引文(参考文献)，是在撰写论文过程中参考过的主要文献目录，是论文作者参阅借鉴前人研究成果的反映，是论文的重要组成部分。引文分析法是一种最科学、可靠而简单的定量分析方法<sup>[1]</sup>。

因为文献被引用是读者阅读了该文献并从中获得帮助的重要标志。而且引文分析的研究方法比其它研究方法简单，数据的获取更迅速、经济而准确<sup>[2]</sup>。

教师和研究生是高校图书馆的重要服务对象，他们的文献需求在很大程度上决定着馆藏建设方向。但研究生对文献的利用，较多依赖本地图书馆收藏。如果本地图书馆馆藏结构不合理，或藏书量不足，都将导致引文研究结果的偏差<sup>[3]</sup>。

本文以农业资源与环境学科博士生导师为检索入口，并以华中农业大学和2010~2012出版年作为限定检索词，获得博士论文34篇，引文7620篇。

值得注意的是，对于文献半衰期的计算是本文的亮点。为了衡量科学文献的老化速度和程度，定量地揭示文献老化的规律，1958年，英国著名的学者J. D. Bernal借用物理学中的放射性物质的衰变性质，生动地描述了科技文献的老化问题，并首次提出了科技文献半衰期的概念，以此作为衡量已发表文献的老化程度指标。所谓文献的半衰期是指在该时间内，已发表的某一学科或领域内尚在被利用的全部文献的一半<sup>[4]</sup>。其计算方法有负指数函数、Burton-Keblner方程以及中值法等，其中中值法计算较为简便，结果较为可靠<sup>[5]</sup>。

中值法计算方法为:

设*i*为引文年令,某学科的引文中第*i*年的引文量为 $a_i$  ( $i = 0, 1, 2, \dots, n$ ), 令

$$A_0 = a_0$$

$$A_1 = a_0 + a_1$$

.

.

.

$$A_n = a_0 + a_1 + \dots + a_n,$$

$A_n$ 为总引文量, 即

$$A_j = \sum_{i=0}^j a_i$$

总存在一个*i*, 使得

$$A_i \leq \frac{A_n}{2} < A_{i+1}$$

则半衰期  $T$  一定落在区间  $[i, i+1]$  内. 假设各年度的引文量是随时间均匀分布的, 令  $T=i+b$  ( $0 \leq b < 1$ ), 经计算半衰期

$$T = i + \frac{\frac{A_n}{2} - A_i}{A_{i+1} - A_i} \dots\dots\dots 1$$

### 第三章 结论与建议

#### 4.1 引文的类型分布

引文类型反映出各种文献载体在科研中的价值、地位、作用情况, 以及科技信息传播渠道的构成和比重<sup>[6]</sup>

由表 1 可以看出, 在全部引文中, 期刊引文平均占引文总数的 95.79%, 这说明期刊具有内容新、信息量大、传递速度快等特点占据被引主导地位, 可以肯定, 在目前和将来, 期刊或期刊数据库将始终是我国科技工作者第一位的文献来源. 因此图书馆应广辟期刊, 并加大对数据库的资金投入, 为教师和学生进行科研提供更便捷的服务. 图书的引文平均占引文总数的 4.21%, 位于第二位. 这说明图书具有理论系统性强、内容完整、观点成熟、应用时间长、便于查阅等特点被引率处于稳定位置. 而学位论文、会议论文和网络信息的利用率不高,

三类文献总和不到1%。究其原因，一方面在于与该类文献的专业性和权威性方面逊色于期刊和图书，使得教师和学生在使用时有考虑；另一方面是由于科研工作者在获得相关资源时常受限制。所以，图书馆应该考虑如何让教师和学生能更为快捷方便地获得高质量的这类几类文献资料。

表 1 引文类型分布

文献类型	期刊	图书	学位论文	会议论文	其它	合计
数量	7299	321	39	3	17	7620
百分比(%)	95.79	4.21	0.51	0.04	0.22	100

## 4.1 引文年代分析

参考文献的时效性问题关系到科研成果的质量和水平，反映了文献对当前研究动态的掌握情况<sup>[7]</sup>。

从表 2 可以看出，不同文献类型中，2005 到 2009 年的引文最多，共 2215 篇，占总引文的 29.1%，2000 到 2004 年的引文次多，共 2167 篇，占总引文的 28.4%，引用其他年代的论文逐年递减。

由表 2 引文年代分布可知，以 20 世纪 80 年代为界限，随着年代的递增，所引用的文献也大幅度的递增，但引用 20 世纪 70 年代前的文献就逐步减少。文献老化的规律基本和前人所做的研究结果相吻合。针对我校图书馆而言，为了更好的服务于我校的教学科研，就要做好文献的剔除工作，腾出宝贵的物理空间，使有用的文献不致淹没在老化的文献群中。对半衰期较短的期刊，要加强文献报道，最好能通过电子邮件通告服务及时告知教师，以提高文献的利用率，并可开展定题服务。对老化的文献则可实行数字化复制或藏入典藏书库中。必要时，还可统计研究具体的老化数据，确定文献开架的年限，从而节约人力和财力。

表 2 引文年代分布

	1970 以前	1970- 1974	1975- 1979	1980- 1984	1985- 1989	1990- 1994	1995- 1999	2000- 2004	2005- 2009	2010- 2012	文献 总数
引文 篇数	66	68	120	210	328	707	1412	2167	2215	327	7620
百分比%	0.9	0.9	1.6	2.8	4.3	9.2	18.5	28.4	29.1	4.3	100

## 4.2 篇均引文数

由表 3 可知，2010 年篇均引文数为 234 篇，2011 为 233 篇，2012 则下降为 200 篇。不难看出，随着年份的增加，篇均引文数有下降的趋势。结合文献类型及语种分析结果，我们可以推断，单篇引文的质量可能有所增加，比如英文文献的引用比例呈增加趋势。

表 3 篇均引文数

年份	2010	2011	2012
总引文量	1644	3974	2002
论文篇数	7	17	10
篇均引文数	234	233	200

### 4.3 引文半衰期

根据公式 1, 因为到 2003 年累积引文相对比例已达 44.95%, 故取  $i=10$ , 代入公式 1 进行计算,  $T=10+(7620/2-3425)/(3850-3425)$ , 计算得出  $T=10.91$ , 即引文半衰期为 10.91 年。

表 4 引文逐年统计数据

年份	引文频次	累积引文频次	累积引文相对比例
2012	11	11	0.00144357
2011	102	113	0.014829396
2010	214	327	0.042913386
2009	402	729	0.095669291
2008	412	1141	0.149737533
2007	440	1581	0.207480315
2006	462	2043	0.268110236
2005	499	2542	0.333595801
2004	436	2978	0.390813648
2003	447	3425	0.449475066
2002	425	3850	0.505249344
2001	423	4273	0.560761155
2000	436	4709	0.617979003
1999	369	5078	0.666404199
1998	301	5379	0.705905512
1997	289	5668	0.743832021
1996	210	5878	0.771391076
1995	243	6121	0.80328084
1994	193	6314	0.828608924
1993	140	6454	0.846981627
1992	129	6583	0.863910761
1991	119	6702	0.879527559
1990	126	6828	0.896062992
1989	99	6927	0.909055118
1988	72	6999	0.918503937
1987	60	7059	0.926377953
1986	51	7110	0.933070866
1985	46	7156	0.939107612
1984	60	7216	0.946981627
1983	39	7255	0.952099738
1982	44	7299	0.957874016
1981	36	7335	0.962598425
1980	31	7366	0.966666667
1979	27	7393	0.970209974
1978	25	7418	0.973490814
1977	26	7444	0.976902887

年份	引文频次	累积引文频次	累积引文相对比例
1976	18	7462	0.979265092
1975	24	7486	0.982414698
1974	12	7498	0.983989501
1973	15	7513	0.985958005
1972	16	7529	0.988057743
1971	13	7542	0.98976378
1970	12	7554	0.991338583
更远	66	7620	1

引文半衰期结果说明，华中农业大学农业资源与环境学科博士毕业生在写作毕业论文时，更注重近 10.91 年的文献，近 11 年来的文献，参考价值最大。这就提示图书馆在针对该学科提供文献服务时，多注重近 11 年来的各种科学出版物，将这方面工作深入细致开展。

#### 4.4 引文文献类型分析

引文可以由不同语种的文献构成，研究引文的语种构成对了解科研人员的外语水平及论文作者对不同语种文献的需求状况<sup>[8]</sup>。

在统计的 7620 篇引文中（具体引文语种分部如表 4），引用语种最高的是英文期刊，合计 4952 篇，占总引文量的 64.99%，其次是中文期刊，共 4952 篇，占总引文量的 64.99%，英文和其他语种的引用文献统称为外文引文，共有 2288 篇，占总引文量的 30.03%。

表 4 图书与文献引文分布表

文献语种	中文			英文			文献总数
	期刊	图书	其它	期刊	图书	其它	
引文篇数	2288	249	37	4952	72	22	7620
百比例 (%)	30.03	3.27	0.49	64.99	0.94	0.29	100

这一结果如实反映出来农业资源与环境学科的研究生培养模式和特点。作为我省重点学科，对于教师和学生的科研要求较高，从事科研需要阅读和参考大量的外文文献，尤其是英文文献，这与国际权威刊物所用文字大多是英文有关。

通过文献语种与博士论文作者毕业年份的交叉列联表分析（见表 5），可以发现，2010 年博士论文引文中引用中文文献 595 篇，占本年总文献的 36%；英文文献 1049 篇，占本年总文献的 64%。2011 年博士论文引文中引用中文文献 1465 篇，占本年总文献的 37%；英文文献 2509 篇，占本年总文献的 63%。2012 年博士论文引文中引用中文文献 514 篇，占本年总文献的 26%；英文文献 1488 篇，占本年总文献的 74%。

表 5 文献语种与毕业年份交叉列联表

		毕业年份			合计
		2010	2011	2012	
文献 语种	中文文献的篇数	595	1465	514	2574
	占本年总文献的百分比	36%	37%	26%	
	英文文献的篇数	1049	2509	1488	5046
	占本年总文献的百分比	64%	63%	74%	
合计		1644	3974	2002	7620
显著性检验		$X^2=79.982$	Df=2	Sig=.000	

从以上数据,不难发现博士论文引文对于英文文献的利用率远远高于中文文献,而且这一趋势在近三年有着更为明显表现。而统计显著性检验结果显示, $X^2$ 值为 79.982,  $p < 0.05$ ,说明博士论文作者的毕业年份与其利用中英文文献的多少存在密切关系,越近毕业的博士生其对英文文献的利用率越高。

调查结果反映出图书馆在外文数据库建设中的具体成效,为教师和学生的科学研究提供了极为便利和有效的学术资源。自然科学研究讲究时效性、新颖性,课题前期需要科技查新,以避免重复劳动,所以,科研人员需要利用国际权威刊物来了解和获取本学科的国内外最新研究动态。因此,图书馆应在馆藏资源建设和资源提供上再接再厉,以教学科研为基础,追踪学科前沿,及时满足、更新电子及纸质外文期刊,并通过大力推广 CASHL、NSTL 合作开展文献传递与馆际互借服务,通过资源共享缓解馆藏外文资源的不足。

表 6 引文中中文出版社分布表

中文图书出版来源	引用数	百分比
科学出版社	77	30.9%
农业出版社	33	13.3%
中国农业出版社	26	10.4%
科学技术出版社	18	7.2%
中国农业大学出版社	11	4.4%
中国农业科技出版社	9	3.6%
高等教育出版社	7	2.8%
中国环境科学出版社	6	2.4%
中国大百科全书出版社	6	2.4%
地质出版社	6	2.4%
中国统计出版社	5	2.0%
中国林业出版社	5	2.0%
其它出版社合计(4次及以下)	40	16.1%
总计	249	100.0%

由表 6可以看出,引用前 3 位的图书出版社为科学出版社、农业出版社及中国农业出版社,分别占比 30.9%、13.3%、10.4%。引用频率高的出版社共 12 种。

这说明,一方面,博士论文引用的中文图书,以学术性最高的科学出版社为主,另一方面,农业方面的出版社排名第二第三说明了鲜明的专业特色。

科学出版社书籍是博士学位引文主要引用图书，图书馆应特别重视该社图书的引进，多引进高质量，精品的，有代表性的科学出版社图书。

表 7 引文中英文出版社分布表

英文图书出版来源	引用次数	百分比
Academic Press	12	16.7%
ASM Press	11	15.3%
China Agriculture Press	11	15.3%
Oxford University Press	5	6.9%
Cambridge University Press	4	5.6%
其它出版社合计（3次及以下）	29	40.3%
总计	72	100.0%

由表 7可以看出，引用前 3 位的英文图书出版社为Academic Press、ASM Press及China Agriculture Press，分别占比 16.7%、15.3%、15.3%。引用频率高的出版社共 5 种。

这说明，一方面，博士论文引用的英文图书，仍旧以有鲜明学术特性的 Academic Press 为主，另一方面，农业出版社排第三也说明了鲜明的专业特色。

Academic Press 书籍是博士学位引文主要引用图书，图书馆应特别重视该社图书的引进，应结合学科方向需要，多引进高质量，精品的，有代表性的，及时更新的 Academic Press 图书。

## 第四章 支撑材料列举（发表的文章、相关照片、系统、软件、获奖等情况列举说明）

### 参考文献

- [1] S. E. Wiberley Jr, "Journal rankings from citation studies: a comparison of national and local data from social work," *The Library Quarterly*, pp. 348-359, 1982.
- [2] M. J. Sylvia, "Citation analysis as an unobtrusive method for journal collection evaluation using psychology student research bibliographies," *Collection building*, vol. 17, pp. 20-28, 1998.
- [3] S. Edwards, "Citation analysis as a collection development tool: a bibliometric study of polymer science theses and dissertations," *Serials review*, vol. 25, pp. 11-20, 1999.
- [4] 邱均平, *文献计量学*: 科学技术文献出版社., 1988.
- [5] 李玉珍 and 郭振平, "科技文献半衰期的计算," *延边大学学报: 自然科学版*, vol. 23, pp. 74-76, 1997.
- [6] 罗玲, "《测绘科学》2000—2004 年引文分析," *测绘科学*, vol. 30, pp. 110-112.
- [7] 安美荣, "高校教师发文的引文实证分析——对中北大学教师近三年发文的调查," *晋图学刊*, vol. 2, pp. 63-65, 2009.
- [8] 王惠翔 and 高凡, "我国情报学主要期刊引文定量分析," *情报理论与实践*, vol. 161, p. 147, 2004.

附录（含专家验收表）

专家验收表

项目名称	基于博士学位论文引文分析的文献需求研究				
主持人	刘远翔	职务/职称	副研究馆员		
所在单位	华中农业大学图书馆（加盖公章）				
专家意见	<p>2013年5月，受CALIS全国农业文献信息中心委托，华中农业大学图书馆组织CALIS全国农业文献信息中心2012年研究项目课题验收专家组对本馆承担的《基于博士学位论文引文分析的文献需求研究》课题进行验收。验收专家组审查了课题承担者提供的技术资料，听取了课题项目负责人的汇报，经过认真讨论，形成验收意见如下：</p> <ol style="list-style-type: none"> <li>1. 课题组提供的技术资料齐全完整，符合验收要求。</li> <li>2. 课题按期完成了预期规定的研究内容，博士学位论文引文分析的文献需求研究意义很大，调研统计方法科学合理，完成的结果及应对措施有利于我馆重点学科期刊资源建设发展和提高读者的利用率。</li> <li>3. 科学性较强，研究思路清晰，工作严谨。</li> <li>4. 课题项目成果达到预期规定要求。</li> <li>5. 建议进一步深入研究，形成系统理论研究成果。</li> </ol>				
专家签名	张俊	李唯	余红	赖星	肖新
职务/职称	研究员	副研	副研	副研	副研

\*如需要可增加页数