

基于改进CARS模型的学术文献被引频次 影响因素分析*

高丹 李秀霞 周娜

(曲阜师范大学传媒学院,日照 276826)

摘要: 为丰富学术文献引言结构的研究,规范学术文献的引言写作,本文探究引言结构对学术文献被引频次的影响,以提高其学术影响力。以Web of Science为数据来源,分析图书情报领域文献的引言结构,将CARS模型简化为“阐述重要性”“文献综述”“研究不足”“研究目的”“预示重要结果”“论文结构”6个语步,并构建引言结构词表。通过Python编程过滤引言结构词表,将学术文献引言部分各语步的词表重合率与对应文献的被引频次进行相关性分析。研究发现,“研究不足”“预示重要结果”“阐述重要性”语步的词表重合率与被引频次呈现较强相关性,表明这3个语步的写作规范度可作为影响被引频次的指标。

关键词: 引言结构;被引频次;CARS模型;引言词表

中图分类号: G312

DOI: 10.3772/j.issn.1673-2286.2018.10.003

被引频次指学术文献发表后的被引用次数,是评价与学术文献有关的期刊、学科(专业)、国家(地区)、单位(个人)的一种方法,是反映学术文献的科学水平和学术影响力的重要指标^[1]。目前,提高被引频次的重要性已在学术界达成广泛共识。

国内外学者从多个角度对被引频次的影响因素进行深入探讨。国外学者如Virgo^[2]验证了被引频次与学术文献重要性的正相关假设;Webster等^[3]指出参考文献的数量与被引频次存在正相关关系;Stanek^[4]验证了学术文献的篇幅长短与被引频次存在相关关系,篇幅较长的学术文献被引频次也相应较多;Maliniak等^[5]研究男性和女性在学术文献被引频次和出版量的差异,发现女性作者的被引频次整体低于男性作者;Kulkarni等^[6]验证医学领域热点话题对被引频次的影响,发现对行业有利的观点,肿瘤学和心脏病学等主题与被引频次有关;Suryani等^[7]通过CARS模型对比分析计算机科学领域中的高被引文献和低被引文献,发现作者需要在引言中阐明研究成果和研究价值,才能使学术文献得到更高的引用频次。国内学者如姜磊等^[8]探讨中文

学术文献的参考文献数量与被引频次的关系,表明参考文献数量与被引频次的相关性不明显,参考文献可以作为一个独立于被引频次的论文评价指标;牛昱昕等^[9]验证了单篇论文的下载频次与被引频次的相关性不显著;苏芳荔^[10]采用符号检验与相关分析的方法,发现合作发表论文的影响力明显高于无合作发表的论文;简琳等^[11]发现关键词个数对文献被引频次存在显著影响;杨利军等^[12]从引用文献的时间、类型、数量3个特征进行分析,得出作者的引用习惯是影响被引频次的重要因素之一。

分析已有研究,发现被引频次影响因素的研究主要涉及参考文献数量、篇幅长短、下载频次、合作模式、关键词个数及作者引用习惯等,通过计量的方法分析各影响因素与被引频次的关系。已有研究大多以外部因素为研究对象,鲜有基于学术文献内容结构的研究。

引言是学术文献的重要组成部分,作者需要通过引言引起读者的阅读兴趣,判定与其研究主题相关后,才会激发深度阅读行为。引言作为学术文献中不可缺少的组成部分,是决定学术文献质量的重要因素之一;而

*本研究得到国家自然科学基金项目“文献内容分析与引文分析融合的知识挖掘与发现研究”(编号:16BTQ074)资助。

被引频次能揭示学术文献的应用价值、参考价值或学术影响力,是进行学术文献评价的重要指标之一。学术文献的引言与被引频次密切相关,两者间的关系值得深度探索。本文拟将引言的内容结构与被引频次结合起来,探讨引言与被引频次的关系,为规范学术文献的引言写作、提高其学术影响力提供帮助。

1 引言结构

1.1 CARS模型介绍

美国体裁分析学家Swalesian学派的代表人Swales在研究各学科的学术文献引言部分时,发现不同学者在写作学术文献这一特殊体裁时,内容结构等方面普遍一致。Swales^[13]认为,语篇具有特定的共同交流目的,这影响并制约文章内容和文体风格。因此,得出“学术文献的写作只有符合其特定的体裁模式,才能实现其交流目的。而这种公认体裁模式是不同研究领域进行学术交际的最好方法”。

Swales在1990年提出引言结构分析模型,即CARS(Create a Research Space)模型^[13],包括确定研究领域、确定研究定位、把握研究契机3个语步(move),以及相应语步的步骤(step)。语步是作者写作目的的总体概况,步骤是为实现语步目的的详细描述。第一语步,确定研究领域,即通过回顾前人研究成果来阐明研究意义与重要性,包括宣称中心活跃度、主题概括、综述以往文献;第二语步,确定研究定位,通过指出前人研究缺陷再次强调研究的必要性,包括反驳已有观点、指出研究不足、提出研究问题和继承前人研究成果4个循环步骤;第三语步,把握研究契机,包括概述研究目的、描述研究特征、预示研究成果、介绍论文结构4个步骤。

Swales的CARS模型为引言提供了普遍意义上的写作思路和框架,使学术文献做到“言之有序”,学术文献遵循的“序”即引言的框架结构。然而,CARS模型中对语步、步骤的概念定义不够细致规范,仅依据于小部分代表性的句法特征,且应用价值较弱,对话类验证只能采取事后分析的方式,对话步的判定仅依赖使用者的直觉,这样很难获得分析的一致性。而且,根据实际写作的需要,CARS模型中一些语步和步骤功能相近,略显冗余,需要进一步整合。

1.2 改进的CARS模型

为改善CARS模型的不足,丰富引言结构,本文尝试提出一种新的引言结构词表,为学者对引言的规范写作提供借鉴和参考。

引言结构词表数据来自Web of Science的科学引文(SCI-EXPANDED)数据库,随机选取图书情报领域的500篇学术文献。通过分析学术文献的引言部分,发现CARS模型可以简化为“阐述重要性”“文献综述”“研究不足”“研究目的”“预示重要结果”“论文结构”6个语步,且语步可自由组合。将文献引言以此分类,在结合Swales等体裁分析学家的语料库积累上,通过自编程得到引言词表(见表1)。

引言词表具有以下优势:①汇总引言部分的高频单词,专注引言体裁的检索需求,为引言的规范写作提供参考与借鉴;②实用性强,可利用重合率来进行语类验证、语步判定,最大程度地满足读者的实际需要;③收词范围较广,取材范围覆盖学术文献引言所使用的正式和非正式语体,收词以高频核心词为基础,同时适当收录了各体裁分析学家认为读者在写作引言时所必须的部分词语;④词性收录全面,对动词、名词、副词、第三人称单数等做了详细归纳。

2 实验与结果分析

引言是学术文献不可或缺的组成部分,起着开篇点题的作用,是决定学术文献质量的重要因素之一;被引频次是反映学术文献的科研水平和学术影响力的重要指标之一。针对两者是否存在相关关系这一问题,本文提出用语步重合率,即各语步与引言词表相重合的数量占各语步总词数的比,与论文被引频次进行相关性分析,探讨CARS模型对论文被引频次的影响。

2.1 数据来源及处理

为分析引言的内容结构与被引频次的关系,基于CARS模型构建引言词表,引言词表的数据来源于图书情报领域的500篇随机学术文献。为保证样本检验数据能正确反映引言词表对学术文献被引频次的影响,样本检验数据源区别于引言词表数据源。样本检验数据来源于Web of Science的科学引文(SCI-EXPANDED)数据库,检索方式采用高级检索,以“Information

表1 引言词表(部分)

阐述重要性	activit (y/ies)、associate、assess (ment)、attention、authority、advantage、change (s/d/ing)、case (s)、company、common (ly)、competitive、care、considerable、concern、detection、detect、develop (ment)、discover (y/ies)、effect (ive)、econom (y/ic)、example、firm、health、increas (e/ed/ing/ingly)、involv (e/ing)、interest、importan (t/ce)、information、individual、key、mak (e/ing)、model、number、organisms、public (ation)、policy、people、progress (ively)、property、primary、publish、perform、provide、quality、recently、research (er)、recovery、severe、reduce、significant、stud (y/ies/ying)、system (s/ic)、support、share、strategic、valu (e/able)、well
文献综述	attention、account、analy (sis/tic/tical)、benefit、better、chemical、chang (e/es)、competence、develop (ed)、different、data、describe、example、explor (e/ing)、establish (ed)、effect (ive/ively/iveness)、expert、general (ly)、however、inform (ation)、impact (s)、includ (e/ing/ed/es)、investigat (e/ion)、literature (s)、measur (e/es/ing/able)、method (s/ology)、number、nevertheless、obtainable、process (es/ing)、problem、present、predict (ive/ion/ing)、propose、practic (ing/ally)、provide、premise、question、research、result (s)、report (s/ed/ing)、reason (s)、review (s/ed/ers)、recently、solve (d)、system (s/atic)、stud (y/ies/ying)、source、suppl (y/ement)、suggest、support、structure、theoretical、tool、use (d)、work
研究不足	although、analy (sis/ses/zing)、assess (ment)、approach、benefit、case、care、construct、despite、discuss (ed/ion)、develop、design、decision、few、focus、formative、factor、however、information、important、impact、literature (s)、less、little、lack、model、make、method、new、organization (al)、paper、policy、approve、publish、problem、provide、performance、potential、question、research (er)、risk、strategies、since、system (s/atic)、study、target、theory、value
研究目的	aim (s/ed)、assess (ing/ment)、address (ing)、across、article、base、build、current、construct、care、discuss (ion)、develop、domain、design、describe、effect (ively)、examine、field、focus、first、factor formative、goal、information、identify (ing)、influence、impac、literature、model、measure、necessary、objective、order、purpose、paper、provide (s/ed)、process (es/ing)、propose、present、predict (ive/ion)、project、practice、relation (ship)、review、research (er)、significant、role、system、stud (y/ies)、therefore、theory、understand、use、various
预示重要结果	address (es/ed)、article、advocate、assess (ing/ment)、accurately、application、between、benefit、current (ly)、critical、consequen (ce/tly)、contribution、competitive、differen (t/ce)、develop、design、effective、future、first、fundamental、highlight、information、important、influence、large、literature、mainly、necessary、paper、provide、polic (y/ies)、potential (ities)、performance、quality、research、result (s)、review、relevant、relate、relationship、study、status、supplier、system、significant、support、topic、theory、theoretical、understanding、various、value、work
论文结构	analysis、article、context、conclude、construct、discuss (ion/ed/es)、data、develop (ed/ment)、describe、explain、follow (s/ed)、future、finally、first、finding、framework、information、include、implication、literature、model、method、next、overview、organize、measure、paper、provide、present (s/ed)、process、propose、research、result、review、section、study、system、suggest、theoretical、theory、use

Science & Library Science”为学科类别，以“knowledge management”为检索词，文献类型为“Article”，限定语言为“English”，发表时间为2014年，检索时间为2018年5月24日，共检索到789篇文献。随机抽取25篇文献作为样本数据。

将25篇文献按照被引频次(排除自引)排序得到表2，每篇学术文献的引言内容按照“阐述重要性”“文献综述”“研究不足”“研究目的”“预示重要结果”“论文结构”6个语步归类，分别统计每篇引言在各语步中出现的总词数 m 。通过Python编程，以引言词表分别过滤引言各语步，得到引言中与词表重合的词数 n ， n/m 为

各语步与引言词表相重合的数量占各语步总词数的比，即重合率。

2.2 结果分析

2.2.1 被引频次与引言各语步重合率的对比分析

为验证学术文献引言结构对被引频次的影响，将引言各语步的词表重合率与被引频次比较。各语步的词表重合率与引言写作的规范程度有关，重合率越高，表示引言与“引言词表”更相符，相应的引言写作用词越规

范。若某篇文献没有个别语步，统计时以错误值“#N/”代替，横轴表示25篇引言，纵轴表示各语步与词表的重合率（见图1、图2）。

由图1发现，“研究不足”“预示重要结果”语步的词表重合率与被引频次曲线走势基本一致，被引频次随着词表重合率的降低而降低，但也存在例外（如序号20

的学术文献引言“预示重要结果”语步的词表重合率骤然升高）。由图2可见，“阐述重要性”“文献综述”“研究目的”语步的词表重合率总体一致，皆与被引频次的曲线走势有较大的不同，说明这3个语步词表重合率对被引频次的影响较弱。“研究目的”语步的平均词表重合率最高为0.11，其次是“阐述重要性”语步，说明这两

表2 引言各语步的词表重合率与被引频次的比较

引言序号	被引频次	阐述重要性			文献综述			研究不足			研究目的			预示重要结果			论文结构		
		n	m	m/n	n	m	m/n	n	m	m/n	n	m	m/n	n	m	m/n	n	m	m/n
1	254	45	6	0.13	0	0	0.00	41	7	0.17	44	7	0.16	33	7	0.21	0	0	0.00
2	121	64	8	0.13	136	8	0.06	42	7	0.17	0	0	0.00	26	5	0.19	0	0	0.00
3	116	29	3	0.11	65	4	0.06	0	0	0.00	47	7	0.15	46	7	0.15	0	0	0.00
4	80	23	3	0.13	89	11	0.12	26	4	0.15	40	7	0.18	0	0	0.00	0	0	0.00
5	73	53	5	0.09	0	0	0.00	43	5	0.11	56	8	0.14	53	7	0.13	0	0	0.00
6	50	41	3	0.08	193	22	0.11	38	4	0.11	35	5	0.14	0	0	0.00	0	0	0.00
7	46	48	5	0.10	175	22	0.13	61	7	0.11	24	3	0.13	66	9	0.14	0	0	0.00
8	40	72	7	0.10	278	29	0.10	39	4	0.10	46	6	0.13	52	7	0.13	0	0	0.00
9	37	70	8	0.11	87	10	0.11	58	6	0.10	55	6	0.10	0	0	0.00	0	0	0.00
10	35	63	5	0.08	102	9	0.09	0	0	0.00	0	0	0.00	32	4	0.13	0	0	0.00
11	31	56	5	0.09	0	0	0.00	0	0	0.00	28	4	0.14	19	2	0.11	0	0	0.00
12	29	47	3	0.06	0	0	0.00	71	4	0.06	24	3	0.13	0	0	0.00	0	0	0.00
13	25	105	9	0.09	144	17	0.12	44	3	0.07	47	6	0.13	0	0	0.00	154	29	0.19
14	22	71	5	0.07	71	7	0.10	66	6	0.09	23	3	0.13	46	5	0.11	86	17	0.20
15	19	73	8	0.11	175	21	0.12	53	4	0.07	58	7	0.12	48	5	0.11	0	0	0.00
16	16	67	7	0.10	137	17	0.12	43	4	0.09	46	5	0.11	42	4	0.10	0	0	0.00
17	14	84	7	0.08	0	0	0.00	58	5	0.09	26	3	0.11	50	4	0.08	86	21	0.24
18	12	77	8	0.10	117	12	0.10	0	0	0.00	0	0	0.00	0	0	0.00	0	0	0.00
19	10	109	11	0.10	0	0	0.00	26	2	0.08	41	5	0.12	40	4	0.10	30	7	0.23
20	8	69	7	0.10	162	19	0.12	52	4	0.08	42	6	0.14	49	10	0.20	0	0	0.00
21	5	53	5	0.09	51	5	0.10	42	3	0.07	20	3	0.15	55	3	0.05	0	0	0.00
22	3	81	6	0.07	101	9	0.09	34	2	0.06	31	5	0.16	0	0	0.00	70	22	0.31
23	2	77	6	0.08	144	5	0.03	22	1	0.05	0	0	0.00	17	1	0.06	58	10	0.17
24	1	69	5	0.07	59	2	0.03	60	3	0.05	69	6	0.09	0	0	0.00	0	0	0.00
25	0	45	4	0.09	0	0	0.00	0	0	0.00	0	0	0.00	0	0	0.00	0	0	0.00

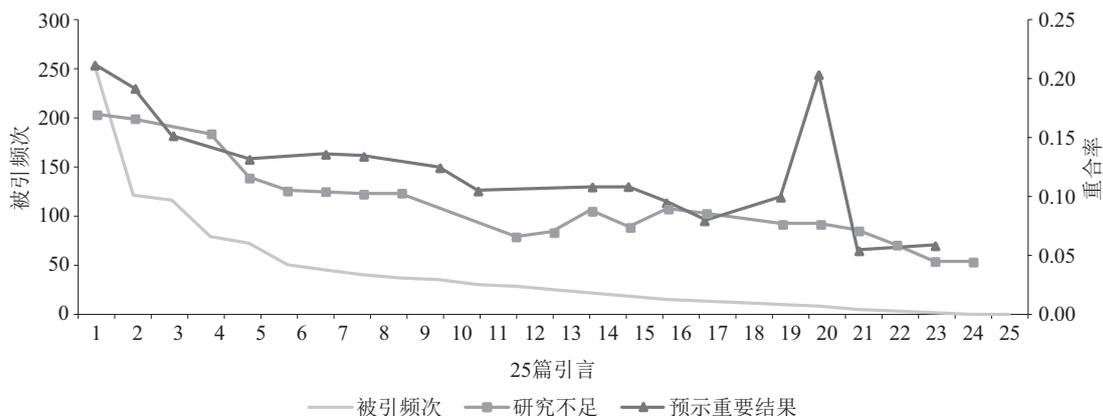


图1 “研究不足”“预示重要结果”语步与被引频次的比较

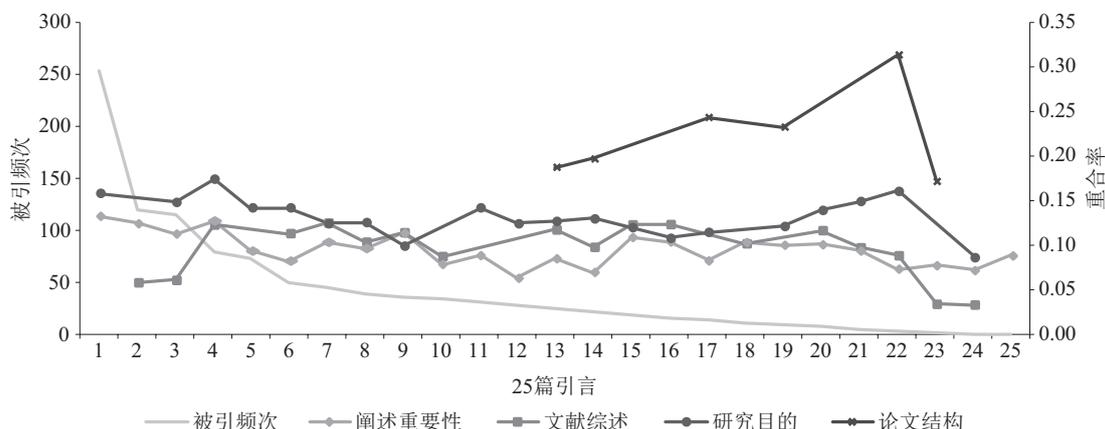


图2 “阐述重要性”“文献综述”“研究目的”“研究不足”“论文结构”语步与被引频次的比较

个语步在引言写作时常用术语具有更高的稳定性。“论文结构”语步的样本太少，数据不纳入分析。

2.2.2 被引频次与引言各语步重合率的相关性分析

为更准确地呈现被引频次与引言各语步间的关联性，本文在SPSS环境下，利用表3的数据计算各指标间的Spearman相关系数。

被引频次与除“论文结构”外的其他所有语步均呈正相关，其中与“研究不足”语步呈强相关性，相关系数为0.552。被引频次与“阐述重要性”“预示重要结果”语步均呈较强相关，相关系数分别为0.451和0.421。由表3还发现，“论文结构”语步与被引频次呈负相关，相关系数为-0.391。“论文结构”语步仅简要概括文献的各章节，并不包含实质学术价值和信息，与被引频次相关性较低。

表3 被引频次与引言各语步间的相关性

	被引频次	阐述重要性	文献综述	研究不足	研究目的	预示重要结果	论文结构
被引频次	1.000	0.451 *	0.104	0.552 **	0.283	0.421 *	-0.391
阐述重要性	0.451 *	1.000	0.354	0.480 *	0.075	0.409 *	-0.415 *
文献综述	0.104	0.354	1.000	0.380	0.132	0.006	-0.177
研究不足	0.552 **	0.480 *	0.380	1.000	0.196	0.242	-0.11
研究目的	0.283	0.075	0.132	0.196	1.000	0.186	0.003
预示重要结果	0.421 *	0.409 *	0.006	0.242	0.186	1.000	-0.203
论文结构	-0.391	-0.415 *	-0.177	-0.110	0.003	-0.203	1.000

注：*相关性在0.05层上显著（双尾），**相关性在0.01层上显著（双尾）。

2.3 引言语步对学术文献被引频次的影响分析

(1) “研究不足”语步对被引频次的影响。“研究不足”语步可以指出前人研究的问题与不足，填补科学研究的缺陷和空白。在前人提出的研究课题中，虽然已有初步的研究成果，但随着科学技术的不断发展，还有待补充、丰富。这种补充性或纠正性的研究课题，强调当前研究的必要性，对社会发展有着深远的科学价值和现实指导意义。“研究不足”语步是引言的重要组成部分，与被引频次联系紧密。

(2) “阐述重要性”语步对被引频次的影响。“阐述重要性”语步的主要作用是声明研究课题的重要意义，阐述研究主题“重大而主要”的学术价值和社会影响，以达到吸引潜在读者的目的。在开篇引言部分，读者就可获知文献的学术价值，从而激发深度阅读行为，被引频次相应提高。

(3) “预示重要结果”语步对被引频次的影响。“预示重要结果”语步说明作者在当前研究中的独创性贡献，推进或解决某一领域的发展。“阐述重要性”语步与“预示重要结果”语步互为关联，但两者侧重点不同。前者强调研究课题进行的必要性与目的，后者着

重研究课题取得的成效和贡献。学术贡献是一篇学术文献的核心价值,“预示重要结果”语步是影响被引频次的因素之一。

被引频次是反映学术文献影响力的重要指标,被引频次有众多的影响因素,本文的实验数据验证了引言语步与被引频次存在关系。上述研究发现,改进后的CARS模型中“研究不足”“阐述重要性”“预示重要结果”3个引言语步的词表重合率与被引频次呈现较强相关性,引言语步的词表重合率越高,代表引言写作越规范,相应的学术文献被引频次和影响力就高。这表明,引言写作规范度与被引频次呈正相关关系,写作规范度对文献的被引频次影响较大,可作为影响被引频次的指标。因此,为提高学术文献的被引频次和影响力,学者应选择合适的引言结构和语步,规范学术文献的引言写作。

3 结论与启示

本文对CARS模型提出改进,并对引言分别建立了“阐述重要性”“文献综述”“研究不足”“研究目的”“预示重要结果”“论文结构”6个语步的词表。本研究丰富了引言结构内容并提高了其应用价值,通过对引言各语步与被引频次的相关分析,发现“研究不足”“预示重要结果”“阐述重要性”语步的词表重合率与被引频次呈现较强相关性,表明这3个语步的写作规范度可以作为影响被引频次的指标,同时表明科学的引言结构对学术文献的被引频次有重要影响。

因此,学者可规范学术文献的引言写作,使用CARS模型构建合适的引言结构,在引言部分重点阐述“原有研究的缺陷”“本课题的重要性”“本课题的学术贡献”,进而提高我国期刊论文的被引频次,提高其学术影响力。

模型构建过程中存在以下不足:①本文选取的数据样本范围单一,只分析了图书情报领域学术文献的引言,使结论分析较片面;②构建引言词表模型时,对语料库的构建具有一定的主观性,使实验数据存在一定误差。后续研究中将对引言词表模型做相应的更新,创建学术文献引言结构语料库,为引言的规范写作提供参考与借鉴,从而提高学术文献的应用价值和学术

影响力。

参考文献

- [1] 杨华,蔡言厚.论提高被引频次的意义、客观条件和主观努力[J].评价与管理,2010,8(2):24-27.
- [2] VIRGO J A. A statistical procedure for evaluating the importance of scientific papers [J]. Library Quarterly, 1977, 47(4): 415-430.
- [3] WEBSTER G D, JONASON P K, SCHEMBER T O. Hot topics and popular papers in evolutionary psychology: Analyses of title words and citation counts in evolution and human behavior, 1979—2008 [J]. Evolutionary Psychology, 2009, 7(3): 348-362.
- [4] STANEK K Z. How long should an astronomical paper be to increase its Impact? [J]. Ads Abstract Service, 2008, 3(9): 1-11.
- [5] MALINIAK D, POWERS R M, WALTER B F. The gender citation gap in international relations [J]. International Organization, 2013, 67(4): 1-10.
- [6] KULKARNI A V, BUSSE J W, LFFAT S. Characteristics associated with citation rate of the medical literature [J]. Plos One, 2007, 2(5): 403-407.
- [7] SURYANI I, YAACOB A, AZIZ N H A. Introduction sections of research articles with high and low citation indices [J]. Pertanika Journal of Social Science & Humanities, 2015, 23(4): 1139-1152.
- [8] 姜磊,林德明.参考文献对论文被引频次的影响研究[J].科研管理,2015,36(1):121-126.
- [9] 牛昱昕,宗乾进,袁勤俭.开放存取论文下载与引用情况计量研究[J].中国图书馆学报,2012,38(4):119-127.
- [10] 苏芳荔.科研合作对期刊论文被引频次的影响[J].图书情报工作,2011,55(10):144-148.
- [11] 简琳,何静,周剑.论文被引的文本因素分析:多学科视角[J].图书情报工作,2011,55(20):32-35.
- [12] 杨利军,万小渝.引用习惯对我国期刊论文被引频次的影响分析——以情报学为例[J].情报科学,2012,30(7):1093-1096.
- [13] SWALES J M. Genre analysis: English in academic and research settings [M]. Cambridge: Cambridge University Press, 1990.

作者简介

高丹, 女, 1994年生, 硕士研究生, 研究方向: 数据挖掘, E-mail: 1172787802@qq.com。

李秀霞, 女, 1971年生, 教授, 硕士生导师, 研究方向: 信息化与信息管理等。

周娜, 女, 1993年生, 硕士研究生, 研究方向: 信息计量。

Analysis of the Influence Factors of the Inciting Frequency to Academic Literature Based on Improvement of the CARS Model

GAO Dan LI XiuXia ZHOU Na
(School of Communication, Qufu Normal University, Rizhao 276826, China)

Abstract: In order to enrich the research of the introduction structure about academic literature, regulate the introduction of academic literatures, the introduction structure dictionary is put forward to further explore the influence of introduction structure to the frequency of academic literature citations, and extend their academic influence. Using Web of Science as data source to analyze the introduction part of literature in the field of library and information science, we simplify the CARS model as "expounding importance" "literature review" "under-researched" "research purpose" "important results" "paper structure" six moves and construct the introduction of structural dictionary. By using python programming to filter the introduction structure dictionary, the dictionary coincidence rate of each move of academic literature introduction and corresponding literature citations were analyzed. Research Findings: the dictionary coincidence rate and times cited of "under-researched" "important results" "expound importance" move have strong correlation, it is indicated that the writing norms of these three steps can be used as the index to influence the citation frequency.

Keywords: Introduction Structure; Citation Frequency; CARS Model; Introduction Dictionary

(收稿日期: 2018-09-27)

书讯

《汉语主题词表》

《汉语主题词表》自1980年问世以后, 经1991年进行自然科学版修订, 在我国图书情报界发挥了应有作用, 曾经获得国家科学技术进步二等奖。为适应网络环境下知识组织与数据处理的需要, 由中国科学技术信息研究所主持, 并联合全国图书情报界相关机构, 自2009年开始进行重新编制工作, 拟分为工程技术卷、自然科学卷、生命科学卷、社会科学卷四大部分逐步完成。目前工程技术卷和自然科学卷已出版。

《汉语主题词表(工程技术卷)》共收录优选词19.6万条, 非优选词16.4万条, 等同率0.84, 在体系结构、词汇术语、词间关系等方面进行了改进创新。《汉语主题词表(自然科学卷)》共收录专业术语12.4万条, 包含数学、物理学、化学、天文学、测绘学、地球物理学、大气科学、地质学、海洋学、自然地理学等学科领域, 收词系统、完整, 语义关系丰富、严谨, 每条词汇都有相应的学科分类号表现其专业属性, 并与同义英文术语对应。同时, 建立《汉语主题词表》网络服务系统, 提供术语查询、文本主题分析、知识树辅助构建等服务。《汉语主题词表》可用于汉语文本分词、主题标引、语义关联、学科分类、知识导航和数据挖掘, 是文本信息处理及检索系统开发人员不可或缺的工具。

《汉语主题词表(工程技术卷)》已于2014年由科学技术文献出版社出版, 分为13个分册, 总定价3 880元。

《汉语主题词表(自然科学卷)》已于2018年5月由科学技术文献出版社出版, 分为5个分册, 总定价1 247元。两卷均可分册购买。