

李红梅. 基于文献检索的医学科研选题三重境界——从查全、查准到验证的方法[J]. 中华医学图书情报杂志, 2018, 27(3): 56-60, 73.

DOI: 10.3969/j.issn.1671-3982.2018.03.011

· 情报研究与方法 ·

## 基于文献检索的医学科研选题三重境界——从查全、查准到验证的方法

李红梅

[摘要] 目的: 阐述信息检索在医学科研选题中的重要作用, 探讨医学科研选题 3 个阶段应遵循的原则和方法。方法: 以一个宽泛的研究主题(生物碱与戒毒)为例, 描述如何通过查全、查准的方法逐步缩小检索范围, 提出假设并进行验证。结果: 遵循上述原则和方法, 可以全面获得生物碱与戒毒的相关研究背景, 准确筛选出“辣椒辣素用于海洛因依赖戒断”的研究课题并进行验证。结论: 通过全面、准确的文献检索可以有效地选题, 通过查阅相关文献和利用非相关文献知识发现方法可以对所提出的假设进行验证, 从而实现科学选题。

[关键词] 查全; 查准; 选题; 验证

[中图分类号] G254.91

[文献标志码] A

[文章编号] 1671-3982(2018)03-0056-05

**Literature retrieval-based recall, precision and verification methods for selection of medical research topics**

**LI Hong-mei**

(Library of Kunming Medical University, Kunming 650500, Yunnan Province, China)

[Abstract] **Objective** To explain the important role of information retrieval and study the recall, precision and verification principles and methods in selection of medical research topics. **Methods** How to gradually narrow down the literature retrieval scope with the recall and precision methods was described with a relatively broad research topic (alkaloids and abstinence from poisons) as an example. A hypothesis was proposed and verified. **Results** The recall, precision and verification principles and methods could comprehensively detect the research background related with "alkaloids and abstinence from poisons", accurately screen and verify the research topic on "capsaicin used in heroin-dependent abstinence from poisons". **Conclusion** Comprehensive and accurate literature retrieval can effectively help to select the research topics, reading the relevant literature and use of the non-relevant literature knowledge discovery method can verify the hypothesis, which can thus help the scientific selection of research topics.

[Key words] Recall; Precision; Topic selection; Verification

医学科研活动是一种以探索真理、解决问题为主要目标的特殊认知活动。从信息学的角度来看, 科学研究是一种通过信息的获取、评价、利用, 进而对信息进行创新的过程。医学文献是医学科研信息

的主要载体, 对医学文献的检索、分析、利用贯穿于医学科研的全过程, 在科研选题阶段体现得尤为突出。科研思维始于发现问题, 准确的选题是科学研究的良好开端, 是科研工作得以展开和深入的基础。爱因斯坦曾经在其论著《物理学上的进化》中说道: “提出一个问题往往比解决一个问题更重要, 因为解决问题也许仅是一个数学上或实验的技能而已, 而提出问题却需要有创造性的想象力, 而且标志着

[基金项目] 昆明医科大学“十三五”校级学科建设项目“图书情报与档案管理”(JB01826)

[作者单位] 昆明医科大学图书馆, 云南 昆明 650500

[作者简介] 李红梅(1968-), 女, 云南大理人, 硕士, 副研究馆员。

科学的真正进步”。本文尝试以一个宽泛的研究主题“生物碱与戒毒”为例,阐述如何通过文献查全、查准的方法逐步缩小检索范围,从而提出假设并进行验证,即从文献检索的角度进行医学科研选题。

### 1 第一重境界——文献的查全阶段

科研选题的首要任务是发现和提出可研究的问题,锁定研究目标。如研究人员欲从生物碱与戒毒领域选题,首先需要弄清楚目前已有的研究涉及了哪些生物碱、针对何种毒品、效果如何、得出了什么结论等。这需要查阅该领域的国内外文献,进行广泛的学术调研,了解该领域研究的现状、研究水平、发展趋势以及存在的问题。因此该阶段检索的目标就是查全,即把所有生物碱与戒毒方面的文献全面检索出来,尽量避免漏检。研究者可以遵循以下 3 个查全原则。

#### 1.1 查全的原则

##### 1.1.1 穷尽检索词的同义词

检索词的同义词是指和检索词具有同等检索意义的词,可以是检索词的学名与俗称、旧称与新称、简称与全称、不同译名或上下位词等。使用同义词检索的目的是尽量全面地把不同作者用不同词汇撰写的文献都检索出来,确保没有遗漏。如生物碱和戒毒都是比较宽泛的概念,检索的时候如果仅用“生物碱 AND 戒毒”检索,命中文献比较少。以中国生物医学文献数据库(CBM)为例,检索时间截至 2018 年 4 月,增加检索词的不同表达,命中结果则相应有所增加(表 1)。

表 1 CBM 数据库常用字段检索结果

序号	检索式	检索结果/篇
#1	生物碱 AND 戒毒	13
#2	生物碱 AND(戒毒 OR 毒品戒断 OR 物质依赖 OR 药物滥用 OR 药物成瘾)	90
#3	(生物碱 OR 东莨菪碱 OR 伊波加因 OR 云南狗牙花) AND(戒毒 OR 毒品戒断 OR 物质依赖 OR 药物滥用 OR 药物成瘾)	108

在检索策略#3 中,对检索词的同义词进行了较多考虑,依然未穷尽生物碱和戒毒的同义词。在实际检索过程中为提高检全率,要全面考虑检索词的

同义词,帮助研究者尽可能地避免漏检。

#### 1.1.2 使用主题词检索

使用主题词检索,可有效提高检索的全面性和准确性。医学主题词为同一概念具有不同表达方式的词语提供了规范、标准的用语,使文献加工处理达到高度的统一和一致,为高质量的文献查询提供了极大的便利<sup>[1]</sup>。目前多数医学数据库采用的主题词来自医学主题词表(Medical Subject Headings, MeSH)。生物碱是一个宽泛的概念,难以根据经验来穷尽所有的生物碱化合物。MeSH 词表却能清晰、明确地呈现不同表达之间的关系,包括与同义词的关系和上下位关系。在 CBM 的主题检索功能下,输入生物碱类查看到生物碱类的树状结构,即生物碱类的所有下位词,囊括了生物碱的所有类型(图 1)。

#### 树形结构 1

##### 杂环化合物

##### 生物碱类

- 乌头碱
- 八角枫碱
- 槟榔碱
- 喜树碱(+1)
- 金鸡纳生物碱类(+2)
- 秋水仙碱(+2)
- 二氢β刺桐定碱
- 依米丁
- 麦角生物碱类(+19)
- 三尖杉甾碱类
- 洛贝林
- 美司卡林
- 毒蕈碱
- 毛果芸香碱
- 吡咯齐定生物碱类(+1)
- 兰尼碱
- 猪毛菜碱类
- 茄生物碱类(+7)
- 司巴丁
- 托烷类(+11)
- 藜芦生物碱类(+5)
- 黄嘌呤类(+5)
- 苦马豆素
- 1-脱氧野尻霉素
- 吗啡生物碱类(+18)
- 苄异喹啉类(+10)
- 石蒜科生物碱类(+1)
- 甜菜拉因类(+2)
- 苯并菲啶类
- 阿片生物碱类(+30)
- 吡啶酮类(+1)

图 1 生物碱类的树状结构

使用主题词检索的扩展检索功能,可把涉及生物碱类主题词下的所有概念的研究全部检索出来。同理,找到戒毒的主题词为物质相关性障碍,可把其下的各种概念一并检出(图 2)。利用主题词检索,查到 4 412 篇文章,检索结果见表 2。

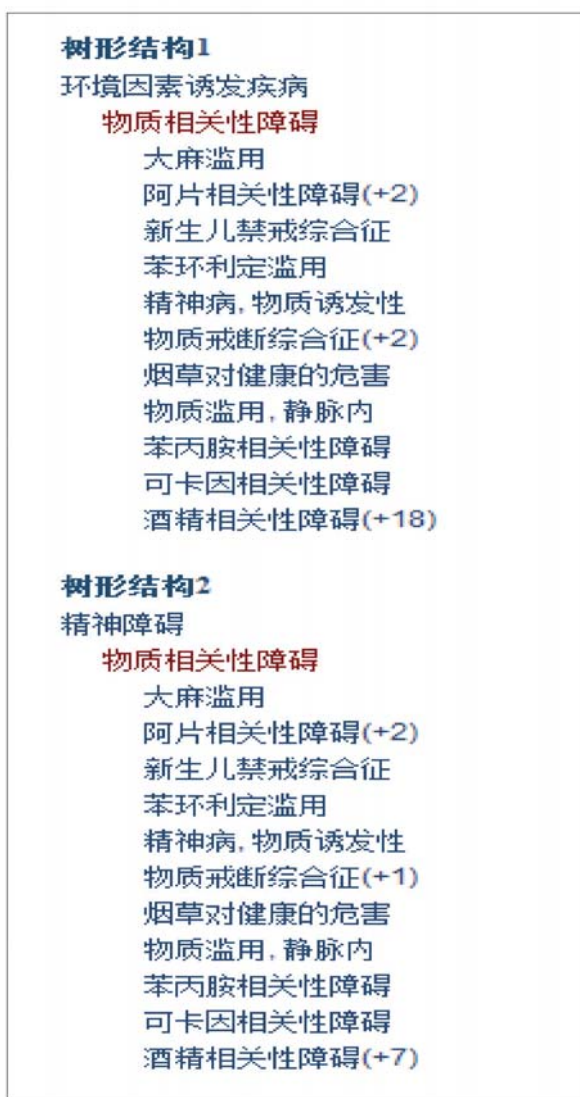


图 2 物质相关性障碍树状结构

表 2 CBM 数据库主题字段检索结果

序号	检索表达式	检索结果/篇
#1	生物碱类[不加权;扩展]	101502
#2	物质相关性障碍[不加权;扩展]	43415
#3	#1 AND #2	4412

### 1.1.3 多个数据库联用

每个数据库的收录范围都有局限,为提高查全

率,研究者需要尽可能地在多个数据库中检索。如分别在 CBM 利用主题词检索,在中国知网学术期刊全文数据库(CNKI)、维普中文生物医学期刊库(CMJD)用一些常用的表达检索,检索结果见表 3。导入文献管理软件 EndNote 中进行去重,得到检索结果为 4 477 篇,高于每一个数据库的单独检索结果。要了解国外的研究现状,还应该选择外文数据库,如 PubMed、Embase 等。在检索词的选择上,也应遵循上述两条原则。

表 3 CBM、CNKI、CMJD 检索结果

数据库	检索表达式	检索结果 /篇
CBM	生物碱类[不加权;扩展] AND 物质相关性障碍[不加权;扩展]	4412
CNKI	(主题=生物碱 OR 主题=茄生物碱 OR 主题=东莨菪碱 OR 主题=伊波加因 OR 主题=云南狗牙花) AND (主题=戒毒 OR 主题=药物滥用 OR 主题=毒品戒断 OR 主题=药物成瘾 OR 主题=物质依赖)(精确匹配)	52
CMJD	(题名或关键词=生物碱 OR 题名或关键词=东莨菪碱 OR 题名或关键词=伊波加因 OR 题名或关键词=云南狗牙花 OR 题名或关键词=茄生物碱) AND (题名或关键词=戒毒 OR 题名或关键词=药物滥用 OR 题名或关键词=毒品戒断 OR 题名或关键词=药物成瘾 OR 题名或关键词=物质依赖)	27

### 1.2 查全在科研选题方面的意义

查全文献的意义在于研究者可以根据检索结果分析了解该领域研究的现状,如有哪些生物碱类用于戒毒治疗、用于何种毒品或物质依赖;国内外研究水平,是基础研究还是临床研究;研究的发展趋势,如当前的研究是否证实生物碱用于戒毒有前景;研究进展,是否有新的生物化学研究成果证实生物碱在戒毒方面的作用;存在的问题,如生物碱用于戒毒有效或无效的原因、目前基础或临床亟待解决的难题等,从而找到研究的空白或突破口。

## 2 第二重境界——文献查准阶段

在查全的基础上进行文献阅读和主题分析,可以了解到研究涉及较多的主题概念,一一排除研究

较多的主题概念,逐步筛选出研究较少的主题概念,从而逐步锁定可能的研究主题。对可能的研究主题概念进行准确性检索,进一步了解是否可以选题。该过程是从广泛检索到准确检索的过渡,是科研人员已有的专业积累和获取的信息的有机结合,是逐步使研究目标清晰化、明朗化的创造性思维过程。

## 2.1 查准的原则

### 2.1.1 检索词尽可能具体、准确

对研究较少的概念进行准确性检索,逐步筛选研究主题。以生物碱为例,在 CBM 中使用检索结果分析功能,按主题排序筛选出排名前 50 的研究,包括阿片生物碱类、颠茄生物碱类等。采用逻辑非运算依次排除这些研究较多的主题概念,找到的文献量较少的主题可能就是目前研究的薄弱环节或空白点。表 4 显示了从大的文献集中逐步排除研究较多的概念,获得较小文献集的过程。在较小的文献集中,科研人员根据自己的学术积累,选择感兴趣的、报道较少的研究主题则较为容易,如文献较少的双氢麦角碱与海洛因依赖、青藤碱与海洛因依赖、辣椒碱与海洛因依赖等主题。直接对通过综述阅读了解到的小而具体的概念进行检索,往往会出现检索结果为零的情况,常常意味着选题的所在。

表 4 在 CBM 数据库使用逻辑非缩小命中文献集

序号	检索表达式	检索结果/篇
#1	“生物碱类”[不加权;扩展] AND “物质相关性障碍”[不加权;扩展]	4412
#2	“阿片生物碱类”[不加权;扩展] OR “东莨菪碱”[不加权;扩展]	30805
#3	“小檗碱生物碱类”[不加权;扩展] OR “乌头碱”[不加权;扩展]	9111
#4	“吸烟”[不加权;扩展] OR “酒精中毒”[不加权;扩展]	28296
#5	#1 NOT (#2 OR #3 OR #4)	439

### 2.1.2 利用副主题词限定主题的研究方向和领域

副主题词是一类限定词,用于对主题词的具体研究方向进行限定,能将同一主题内容的文献按不同研究领域进行划分,以提高主题词的专指性。用户根据需要选择一个或多个副主题词对主题词进行限定,能提高文献的查准率<sup>[2]</sup>。将表 4 检索策略#1

加副主题词限定,检索式为“物质相关性障碍/药物治疗 AND 生物碱类/治疗应用”,文献量从 4 412 篇缩减到 1 464 篇。

### 2.1.3 在准确检索的基础上达到“全”和“准”的平衡

检索结果为零,并不意味着所选主题的新颖性,也有可能是选词不当或检索字段不当造成的漏检。所以在查准时,也要考虑检索词选择的全面性,以达到“全”和“准”的平衡。用选定的某一较小概念进行检索时,也应遵循前述 3 条查全原则,尽量全面地检索出该较小领域的全部研究。如当检索范围缩小到“辣椒碱与海洛因依赖”时,要考虑辣椒碱的同义词如“辣椒素”“辣椒辣素”等,海洛因依赖的同义词如“海洛因成瘾”“海洛因滥用”“海洛因吸食”等,以及利用各自的主题词“辣椒辣素”“海洛因依赖”进行扩展检索,同时也要在多个数据库中进行检索。

## 2.2 查准在科研选题方面的意义

一个好的科研选题,必须是同时具备新颖性和可行性的课题,即未见相同文献报道,又有足够多的相关文献和背景文献可供参考和借鉴。查准的意义在于准确地筛选出可以进行研究的主题及其研究方向并提出研究假设。如通过不断排除研究较多的主题后,提出辣椒碱用于海洛因依赖戒断的假设。此时使用较准确的概念“辣椒碱”及其同义词“辣椒素”“辣椒辣素”等和“海洛因依赖”及其同义词“海洛因成瘾”“海洛因滥用”“海洛因吸食”等分别检索,检索结果有较多文献。两个检索结果逻辑与检索结果为零,即初步可以认为该假设是新颖的,并且是有一定研究基础的。

## 3 第三重境界——选题验证阶段

经过查全到查准的过程,初步明确了自己的主题或假设是否成立,是否可行,是否还需要验证,此为选题之第三重境界。选题验证的过程就是一个证明新颖性、找寻假设的合理性和科学性的过程。验证可以依据前序研究结果和相关研究结论,也可以采用文献学非相关文献知识发现法,从而达到科学选题。

### 3.1 前序研究的结果指示该假设可能成立

前序研究结果常常为后序更深入的研究提供研究基础和可能条件。如通过对筛选出的较小概念

“青藤碱”进行检索,可以了解到基础研究证实青藤碱对阿片类依赖具有防治作用且没有成瘾性,临床上尚未见青藤碱单方制剂作为戒毒药物的相关研究,提示研究者青藤碱对阿片类依赖临床干预的安全性和有效性的假设可能成立。

### 3.2 基础研究的结论作为假设的理论基础

在课题非常新颖未查到任何前序研究的情况下,分别检索相关概念,找到的基础研究的结论也可作为假设的理论基础。如上述辣椒碱用于海洛因依赖戒断的假设未查到任何文献,分别查找辣椒碱和海洛因依赖的文献,了解到辣椒碱具有消炎镇痛、心血管保护、抗癌以及消化系统保护等药理作用,而海洛因依赖者主要表现为哈欠喷嚏、流泪流涕、怕冷出汗、寒热交替、寒战震颤、恶心呕吐、厌食、腹痛、腹泻、肌肉疼痛、骨和关节痛、心悸、烦躁焦虑、怠倦乏力、失眠等症状。检索分析提示该假设逻辑关系成立,具有科学性和合理性。

### 3.3 非相关文献知识发现法验证假设的合理性和科学性

非相关文献知识发现法是由美国芝加哥大学情报学教授 Don R. Swanson 于 1985 年创立的一种科学研究方法。该方法的基本思想是两组非相关文献集 A 和 C 可以通过某中间概念或中间文献 B 建立关联,通过 A、B、C 之间的关系挖掘文献的隐含知识,形成合理的假设指导科学实践<sup>[3-4]</sup>。基于该方法需要找到中间概念 B,从而发现非相关文献的潜在联系,帮助我们提出有价值的假设。

此时可运用数据库的检索结果聚类分析功能,聚类分析是数据库将返回的检索结果进行聚类/分类等处理,使结果变得更加有条理的一种方法,方便用户查看感兴趣的主题或者根据聚类情况提出更准确的查询,缩减用户浏览的文献数量,缩短用户查询时间<sup>[5-8]</sup>。

基于前期研究已筛选出了 A、C 两个非相关文献集,如辣椒碱和海洛因戒断、双氢麦角碱与海洛因依赖、青藤碱与海洛因依赖等。以 CBM 为例,采用辣椒碱的主题词“辣椒辣素”进行主题字段检索,获得 A 文献集进行主题聚类分析,在疾病-病理状态、体征和症状-体征和症状类目下涉及的主题概念有疼痛、神经病学表现等;采用相同操作检索海洛因依

赖获得 C 文献集,主题分析在体征和症状类目下涉及的主题也有疼痛。由此推想辣椒辣素 A 和海洛因依赖 C 之间可能存在概念 B,即疼痛。表 5 显示了验证疼痛确实为概念 B 的过程,可认为辣椒辣素和海洛因依赖可能存在隐含关联,是一个有价值的假设。

表 5 在 CBM 中验证“疼痛”为概念 B

序号	检索表达式	检索结果/篇
#1	“辣椒辣素”[不加权;扩展]	829
#2	“疼痛”[不加权;扩展]	299383
#3	“海洛因依赖”[不加权;扩展]	4157
#4	#1 AND #2	120
#5	#2 AND #3	38
#6	#1 AND #3	0

## 4 小结

选题是科学研究的起点,是在有效利用文献基础上的继承和创新,对科学研究起着关键性、方向性、战略性和决定性作用<sup>[9-10]</sup>。只有全面准确地检索、分析和利用国内外文献,才能准确把握本专业领域的研究前沿动态,提出新颖的有价值的研究假设,形成新的思维指向,构成新的研究起点。利用聚类分析、战略坐标分析、社会网络分析和高频关键词分析等可视化知识图谱方法进行选题<sup>[9]</sup>受科研人员青睐,但操作相对复杂,需要信息分析人员与科研人员相互配合才能完成。本文提出的基于文献检索的医学科研选题是通过文献的查全阶段全面了解相关研究现状、查准阶段提出科研假设,并通过验证阶段检验所提出的假设的选题方法。该方法强调充分利用数据库现有功能,遵循查全、查准以及新颖性原则,是一种适合普通科研人员或刚入行的研究人员的、可以轻松把握且简单有效的选题方法。

### 【参考文献】

- [1] 李丹亚,胡铁军,诸文雁,等. 中文医学主题词表检索系统[J]. 中华医学图书馆杂志,2001(4):1-2,9.
- [2] 李纯芳,谈伟文. CBMdisc 检索词的选择及使用[J]. 医学情报工作,2004(1):20-22.

(下转第 73 页)

医学信息资源的编目工作更加专业化、集成化及智能化,更好地满足信息化时代用户个性化、多元化的需求。

#### 【参考文献】

- [1] 李欣宇. 利用图书分类编目提高图书管理效率的探索与实践[J]. 图书情报与档案管理, 2017(28): 70.
- [2] 张秀兰, 潘丹. 我国文献编目规则的历史沿革[J]. 图书馆论坛, 2016, 36(10): 32-40.
- [3] 翟晓娟. 编目的过去、现在和未来[J]. 大学图书馆学报, 2012, 30(2): 49-55.
- [4] 耿晓红, 周国华. 图书馆编目的边缘化与编目馆员的转型[J]. 大学图书馆学报, 2018, 36(1): 62-64.
- [5] 刘细文, 贾苹, 王保成. 中国科学院国家科学图书馆阵地服务的转型探索[J]. 图书情报工作, 2013, 57(18): 6-10.
- [6] 袁宝龙. 图书馆文献编目业务外包的现状审视及其发展前景预测[J]. 图书馆研究, 2016(3): 49-52.
- [7] 柴荣, 任慧玲. 医学图书馆面向公众健康的科普信息服务[J]. 中华医学图书情报杂志, 2016, 25(12): 53-56.
- [8] 张蕾. 《中国图书馆分类法(第五版)》实践中常见的若干问题及对策[J]. 大学图书馆学报, 2014, 32(6): 71-74.
- [9] 阮学平, 刘妮波. 医药学发展与医学文献分类[J]. 热带医学杂志, 2012, 12(4): 494-496.
- [10] 朱晓东, 管会君. 浅谈专科医院图书编目质量的提高[J]. 科技创业月刊, 2013, 26(4): 91-92.
- [11] 杜建, 唐小利. 精准医学的内涵演化、重点领域与我国发展对策[J]. 中国科学基金, 2016(1): 20-26.
- [12] 梅甜, 张洋, 胡珊, 等. 精准医学体系的构建及其面临的挑战[J]. 中国数字医学, 2016, 11(1): 44-48.
- [13] 顾彝. 十字路口的图书馆编目工作[J]. 国家图书馆学报, 2015, 24(6): 9-11.
- [14] 程焕文. 文献编目: 图书馆的最后专业技术领地[J]. 国家图书馆学报, 2015, 24(6): 6-9.
- [15] 葛红梅, 徐晶晶, 董鹏. 资源描述与检索(RDA)书目数据在图书馆编目中的应用[J]. 中华医学图书情报杂志, 2015, 24(2): 38-42.
- [16] 李依林, 邵雪梅. 浅谈医学实验记录规范化及其评价标准[J]. 中华医学科研管理杂志, 2015, 28(6): 468-470.
- [17] 王新娅, 张迪, 弭凤玲, 等. 某传染病科院所医学实验记录存在的问题和规范化举措探讨[J]. 中国医刊, 2017, 52(3): 104-107.
- [18] 王景侠. 21世纪图书馆资源编目体系研究[J]. 数字图书馆论坛, 2018(1): 17-23.

[收稿日期: 2018-01-20]

[本文编辑: 刘娜]

(上接第 60 页)

- [3] Swanson DR. Fish oil, Raynaud's syndrome, and undiscovered public knowledge[J]. *Perspect Biol Med*, 1986, 30(1): 7-18.
- [4] 张晗, 崔雷, 姜洋. 运用非相关文献知识发现方法挖掘科研机构潜在的合作方向[J]. 现代图书情报技术, 2006, (4): 45-48.
- [5] 刘远超, 王晓龙, 刘秉权, 等. 信息检索中的聚类分析技术[J]. 电子与信息学报, 2006, 28(4): 606-609.
- [6] Lin Y, Li W, Chen K, *et al.* A document clustering and ranking system for exploring MEDLINE citations[J]. *Journal of the American Medical Informatics Association*, 2007, 14(5): 651-661.
- [7] 王秀艳, 崔雷. 基于文本主题相似性的专题文献检索结果的聚类分析[J]. 情报学报, 2011, 30(5): 456-463.
- [8] 王志梅, 张俊林, 李秋山. Web 检索结果快速聚类方法的研究与实现[J]. 计算机工程与设计, 2004, 25(12): 2231-2233.
- [9] 李雅, 侯海燕, 朱建春, 等. 知识图谱方法在科研选题中的应用研究: 以乙酰甲喹纳米乳的研制及药效评价研究为例[J]. 图书情报工作, 2013, 57(9): 84-91.
- [10] 李雅, 侯海燕, 朱建春, 等. 新形势下医学科研选题的原则和方法[J]. 科技与出版, 2012(4): 34-36.

[收稿日期: 2018-02-01]

[本文编辑: 施沅坤]