

浅析张琪玉教授情报语言学研究方法 在其研究中的应用

曹树金¹, 罗春荣², 马翠嫦²

(1. 中山大学资讯管理学院, 广州 510006; 2. 中山大学图书馆, 广州 510275)

摘要: 对张琪玉教授在《情报语言学基础》和《情报检索语言实用教程》等论著中系统构建的情报语言学研究方法体系在其研究中的实际运用进行分析和总结, 从情报语言学研究方法应用的角度深入体会张琪玉教授的学术精神, 探讨研究方法对于理论和实践的意义及其对图书情报学研究的启示。

关键词: 张琪玉; 情报检索语言; 情报语言学

中图分类号: G350

DOI: 10.3772/j.issn.1673—2286.2014.08.006

1 引言

张琪玉教授是中国图书馆学界、情报学界最负盛誉的大师之一。上世纪八十年代初张琪玉教授出版的《情报检索语言》^[1]一书, 首次提出情报检索语言基本原理一致性的观点, 开启了张琪玉教授对各种情报语言的统一研究, 是探索情报语言影响检索效率的共同规律和有效改进检索途径的第一次实践^[2]。此后, 他以自己卓尔不群的学识建构起了具有中国特色的情报语言学理论体系, 并在众多的实践应用领域取得了可观的成就, 开辟了我国情报语言学理论研究的新天地^[3]。

张琪玉教授之所以能够取得如此辉煌的学术成就, 我们体会到的原因主要有三个, 一是他非常热爱图书馆学情报学, 几十年如一日, 不论时节, 不废分秒, 执着耕耘; 二是他目光敏锐, 能够把握学科发展的大势, 能够洞察研究对象的本质; 三是他注重研究方法, 不仅能够发现方法、构建方法, 而且切实应用合适的研究方法于自己的探索中。本文仅就张琪玉教授构建的情报语言学研究方法体系在其研究中的实际运用进行分析和总结。

张琪玉教授在《情报语言学基础》^[4]中首次提出明晰情报语言学的一般问题, 包括情报语言学的学科地位、研究范围、产生发展历史、研究对象、研究方法和相关学科等。其中, 情报语言学的研究方法, 可概括为结构功能分析法、历史演进研究法、比较研究法、调查

整理法、归纳和演绎法、原理或方法移植、理想语言设计法、现用语言改进法和其他方法等^[4-6]。情报语言学研究方法的提出不仅奠定了情报语言学研究的方法基础, 也贯穿于张琪玉教授三十多年来情报语言学的研究当中。以下回顾张琪玉教授在实际研究中对于情报语言学研究方法的具体运用, 从而加深我们对研究方法对于理论和实践意义的理解, 以期对我们今后的学术研究有所启发。

2 作为“元方法”的结构功能分析法

张琪玉教授认为, 在情报语言学所有研究方法中都含有结构功能分析的内容, 都是从结构功能分析衍生出来的^[4-6], 因而我们称之为研究方法中的“元方法”。

2.1 结构功能分析法的基本原理

张琪玉教授一开始就认识到, 提高检索效率是情报语言学研究根本目的和核心问题, 情报检索语言的检索效率是其功能决定的, 而其功能则是由其结构决定的。所以, 研究情报检索语言的性质, 主要就是要分析解剖其结构。一种情报检索语言是否能较充分地发挥其功能, 主要取决于它的结构的情况, 即: 一种情报检索语言的检索效率, 实际取决于它的宏观结构是由

哪些种微观结构构成的,以及每一种微观结构的编制质量^[2,7-8]。情报检索语言功能的完善,都是通过其结构的增加和改进来达到目的的。所以,从宏观和微观两个层次解剖分析其结构与功能的关系,弄清某种结构产生什么功能,某种功能是由哪些结构产生的,产生同一种功能的不同结构孰优孰劣,以及某种结构的适用范围和条件等,是极重要的一种研究方法^[4-6]。

而发挥情报检索语言的四项基本功能所采用的各种方法,其绝大多数可以认为是构成各种情报检索语言的微观结构。一种情报检索语言到底可采用哪些种微观结构作为它的构成因素,应依据那些微观结构是否符合特定的需要以及是否能很好地互相配合来决定。但是,由于一些微观结构具有相同的功用,并且各有长处和短处,同时,这种类型的情报检索语言所采用的某种微观结构原理,往往也可以应用于那种类型的情报检索语言,这样,就为情报检索语言的创制者留下了选择和创造的余地,留下了相当多的设计自由。情报检索语言类型的多样化,同一类型语言结构上的多样化,各种类型语言之间在方法上互相渗透、互相吸取的现象,正是说明各种情报检索语言在基本原理上的一致性和在构成方法上的差异性。情报检索语言创制者的设计思想和设计水平不同,是造成各种语言构成方法上差异的重要原因。这也说明,情报检索语言是可以自由设计,它的宏观结构和微观结构是可以不断创新,而逐步接近最佳功能的^[9]。

在此基础上,张琪玉教授针对不同的情报语言构造,提出实现不同情报语言功能的基本方法,包括:

(一)对于“对文献情报内容及外表特征标引”这一功能,可采用标识构造、标识规范化处理和标识转制度控制三种方法;(二)对于“对内容相同及相关的情报加以集中或揭示其相关性”的功能,可采用等级结构显示法、参照系统显示法、轮排聚类显示法、范畴聚类显示法、概念关系图示法、超文本显示法和类目注释及交替类目等方法;(三)对于“对大量情报加以系统化或组织化”的功能,可采用分类排列法、字顺排列法和分类与字顺结合的排列法等三种方法;(四)对于“便于将标引用语和检索用语进行相符性比较”这项功能,则可采用分类表、词表及其使用规则来加以严格控制,保证标引用语和检索用语的形式都以标准化的方法来实现,从而在结构功能分析的基础上,提出了情报语言实现的基本方法,从而为情报语言编制和改良提供了理论与实践基础^[6,10]。

2.2 结构功能分析法的主要应用

结构功能分析法是研究情报检索语言最为有效的方法。张教授在《情报检索语言》、《情报语言学基础》和《情报语言学实用教程》等专著^[4-6]中,以及在《情报检索语言大纲》^[11-12]、《论情报检索语言的研究、创制与普及》^[7]和《情报语言学的若干研究心得和收获——张琪玉学术思想自述》^[2]等研究中论述情报检索语言构造基本方法时,就是遵循结构功能分析方法的思路进行的。除了在情报语言学基础理论的应用外,对于情报语言学的一些具体理论问题也充分体现了结构功能分析法应用的具体思路:

《情报检索语言语法体系初探》一文正是采用结构功能分析法,探讨情报检索语言语法体系的构建问题的:(1)宏观结构分析方面,“情报检索语言由词汇和语法组成;词汇是指登录在类表、词表中的全部标识,一个标识(分类号、检索词、代码)就是它的一个语词,而分类表、词表则是它的词典;语法是指如何创造和运用哪些标识来正确表达文献内容和情报需求,以有效地实现情报检索的一套规则”。(2)微观结构分析方面,标识的语法套规则包括:规范造词材料的选择、造词方法、区分词类并规定各种词类的不同职能等词法问题;“句子”问题、基本造句方法—匹配、造句手段—组配符号、词序控制、句式转换、布尔逻辑造句法和加权造句法等句法问题。(3)结构—功能分析方面,如对微观结构“词法问题”中的“造词方法”进行探讨时,是以提高检索效率为目标,围绕“检索—标引用词”这一功能进行的:所谓检索词的造词方法,是指对自然语言中的词和词组(即造词材料)加以规范化处理,使之成为(改造成)符合一定要求的标引—检索用词;而所谓合格的标引—检索用词,是指那些词符合唯一性、规律性、定型性、通用性、准确性以及政治思想上的正确性等质量要求,以保证在检索中有较高的效率。

在此基础上,张教授提出从词类规范、词形规范和词义规范三个方面,对自然语言词和词组的规范化处理,从而实现以下功能:词类规范:构造具有检索意义的名词和名词性词组;词形规范:排除自然语言中的“多词—义”的现象,从而使表达某一概念的标引—检索用词只有一种词形,避免标引分散而漏检,同时也可达到使之符合通用性等要求;词义规范:排除“一词—多义”现象,从而使一个标引—检索用词只表达一个概

念,可避免误检,从而系统梳理和归纳各类情报语言的语法体系^[13]。如图1所示。

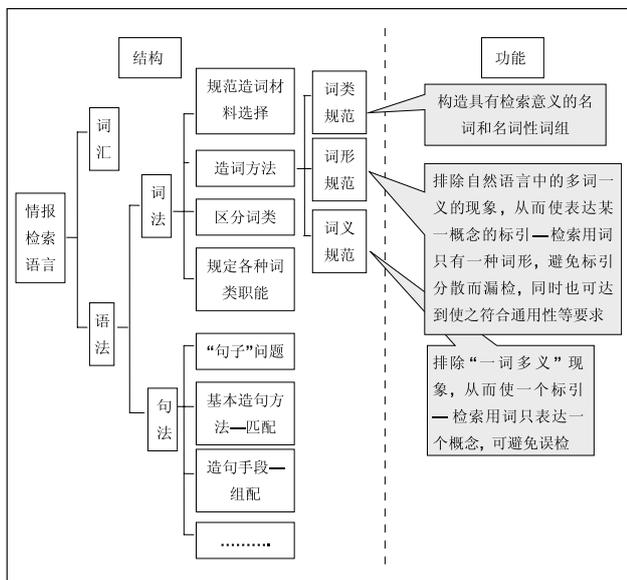


图1 《情报检索语言语法体系初探》中的结构—功能分析图解

《情报检索语言中聚类的原理和方法》则围绕情报检索语言的“聚类”功能,结合各类情报检索语言的结构体系对其原理和方法进行探讨。张琪玉教授提出,“聚类”的目的是使情报检索语言具有“对内容相同及相关的情报加以集中或揭示其相关性”这项功能。各种情报检索语言所采用的显示概念之间关系的一切方法,都是聚类方法。聚类原理,也正在于通过一定形式揭示情报检索语言所表达的文献主题之间的相同性、相似性和相关性^[14]。

《文献主题的构成因素及层次》一文则采用结构功能分析法,从文献标引和检索的角度探讨文献主题因素的构成因素及层次问题。张琪玉教授指出,文献主题就是概括文献中关于某一事物的情报内容的概念,文献标识则是文献主题的书面表达形式。绝大多数文献主题是由多个主题因素构成的,各个主题因素在文献主题中都按其描述功能处于一定的位置,从而形成一个具体性递增的层次结构。每一个主题因素在文献检索中都起着一定的作用,但其重要性又是各不相同的。因此,弄清楚文献主题的构成因素及层次,是情报检索方面的基础研究课题之一。现代情报检索语言的重要理论——分面组配理论,以及应用此理论的组配分类法和组配主题法,就是根据概念的分析与综合原

理,在对文献主题的构成因素及层次结构进行分析研究的基础上形成和创造出来的^[15]。

《汉语检索词词素轮排索引编制法探索》从“轮排索引的功用”明晰其功能,从“轮排索引的编制原则”探讨轮排索引的原则和方式,明晰轮排索引的结构和方法,从而归纳出轮排索引编制的方法和步骤^[16]。

《人-机结合的题内关键词索引可回避汉语分词难题》在明晰题内关键词索引工作原理的基础上,根据其结构特征探讨功能改进的设想和方法^[17]。

《论后控制词表》论述了后控制词表的控制机理、控制程度、编制特点、各种编制方法及其在控制上的差别,提出一种“分类词表+字顺/轮排表”的结构模式,并说明利用后控制词表检索文献的各种方法^[18]。

3 作为“通用方法”的比较研究法和归纳演绎法

如果说结构功能分析法衍生出其他研究方法,我们称之为“元方法”,那么比较研究法和归纳演绎法作为研究多类型情报语言的情报语言学研究中的常用方法,是结构功能分析法等各种研究方法实现的主要手段,我们将其称为情报语言学研究中的“通用方法”。

3.1 比较研究法

3.1.1 比较研究法的基本原理

比较研究法,即将两种或多种情报检索语言进行比较,或将产生同一种功能的不同结构、方法进行比较,看出它们的异中之同或同中之异,看出它们各自的长处和短处^[4-6]。

张琪玉教授在《情报检索语言原理的一致和方法的差异》^[9]一文中对于比较研究法的原理有较深入的阐释,他指出:由于各种类型的情报检索语言都是在寻求更佳的检索效果中创制出来的,所以,它们的基本原理是一致的;但是,它们在表达各种概念及其相互关系时和在解决对它们提出的那些共同要求时所采用的方法不同,因而形成了不同的类型和语种。这是研究情报检索语言应有的一个基本观点。以这一观点作指导,对各种类型的情报检索语言作统一研究,有助于理解情报检索语言的本质及各种情报检索语言之间的联系与区别,有助于掌握研究情报检索语言的一

种重要方法——比较研究法，就是“异中见同，同中见异”，“既见树木，也见森林”。这实际上也是张琪玉教授从结构功能分析法分化出来的比较研究法的具体体现^[6]。

在此基础上，张琪玉教授提出了情报检索语言在方法上的差异性，并以构造概念标识和显示概念关系的方法为例子研究情报语言在符号、结构、类目、系统、排列、复分等方面采用不同的方法而引起功能的差异。张琪玉教授指出，可以说没有两种原理相同的方法在性能上是完全相同的。弄清不同方法之间的差异，比之弄清不同方法之间的原理和作用上的一致性，是更为深入的研究和更加深刻的认识。可见，采用比较研究的方法研究各种情报检索语言在方法上的差异性，是对情报检索语言结构与功能关系的更深入的认识，是情报检索语言结构的改进完善和创新的重要依据^[9]。

3.1.2 比较研究法的应用

除了在《情报检索语言》、《情报语言学基础》和

《情报语言学实用教程》等专著^[4-6]中广泛采用比较研究法外，张教授对于各种专门问题的研究也大量采用比较研究的方法：

在《用〈汉语主题词表〉可编制什么样的检索工具》一文中，张琪玉教授分析传统标题目录式检索工具、比号卡组配检索式工具、关键词索引式工具和表式索引式工具等四种检索工具的检索性能时，就用到比较研究的方法，从而对采用《汉语主题词表》编制的各类检索工具的功能优势进行比较，为实践提供指南^[19]，如表1所示。

《试论隐含主题》一文通过将隐含主题与人们较为熟悉的显见主题进行比较，从而分析隐含主题的特征和分析方法，并探讨隐含主题的标引方法，并得出“隐含主题和显见主题的相对性”和“参照系统注释的完善性影响隐含主题分析”两个结论^[20]。

《题名关键词与正文关键词检索性能的差别》针对采用关键词进行题名关键词检索的索引数据库和采用关键词进行全文检索的全文数据库的检索性能差异进行比较研究^[21]。

《自动抽词与自动分词》一文在明晰自动抽词与自

表1 《用〈汉语主题词表〉可编制什么样的检索工具》中对于四种检索工具的比较研究

检索工具种类 比较项目	传统标题 目录式	比号卡组配 索引式	关键词 索引式	表式 索引式
多途径检索功能	主题词轮排时好，否则差	很好	很好	很好
自由扩检、缩检性能	不好	好，能精确检索	好，能精确检索，且方便	可以，能精确检索，但不方便
标识专指度	不高	高	高	高
检索的直接性	好	不好	稍好	稍好
检索速度	一般较快	相当快	相当快	极慢
编制检索工具所费人力	最多	较少	不专门编文献目录时少，否则较多	最少
适用范围	综合性单位，读者较多	专业单位，读者较少	专业单位，读者较多	专业单位，文献不超过一万篇，读者少
其它		可用少量人力加一套著者索引，标引错误易改	标引错误较难改	标引错误易改；索引可复制，既可作报导用，也可作检索用
总评：从充分发挥叙词法的检索效率看	差，但对读者较习惯	较好，但检索的直接性差	好，但编制检索工具所费人力较比号卡组配式稍多	尚可，但检索速度太慢

动分词内涵的基础上, 探讨二者的区别与联系, 并结合自然语言的已有理论提出提高自动抽词检索效率的设想^[22]。

《自然语言与人工语言对应转换——情报检索语言走向自动化之路》一文, 通过对19种标引模式从抽取主题概念、所用语言工具、转换方式和检索标识等方面进行比较研究^[23]。

《网络信息检索用语言的发展趋势》一文运用比较研究的方法, 对情报检索语言和自然语言的特征进行比较, 提出情报检索用语发展的大趋势是情报检索语言与自然语言从两者的初级结合到完全融合的过程。在网络信息检索中, 采用自然语言取代人工语言的语言是否会实现还有待历史证实, 在网络检索工具中人工语言与自然语言并用是现阶段较合理的模式^[24]。

3.2 归纳法和演绎法

3.2.1 归纳法和演绎法的基本原理

归纳法和演绎法是反映从个别到一般和从一般到个别的两种思路相反的方法, 因而是情报检索语言研究中理论形成与演进的常用方法, 通常一起使用。与比较法一样, 它是结构功能分析方法实现的主要手段。

归纳法, 如根据“叙词的轮排能产生多向成族和多途径检索的效果, 组配分类号的轮排能产生多向成族和多途径检索的效果, 关键词的轮排能产生多向成族和多途径检索的效果”, 而推出“标识的轮排都能产生多向成族和多途径检索的效果, 标题词的倒置加正装标题见倒置标题的参照相当于标识的轮排”, 而推出“标题词的倒置加正装标题见倒置标题的参照也能产生多向成族和多途径检索的效果”的结论^[4-6]。

演绎法如, “标识的轮排都能产生多向成族和多途径检索的效果, 标题词的倒置加正装标题见倒置标题的参照相当于标识的轮排”, 而推出“标题词的倒置加正装标题见倒置标题的参照也能产生多向成族和多途径检索的效果”的结论^[4-6]。

3.2.2 归纳法和演绎法的主要应用

(1) 归纳法与演绎法的单独使用

《情报检索语言的国家特点、时代特点和自然语言特点》一文采用归纳的方法, 从国家特点、时代特点和

自然语言特点三个方面对各种情报检索语言的表现形式、特点成因进行归纳, 最终提出“情报检索语言的国家特点、时代特点和自然语言特点是客观形成的”, “对带有国家痕迹和时代痕迹的外国的和本国过去时代的分类表、词表的苛求, 是违反历史唯物主义的”和“外国和本国过去时代的分类表、词表的确定, 都始终和读者问题联系在一起”的三个结论^[25]。

《情报检索全过程中概念与标识的对应转换》一文采用归纳法, 在明晰事物、概念和语词标识及三者关系的基础上, 提出概念与标识的对应转换是过程、是原理、是方法的主体, 并通过归纳的方法讨论概念与标识对应转换的类型, 和概念与标识对应转换的模式、方法、工具和对应转换损耗的问题^[26]。

《体系分类法的准则和惯例》采用归纳法归纳出体系分类法的准则和惯例。张琪玉教授指出, 任何一部有一定质量的体系分类法, 在其创制过程和使用过程中, 都是遵循着许多准则和惯例的。那些准则和惯例, 可以说都是从实践中总结出来的具有普遍适用性的原理和方法^[27,28]。

《分类标记原理与方法概述》采用归纳的方法, 对分类标记的功能、对分类标记的要求、标记符号、标记制度、标记方法、分类标记的类型、分类标记的某些问题等进行一一概括, 并提出对未来发展方向的猜想^[29]。

《关于自然语言检索问题》在对自然语言检索已有研究进行归纳和综述的基础上, 重点分析了全文检索所适应和不适应的检索要求及其评价, 通过对关键词检索、全文检索、搜索引擎、自动标引和分类等的分析说明了当前关于自然语言检索的主流观点的不切实际, 指出自然语言检索的发展方向是与情报检索语言结合和融合^[30]。

《文献标引中人与计算机的分工协作》在明晰半自动标引的必要性的基础上, 通过归纳的方法总结人与计算机分工协作的主要模式^[31]。

《文献标引是需要智慧的近乎艺术创造的处理过程》一文采用演绎的方法, 从优秀“标引作品”的质量标准和对标引人员的要求两方面, 对“文献标引是需要智慧的近乎艺术创造的处理过程”这一主题进行阐释, 更从标引人员的知识功底和专业视野、标引方略、标引的方法和技巧等方面对“艺术创造”的含义进行分解, 从而得出自动标引不能完全取代人工标引的结论^[32]。

《论自由标引》一文主要采用归纳和演绎法, 论述了自由标引的概念、优点及适用范围、种类、基本

方法和要点,以及自由标引系统配备后控制词表的必要性^[33]。

(2) 归纳法与演绎法的同时使用

很多时候,情报语言学的研究更多的是同时使用归纳法与演绎法的。如张琪玉教授在《人工语言与自然语言、先控制与后控制的界限在计算机系统中可淡化或取消》中,通过演绎法论述其基本原理,并通过归纳法得出论文主旨作为结论:(1)演绎1:“具有大量自然语言入口词的人工语言词表,其主要功能本来就是将自然语言词转换成人工语言词。而作为人工语言系统的自然语言接口的自然语言—人工语言对应表,其功能也是在检索时将用户所使用的自然语言检索用语转换成检索系统所使用的人工语言标引用语,以便使两者能进行相符性比较。所以,在自然语言系统中,将属于真正等同关系的那部分自然语言词通过这两种工具使其自动转换成控制词用于标引文献是完全可以的”。(2)演绎2:“后控制词表具有人工语言和自然语言的双重性质,其控制词是人工语言,其被控制词是自然语言。控制词与被控制词的关系包括等同、等级、相关三种关系,其中等同关系的词(包括可视作等同关系的词)占大部分。属于真正等同关系的那部分自然语言词,将其转换成控制词对文献进行标引也是完全可以的”。(3)加上“在计算机系统中采用人工语言与自然语言混合模式,标引用词既有控制词,也有自然语言词……有灵活性,无疑具有更多的优点”的论点。(4)在上述两个演绎和一个论点的基础上,归纳出“人工语言与自然语言、先控制与后控制的界限在计算机系统中淡化或取消,实际并没有削弱对标引、检索过程的控制,相反可提高检索率”的结论^[34]。

《体系分类法中“集中与分散”的矛盾》则采用归纳和演绎的方法,提出体系分类法中在集中与分散问题的基础上,探讨这一矛盾的成因和补救方法,并进行效果分析,最后提出进一步探讨的必要性^[35]。

《主题标引的原理和方法》的研究采用归纳法和演绎法,从文献主题分析、用主题词表达文献、主题标引过程对主题目录结构和功能的影响三个方面阐述主题标引的原理和方法^[36,37]。

《自然语言检索中各种因素对检索效率的影响》从情报语言学角度对自然语言检索效率的各种影响因素作深入研究。通过归纳和演绎的方法,分别对文本类型、检索用语的专指度、在句、段、节、篇不同范围内进行组配检索、文本用词的不规范性、不同标引方法和对

自然语言进行词表控制的程度等方面对检索效率的影响进行分析,从而归纳出可提高自然语言检索效率的有益方法^[38]。

《积极为自然语言与情报检索语言的结合创造条件——建议大量编制自然语言词表》指出,情报检索语言与自然语言的结合,是文献标引—检索用语言发展的大趋势。自然语言词表是两者结合的必要条件,在文献标引—检索用语言的进一步发展中起着关键作用。张琪玉教授综合采用归纳和演绎的方法,指出当前所需的自然语言词表的类型,并阐述当前需要的自然语言词表的类型和自然语言词表编制中的方法和技术问题^[39,40]。

《全文检索系统的检索性能》一文主要采用归纳和演绎法,分析了全文检索系统适应的检索要求和不适用的检索要求,指出一个好的全文检索系统应是一个集成系统,是全文数据库和文献目录数据库的有机结合体^[41]。

《文献题名初步研究》一文主要采用归纳和演绎法探讨了文献题名的情报价值、题名用词的规律、各类型文献题名的特点、题名对自动标引和自然语言检索的价值等问题^[42]。

《网络信息检索工具增强关键词检索功能的措施》主要采用归纳和演绎的方法对国外网络信息检索工具增强关键词检索功能的各种措施进行综述。在此基础上,张琪玉教授提出,这些措施对改善关键词检索功能起到了一定作用,值得我们结合国情进行吸取。但这些措施还不可能彻底消除关键词检索的缺陷,甚至不可能基本消除其缺陷,因而也不可能否定人工标引在现阶段必要性和合理性的观点^[43]。

(3) 归纳演绎法与其他方法的同时使用

张琪玉教授在《情报检索语言原理的一致和方法的差异》一文中,综合采用结构功能分析法、比较法、归纳和演绎方法,概括出情报检索语言原理宏观和微观上的一致性。张琪玉教授指出,宏观上各类情报检索语言在目标、本质、主题表达和功能等方面是一致的;微观上,虽然各种类型情报检索语言所采用的方法虽然不同,但其原理和作用一致。因而,弄清各种类型情报检索语言所采用的不同方法在原理和作用上的一致性,就可以抽象、概括出一般的原理。这一般的原理,高于对个别类型情报检索语言的认识,对情报检索语言的研究、创制和使用,都具有指导意义^[9]。

正是由于各种类型的情报检索语言在基本原理上

是一致的,所以在方法上就有“互相渗透”、“互相吸取”的基础,它们的结构就可以互相移植或“仿制”。情报检索语言结构的移植或“仿制”,即不同情报检索语言的组合,能够增强、完善某种情报检索语言的功能,乃至设计出新颖的情报检索语言。例如将分面分类表与叙词表的结构组合在一起,结果创造出了新颖的分面叙词表。正是基于各种情报检索语言在原理上的一致性,我们在研究情报检索语言时才有可能采用比较方法、归纳方法和演绎方法,才能使不系统的认识上升为系统的理论^[9]。

《分类语言、主题语言与自然语言一体化检索系统与〈中国财经报刊数据库〉的实践》则综合采用归纳演绎和比较的方法论述分类语言、主题语言与自然语言一体化检索系统的原理、作用、实现的核心技术、模式,并列了一个实例^[44]。

《主题标引过程对主题目录结构和功能的影响》一文通过综合采用结构功能分析法、归纳和演绎法,从检索途径、文献聚类、词串的易理解性、款目群的协调性、款目群排列的系统性、款目冗余和记录重要性等7个方面研究了主题标引过程对于主题目录结构和功能的影响^[45]。

《因特网大众分类法——一种独创的分类法》一文综合采用比较法和归纳演绎法,从因特网大众分类法的产生背景、结构、适用人群、使用环境、所采用的方法等方面分析,说明它是一种独立创造的分类法,而不是对传统分类法改进和发展的结果,因此不宜用传统分类法的准则来对它进行框范,也不宜用它的准则来框范传统分类法^[46]。

4 情报语言学研究的“一般方法”

情报语言学研究的一般方法是指情报语言学研究领域中较常使用的方法,可包括历史演进研究法、调查整理法、原理或方法移植、理想语言设计法、现用语言改进法等。

4.1 历史演进研究法

情报检索语言一直在不断进化中,所以要了解所采用的各种原理和方法孰优孰劣,可以从历史演进过程中看哪些原理和方法是如何被扬弃的^[4-6]。

在《组配及其演变》一文中,就采用了历史演进研

究法对组配问题进行研究。张琪玉教授在介绍组配的原理和作用的基础上,从体系分类法和标题法中的组配法开始到组配分类法和组配主题法,再到组配次序的规定、组配句式的转换、自由组配检索与标识单元检索方式系统、组配假象与对策、标引频率、检索试探与扩缩改检、文献数据库与多特征检索及布尔检索和加权检索等方面,通过历史演进分析的方法对组配法的应用、发展和演变进行介绍,并最终提出组配化是情报检索语言的一种发展趋势的重要结论^[47]。

4.2 调查整理法

在情报语言学领域,有不少已经使用的原理和方法,那些尚未见之于文献,或虽有文献记载但十分零散和不全面、不系统的,可采用这种研究方法了解和掌握^[4-6]。

如张琪玉教授在《论情报检索语言的研究、创制与普及》一文中就引用了张蕾关于我国情报检索语言的现状的调查,“我国在1949-1981这32年间发表的情报检索语言文献中,关于分类法的文献占73.48%,关于标题法、单元词法、叙词法、关键词法和引文索引法等文献共占12.71%。在分类法文献中,关于分面分类法的文献仅占1/50(专文不足10篇,分类表仅见1种)。由此可见,我们几乎只注意了体系分类法这一种情报检索语言类型。关于体系分类法的文献,大体可分为理论性文章(约2/5)和普及性文章(约3/5)两类。在普及性文章一类中,主要是关于各种常用分类法的一般介绍、图书分类方法、图书改编(重行分类)方法的文章。在理论性文章一类中,则主要是关于体系分类法的‘三性’(思想性、科学性、实用性)及其相互关系、部类设置、大类序列、科学分类与图书分类的关系等的讨论,以及对旧分类法的批判。关于《国际十进分类法》这种重要的情报检索语言,批判性文章居多(该法确有严重的错误和缺点,如关于台湾的类目隶属关系等),从检索效率角度对它作分析研究的文章仅有极少几篇。总之,可以说除了丁珂、刘国钧、曾世荣等的为数极少的文章、教材以及一些译文,曾论及或涉及检索效率问题外,就基本上没有较为明确地提出过‘情报检索语言的检索效率’这个概念的文献了^[48]”。在此调查和整理的基础上,张琪玉教授提出我们的情报检索语言理论研究工作必须开辟新路的观点^[7]。

《网络信息检索工具的分类体系——网络信息检索工具发展的方向与提高竞争力的途径(连载三)》主

表2 《网络信息检索工具的分类体系——网络信息检索工具发展的方向与提高竞争力的途径(连载三)》中调查整理法的使用

概念	一级类目	二级类目	专题栏目	概念	一级类目	二级类目	专题栏目
当日新闻, 新闻, 时事, 社论 热点聚焦, 社会热点, 焦点, 焦点新闻, 聚焦信息 活动与议题, 动态信息, BBS 酷站, 新站 分类广告	3 3 1	2 3 2	2 2	城市地图, 地图检索, 地图信息 时刻表	1	4	1
政治、政府、中央政府、各地 政府、管理机构、机关、行政、 政策 法律、立法、法规、司法 国防、军事、军校	28 7	11 6 9		服务, 生活服务, 公共服务, 公 众服务 人才市场, 人才服务, 人力资 源, 就业, 招聘 电话簿, 黄页检索, 白页检索, 邮政编码 天气, 气象, 天气检索, 日历 资讯, 资讯服务, 资讯业, 资 讯公司, 生活资讯	12 4 2 1 7	2 7 5 2 3	1 1
经济、商品经济、财经 工业, 纺织 农业, 林业, 牧业, 渔业 商业, 商务, 贸易, 买卖, 商场, 商店, 供求信息 房地产, 地产, 房屋资讯, 物业 法商管理, 财务	22 13 1 29 2 2	2 4 5 4 5		儿童, 儿童世界, 少年儿童	3		
金融, 银行, 投资 证券, 股票, 股市, 股市检索, 期货 保险 理财	11 3 1 1	12 6 2	2	社会, 社会生活, 社会学 人文, 人类 家庭, 交友 节日, 民俗 宗教 血型星座论命	21 13 6	2 6 4 2	
教育, 教学, 进修, 学习, 补 习, 专题教育, 升学考试 大学, 大专院校, 独立院校, 学院, 大中专院校, 专科 国外院校, 留学 中学, 小学, 幼教, 学校 教育法规, 教育查询, 校园刊物 资讯教育 科研, 研究, 成果, 发明, 专利, 项目, 科研基金	37 6	8 17 7 8 3 1 8		人物, 自传, 政治家, 科学家 作家, 艺术家, 艺人, 音乐家, 画家, 漫画家 运动家, 运动员, 运动明星 机构、组织 组织 教育机构, 科研机构, 研究所, 出版机构, 新闻机构, 体育组织 医院, 保健院所 工商机构, 公司, 企业, 贸易 机构, 服务机构, 企业集锦 社会团体, 社团, 中央团体, 同 学会	2 4 2 2 4 3 3	3 11 3 5 1 8 4 15 3	1 1
媒体, 大众传播, 大众媒体, 多媒体, 广播, 电视, 电影, 影视 报刊, 报纸, 期刊, 学刊通讯, 科技期刊, 杂志社 网络书籍杂志, 在线杂志, 新媒 体杂志, 电脑杂志 出版, 图书, 图片, 摘要, 出版社 新闻, 新闻天地, 新闻检索, 新闻组检索	21 1 3 4 19	26 16 3 4 14	1 1	国际, 国家, 区域, 省市, 地区 中国, 中国大陆, 北京, 上海, 深圳, 香港, 台湾, 美国 各地新闻, 中国大陆新闻, 香 港新闻, 台湾新闻 美国新闻, 日本新闻, 韩国新闻 世界地理和文化, 历史 外国资源, 地区性资料 个人主页 个人兴趣	4 3 1 3 1 1	3 10 4 3 5 1	
文化 艺术, 艺林文苑 文学, 小说, 散文, 幽默 美术, 绘画, 国画, 漫画, 动画, 摄影	25 28 8 1	3 4 13 14		图书馆 综合参考, 参考, 参考资料, 工 具书, 参考工具 词典, 词汇表, 缩略语, 引语	1 12	1 8 3 5	

要采用调查整理法进行研究。张琪玉教授在明晰分类体系对网络信息检索工具质量影响的基础上, 对网络信息检索工具分类体系的现状进行调查分析, 调查了国内外和港台地区共43种综合型网络信息检索工具, 最终提出网络信息检索工具设计方案的改进方式^[49], 如表2所示。

4.3 原理或方法移植

原理或方法移植, 例如, 用自由词作补充标引, 是原来的叙词法所用的方法, 这种方法可提高检索效率, 并且使用比较自由, 是一种好方法, 那么是否可以将这

种方法(或其原理)移植到其他类型的情报检索语言中去呢^[4-6]?

张琪玉教授在《字面相似聚类法辅助构造词族表、分面类表和自动标引》中借鉴字面相似聚类的原理构造词族表、分面类表和自动标引。张琪玉教授在明晰相关概念的基础上, 阐述了词和词间聚类的现象和原理: 具有相同词素的词和词组之间, 必然绝大多数在意义上有某种联系, 存在着聚类现象——字面成族现象, 即可以构成一个粗泛的词族的现象。将此原理为基础提出字面相似聚类法及其实现方法, 即: 通过字顺序列、词素模糊匹配和词素轮排。在此基础上, 进一步探讨这一方法在辅助构建词族表、分面类表和自动标引中的应用^[52]。

张琪玉教授在《学科-事物概念组配型检索语言——关于情报检索语言的遐想与求索》中提出的学科-事物组配模式,其本质属性可归纳为:(1)学科聚类系统与事物聚类系统的结合;(2)先组式语言与后组式语言的结合、体系分类法与组配分类法的结合;(3)人工语言与自然语言的结合;(4)号码标识与语词标识的结合、系统序列与字顺序列的结合;(5)不变概念代码与可变概念体系的结合。其主要实现方法可概括为“分面分析+概念代码+概念对应转换+数据库技术”。这种语言的性能可概括为:(1)分类法与主题法彻底一体化;(2)充分发挥情报检索语言对知识进行系统组织和对自然语言进行规范控制的功能;(3)用户可十分方便地进行标引和检索;(4)概念可不断增补及概念的代表词可进行更换;(5)用户区别不出是自然语言还是人工语言,而其实是由严密的人工语言控制的;(6)修订不受已标引文献所牵制,故分类体系可逐步完善;(7)可以挂接英文索引、分子式索引等以及可用于机助标引^[50]。

《概念分面组配型自动分类系统》主要是采用原理或方法移植的方法进行研究。张琪玉教授在概念分面组配理论的基础上,提出概念分面组配型自动分类系统的构想,认为概念分面组配理论具备“能较快付诸实现、较高的检索效率、适当适用人工辅助抽词、可提供多功能检索”等特征,并从质量控制措施、增强检索功能措施、检索系统功能分析等方面对概念分面组配型自动分类系统进行详细论述^[51]。

《因特网大众分类法的本质属性》一文采用归纳演绎与原理或方法移植的方法,针对当时热门的网络分类法问题,归纳出因特网大众分类法的本质属性及其具体体现,最后提出借鉴文献分类法的基本原理,按照它的本质属性来对它作某些改进则是完全可能的结论^[53]。

4.4 理想语言设计法

理想语言设计法,即设想一种功能较完善的理想情报检索语言,然后努力寻找达到这些理想标准的具体方法,进行编制—试用—优化,最后成为一个可以实际应用的结构形式^[4-6]。

张琪玉教授在《对未来分类法的憧憬》一文中提出了理想分类法的结构模式:(1)学科-事物概念组配型的分类法;(2)多聚类中心的、线性结构与网状结构

相结合的,学科和事物概念全向聚类的分类法;(3)分类体系的建立更侧重于揭示各种知识的内在联系,重点不在于构建表示某种信仰的宏观框架;(4)国际化与民族化妥善结合的分类法;(6)在一个整体框架下由众多专业分类法集成的分类法;(7)分类体系和类目可不断革新的分类法;(8)能与世界上现有主要分类法通过类目对应转换达到基本兼容的分类法;(9)伸缩性很好的分类法;(10)分类法主题法一体化;(11)有详细的、完善的自然语言入口的分类法,并可用于人机结合赋号标引;(12)多语种的分类法;(13)计算机化并可在网络上应用的分类法;(14)有充分文献保证的分类法,但不是仅限于专著文献保证。据此,张琪玉教授设想,未来分类法的编制可能是先构筑部分(专业分类法),再形成整体。事实上,未来分类法的构成原理和方法大多已存在于现有情报检索语言中,只是找出它们完善结合(实现)的方案还需要作出很大的努力^[54]。

4.5 现用语言改进法

现用语言改进法,即对一种现用的情报检索语言,如《中图法》或《汉语主题词表》进行分析,查明其存在的问题,寻求进一步完善化和合理化的方法^[4-6]。

情报检索语言的检索效率由其各种检索功能决定,而情报检索语言的各种检索功能由其结构产生。产生检索功能的是情报检索语言的微观结构。若干微观结构的有机结合才能构成一种情报检索语言。各种类型的情报检索语言都是一种结构模式,即其宏观结构。微观结构的数量、种类及结合模式,形成情报检索语言检索效率的整体水平。正是这一原理,张琪玉教授提出寻求理想情报检索语言的结构模式,从而发现新功能,去探索创新之路^[2]。

如张琪玉教授在《关于〈中图法〉增加组配成分的可能性和方法的探索》一文就采用现有语言改进法为《中图法》寻求进一步完善化和合理化的方法。张琪玉教授从体系分类法检索和排架的两种职能出发,认为体系分类法既有“鸟瞰全貌”、“触类旁通”的优点,又具有“集中与分散”矛盾的缺点,因此,在保持科学系统性的前提下尽量增加组配成分是体系分类法现代化改造的主要方向。在此基础上,张琪玉教授通过明晰组配的原理和方法,对多重列类法的功能、缺点和改进方法进行探讨,并对改进的可能性进行评估^[55]。

5 其他研究方法

除了上述方法外,张教授还提及其他研究方法,如数学方法、统计方法、实验方法等^[4-6]。

张琪玉教授撰写的《自然语言检索中各种因素对检索效率的影响》一文就用到实验和统计法。为了分析用语专指度对于检索结果质量的影响,张琪玉教授在《情报语言学文献库》中分别以不同专指度的关键词检索关于“分类标引”的文献,从而调查专指度的影响:张琪玉教授用“分”字直接在题录文本中进行匹配查找,检索结果为4647篇,用泛指词“分类”进行匹配查找,检索结果为3680篇,用专指词“分类标引”进行匹配查找,检索结果仅为54篇;而用人工语言(分面分类法的分类号,相等于检索词)进行查找,检索结果为1394篇。通过这一实验和统计,张琪玉教授说明了用语专指度对于检索结果带来的差异^[37]。

《文献题名自动抽词——分类标引系统》一文指出,从题名抽取关键词作为文献检索标识的方法在自然语言检索中具有重要意义,如果对从题名抽出的关键词进行粗略分类,也可具备一定的分类检索功能。据此,张琪玉教授设计出文献题名自动抽词——分类标引系统并报告了系统的架构、功能与主要算法。通过实验研究发现该系统虽然比不上用人工语言标引准确率高,但比对题名的模糊检索好得多,既可进行关键词的任意一致匹配,也可进行前方一致匹配和完全一致匹配;分类检索时则可进行前方一致匹配和完全一致匹配,所有的检索结果都可将相同关键词集中并排序后输出到屏幕或检索结果文件^[56]。

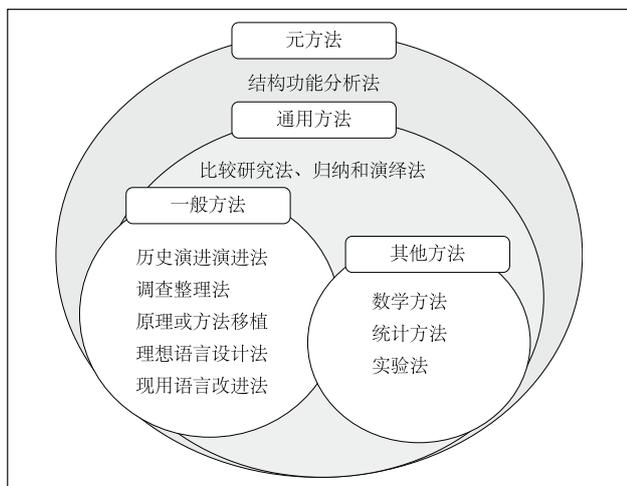


图2 张琪玉教授情报语言学研究方法论体系

通过对于情报语言学的研究方法及其在张琪玉教授研究中应用情况的回顾,不难发现:情报语言学的研究方法是科学研究中通用研究方法在情报语言学这一领域的具体运用,张琪玉教授在情报学语言理论研究的过程中,不仅通过研究方法的具体应用,取得了情报语言学领域的巨大成就,还通过自身体会和感悟,将若干通用的研究方法归纳和提炼成情报语言学的研究方法,形成情报语言学领域的特有的方法论体系,如图2。因而,张琪玉教授的理论成就、学术创新能力和治学精神都是值得我们后辈仔细体味和继承发扬的。

参考文献

- [1] 张琪玉.情报检索语言[M].武汉:武汉大学出版社,1983.
- [2] 张琪玉.情报语言学的若干研究心得和收获:张琪玉学术思想自述[J].图书情报工作,2009(20):5-9,29.
- [3] 张欣毅,张京生.走向自然语言与情报检索语言结合之路:与我国著名情报语言学家张琪玉教授的通讯访谈[J].图书馆理论与实践,2001(2):3-5.
- [4] 张琪玉.情报语言学基础[M].武汉:武汉大学出版社,1987.
- [5] 张琪玉.情报语言学基础[M].增订2版.武汉:武汉大学出版社,1997.
- [6] 张琪玉,侯汉清.情报检索语言实用教程[M].武汉:武汉大学出版社,2004.
- [7] 张琪玉.论情报检索语言的研究与创制与普及[J].图书情报知识,1983(4):10-15.
- [8] 张琪玉.提高情报检索效率是情报检索语言研究的根本目的:答黄立军同志[J].图书情报知识,1985(3):10-13.
- [9] 张琪玉.情报检索语言原理的一致和方法的差异[J].图书馆建设,1994(6):11-13.
- [10] 张琪玉.情报检索语言方法综述[J].图书情报知识,1984(2):7-13,55.
- [11] 张琪玉.情报检索语言大纲[J].图书馆学刊,1981(3):54-66,45.
- [12] 张琪玉.情报检索语言大纲:续完[J].图书馆学刊,1981(4):8-20,50.
- [13] 张琪玉.情报检索语言语法体系初探[J].图书馆理论与实践,1986(3):12-16.
- [14] 张琪玉.情报检索语言中聚类的原理和方法[J].北京图书馆馆刊,1997(1):18-25.
- [15] 张琪玉.文献主题的构成因素及层次[J].图书情报知识,1985(1):39-41.
- [16] 张琪玉.汉语检索词词素轮排索引编制法探索[J].图书与情报,1992(4):16-19.
- [17] 张琪玉.人-机结合的题内关键词索引可回避汉语分词难题[J].

- 图书馆杂志,1993(4):14-15.
- [18] 张琪玉.论后控制词表[J].图书情报工作,1994(1):1-4.
- [19] 张琪玉.用《汉语主题词表》可编制什么样的检索工具[J].图书馆学刊,1981(1):26-29.
- [20] 张琪玉.试论隐含主题[J].图书馆理论与实践,1993(2):25-27.
- [21] 张琪玉.题名关键词与正文关键词检索性能的差别[J].中国索引,2004(4):58-59.
- [22] 张琪玉.自动抽词与自动分词[J].图书馆杂志,2002(3):13-14.
- [23] 张琪玉.自然语言与人工语言对应转换:情报检索语言走向自动化之路[J].中国图书馆学报,1996(1):37-40.
- [24] 张琪玉.网络信息检索用语言的发展趋势[J].图书馆杂志,2001(3):5-7,22.
- [25] 张琪玉.情报检索语言的国家特点、时代特点和自然语言特点[J].图书馆理论与实践,1989(4):9,8.
- [26] 张琪玉.情报检索全过程中概念与标识的对应转换[J].图书与情报,2002(2):33-34.
- [27] 张琪玉.体系分类法的准则和惯例[J].晋图学刊,1992(4):3-9.
- [28] 张琪玉.体系分类法的准则和惯例(续)[J].晋图学刊,1993(1):12-17,54.
- [29] 张琪玉.分类标记原理与方法概述[J].图书馆,1993(1):17-21,49.
- [30] 张琪玉.关于自然语言检索问题[J].图书馆论坛,2004(6):211-213,145.
- [31] 张琪玉.文献标引中人与计算机的分工协作[J].中国索引,2004(4):58.
- [32] 张琪玉.文献标引是需要智慧的近乎艺术创造的处理过程[J].图书馆杂志,2004(3):24-25.
- [33] 张琪玉.论自由标引[J].图书馆学刊,1995,17(5):35-37.
- [34] 张琪玉.人工语言与自然语言 α 先控制与后控制的界限在计算机系统中可淡化或取消[J].图书馆杂志,1997(5):6-7,5.
- [35] 张琪玉.体系分类法中“集中与分散”的矛盾[J].图书馆志,1982(1):14-16.
- [36] 张琪玉.主题标引的原理和方法(一)[J].图书馆学刊,1996(1):26-28.
- [37] 张琪玉.主题标引的原理和方法(二)[J].图书馆学刊,1996(2):30-35.
- [38] 张琪玉.自然语言检索中各种因素对检索效率的影响[J].情报理论与实践,1997(5):2-4.
- [39] 张琪玉.积极为自然语言与情报检索语言的结合创造条件:建议大量编制自然语言词表(上)[J].图书馆杂志,1999(9):7-9.
- [40] 张琪玉.积极为自然语言与情报检索语言的结合创造条件:建议大量编制自然语言词表(下)[J].图书馆杂志,1999(10):8-10.
- [41] 张琪玉.全文检索系统的检索性能[J].江西图书馆学刊,2004(3):1-2.
- [42] 张琪玉.文献题名初步研究[J].江西图书馆学刊,2006(3):4-5.
- [43] 张琪玉.网络信息检索工具增强关键词检索功能的措施[J].图书馆杂志,2001(1):7-10.
- [44] 张琪玉.分类语言、主题语言与自然语言一体化检索系统与《中国财经报刊数据库》的实践[J].现代图书情报技术,2002(01):66-68.
- [45] 张琪玉.主题标引过程对主题目录结构和功能的影响[J].情报理论与实践,1994(5):29-30,44.
- [46] 张琪玉.因特网大众分类法:一种独创的分类法[J].江西图书馆学刊,2005(1):4-10.
- [47] 张琪玉.组配及其演变[J].情报理论与实践,1995(1).
- [48] 曾蕾.我国三十二年来检索语言文献的初步调查与简略分析[J].图书情报知识,1982(2):29-33.
- [49] 张琪玉.网络信息检索工具的分类体系:网络信息检索工具发展的方向与提高竞争力的途径(连载三)[J].江苏图书馆学报,2002(4):7-11.
- [50] 张琪玉.学科-事物概念组配型检索语言:关于情报检索语言的遐想与求索[J].图书馆杂志,1997(2):9-12.
- [51] 张琪玉.概念分面组配型自动分类系统[J].图书馆学刊,2002(6):9-10.
- [52] 张琪玉.字面相似聚类法辅助构造词族表、分面类表和自动标引[J].图书馆论坛,2002(5):95-96.
- [53] 张琪玉.因特网大众分类法的本质属性[J].图书馆杂志,2002(11):21-22.
- [54] 张琪玉.对未来分类法的憧憬[J].情报理论与实践,2003(1).
- [55] 张琪玉.关于《中图法》增加组配成分的可能性和方法的探索[J].国家图书馆学刊,1985(3):22-38.
- [56] 张琪玉.文献题名自动抽词:分类标引系统[J].图书馆杂志,1998(4):8-11.

作者简介

曹树金,男,博士,中山大学资讯管理学院教授,研究方向:信息组织、用户信息行为、信息检索等, E-mail: caosj@mail.sysu.edu.cn。
罗春荣,女,中山大学图书馆研究馆员,研究方向:电子馆藏发展与管理、数字图书馆建设等。

Prof. Zhang Qiyu's Information Retrieval Language: Methods and Application

CAO ShuJin¹, LUO ChunRong², MA CuiChang²

(1. School of Information Management, Sun Yat-sen University, Guangzhou 510006, China;

2. Library, Sun Yat-sen University, Guangzhou 510006, China)

Abstract: This paper analyzes the methodology and applications of professor Zhang Qiyu's information retrieval language system, deeply appreciates his academic spirit, and explores its implications for library and information science.

Keywords: Zhang Qiyu; Information retrieval language; Information linguistics

(收稿日期: 2014-08-04)

《数字图书馆论坛》2014年征稿启事

《数字图书馆论坛》是由科学技术部主管、中国科学技术信息研究所主办的专业性学术刊物(月刊),国际标准刊号ISSN: 1673-2286,国内统一刊号: CN:11-5359/G2。本刊是“中国科技核心期刊”统计源刊,是CSSCI扩展版来源期刊。

本刊是我国唯一一本以“数字图书馆”命名的刊物,一直关注国内外数字图书馆领域的相关研究和实践,设有:专家访谈、专家视点、专题研究、技术前沿、应用案例、业界动态等栏目,报道主题涵盖信息检索、数字资源、知识组织、语义技术、开放获取、用户服务等,侧重反映数字图书馆领域在资源建设、技术应用和产品服务等方面的新趋势、新发展和新变革。

本刊注重稿件的学术水准、研究内容和研究特色,来稿需要满足以下基本要求:①未发表过、未一稿多投的原创性论文;②主题鲜明、数据可靠、文字通顺、引用规范;③来稿应包含以下项目:中文和英文的标题、作者姓名、单位、摘要、和关键词,以及中图分类号、参考文献和作者联系方式;④本刊在30天内将通知作者是否录用,30天后作者可另行投稿。请登录本刊网站(<http://www.DLF.net.cn>)进行在线投稿。

本刊既厚名家、更重新人。欢迎国内外作者赐稿,一经发表即向作者寄送稿酬和样刊。本刊开放出版(网址: <http://www.DLF.net.cn>),也期待着相关专家在阅读或利用后提出宝贵意见和建议。