

学术论文问题知识元的类型与描述规则

索传军 赖海媚

摘要 学术论文是关于科学问题的论证与求解,读者通常需要花费大量的时间和精力去研读文献才能找到其所研究的问题,本文拟分析、归纳学术论文所述研究问题的类型及其描述结构。在一般的学术论文结构中,作者通常在引言部分对研究问题进行阐述,因而本文选取学术论文引言部分作为语料,采用语言学中的语步分析法对引言中表述研究问题的句子进行分析,依据问题的求解目标将研究问题的知识元划分为理解型问题知识元、解决问题知识元和探究型问题知识元三种类型,并进一步对表述问题知识元的句子的线性结构进行分析,从而归纳出不同类型问题知识元的描述规则。图2。表10。参考文献35。

关键词 学术论文 问题知识元 描述规则

分类号 G202

Types and Description Rules of Problem Knowledge Units in Academic Papers

SUO Chuanjun & LAI Haimei

ABSTRACT

The research problem of an academic article usually involves multiple related problems. The research problems themselves are different in type and structure. The research problems discussed in each paper are different, and the types and structures are also different. Because academic retrieval is still at a coarse-grained stage, and the complexity and randomness of natural language expression, users cannot quickly understand the issues that academic papers need to study.

The purpose of this study is to explore the application of knowledge unit in the content organization and retrieval of academic literature with the help of the theory of knowledge units, which has a promising application prospect, and linguistic related knowledge. This research only analyzes the problem knowledge units in academic papers, and summarizes the types, description rules and logical structures of the problem knowledge units in academic papers, with the view that the current academic retrieval method cannot meet the users' fine-grained needs and the description of the content structure of academic papers. The method restricts the machine to provide help in solving problems such as knowledge interpretation and mining.

This research mainly uses content analysis and inductive deduction. According to the IMRD structure model, the writers of the thesis mainly arrange the research questions in the introduction part. Therefore, this study selects the introduction part of the academic paper as the corpus basis to discuss the problem knowledge unit in the introduction part of the academic paper. The description of problem knowledge units

通信作者:索传军,Email:suocj@ruc.edu.cn,ORCID:0000-0002-7416-1531(Correspondence should be addressed to SUO Chuanjun,Email:suocj@ruc.edu.cn,ORCID:0000-0002-7416-1531)

in academic papers is mostly in the form of sentences. Different types of problem knowledge units have different description methods, and the structure and complexity of sentences are also quite different. Therefore, this article adopts the interpretation content analysis method to analyze the problem statement sentences, summarize the types and language description rules of problem knowledge units, and then the logic structure of the problem knowledge units is further analyzed based on the description rules.

The main conclusions of this study are as follows. First, according to the solving goal of problem knowledge units in academic papers, this study divides problem knowledge units into three types: the problem knowledge unit of comprehension, the problem knowledge unit of problem-solving, the problem knowledge unit of exploration. Among them, the problem knowledge unit of comprehension is about the identification and research of the nature and state of the research object; the problem knowledge unit of problem-solving is the study of the specific solution of a certain problem; the problem knowledge unit of exploration is the research on the internal mechanism and regularity of things. Second, problem knowledge units of various types have more general language description rules and logical structures. Third, the sentences used to express the problem knowledge units can be divided into three types: general interrogative sentences, special interrogative sentences, and declarative sentences. Most of them are declarative sentences.

This research involves a single subject area, there are few periodicals covered, the analysis results are not representative enough, and lack of verification of the description rules. The problem knowledge units' description rules and logical structure are both artificially constructed, and there is a large subjectivity. It is not known whether it is versatile in other subject areas. 2 figs. 10 tabs. 35 refs.

KEY WORDS

Academic paper. Problem knowledge unit. Describe rules.

科学问题是科学活动的逻辑起点,推动着科学研究的深入和发展。通常,用户会带着一个具体问题去查找文献,但是现有的数据库并不能按照“问题”进行检索。同时,由于自然语言表达的复杂性,用户无法快速了解学术论文所研究的问题,往往需要花费大量的时间和精力去研读文献,才能找到与自己需求相关的文献和问题。如果能够有效揭示学术论文所研究的问题,必将加快用户对相关问题和文献的发现、获取和理解。甚至,用户不仅能够获取与该问题相关的文献,还可以直接获取相关问题的知识元。然而,目前学界对学术论文所包含的科学问题和知识元的研究还很少,尚不清楚学术论文中的研究问题是如何表达的,对知识元的研究还停留在概念和结构等理论分析层面。为此,本文从语言学和问题学的视角,深入到学术论文的引言部分,利用语步分析法标注出作

者在学术论文引言部分表述的与研究问题相关的语句,找出表述研究问题的句子中比较共性的搭配形式,归纳出学术论文中问题知识元的类型和描述规则,以期为机器对问题知识元的识别、挖掘与学习奠定一定基础,从而实现用户基于“研究问题”途径对学术论文及其问题知识元的快速检索和获取。

1 科学问题与知识元相关研究

在 20 世纪末和 21 世纪初,我国图书馆学界围绕“研究对象和理论体系”进行过广泛讨论,提出了“知识说”,即以知识为研究对象、以知识管理为基础来构建图书馆学理论体系的思想。较有代表性的成果和观点包括:蒋永福在《图书馆与知识组织——从知识组织的角度理解图书馆学》中指出,图书馆是知识组织的社会组织,

图书馆知识组织的实质是为客观知识主观化提供社会保障,知识组织理论可成为图书馆学的理论基础^[1],在《客观知识与图书馆——从客观知识角度理解图书馆学》^[2]中认为,客观知识应成为图书馆学的逻辑起点。王子舟在《知识集合初论——对图书馆学研究对象的探索》^[3]和《知识集合再论——对图书馆学研究对象的阐述》^[4]中强调,图书馆学研究对象应转向知识集合;柯平在《以知识管理为基础的图书馆学》^[5]和《我国图书馆学的知识学派建设研究》^[6]两篇论文中强调,在数字时代,图书馆学应以知识管理为基础,并呼吁建设“图书馆学的知识学派”。与此同时,还有一些学者相继提出了知识元和知识单元等概念,如赵红州等在《知识单元与指数规律》^[7]一文中,从知识计量角度提出了“知识单元”的假说;马费成在《情报学的进展与深化》^[8]中提出“知识表达和组织必须从物理层次的文献单元向认知层次的知识单元或情报单元转换”;温有奎等在《知识元链接理论》^[9]中提出“知识元是一个由若干个必要的要素组成的实体,由向导信息和知识组成”。随后,图书情报学科的一些学者进一步推动图书馆学研究对象由文献到文献内容的深入和发展。近些年,随着大数据、人工智能和知识服务业的快速发展,知识挖掘、知识表示与知识问答等成为新的热点,因而知识元与问题知识元再次引发多学科学者的讨论。

1.1 问题理论相关研究

问题是一种常见的自然语言现象,对问题的内涵、分类以及结构形式等的分析,可以帮助人们更好地理解和探索问题,并进行创新。林定夷在《问题学之探究》一书中指出,“问题”是科学研究的真正灵魂,贯穿于科学研究的始终^{[10]1}。科学问题作为科学研究中的核心,对于科学研究起着引导和指示作用。在科学研究过程中,由“科学问题的抽象表达”到“研究问题的具体表述”的过程,是一个对研究问题进行分析与解决的过程。学术论文是科学研究成果的一

种主要表现形式。实质上,学术论文就是关于科学问题求解过程的记录和描述。

科学研究从问题开始,问题推动研究、指导研究;研究固然是为了解决问题,但同时也会发掘出更加深入的问题^{[10]18}。以往,问题研究主要集中在语言学领域、机器学习领域以及哲学领域。语言学着重探讨“问题”的语法、语义特征,强调“问题”在人际交往中的功能,依据中心词将“问题”分为人物、时间、地点、实体、数字、描述等几个大类。在机器学习领域,问答系统侧重分析“问题”的语言表达形式,这里的“问题”主要是以疑问句的形式存在,对“问题”进行分类也就是对疑问句的句法结构进行分类。在哲学领域,逻辑学的分支问题逻辑学探讨问题的分类、问题的逻辑本质、问题之间的逻辑关系以及问题与解答之间的相互关系。作为问题逻辑学的研究对象,“问题”最为根本的涵义在于它是一种思维形式。不同的逻辑学家对问题的分类不同,概括来说,一般可分为基本问题和复合问题。在科学哲学领域,美国著名科学哲学家L·劳丹在其1977年出版的《进步及其问题——一种新的科学增长理论》一书中提出“科学进步的问题解决模型”,强调科学研究的目的是“解决问题”,并将问题分为两大类——经验问题和概念问题^[11]。张华夏则将科学问题划分为解释问题、预言问题、检验问题、理论寻求问题和探索新理论问题五类^[12]。

此外,在图书情报领域,相关成果正在推动图书馆学研究对象由文献转向文献内容。知识工程领域通常利用计算机自然语言处理技术通过语词和句子来分析文本语义内容,如温有奎等的《基于创新点的知识元挖掘》^[13]从创新点或创新知识元挖掘等角度研究学术论文的内容。但图书情报学科重在文献或知识的描述,旨在帮助用户对文献或知识的发现与获取。因而,图书情报领域对文献内容的研究,重点不是(像知识工程领域那样)对其语义内容如何表达,而是对其研究内容和问题的揭示和描述,从而实现用户基于“问题”途径的检索,节省用户

发现、获取研究问题的时间。

近几年,人工智能领域利用知识图谱技术开展的“问答知识系统”研究成为新的研究热点。但与人工智能领域不同的是,图书情报领域长期关注的是对文献的描述。尽管已经认识到对文献内容研究的重要性,但如何分析学术论文等文献的语义内容,如何描述其研究问题还很欠缺。本质上,文献是记录知识的载体和容器,学术论文内容是关于科学问题的论述和求解,核心内容就是关于科学问题的记录、描述与回答。因而,从研究问题的揭示与描述视角探究学术论文的内容,是图书情报领域研究的一个新视角。

1.2 知识元相关研究

近些年,知识元已成为图书馆学、情报学和出版学等多学科的研究热点。从相关文献看,国内外专家学者主要对知识元的概念、类型与结构,以及知识元挖掘等内容进行了分析与研究。

(1) 知识元概念

早在20世纪70年代后期,弗拉基米尔·斯拉卡就指出,知识的控制单位将从文献深入到其中的数据、公式、事实、结论等最小的独立的“知识元”(当时他称之为“数据元”),这是有记载的关于知识元概念最早的定义^[14]。Bergocist在1987年将问题求解系统看作是一个知识元系统(A system of knowledge units),提出了建立基于知识的问题求解系统和面向类型编程(TOP)的方法,他认为知识元(Knowledge Unit, KU)是用于模拟现实世界的知识件(A piece of real-word knowledge)^[15]。Harith Alani等集成了通用句法分析模块,采用语义驱动方法,通过全文本分析提取知识元^[16]。

朱晓芸等提出“广义知识元”概念,将这个概念作为数据、模型、方法、知识的统一模式,认为广义知识元由知识元名称和一组属性组成,不能再分的知识元称为原子知识元^[17]。孙成江等认为,知识元的构成为:知识元=信息元+经验

+智慧+问题的解决^[18]。温有奎将知识分解为可独立使用的最小的单位——知识元,由名称、属性、操作和导航四个要素组成^[9]。文庭孝认为,知识元是文献中最小的、不可分割的、相对独立的一个元素^[19]。索传军认为,知识元是指语义上相对完整地表达特定知识的最小的内容单元^[20]。戎军涛等将知识元界定为在一定语境中相对独立的、具有完备知识表达的、最小粒度的语义单元^[21]。

通过对上述文献的梳理可知,国内外多数学者一致认为,知识元是知识的基本组成单位,是具有完备语义结构和独立的不能分割的最小知识单元。不过,对知识元的理解也存在明显的差异,如关于知识元的粒度认识意见不一;另外,知识元被认为是最小的知识单元,但已有研究对“最小”的程度没有明确的描述,对知识元的类型和结构也没有形成统一认识。

(2) 知识元结构

温有奎等对知识元的内容进行了定义与分类,描述并实现了针对文献资源的抽取方案,认为知识元的结构包括“类型、名称、内容”三种元素^[22]。化柏林对文献进行内容分析,归纳出学术论文中方法知识元的类型与描述规则,为后续的方法知识元抽取提供了支撑^[23]。王宇等人用“来源—导航—名称—内容—主题领域”五元组形式来表示期刊文献的知识元^[24]。刘森等针对期刊文献资源提出了一种利用六元组(编号、导航、来源、类型、特征词、内容)表示期刊文献知识元的方法,并将期刊文献的主题句作为知识元内容的表示,实现了句子级别的知识元抽取^[25]。五元组、六元组的知识元表示方式显得冗长,知识元之间关系的描述工作量大,容易导致推理不完备,同时也增加了计算机对自然语言处理的难度,使得知识元抽取的准确率下降。

1.3 问题知识元相关研究

学术问题在科学研究中占据着重要的地位,科学发现的过程就是提出问题和解决问题的过程。一方面,科学问题从相应的科学研究

活动中产生,另一方面,科学研究活动又始于科学问题。另外,在解决科学问题的同时,又会产生新的有待研究的科学问题。因此,问题的提出和解决是科学创新和发展的关键环节,而学术论文是科学问题求解方法、过程和结论的记录。因而,要对一篇学术论文进行阅读和分析,首先要明确这篇论文是针对何种学术问题进行的研究,亦即这篇论文要解决的研究问题是什么。图书情报学领域的专家学者通过引入知识元理论来分析学术论文中的研究问题,用问题知识元来表示学术论文中的研究问题。

目前,不同学者对问题知识元的定义和结构的理解有所差异。高国伟等提出了一种基于“概念、关系、问题”的知识元表示方法,该方法将知识按照属性差异分为概念知识元集合、关系知识元集合和问题知识元集合^[26],他对问题的分类借鉴了温有奎的分类方法。温有奎按照属性将问题划分为描述型问题和过程型问题,前者是指通过名词解释、数值、特征等对问题的初试状态和目标状态予以描述,后者指的是通过步骤、方法、原理、定义等对问题的初试状态和目标状态予以描述^[13]。秦春秀构建了基于知识元的科技文本内容描述框架,将问题知识元界定为在特定知识背景下提出的关于科学知识和科学实践中需要解决而尚未解决的问题,并总结出问题知识元常用的几种描述规则:“存在/关于……问题”“分析……问题”“注意……问题”“涉及……问题”“处理……问题”“解决……问题”等^[27]。虽然,这些学者对问题知识元的描述规则开展了一定程度的探讨,但对问题知识元的定义还不够科学严谨,对问题知识元描述规则的分析还不够透彻。

总之,问题知识元能够帮助人们发现研究问题之间潜在的、隐含的关系,帮助用户快速有效地了解一篇学术论文所要研究的问题。当前,国内外学者对问题、知识元和问题知识元开展了一定程度的研究,国内学者中图书情报学领域的专家学者是知识元和问题知识元研究的主力军。但是,多数学者主要对知识元

和问题知识元的概念和结构进行了分析研究,对问题知识元的语言学特征和描述规则等研究很少。

2 引言部分问题句分析与数据标注

引言又称前言或绪论,其作用是向读者初步介绍文章内容。通常,作者在撰写引言时,首先会引入主题,紧接着回顾前人研究取得的成果和提出存在的问题,最后阐述论文所要探究的问题。引言部分的结构和用词,通常具有较为鲜明的语言学特征。

2.1 引言的结构分析

编辑出版学和语言学领域的专家学者对学术论文引言部分的结构进行了分析。王小唯等从编辑与写作的角度分析了引言的结构,认为引言的结构模型可归纳为层次模型和要素模型,即将引言分成两个层次(指出论文创新性和体现论文科学性)和五个要素(总结和分析相关研究成果、找出研究中的未解问题、提出论文要解决的问题、阐明研究问题的思路、简述研究问题的方法)^[28]。从语言学角度分析学术论文的引言结构模式,还可以参考 Swales 的 CARS(Creating A Research Space)^[29] 语步/步骤模型(见表 1)。CARS 模型是对应于段落层次的研究,主要内容是在论文结构 IMRD(Introduction- Methods- Results-Discussion)下,将引言部分的内容划分成多个语步(move),再探究各个语步内的步骤(step),从而形成对引言整体结构的概括。

综上所述,论文作者对研究问题的介绍与阐述主要安排在引言部分,目的是指明当前研究未解决的问题和自己所要解决的问题。从逻辑思维的角度和写作程式看,一篇学术论文一般解决一个科研问题,作者通常需要先介绍自己所要解决的问题,再给出相对应的解决方法。这些有关“研究问题”的信息都会最先反映在引言中。故本研究对问题知识元的探讨,以引言文本作为语料基础。

表1 CARS 语步/步骤模型

Move(语步)	Step(步骤)
(1) Establishing a territory (确立研究领域)	① Claiming centrality and/or(建立中心议题,并/或)
	② Making topic generalization and/or(概括论题内容,并/或)
	③ Reviewing items of previous research(回顾前期研究成果)
(2) Establishing a niche (设置合适的研究地位)	①a Counter-claiming or(反面论证,或)
	②b Indicating a gap(指出研究差距)
	③c Question-raising or(提出问题,或)
	④d Continuing a tradition(继承前期研究传统)
(3) Occupying the niche (占据研究地位)	①a Outlining purposes or(概述研究目的,或)
	①b Announcing present research(通报当前研究状况)
	② Announcing principal findings(通报主要发现)
	③ Indicating RA structure(介绍科研论文结构)

2.2 研究数据的来源

本研究选取2018、2019年《图书情报工作》刊登在“理论研究”“工作研究”“情报研究”和“知识组织”等栏目上的论文,将引言部分提取出来作为数据来源。《图书情报工作》在2018、2019年发表在以上四个栏目的论文共计440篇。其中,48篇论文的作者没有撰写引言或在引言部分没有将研究问题表述出来,考虑到本文的研究目的,将这一部分论文从样本中剔除。最终,标注有效文本共计392篇。首先将每篇论文的引言部分抽取出来并进行编号,然后对每篇的引言内容进行阅读分析,找出表述“研究问题”的句子,形成本研究的基础数据。

2.3 研究数据的标注

(1) 研究数据的标注流程。通过梳理《中外应用语言学论文引言部分语步对比研究》^[30]和《中美博士论文摘要语步结构的比较研究》^[31]等文献发现,现有的语步标注方法大多是通过人工标注。为保证语步标注的准确性,本研究亦采用人工标注方法对引言部分进行语步标注。但语步之间、步骤之间的界限模糊,会导致语步分析具有一定的主观性。为了提高文本标

注和分析的准确度及信度,本研究采用两名标注者共同参与标注的方法,以减少主观因素的影响。

本研究参照Swales的CARS语步/步骤模式,首先分析学术论文引言部分的语步结构,以句子作为语步标注的基本单位,然后通过试验标注减少争议,最终确定引言部分的语步标注方法。在标注过程中,先对所收集的语料作语步标注,再探寻各语步内的步骤,继而标注出论文中“提出问题”的句子(见图1)。

(2) 研究数据的标注方法。首先,由两名标注者一起讨论CARS语步/步骤模型,熟悉每个语步、步骤的概念和功能;然后,通过分析样本来进一步熟悉标注步骤和方法;最后,进入语步标注验证阶段,通过标注相同的文本、对比分析结果来检验标注的一致性。凡是标注者之间存在争议之处,都要进行讨论、协商,最后达成一致。当确认标注者熟悉方法和步骤之后,分别进行独立的语步标注,然后再比较标注结果。采用Holsti的信度系数C. R. (Coefficient of Reliability)测算法来判断标注者之间分析结果的一致性。本次标注的C. R. 为0.946,高于0.75,说明信度很好。

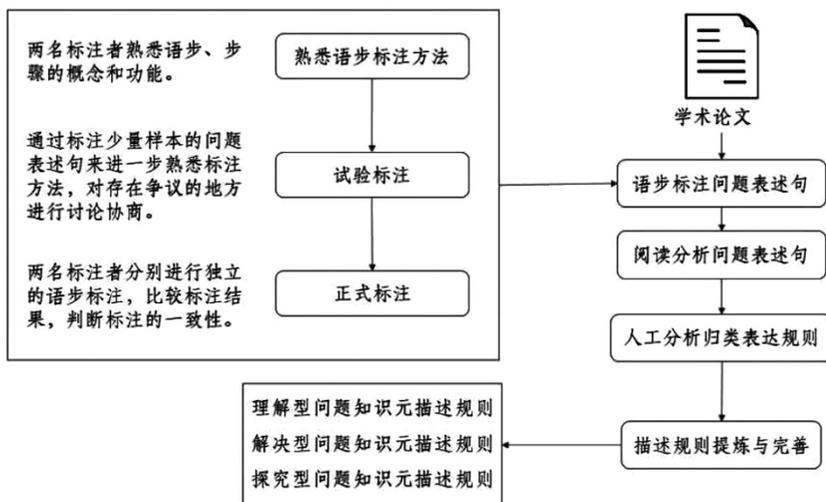


图 1 研究数据标注流程

(3)“研究问题”句子的标注数据样例。依据上述数据标注方法,标注出论文引言中“提出问题”的句子,然后剔除掉句子中的学术名词以及其他概念术语,形成学术论文“研究问题”句子的线性结构,如图 2 所示。

123号引言

作为图书馆的数字化服务形式之一,移动图书馆能够集成Web2.0环境下的相关信息应用技术[1],从而更加有利于满足个性化阅读、内容生成与分享、社交互动、自主参与等用户需求,但国内外目前针对移动图书馆社群化服务的研究成果相对较少。【本研究针对目前移动图书馆社群化服务研究的不足,从用户参与(user engagement)角度入手,试图在研究中解决如下问题:①移动图书馆社群化服务的内涵是什么?②移动图书馆社群化服务用户参与的影响因素有哪些?③这些影响因素对移动图书馆社群化服务用户参与的作用路径是怎样的?】本研究以半结构式访谈的调研方式获取了大量研究材料,通过扎根理论方法进行探索式分析,以深入分析移动图书馆社群化服务用户参与的影响因素,以及各因素之间的作用路径。本研究可帮助移动图书馆服务提供商从用户角度深入分析移动图书馆社群化服务构建的方向,为移动图书馆未来在资源组织、服务更新等方面完善自身功能提供一定的参考和指导。

123号引言中描述问题的句子

本研究针对目前移动图书馆社群化服务研究的不足,从用户参与(user engagement)角度入手,试图在研究中解决如下问题:①移动图书馆社群化服务的内涵是什么?②移动图书馆社群化服务用户参与的影响因素有哪些?③这些影响因素对移动图书馆社群化服务用户参与的作用路径是怎样的?

123号引言问题句子的线性结构

本研究……试图……解决……问题:是什么/哪些/怎样的?

图 2 表述“研究问题”的句子及句子的线性结构样例

3 学术论文问题知识元的含义与类型

关于问题知识元的概念,目前还没有统一明确的定义。本研究在以往研究(详见本文 1.3“问题知识元相关研究”部分)的基础上,对问题知识元的含义和类型作进一步的探讨。

3.1 学术论文问题知识元的含义

问题知识元的核心是“问题”。问题是以一定的语句形式揭示认识主体的知识相对于一定的认识目标来说所存在的差距、不足或矛盾的状态^[32]。研究问题是对事实的挖掘、原因的探索和过程的阐释,目的在于求真;学术论文作为一类重要的研究成果,所包含的问题知识源于

作者已有的知识体系,通过对其进行检验、批判或深化,在推动知识更新上做出具体且有原创性的贡献^[33]。

学术论文的研究问题实质上是指在一定时代和知识背景下,科学活动主体提出的需要在认识和实践加以解决而尚未解决的疑难、矛盾,具有客观性和唯一性。从形式上看,研究问题通常是用语言概括表述出来的语句(如问句等)。从知识元角度分析,学术论文中的问题知识元,可以看作是科研主体基于语言文字对研究问题的表述形式,也是研究问题的反映形式。在科学研究过程中,同一个研究问题可以用多个问题知识元进行表述。科研主体的知识背景不同,对问题的理解、看问题的角度以及解决问题的方式就会不同,对研究问题的表述及问题知识元的呈现形式也就不同。所以,相比于研究问题,问题知识元的表现形式具有一定的主观性和多样性。由此可知,问题知识元与研究问题之间是反映与被反映、多对一的关系。因而,依据笔者对知识元的认识——知识元是指语义上相对完整地表达特定知识的最小的内容单元,那么问题知识元就是指语义上相对完整

地表达特定研究问题的最小的内容单元。

概括地说,问题知识元是以问题形式存在的,用以描述研究问题的知识单元。问题知识元来源于科研工作者对研究问题的理解、表达与输出,是科研主体对研究问题的表述形式。问题知识元的描述带有明显的科研主体的主观性。问题知识元是一个过程性概念,它是发现问题、分析问题、提出问题和解决问题的统一,而且不同的问题知识元的侧重点及描述规则有所不同。问题知识元在学术论文中则表现为与所要研究的问题及相关问题有关的表述语句。

3.2 学术论文问题知识元的类型

问题知识元是研究问题的表述形式,在学术论文中表现为描述文献所要研究的问题及相关问题的语句。本文从语句层面分析问题知识元的描述规则,即分析问题表述句的语言表达规则。在对问题表述句的分析过程中,剔除句子中的学术名词以及其他概念术语,形成句子的线性结构(见表2),分析句子如何通过语调和特征词等外部形态来表达意义,从而判定学术论文问题知识元的类型。

表2 对表述研究问题句子的标注样例

-
- 09 本文利用……对……进行分析,对……进行比较分析,为……提供参考。
 10 笔者认为,有必要对……进行全面梳理,分析……,挖掘……并提出建议。
 11 因此,将……应用到……中具有实际意义,值得探究。
 12 本文力图在……背景下探讨……,……为……提供经验支持。
 13 因此,本文尝试从……视角对……进行梳理总结,提出……模式。
 14 本研究分析……,检验差异是否存在、……如何,为……提出建议。
-

通过对392篇引言中表述研究问题的句子进行标注和分析,本研究得到表述研究问题的句子的线性结构(说明:并不一定是问句),其中2018年233个,2019年159个。依据学术论文中问题知识元求解的目标,结合问题学对问题类型的划分,通过统计分析392个表述研究问题的句子的语义,本研究将学术论文问题知识元归纳为三种类型:理解型问题知识元、解决型问题知识元和探究型问题知识元。探究型问题知

识元可进一步细分为原因探究型、现状探究型、规律探究型、关系探究型和异同探究型五种类型(见表3)。

理解型问题知识元是关于研究对象的性质、状态等的识别和研究,在整个样本中占比为11.5%,从问题的内容和类型上看,对应于“是什么”的问题。解决型问题知识元是关于某一问题的具体解决方法的研究,目的是找到解决特定科学问题的具体方法,在整个样本中占比最

大(45.4%),这与图书情报学归属于应用型学科的学科性质是比较一致的,从问题的内容和类型上看,对应于“怎么办”的问题。探究型问

题知识元是关于事物内在作用机理和规律性的研究,在整个样本中占比为43.1%。从问题的内容和类型上看,对应于“为什么”的问题。

表3 学术论文问题知识元的类型

问题知识元的类型	问题知识元的子类型	问题类型	百分比(%)
理解型问题知识元	-	是什么	11.5
解决型问题知识元	-	怎么办	45.4
探究型问题知识元	原因探究型	为什么	43.1
	现状探究型		
	规律探究型		
	关系探究型		
	异同探究型		

4 学术论文问题知识元的描述规则

由上述分析可知,问题知识元可分为理解型问题知识元、解决型问题知识元和探究型问题知识元,下文对每一类表述问题知识元的句子进行分析,找出问题表述句子中比较共性的搭配形式,得出这些句子在语言表达上的规律,然后总结形成问题知识元的描述规则。从语言学角度看,学术论文问题知识元描述规则实质上就是相关特征词间具有的共性搭配规则。需要说明的是,在标注的问题知识元句子的线性结构中,有一部分句子的线性结构样本没有明显的共性或实际样本数太少(<5)。也就是说,问题知识元描述规则对应的线性结构实例数小于实际样本数,在本研究中约占总样本数的72.7%。

4.1 理解型问题知识元及其描述规则

(1)理解型问题知识元的含义。通过梳理《基于语义摘要的农业方式型问句答案的抽取》^[34]和《问答系统命名实体识别改进方法研究》^[35]发现,科学问题最一般的表述形式是问句,故本研究首先对表述问题知识元的问句形

式进行分析。疑问句分为一般疑问句和特殊疑问句,以本研究的样本为例,多数为特殊疑问句,少数为一般疑问句。这一类问题知识元可称为理解型问题知识元,包含的信息以“研究问题”为主,较少涉及“研究方法”。它的基本特征是说明目前认知状态和理想认知状态之间的差距是什么,至于如何缩小这个差距则较少涉及。作者对于这类问题通常没有可供选择的答案,通过疑问词传达出研究者对某研究领域某一对象的性质、状态、原因等的思考和解答。

(2)理解型问题知识元的描述规则。对理解型问题知识元的代表性句子进行分析,可总结出以下两种描述规则。

①第一类句子重点关注研究问题本身,只包含“研究问题”。

描述规则为:本研究\本文+试图+关注\解决\回答\探索+问题;是否\是什么\为什么\怎么样?

②第二类句子包含“研究问题”和“研究意义”两个信息项。

描述规则为:是否\是什么\为什么\怎么样?旨在\具有\促进+意义\价值。具体样例见表4。

表4 理解型问题知识元描述规则及举例

描述规则	实例数(37)	样例
本研究\本文+试图+关注\解决\回答\探索+问题:是否\是什么\为什么\怎么样?	23	本研究针对目前移动图书馆社群化服务研究的不足,从用户参与角度入手,试图在研究中解决如下问题:①移动图书馆社群化服务的内涵是什么?②……?③……?
是否\是什么\为什么\怎么样?旨在\具有\促进+意义\价值。	14	那么用户何时会使用电子健康网站?什么情况下,用户会在电子健康网站中披露信息……本研究以上述问题为切入点,对电子健康网站用户信息披露行为进行研究,旨在确定信息披露行为影响因素,进而针对性地优化电子健康网站服务,促进电子健康网站良性发展,间接助力“健康中国2030”战略目标的实现。

4.2 解决型问题知识元及其描述规则

(1)解决型问题知识元的含义。这一类问题知识元通常包含“研究问题”和“研究方法”两个信息项,主要强调问题是什么以及问题应如何解决,本质上是要解决“怎么办”,可以将这类问题称为解决型问题知识元。解决型问题知识元的起因是研究者察觉到需要探索某种方法来改变某个领域中的某些状况,目前的具体状况便是研究者面临的当前状态,而目标状态则是要填补当前状态中某个领域的空白或不足,克服缺陷,从而获得更大程度满足的理想状态。任何解决型问题知识元,其实质都是要寻找解决问题的方法。在任何解决型问题知识元中,问题指向可以是具体的,而问题的解答都是以“方法”为内容。对于当前状态与目标状态是否构成真正的问题,这里不予探究。

(2)解决型问题知识元的描述规则。对于科学研究而言,最重要的两个要素是研究问题和解决方法。为解决某个科学问题,科研工作者通常需要借鉴已有的某种方法或是给出具有针对性的解决方法。为了传播科研成果,这些关于问题和方法的信息都会反映到科学论文中,并且为了使读者更好地阅读和理解论文,作者也会在“研究问题”的阐述上表明这些信息。因此,问题和方法构成了科研工作的重要内容,会在不同程度上表现在论文中,固化为论文中的某些词汇或词汇组合。

经统计分析,在问题表述句中,名词类特征

词主要有“问题”“方法”“策略”“方案”“模型”和“模式”等,除“问题”一词属于“问题类名词”外,我们将其余的名词统称为“方法类名词”。带有“方法类名词”的问题知识元根据“研究方法”是否存在,可分为两类。

①第一类是“问题→方法”型,即针对某一问题提出相应的解决方法、策略、方案或模型。其研究目的是探究解决问题的新方法。

描述规则为:针对\为了解决+问题\不足+提出\构建+方法\模型。

②第二类是“方法→问题”型,研究者在遇到新问题时,需要借鉴或采用已有方法去解决某一问题。其研究目的是借鉴已有方法解决新问题。

描述规则为:本文\本研究\笔者+[拟\试图]+采用\运用\基于+方法\技术+探讨\探究\分析+问题。具体样例见表5。

4.3 探究型问题知识元描述规则

问题探究是扩展知识范围、获得新知识的一种重要手段。探究型问题知识元注重对某一研究问题的一个或几个方面的探索,如研究对象、研究价值等。探究型问题的解就是问题解答中可能要达到的目标状态,从而缩小或消除问题的当前状态和目标状态之间的差距。问题探究的不同对象决定了问题细分的类型。本文将探究型问题知识元分为原因探究型、现状探究型、规律探究型、关系探究型和异同探究型五种类型。

表 5 解决型问题知识元描述规则及举例

	描述规则	实例数(133)	样例
①问题→方法	针对\为了解决+问题\不足+提出\构建+方法\模型。	52	为了解决领域跨度环境下的概念漂移(特征漂移)问题,本文选择亚马逊在书籍、DVD和音乐类目下的中文评论作为实验数据,情感分析作为研究任务,提出一种面向跨领域的深度循环神经网络模型,以实现不同领域环境下的知识迁移。
②方法→问题	本文\本研究\笔者+[拟\试图]+采用\运用\基于+方法\技术+探讨\探究\分析+问题。	81	①笔者采用网络调研法对部分城市图书馆年报制度及年报编制发布情况进行调查,分析……。 ②因此,本文试图运用扎根理论方法,探究电子舞曲听众的音乐信息检索需求……

(1) 原因探究型问题知识元描述规则

原因探究型问题知识元的目的是要探求某种事实或规律背后的原因,其描述规则总结

为:本文\本研究\笔者+通过\采用\运用+方法+识别\探究\分析\挖掘+[影响]因素,具体样例见表6。

表 6 原因探究型问题知识元描述规则及举例

	描述规则	实例数	样例
原因探究型	本文\本研究\笔者+通过\采用\运用+方法+识别\探究\分析\挖掘+[影响]因素。	17	①本文通过构建公共图书馆创客空间用户使用意愿影响因素模型和问卷调查,探究公共图书馆创客空间用户使用意愿影响因素…… ②本研究拟在对部分高校图书馆相关人员进行访谈的基础上,运用扎根理论分析高校图书馆参与公共文化服务体系的影响因素……

(2) 现状探究型问题知识元描述规则

现状探究型问题知识元的目的在于对某一种情况或现象进行调研,以了解某种现状,其

描述规则为:本研究\本文\笔者+对……+展开\进行+调研\调查+[研究目标],具体样例见表7。

表 7 现状探究型问题知识元描述规则及举例

	描述规则	实例数	样例
现状探究型	本研究\本文\笔者+对……+展开\进行+调研\调查+[研究目标]。	74	①为了解和比较中美一流大学图书馆移动服务的发展情况,本研究对中国的137所“双一流”大学和美国排名前100名大学的图书馆进行调研,以期为中国大学图书馆的移动服务发展提供参考。 ②基于此本研究在文化需求精准识别视角下对不同文化水平、不同年龄段的安徽省农村居民的文献信息需求进行实证调查,以期为县级公共图书馆精准服务提供借鉴和参考。

(3) 规律探究型问题知识元描述规则

规律探究型问题知识元的对象为某类现象的

规律,研究者通过分析得出发展变化的特点,据此去揭示现象的本质规律或特征。规律探究型问题

知识元的描述规则为:本文\本研究+[方法]+探讨 \揭示+规律\作用机理,具体样例见表 8。

表 8 规律探究型问题知识元描述规则及举例

	描述规则	实例数	样例
规律探究型	本文\本研究+[方法]+探讨\揭示+规律\作用机理。	8	① 本研究 将从用户视角入手,探讨图书馆使用行为与体验对消减用户信息搜寻焦虑的 作用机理 …… ②基于非正式网络学术信息交流模式, 本文将探讨 科研用户网络学术信息搜索与利用中信息选择评价关键因素间的内在作用关系,揭示其相互影响 规律 ……

(4)关系探究型问题知识元描述规则 描述规则可总结为:本研究\本文\笔者+[方法]+揭示\探索+关系,具体样例见表 9。
关系探究型问题知识元的目的在于对两种或两种以上不同主体之间的关系进行探索,描

表 9 关系探究型问题知识元描述规则及举例

	描述规则	实例数	样例
关系探究型	本研究\本文\笔者+[方法]+揭示\探索+关系。	5	① 本文 以南京农业大学图书馆 2007—2016 年近 10 年间馆员科研产出时馆员和合作者的状况进行多样性 层次分析,探索 馆员多样性与馆员合作研究的相互关系。 ②为此, 本研究 以图书为研究对象,针对图书的网络口碑传播渠道进行类别划分……,以深入 揭示 网络口碑、传播渠道和消费者图书购买意愿三者之间的 作用关系 ,为图书出版行业相关主体提供决策借鉴。

(5)异同探究型问题知识元描述规则 特征或不同点,描述规则可总结为:本研究\本文+[方法]+揭示\探讨+差异\异同,具体样例见表 10。
异同探究型问题知识元的目的在于探讨不同主体之间的属性,通过分析归纳来发现共同

表 10 异同探究型问题知识元描述规则及举例

	描述规则	实例数	样例
异同探究型	本研究\本文+[方法]+揭示\探讨+差异\异同。	11	①因此, 本研究 基于 定量分析方法 ……,揭示其产生的成果规模和学术影响力的学科 差异,探究 …… ②因此, 本研究 基于 图书馆治理理论 对中英两国公共图书馆法进行全面系统的 分析,揭示 两国公共图书馆法的 异同 ,为我国公共图书馆法的实施与完善提供参考。

5 研究结论

本研究选取《图书情报工作》在 2018 和

2019 年的“理论研究”“工作研究”“情报研究”和“知识组织”四个栏目刊载的 392 篇具有引言的论文,利用语言学的语步分析法,对引言中所述研究问题的句子进行标注。对标注出的描述

研究问题的句子进行分析,共获得 392 个描述研究问题的句子的线性结构。然后再对这些句子的语义和线性结构进行分析,得出如下结论:第一,依据学术论文中问题知识元求解的目标,问题知识元可以分为三种类型——理解型问题知识元、解决型问题知识元和探究型问题知识元,探究型问题知识元又可以细分为原因探究型、现状探究型、规律探究型、关系探究型和异同探究型五种类型。其中,理解型问题知识元是关于研究对象的性质、状态等的识别和研究;解决型问题知识元是关于某一问题的具体解决方法的研究;探究型问题知识元是关于事物内在机理和规律性的研究。第二,各种类型的问题知识元具有较为通用的逻辑结构和描述规则。本文通过分析得到,理解型问题知识元有两种描述规则,解决型问题知识元有两种描述规则,探

究型问题知识元对应于原因探究型、现状探究型、规律探究型、关系探究型和异同探究型五种类型,具有五种描述规则。

本文关于问题知识元的类型与描述规则的研究,采用语言学的语步分析法,通过人工阅读与标注,对《图书情报工作》发表的 392 篇论文进行了分析,涉及的学科领域单一,覆盖的期刊种类少,对问题知识元描述规则的归纳可能存在一定的主观性。但是,本研究一方面为机器识别与挖掘学术论文中问题知识元的语料建设提供了一定的理论基础;另一方面,对于广大读者快速有效地了解学术论文研究问题具有一定的指导作用,总体上已经达到了预期的研究目的。此外,本文中的研究方法与其他学科领域以及其他期刊论文是否具有通用性,有待后续进一步研究和验证。

参考文献

- [1] 蒋永福. 图书馆与知识组织——从知识组织的角度理解图书馆学[J]. 中国图书馆学报,1999(5):19-23. (Jiang Yongfu. Library and knowledge organization: understanding library science from the perspective of knowledge organization[J]. Journal of Library Science in China, 1999(5):19-23.)
- [2] 蒋永福. 客观知识与图书馆——从客观知识角度理解的图书馆学[J]. 中国图书馆学报,2000(5):46-51. (Jiang Yongfu. Objective knowledge and library: understanding library science from the perspective of objective knowledge[J]. Journal of Library Science in China, 2000(5):46-51.)
- [3] 王子舟. 知识集合初论——对图书馆学研究对象的探索[J]. 中国图书馆学报,2000(4):7-12. (Wang Zizhou. On knowledge sets: exploring on objects of library science[J]. Journal of Library Science in China, 2000(4):7-12.)
- [4] 王子舟. 知识集合再论——对图书馆学研究对象的阐释[J]. 图书情报工作,2000(8):5-11. (Wang Zizhou. Again on the sets of knowledge: elucidation about the object of library science studies[J]. Library and Information Science, 2000(8):5-11.)
- [5] 柯平. 以知识管理为基础的图书馆学[J]. 郑州大学学报(哲学社会科学版),2003(1):10-15. (Ke Ping. Library science based on knowledge management[J]. Journal of Zhengzhou University(Philosophy and Social Sciences Edition), 2003(1):10-15.)
- [6] 柯平,洪秋兰. 我国图书馆学的知识学派建设研究[J]. 图书情报工作,2008(24):20-23,46. (Ke Ping, Hong Qiulan. Research on the construction of intellectual school of library science in China[J]. Library and Information Service, 2008(24):20-23,46.)
- [7] 赵红州,蒋国华. 知识单元与指数规律[J]. 科学学与科学技术管理,1985(4):39-41. (Zhao Hongzhou, Jiang Guohua. Knowledge unit and index law[J]. Science of Science and Management of S. & T, 1985(4):39-41.)
- [8] 马费成. 情报学进展与深化[J]. 情报学报,1996,15(5):338-344. (Ma Feicheng. Progress and deepening of

- information science[J]. Journal of the China Society for Scientific and Technical Information, 1996, 15 (5): 338-344.)
- [9] 温有奎, 徐国华. 知识元链接理论[J]. 情报学报, 2003, 22 (6): 665-670. (Wen Youkui, Xu Guohua. Knowledge element linking theory[J]. Journal of the China Society for Scientific and Technical Information, 2003, 22(6): 665-670.)
- [10] 林定夷. 问题学之探究[M]. 广州: 中山大学出版社, 2016. (Lin Dingyi. Exploration of problem-ology[M]. Guangzhou: Sun Yat-sen University Press, 2016.)
- [11] L·劳丹. 进步及其问题——一种新的科学增长论[M]. 刘新民, 译. 北京: 华夏出版社, 1990. (Larry Laudan. Progress and its problems: a new theory of scientific growth[M]. Liu Xinmin, trans. Beijing: Huaxia Press, 1990.)
- [12] 张华夏. 论科学问题的逻辑结构(下)[J]. 社会科学战线, 1992(2): 36-41. (Zhang Huaxia. On the logical structure of scientific problems(II) [J]. Social Science Front, 1992(2): 36-41.)
- [13] 温有奎, 温浩, 徐端颐, 等. 基于创新点的知识元挖掘[J]. 情报学报, 2005(6): 663-668. (Wen Youkui, Wen Hao, Xu Duanyi, et al. Knowledge metadata mining based on innovation points [J]. Journal of the China Society for Scientific and Technical Information, 2005(6): 663-668.)
- [14] 徐如镜. 开发知识资源, 发展知识产业, 服务知识经济[J]. 现代图书情报技术, 2002(s1): 4-6. (Xu Ru-jing. Developing knowledge resources, developing knowledge industry and serving knowledge economy[J]. New Technology of Library and Information Service, 2002(s1): 4-6.)
- [15] Bergocist J T. A type-oriented approach to knowledge-based system; acta polytechnica scandinavica [D]. Helsinki: Heisinki University of Technology, 1987.
- [16] Alani H, Sanghee Kim, Millard D E. Automatic ontology-based knowledge extraction and tailored biography generation from the web[J]. IEEE Intelligent Systems, 2003, 18(1): 14-21.
- [17] 朱晓云, 陈奇, 杨彬, 等. 决策支持系统中的广义知识元及模型库[C]//《控制与决策》编辑部. 中国控制与决策学术年会论文集: 1993年卷. 沈阳: 东北大学出版社, 1993: 33-36. (Zhu Xiaoyun, Chen Qi, Yang Cheng, et al. General knowledge unit and model library of decision support system[C]//Editorial Department of *Control and Decision*. Proceedings of the Annual Conference on Control and Decision in China; vol 1993. Shenyang: Northeastern University Press, 1993: 33-36.)
- [18] 孙成江, 吴正荆. 知识服务战略: 创建增值联盟[J]. 情报科学, 2002, 20(10): 1028-1030. (Sun Chengjiang, Wu Zhengjing. Knowledge service strategy: creating value-added alliance [J]. Information Science, 2002, 20(10): 1028-1030.)
- [19] 文庭孝. 知识单元的演变及其评价研究[J]. 图书情报工作, 2007(10): 72-76. (Wen Tingxiao. Research on the evolution of knowledge unit and its evaluation[J]. Library and Information Service, 2007(10): 72-76.)
- [20] 索传军. 知识转移视角下的学术论文老化与创新研究[J]. 图书情报工作, 2014, 58(5): 5-12. (Suo Chuanjun. Study on obsolescence and innovation of academic papers from perspective of knowledge transfer[J]. Library and Information Service, 2014, 58(5): 5-12.)
- [21] 戎军涛. 学术论文内容知识元语义描述模型研究[J]. 情报科学, 2019, 37(7): 30-35. (Rong Juntao. Research on the semantics description model of knowledge unit in academic papers[J]. Information Science, 2019, 37(7): 30-35.)
- [22] 温有奎, 温浩, 徐端颐, 等. 基于知识元的文本知识标引[J]. 情报学报, 2006, 25(3): 282-288. (Wen Youkui, Wen Hao, Xu Duanyi, et al. Text knowledge indexing based on knowledge unit[J]. Journal of the China Society for Scientific and Technical Information, 2006, 25(3): 282-288.)
- [23] 化柏林. 学术论文中方法知识元的类型与描述规则研究[J]. 中国图书馆学报, 2016, 42(221): 30-41.

- (Hua Bolin. Types and description rules of knowledge elements about method in academic papers[J]. Journal of Library Science in China, 2016, 42(221): 30-41.)
- [24] 王宇, 李秀秀. 基于知识元的期刊文献知识组织研究[C]//2017第二届社会科学与人文国际会议论文集, 2017:9-14. (Wang Yu, Li Xiuxiu. Research on periodical literature knowledge organization based on knowledge element[C]//Proceeding of 2017 International Conference on Social Sciences and Humanities. 2017:9-14.)
- [25] 刘淼, 王宇. 基于主题句的期刊文献知识元库构建[J]. 情报杂志, 2012(11): 145-149. (Liu Miao, Wang Yu. Building of periodical literature knowledge element database based on topic sentence[J]. Journal of Intelligence, 2012(11): 145-149.)
- [26] 高国伟, 王亚杰, 李永先. 知识元表示方法研究[J]. 现代情报, 2015, 35(3): 15-18, 33. (Gao Guowei, Wang Yajie, Li Yongxian. Research on the method of knowledge element representation[J]. Journal of Modern Information, 2015, 35(3): 15-18, 33.)
- [27] 秦春秀, 刘杰, 刘怀亮, 等. 基于知识元的科技文本内容描述框架研究[J]. 图书情报工作, 2017, 61(10): 116-124. (Qin Chunxiu, Liu Jie, Liu Huailiang, et al. Research on content description framework of science and technology text based on knowledge unit[J]. Library and Information Service, 2017, 61(10): 116-124.)
- [28] 王小唯, 吕雪梅, 杨波, 等. 学术论文引言的结构模型化研究[J]. 编辑学报, 2003, 15(4): 247-248. (Wang Xiaowei, Lü Xuemei, Yang Bo, et al. Research on structural modeling of academic paper introduction[J]. Acta Editologica, 2003, 15(4): 247-248.)
- [29] 于晖. 功能语篇体裁分析理论与实践[M]. 北京: 外语教学与研究出版社, 2018: 22-23. (Yu Hui. Theory and practice of genre analysis of functional discourse[M]. Beijing: Foreign Language Teaching and Research Press, 2018: 22-23.)
- [30] 钱丹凤, 潘璠. 中外应用语言学论文引言部分语步对比研究[J]. 外语教育(2014年年刊), 2015: 210-218. (Qian Danfeng, Pan Fan. A comparative study of introductions in Chinese and foreign applied linguistics[J]. Foreign Language Education (2014 Annual Review), 2015: 210-218.)
- [31] 杨思悦. 中美博士论文摘要语步结构的比较研究[D]. 临汾: 山西师范大学, 2018. (Yang Siyue. A comparative study of moves in English abstracts of doctoral dissertations of Chinese and American universities[D]. Linfen: Shanxi Normal University, 2018.)
- [32] 张华夏. 论科学问题的逻辑结构(上)[J]. 社会科学战线, 1992(1): 54-61. (Zhang Huaxia. On the logical structure of scientific problems(I)[J]. Social Science Front, 1992(1): 54-61.)
- [33] 刘春燕. 学术问题的基本特征: 分析转基因态度研究中的提问方式[J]. 华中科技大学学报(社会科学版), 2016, 30(5): 107-115. (Liu Chunyan. Fundamental characteristics of research questions: analysis of raising question on attitudes toward GMOs researches[J]. Journal of Huazhong University of Science and Technology (Social Science Edition), 2016, 30(5): 107-115.)
- [34] 张军亮, 朱鹏. 基于语义摘要的农业方式型问句答案[J]. 情报学报, 2014, 33(1): 77-83. (Zhang Junliang, Zhu Peng. Answers extraction of agricultural how-to questions based on semantic summary[J]. Journal of the China Society for Scientific and Technical Information, 2014, 33(1): 77-83.)
- [35] 鲍静益, 于佳卉, 徐宁, 等. 问答系统命名实体识别改进方法研究[J]. 数据采集与处理, 2020, 35(5): 930-941. (Bao Jingyi, Yu Jiahui, Xu Ning, et al. Research on improved method of named entity recognition in question answering system[J]. Journal of Data Acquisition and Processing, 2020, 35(5): 930-941.)

索传军 中国人民大学信息资源管理学院教授, 博士生导师。北京 100872。

赖海媚 中国人民大学信息资源管理学院硕士研究生。北京 100872。

(收稿日期: 2020-05-22; 修回日期: 2020-08-15)