

● 常霞, 魏绪秋, 张以迪, 徐雨薇 (山东理工大学信息管理学院, 山东 淄博 255049)

## 基于知识单元属性特征的学术论文创新性评价研究\*

**摘要:** [目的/意义] 创新是科学研究的本质。科学评价学术论文创新性有助于营造良好的科研氛围, 正确引导科研方向, 推动学术进步。[方法/过程] 基于知识单元发展属性特征与突变属性特征, 综合考虑学术论文中知识单元的动态发展过程, 构建知识单元新颖性评价指标与学术论文创新性综合评价模型, 并以 2023 年度“数字人文”领域中高被引学术论文为研究对象进行实证研究。[结果/结论] 所构建模型能够较好地识别出当前研究发展的热点与前沿性知识单元; 基于知识单元属性视角筛选出的高创新性学术论文符合当前研究的热点, 对今后相关研究具有借鉴意义。[局限] 未对知识单元进行具体分类, 如问题创新或方法创新等。在未来的研究中, 将进一步细化知识单元的种类, 探究不同种类知识单元组合的特征。

**关键词:** 学术评价; 创新性测度; 知识单元属性; 发展特征; 突变特征; 数字人文

**DOI:** 10.16353/j.cnki.1000-7490.2024.11.008

**引用格式:** 常霞, 魏绪秋, 张以迪, 徐雨薇. 基于知识单元属性特征的学术论文创新性评价研究 [J]. 情报理论与实践, 2024, 47 (11): 71-80.

### Research on Innovative Evaluation of Academic Papers Based on Knowledge Unit Attributes

Chang Xia, Wei Xuqiu, Zhang Yidi, Xu Yuwei

(School of Information Management, Shandong University of Technology, Shandong Zibo 255049)

**Abstract:** [Purpose/significance] Innovation is the core of scientific research. The scientific evaluation of academic papers plays a crucial role in fostering a conducive environment for scientific research, guiding accurate research directions, and promoting academic advancement. [Method/process] This study comprehensively investigates the dynamic development process of knowledge units in academic papers based on two dimensions: developmental attribute characteristics and mutation attribute characteristics. It further constructs an evaluation index for the novelty of knowledge units and an innovative comprehensive evaluation model for academic papers. Empirical research is conducted on highly cited academic papers in the field of “digital humanities” in 2023. [Result/conclusion] The findings demonstrate that this model effectively identifies current research hotspots and cutting-edge knowledge units. The selection of highly innovative academic papers based on knowledge unit attributes aligns with existing research trends and holds significant implications for future related studies. [Limitations] This study does not specifically categorize knowledge units into problem innovation or method innovation, among others. Future studies will refine the classification of knowledge units to explore the distinctive characteristics associated with different types of combinations.

**Keywords:** academic evaluation; innovative measurement; knowledge unit attribute; development characteristics; mutation characteristics; digital humanities

## 0 引言

党的二十大报告指出, “创新是国家发展的第一动力, 必须坚持创新在我国现代化建设全局中的核心地位”。早在 2016 年, 我国就明确提出建设创新型国家三步走战略目标, 即到 2020 年进入创新型国家行列, 到 2035

年跻身创新型国家前列, 到中华人民共和国成立 100 年时成为世界科技强国<sup>[1]</sup>。学术创新作为科学研究的结晶, 是创新的重要组成部分, 亦是驱动创新型国家建设战略实施的动力来源<sup>[2]</sup>。以国家战略需求为导向, 加强基础研究、突出原创, 对驱动国家创新发展具有重要意义。

学术论文具有多元价值的特点。随着科学技术的进步, 学术成果数量和丰富度不断增加, 科研人员从海量文献中发现创新性学术论文难度增大。2021 年 7 月, 国务院办公厅发布《关于完善科技成果评价机制的指导意

\* 本文为国家社会科学基金青年项目“基于文本挖掘的学术论文创新性识别研究”的成果, 项目编号: 21CTQ022。

见》,强调要以破除“唯论文”和“SCI至上”为突破口,科学确定评价标准,坚持开展多层次差别化评价<sup>[3]</sup>。以同行评议为代表的学术评价,是最能接近评价对象真实状态的方法,但不断增长的科研文献,使得现有评审系统难以满足当前评审需求<sup>[4]</sup>。因而,基于文献计量指标与语义内容的学术论文创新性评价理论与方法不断发展完善。但以文献计量指标为主的评价方法过度强调数量、被引频次等外在指标,忽视了评价的价值性<sup>[5]</sup>。以文本挖掘技术为背景,新颖评价由外在指标转向语义内容评价,但面对学术论文复杂的知识结构和语义信息,其实现过程难度较大<sup>[6]</sup>。

基于此,一方面,从主题内容视角出发,抽取学术论文中具有代表性的知识单元,基于知识单元发展属性特征,对学术研究的核心知识进行识别与评估;另一方面,构建知识单元共现网络,基于其在网络中的属性,反映知识单元突变属性特征。综合知识单元发展属性特征与突变属性特征评价学术论文创新性。

## 1 相关研究

从学术论文内在属性来看,学术论文评价的本质体现为学术论文创新性的测度<sup>[7]</sup>。当前主要针对学术论文的内部和外部特征,采用定量、定性的方法和科学计量、数据挖掘、深度学习等技术,开展学术论文创新性评价的相关研究。

### 1.1 基于引证数据的学术论文创新性评价

颠覆性指数是近几年被提出的可以直接测度论文颠覆性创新程度的计量指标,基于网络中论文节点的深层引用关系,高创新性的论文往往会对既有研究领域学科范式产生颠覆性影响<sup>[8]</sup>。颠覆性指数被广泛应用于评估学术论文创新性。Wu等<sup>[9]</sup>基于论文局域引用结构对以往知识的替代作用等来衡量论文的颠覆性。Jiang等<sup>[10]</sup>和C. Meyer等<sup>[11]</sup>分别利用颠覆性指数识别了能源安全和合成生物学的创新性论文。在引文网络、引文内容、引证指标等三个方面,宋歌<sup>[12]</sup>基于文献在引文网络中的产生和扩散,提出了“S指数”来评价科研成果的创新力;魏绪秋等<sup>[13]</sup>从引证意图、引证内容、引证位置等视角构建模型来评价学术论文创新性;杨思洛等<sup>[14]</sup>从被引次数、同被引指标等角度来衡量专利技术的价值。

### 1.2 基于文本内容的学术论文创新性评价

基于文本内容展开的学术论文创新性评价,主要包括以下三个方面。①词频统计方法是利用知识单元累计出现次数、交叉出现频次和词频差异度等来评估知识单元创新性或新颖性。Yang等<sup>[15]</sup>利用知识单元的相对累计频率和累计频率的相对增长率来分别构建知识单元增长指数和新

颖性指标。②主题探测追踪法是针对目前研究领域较为前沿的热点建立模型,深入文章主题内容,强调关键词、文章主题等具有语义信息的新颖性评价<sup>[6]</sup>。曹树金等<sup>[16]</sup>、K. Matsumoto等<sup>[17]</sup>分别使用LDA主题模型、SVM机器学习算法和有序逻辑、最小二乘回归模型,基于主题演化视角,提出了一种量化焦点论文主题与同领域论文间相似度的新颖性指标,动态评价学术论文创新性。③语义相似度方法通常基于相似性度量理论,通过计算词语与概念间语义距离来衡量相似度。汪雪峰等<sup>[18]</sup>将非结构化文本内容转化为向量空间模型,运用余弦相似度测度文本内容相似度。侯剑华等<sup>[19]</sup>提出了一种基于SAO-ADV模型及语义相似度的论文创新性评价方法。

在共词网络背景下,主题探测建立在连接模式上,新颖独特的知识嵌入成为学科新型主题的发现逻辑<sup>[20]</sup>。基于网络视角识别创新性学术论文引起了学者广泛关注。潘现伟等<sup>[21]</sup>、S. Shibayama等<sup>[22]</sup>等利用复杂网络分析法,分别从文献相似性网络与引文网络视角构建了论文重要性评价模型。R. K. Amplayo等<sup>[23]</sup>基于网络特征提取方法,构建了一个自动编码器神经网络,用以检测论文新颖性。当前研究主要以学术论文为网络节点,探究学术论文在网络中的局部特征或整体特征,较少考虑学术论文本身内容信息。本文提取学术论文中知识单元,以知识单元属性特征为研究对象,构建知识单元共现网络,并基于知识单元发展属性特征与突变属性特征测度学术论文创新性。

## 2 基于知识单元属性特征的指标构建

从知识进化理论视角探究知识的发展过程。英国哲学家卡尔·波普尔认为知识的发展过程类似于达尔文的自然选择过程,在知识发展过程中,符合当前科学价值的知识得以保存、分享和传承;反之,则会被淘汰。知识进化理论提出知识系统进化的两种方式:知识继承性,使重要性知识在时序下能稳定存在;来自外部因素对于知识系统的作用,使系统内的知识产生突变<sup>[24]</sup>。

知识单元属性具有多样性、关联性和动态性等特点。随着时间变化,有些知识单元是稳定存在并不断发展的,而有些知识单元可能随时间而衰退。知识的结构属性是知识所具有的层次和组织结构相关的特征:知识可以由不同的组成单元构成,知识之间的连接方式和强度可以影响知识的获取、交流和共享。

共词分析是常用的知识网络工具。共词分析的词源包括知识单元、主题词、标题摘要提取词。知识单元之间的共现强度可以揭示知识关联关系及知识结构<sup>[24]</sup>。有学者将共词分析法应用于学术论文新颖性的评价中,以文献中知识单元、主题词的共现频次等相关属性,评价学术论文

相关性。沈阳<sup>[25]</sup>通过不同时期知识单元频度及检索历史中知识单元出现频次与时间等其他参数评价知识单元创新度。

## 2.1 “发展属性特征”测度指标

创新性论文通常会基于已有研究,发展和推进现有理论与观点。知识单元的发展属性特征体现了学者对最新研究的关注和对已有知识的理解、推进以及对实际应用的关注。鉴于此,本文从知识单元发展规模和知识单元关联强度两个维度构建了知识单元发展属性特征模型,如图1所示。

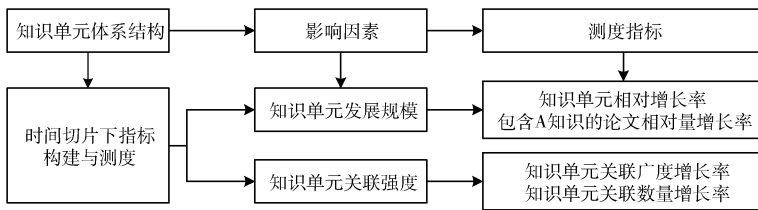


图1 知识单元发展属性特征模型

Fig. 1 Model for characterizing the developmental attributes of knowledge units

1) 知识单元发展规模。知识单元发展规模反映了其在该领域的相对重要程度,知识单元规模的差异性会随时间变化而产生较大的势差,对推动知识传播、利用与增长具有重要意义。F. Pest 等<sup>[26]</sup>采用生存分析方法对知识单元演化过程进行详细分析,发现通过测量知识单元的出现和消失可以阐明研究课题演化的一些相关方面。

本文以知识单元相对增长率和包含该知识单元论文相对发文量的增长率作为知识单元发展规模的测度指标,见公式(1)~公式(3)。具体来说,在新兴研究领域或研究主题中可以观察到更高的科学生产增长率,通过观察知识单元的增长可以衡量研究课题的变化。

$$v_i = \frac{\alpha_t - \alpha_{t-1}}{\alpha_{t-1}} \quad (1)$$

式中, $v_i$ 表示某一时间段知识单元*i*的相对增长率; $\alpha_t$ 表示在*t*时间,知识单元*i*出现的频次; $\alpha_{t-1}$ 表示在*t-1*时间知识单元*i*出现的频次。

$$f_i = \frac{\beta_t - \beta_{t-1}}{\beta_{t-1}} \quad (2)$$

式中, $f_i$ 表示某一时间段包含知识单元*i*的论文相对发文量的增长率; $\beta_t$ 表示在*t*时间包含知识单元*i*的论文相对发文量; $\beta_{t-1}$ 表示在*t-1*时间包含知识单元*i*的论文相对发文量。

$$O_{pi} = \sum (v_i + f_i) \quad (3)$$

式中, $O_{pi}$ 表示知识单元*i*的发展规模。

2) 知识单元关联强度。新颖性可以视为已有知识单元与新知识单元的关联与关联程度。新知识的产生是对原

有知识体系内容的优化和补充,因而新知识与原有知识之间在内容属性上存在较强的关联关系<sup>[24]</sup>。节点度作为复杂网络中节点个体特征指标,能够在一定程度上衡量节点个体在整个网络中的地位,在加权网络中扩展为节点强度<sup>[27]</sup>。

本文在知识单元关联度时序变化背景下,以知识单元关联广度与知识单元关联数量两个维度作为知识单元关联强度的测度指标,见公式(4)~公式(6)。知识单元关联广度基于其在知识网络中的关联属性特征,即点度中心性来表示,如果一个知识单元与其他较多知识单元之间存在

在直接联系,那该知识单元就居于中心地位,在整个知识网络中具有较大影响力。知识单元关联数量基于其在加权知识网络中的关联属性特征,进一步考虑了知识单元连接相应边的权值,能够更加科学地衡量知识单元的局部特征。

$$c_i = \frac{m_t - m_{t-1}}{m_{t-1}} \quad (4)$$

式中, $c_i$ 表示知识单元*i*的关联广度增长率; $m_t$ 表示在*t*时间知识单元*i*的点度中心性; $m_{t-1}$ 表示在*t-1*时间知识单元*i*的点度中心性。

$$s_i = \frac{n_t - n_{t-1}}{n_{t-1}} \quad (5)$$

式中, $s_i$ 表示知识单元*i*的关联数量增长率; $n_t$ 表示在*t*时间知识单元*i*与其他知识单元直接相连的加权点度中心性; $n_{t-1}$ 表示在*t-1*时间知识单元*i*与其他知识单元直接相连的加权点度中心性。

$$O_{qi} = \sum (c_i + s_i) \quad (6)$$

式中, $O_{qi}$ 表示知识单元*i*的关联强度。

CRITIC法是由Diakoulaki提出的一种客观权重赋权法,综合考虑了指标对比强度和冲突性,适用于多对象、多指标的综合评价方式<sup>[28]</sup>。其算法步骤如下。

第一步 对数据标准化处理:

$$x'_{ij} = \frac{x_j - x_{\min}}{x_{\max} - x_{\min}} \quad (7)$$

第二步 以标准差的形式展现指标变异性。假设研究对象数量为*n*,则评价指标变异性系数计算公式为:

$$\begin{cases} \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x'_{ij} \\ Z_j = \sqrt{\frac{\sum_{i=1}^n (x'_{ij} - \bar{x}_j)^2}{n-1}} \end{cases} \quad (8)$$

第三步 计算指标之间冲突性。使用相关系数来表示指标间的相关性:

$$T_j = \sum_{i=1}^j (1 - t_{ij}) \quad (9)$$



第四步 计算权重系数:

$$W_j = \frac{Z_j \times T_j}{\sum_{j=1}^4 Z_j \times T_j} \quad (10)$$

基于 CRITIC 权重法确定  $v_i$ 、 $f_i$ 、 $c_i$ 、 $s_i$  的权重, 分别为  $l_1$ 、 $l_2$ 、 $l_3$ 、 $l_4$ 。综合知识发展规模与知识单元关联强度, 构建发展特征测度指标:

$$O_i = \frac{\sum (O_{pi} + O_{qi})}{k} = \frac{\sum (v'_i l_1 + f'_i l_2 + c'_i l_3 + s'_i l_4)}{k} \quad (11)$$

式中,  $O_i$  表示文章知识单元发展属性特征;  $O_{pi}$  表示知识单元发展规模;  $O_{qi}$  表示知识单元关联强度;  $v'_i$  表示标准化后的指标, 其他同理;  $k$  表示待评价学术论文知识单元数量, 若计算单个知识单元发展属性特征, 则  $k = 1$ 。

## 2.2 “突变属性特征”测度指标

突破性特征主要面向两个方面: 一方面, 新知识的产生使得原有知识系统发生改变, 在内容上表现为产生新的内容, 象征着知识具有一定的“新颖性”; 另一方面, 新知识的增长特征是突破性创新的重要表现特征, 即新知识的扩散速度越快, 表明主体对于新知识的接受程度越大, 越能够被主体吸收并利用, 产生突破性创新的潜力越大。知识的新颖性越高, 产生突破性创新的概率越大<sup>[29]</sup>。Yu 等<sup>[30]</sup>认为随着研究领域的发展, 新知识单元出现程度会迅速上升。因此, 本文围绕上述两个方面从知识单元突变种类与知识单元突变强度两个维度构建知识单元突变属性特征测度指标。同时, 为了减轻时间因素干扰, 本文将研究过程划分为 4 个时间段, 并赋予相应的数据表示。

1) 知识单元突变种类。在共词网络的基础上, 不同时间窗口内, 知识单元的节点中心性及其关联的知识单元内容会发生一定变化, 如图 2 所示。本文结合知识单元节点中心性及内容两个方面, 将数据按照时间顺序进行划分, 具体展开如下:

$$N_{pi} = \sum \frac{1 - \frac{D_{i\pi}}{D_{ic}}}{t_x - t} \quad (12)$$

式中,  $\frac{D_{i\pi}}{D_{ic}}$  表示在  $t$  时间段已有知识单元在知识单元  $i$  所关联知识单元中的占比;  $1 - \frac{D_{i\pi}}{D_{ic}}$  表示在  $t$  时间段新出现知识单元占比;  $t_x$  为固定值, 本文将其设置为 4,  $t \in \{0, 1, 2, 3\}$  (2017—2018 年时间设置为 0, 其他时间段依次类推)。

2) 知识单元突变强度。亦称为新知识的涌现程度。新知识产生之后扩散程度有所差异, 知识突变程度有所不同, 如图 3 所示。因而, 本文结合知识单元加权节点中心性及内容两个方面具体展开:

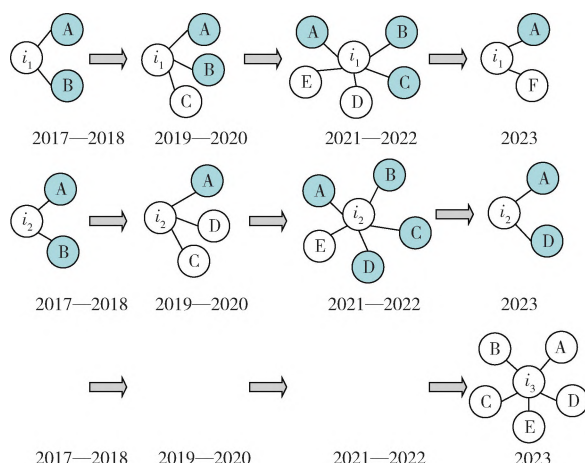


图2 知识单元关联种类发展变化

Fig. 2 Development changes in knowledge units linkage categories

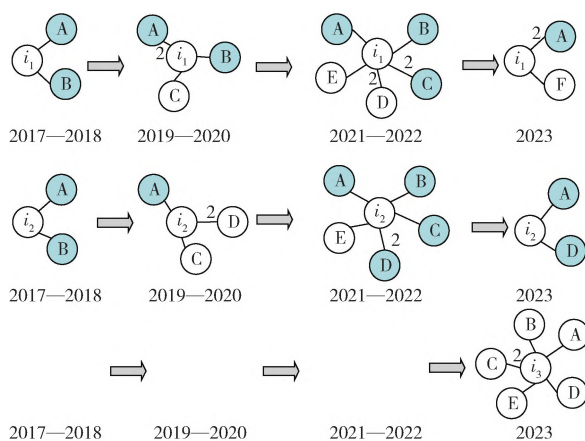


图3 知识单元关联强度发展变化

Fig. 3 Development changes in the strength of association of knowledge units

$$N_{qi} = \sum \frac{1 - \frac{L_{i\pi}}{L_{ic}}}{t_x - t} \quad (13)$$

式中,  $\frac{L_{i\pi}}{L_{ic}}$  表示在  $t$  时间段已有知识单元在知识单元  $i$  所关联知识单元中的占比程度;  $1 - \frac{L_{i\pi}}{L_{ic}}$  表示在  $t$  时间段新出现知识单元的占比程度;  $t_x$  为固定值;  $t$  同公式 (12)。

综合知识单元突变种类与突变强度, 构建知识单元突变属性特征测度指标 (见公式 (14)), 并对相应指标进行标准化计算与赋予相应权重。

$$N_i = \frac{\sum (N_{pi} + N_{qi})}{k} \quad (14)$$

式中,  $N_i$  表示知识单元突变属性特征;  $N_{pi}$  表示知识单元突变种类;  $N_{qi}$  表示知识单元突变强度;  $k$  表示待评价学术论文知识单元数量, 若计算单个知识单元突变属性特征, 则

$k = 1$ 。

### 2.3 学术论文创新性综合评价模型

创新性测度是对学术论文创新情况的定量描述，通常是指对学术论文整体创新所达到的程度或水平的测度<sup>[31]</sup>。学术论文创新性体现为其所包含知识单元的新颖性，本文在综合考虑论文知识单元发展属性特征和突变属性特征的基础上探究学术论文创新性，构建综合评价模型：

$$I = Q_i + N_i \quad (15)$$

式中， $I$ 表示学术论文创新性； $Q_i$ 表示学术论文知识单元发展属性特征； $N_i$ 表示学术论文知识单元突变属性特征。

## 3 实证分析

### 3.1 数据来源及数据预处理

3.1.1 数据来源 数字人文也被称为人文计算，是借助计算机技术与人文学科进行交叉而产生的新兴学科。与传统人文学科不同，数字人文兼具跨机构性、跨学科属性，其最终目的是发现知识、展现知识，进而创造知识<sup>[32]</sup>。其研究的跨学科领域属性要求数字人文领域学者必须更加充分、迅速和精准地了解领域内研究进展<sup>[33-34]</sup>。因此，本文以“数字人文”领域相关文献为例，从知识单元发展属性特征和突变属性特征维度评价学术论文创新性，以帮助研究者掌握、发现该领域研究进展。

本文实证研究来源为中国知网数据库，检索主题为“数字人文”，检索范围为2017—2023年，检索类型为学术期刊。经检索，共获得2079条有效文献题录，并从中提取知识单元、发表时间、引证数据等信息作为基础数据集。本文以2023年度学术论文为研究对象，测度2023年度被引量排序前6名的12篇学术论文的创新性。

#### 3.1.2 数据处理

第一阶段数据清洗：删除知识单元空缺、知识单元不完整等无效文献题录，共获得2079条有效文献题录。对2079篇文献题录的知识单元进行抽取，统一知识单元格式并合并同义词（如“citespace”统一为“CiteSpace”，“大学图书馆”统一为“高校图书馆”等）。

第二阶段词集构建：首先按照年份分别提取学术论文知识单元，统计不同年份发表学术论文数量和知识单元频次，进而对不同年份提取的知识单元构建共词矩阵，利用Ucinet软件和VBA程序代码计算知识单元点度中心性和加权点度中心性。

本文以2023年478篇中1309个知识单元为主要研究对象，以被引量排序前6名的12篇学术论文为待评价文献。随着信息技术的发展，数据挖掘技术、可视化技术、深度学习等技术在人文社会学科领域得到了广泛应用，推动了人文社会学科的研究，数字人文也应运而生，并从

2017年开始，该领域发文数量呈明显增加趋势。

### 3.2 结果分析

3.2.1 发展属性特征测度结果分析 通过对1309个知识单元在近7年增长率、所在论文占比数量增长率以及关联广度和数量增长率进行量化并比较，得出知识单元总体发展属性特征排名，见表1。鉴于文章篇幅有限，表1只展示了前10名知识单元发展属性特征得分。

表1 知识单元发展属性特征指标  
Tab. 1 Development attributes of knowledge units

知识单元	知识单元发展规模		知识单元关联强度		知识单元发展属性特征
	知识单元增长率	知识单元论文相对增长率	知识单元关联广度增长率	知识单元关联数量增长率	
信息资源管理	0.2189	0.2658	0.2393	0.2302	0.9542
新文科建设	0.2095	0.2404	0.1692	0.1966	0.8158
人才培养	0.1950	0.2617	0.1543	0.1751	0.7861
ChatGPT	0.1767	0.2606	0.1417	0.1879	0.7670
档案开发	0.2048	0.2762	0.1303	0.1551	0.7664
文化数字	0.1486	0.2554	0.1677	0.1638	0.7355
信息技术	0.1345	0.2502	0.1526	0.1658	0.7031
CiteSpace	0.1814	0.2565	0.1285	0.1335	0.6999
大语言模型	0.1486	0.2554	0.1331	0.1557	0.6928
语义	0.1158	0.2400	0.1616	0.1660	0.6835
...	...	...	...	...	...

以“信息资源管理”知识单元各属性数据变化为例，具体变化过程，如图4所示。该知识单元在2019年出现了1次，与3个知识单元之间存在关联，而在2023年出现了7次，与25个知识单元之间存在关联。其在2023年全部知识单元出现频次总体排名中较为靠前，增长速度较快，是当前研究热点，并具有一定的发展潜力。鉴于信息资源的数字化与网络化趋势愈加明显，且数字人文领域的研究需要大量信息资源，越来越多的学者开始探究利用信息资源管理实现信息资源的共享和利用。

随着数字化时代的发展，自然语言处理、深度学习、ChatGPT等信息技术被广泛应用于数字人文领域。越来越多的文本、图像、音频等数据被数字化保存和传播，为数字人文研究提供了大量数据资源。大型语言模型可以通过学习这些数据来获取文本语言规律和知识。同理，知识单元“ChatGPT”与“信息资源管理”具有一定的相似性，出现时间短，发展速度快，具有较高的发展特征分数。

3.2.2 突变属性特征测度结果分析 本文选取数字人文领域2023年待评价论文中被引量排序前6名的12篇文献进行具体研究，共提取51个知识单元。为了探测知识单元突变属性特征，将时间窗口定为2019—2020年、2021—2022年和2023年，并以2017—2018年数据为基础，探究2023年知识单元突变属性特征，见表2。鉴于文章篇幅有限，表2只展示了部分知识单元突变属性特征分数。

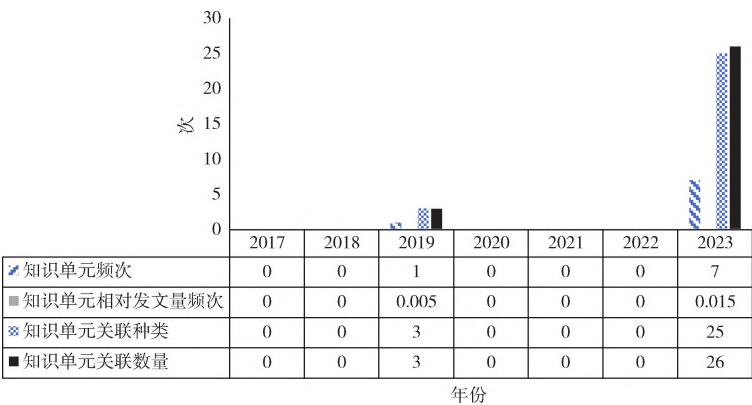


图4 “信息资源管理”发展属性特征  
Fig.4 Development attributes characteristics of “Information Resource Management”

知识单元突变属性特征是表示某时间段知识单元与前一时间段相比,该知识单元与新知识单元组合度。若数值越大,表明知识单元与较多以前未出现过的知识单元进行

组合;若出现在同一篇论文中,表明该篇论文探讨或研究的问题是当前新出现的热点或前沿。以“学科建设”为例,在2017—2018年,“学科建设”与“共建共享”“数字人文”“信息服务”等13个知识单元之间存在关联,而在2019—2020年,“学科建设”与“功能定位”“学科发展”和“情报学”三个知识单元之间存在关联,均为新出现知识单元组合。因而,“学科建设”在2019—2020年中新组合占比是1,为了减少时间因素的干扰,时间越靠近2023年,时间因素干扰越小,本文将2019—2020年设置为1,总时间段设置为4,所以2019—2020年阶段“学科建设”的突变种类为0.3333。同理,突变强度考虑了知识单元关联数量总和,原理同知识单元关联种类。对知识单元突变种类与突变强度不同阶段数据统计汇总并赋予权重,得到知识单元突变属性特征得分,见表3。

表2 知识单元突变属性特征指标

知识单元	知识单元突变种类			知识单元	知识单元突变强度		
	2019—2020年	2021—2022年	2023年		2019—2020年	2021—2022年	2023年
学科建设	0.3333	0.5000	1.0000	学科建设	0.3333	0.5000	1.0000
智慧图书馆	0.3333	0.3846	0.8929	智慧图书馆	0.3333	0.4250	0.8929
图书馆员	0.2917	0.2500	1.0000	档案馆	0.3333	0.3750	0.9286
非物质文化遗产	0.3030	0.3919	0.7857	人工智能	0.2917	0.4333	0.9091
档案馆	0.3333	0.3000	0.8235	文化遗产	0.2963	0.4583	0.8667
...	...	...	...	...	...	...	...

表3 知识单元突变属性特征分数  
Tab.3 Knowledge units mutation attribute characterization score

知识单元	知识单元突变种类	知识单元突变强度	知识单元突变属性特征
学科建设	0.5312	0.4688	1.0000
智慧图书馆	0.4307	0.4222	0.8529
图书馆员	0.3984	0.3942	0.7926
档案馆	0.3598	0.4186	0.7784
非物质文化遗产	0.3706	0.3968	0.7674
人工智能	0.3153	0.4179	0.7332
学科体系	0.3288	0.3552	0.6840
古籍保护	0.3144	0.3470	0.6614
...	...	...	...

以“智慧图书馆”知识单元突变种类为例,如图5所示。在2017—2018年,“智慧图书馆”与“馆员”“人文智慧”等5个知识单元存在关联;

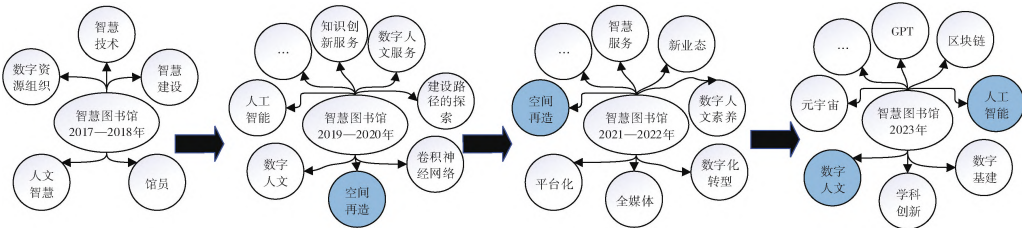


图5 “智慧图书馆”突变属性特征  
Fig.5 “Smart Library” mutation attribute feature map



出现频次、关联广度、关联数量和被引频次。由图6可知,“知识图谱”知识单元的频次、关联广度、关联数量均呈现增长趋势。本文基于“知识图谱”所在的两篇文章,即《国际数字人文研究特征与知识结构》与《我国数字人文研究脉络及其在图书馆学情报学领域的典型应用》,探究该知识单元的被引频次特征,虽然该知识单元在2021年达到被引高峰,但在2022年与2023年仍保持相对稳定的被引频次,同时该知识单元被引频次平均增长率在32%左右。2017年“知识图谱”出现频次为2,总被引为5次,篇均被引为2.5次。随后,该知识单元分别与不同知识单元组合出现在新的文章中,并不断扩散累积。基于该知识单元各属性与被引频次的变化,可以看出“知识图谱”知识单元在2017年时具有较高的新颖性和创新性,并随着趋势的变化,其影响力具有一定的持续性。同理,以2023年待评价学术论文为例,新出现知识单元或具有持续影响力的知识单元可以作为待评价文献创新性的衡量标准。

表4 学术论文创新性测度结果排名

Tab. 4 Ranking of innovative measurement results of academic papers

题名	知识单元	知识单元突变属性特征	知识单元发展属性特征	创新性
档案学科建设与人才培养的数字转型——基于图书情报与档案管理一级学科更名为信息资源管理的思考	档案学;学科建设;人才培养;信息资源管理;数字转型	0.6217	0.6706	1.2934
古籍传承性保护再认识	古籍保护;传承性保护;非物质文化遗产;学科建设	0.7091	0.4769	1.1859
我国文化遗产数字人文研究的推进策略分析	文化遗产;传承与保护;数字人文;学科体系	0.5932	0.3851	0.9783
基于数字叙事理论的档案服务创新研究	数字叙事;数字人文;档案馆;档案服务	0.5321	0.4307	0.9627
数字人文视域下文化记忆机构价值共创研究及实践述评	文化记忆机构;GLAM;价值共创;数字人文;研究述评	0.5110	0.4491	0.9600
图书馆IT应用十大趋势	智慧图书馆;图书馆服务平台;区块链;细颗粒度加工;元宇宙;数字人文;网借服务;图书馆联盟	0.5329	0.4208	0.9536
...	...	...	...	...

如表4所示,杨文等在《图书情报工作》上发表的《档案学科建设与人才培养的数字转型——基于图书情报与档案管理一级学科更名为信息资源管理的思考》一文新颖性最高,其认为图书情报与档案管理一级学科更名为信息资源管理对档案学科建设和人才培养来说是一个巨大的机遇,急需通过明确学科建设定位等举措实现档案学科建设和人才培养的动态化调整与优化。该论文共被引13次,论文知识单元发展特征为0.6706,知识单元突变特征为0.6217,其包含的主要知识单元为档案学、学科建设、人才培养、信息资源管理和数字转型。

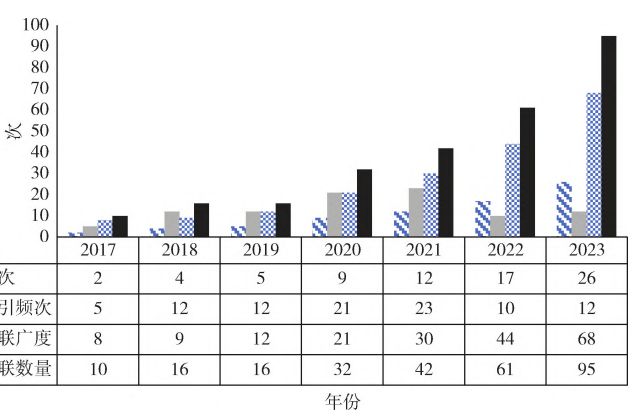


图6 “知识图谱”各属性与被引频次的变化

Fig. 6 Changes of attributes and citation frequency of “Knowledge Graph”

3.2.3 学术论文创新性测度结果分析 本文从待评价的12篇论文中共提取了51个知识单元,利用公式(14)测度每个知识单元的发展属性特征与突变属性特征,最终综合知识单元的属性特征测度文章创新性。对12篇依据学术论文创新性得分依次排序,见表4。鉴于文章篇幅有限,表4只展示了部分学术论文。

该篇论文知识单元发展属性特征和突变属性特征都较为突出,表明在该篇文章中知识单元发展创新能力较强,如“档案学”在2017—2023年期间,其频次、关联程度整体呈上升趋势,在数字人文领域,其在2017年出现频次为0,但在2023年出现频次为9,整体排名中处于第8位。该知识单元关联种类同样由2017年的0次增加到2023年的31次,整体排名第8位。数字化技术的普及使得档案材料的存储、管理、检索和分析等变得更加高效和便捷,极大地提高了档案学研究效率和水平,研究者可以方便地收集数据与处理大量数据,开展更广泛和深入的

研究。

知识单元“信息资源管理”在2019年仅与“数字转型”“数字人文”“图书情报与档案管理”三个知识单元存在关联,是学者们基于此知识单元的最初探索,并未引起学界的关注。2023年图书情报与档案管理一级学科正式更名为信息资源管理,图书情报与档案管理领域与数字人文领域存在着密切联系,二者相辅相成,互相影响。学者们围绕此次更名事件从不同角度、结合不同领域知识发表自己的见解,成为当前学术界讨论的热点之一。2023年,学术界围绕“信息资源管理”与“档案学”“数字转型”“新文科”“图书情报与档案管理”多个知识单元组合开展广泛研究,该知识单元的发展仍呈现增长趋势。

姚伯岳等在《中国图书馆学报》上发表的《古籍传承性保护再认识》一文新颖性次之,其在理念方面进行了进一步创新。学术界在总结出古籍原生保护与古籍再生性保护概念之后,又提出了古籍传承性保护概念,该篇文献对古籍传承性保护概念进行了进一步梳理。该篇论文共被引10次,论文知识单元发展特征为0.4769,知识单元突变特征为0.7091,包含主要知识单元为古籍保护、传承性保护、非物质文化遗产和学科建设。“古籍保护”与“传承性保护”在2022年受到了学者广泛关注,随着学者研究不断深入,古籍传承性保护理念被提出,并对古籍再生性保护与古籍原生性保护产生了重要意义。“古籍传承性保护”的提出进一步完善了古籍保护事业,为进一步构建合理的古籍保护学科知识体系创造了重要条件。

#### 4 结束语

随着学术界的发展和变化,学术论文评价成为学术研究的热点问题,学术论文作为学术研究的主要成果之一,对其进行评价涉及学者学术声誉以及成就的发展,是学术界的重要环节。评价的本质是为了推动学术进步、促进学术合作和提高学术质量。在当前学术论文评价研究中,学者们主要通过机器学习、自然语言处理等方法,通过比较待评价论文所包含知识与已有论文知识相似度、收集学术论文外部数据等,从学术论文新颖性、价值性、影响性等方面进行评价。

本文基于知识单元的动态发展过程,考虑知识单元出现、发展、消失或突变特征与时间因素对知识单元的影响,综合知识发展的局部特征、整体特征及知识单元在网络中的属性特征等,构建了知识单元新颖性测度指标,从学术论文知识单元视角评价学术论文创新性。但此方法仍存在一些不足:①只提取了能够概括文章中心的知识单元,未对知识单元进行具体分类,如问题创新或方法创新等;②仅以数字人文领域部分学术论文为研究对象,模型

适用性具有一定的局限性,如未充分考虑该领域的跨学科性质,评价结果可能无法准确反映不同学科间创新性的差异。在未来的研究中,将会进一步细化知识单元的种类,探究不同种类知识单元组合的特征,并聚焦于某一传统学科,以更好地评估学科研究方法和理论的创新性。□

#### 参考文献

- [1] 谢梅. 科技创新为新质生产力“蓄势赋能”[J]. 人民论坛, 2024(5): 66-68. (XIE Mei. Scientific and technological innovation “building momentum and empowering” for new quality productivity [J]. People's Tribune, 2024(5): 66-68.)
- [2] 周海晨. 面向学术评价的学术创新知识挖掘研究[D]. 南京: 南京农业大学, 2021. (ZHOU Haichen. Research on text mining of academic innovation knowledge for academic evaluation [D]. Nanjing: Nanjing Agricultural University, 2021.)
- [3] 国务院办公厅关于完善科技成果评价机制的指导意见[EB/OL]. [2024-03-12]. [https://www.gov.cn/gongbao/content/2021/content\\_5631817.htm](https://www.gov.cn/gongbao/content/2021/content_5631817.htm). (The Guiding opinions of the General Office of the State Council on improving the evaluation mechanism of scientific and technological achievements [EB/OL]. [2024-03-12]. [https://www.gov.cn/gongbao/content/2021/content\\_5631817.htm](https://www.gov.cn/gongbao/content/2021/content_5631817.htm).)
- [4] 秦成磊, 章成志. 大数据环境下同行评议面临的问题与对策[J]. 情报理论与实践, 2021, 44(4): 99-112. (QIN Chenglei, ZHANG Chengzhi. Problems and responses strategy of peer review in the context of big data [J]. Information Studies: Theory & Application, 2021, 44(4): 99-112.)
- [5] 曾建勋. 推动科研论文语义评价体系建设[J]. 数字图书馆论坛, 2021(11): 1. (ZENG Jianxun. Promote the construction of semantic evaluation system for scientific research papers [J]. Digital Library Forum, 2021(11): 1.)
- [6] 吴欣雨, 李涵昱, 张智雄, 等. 科技文献评价中语义新颖性研究综述[J]. 数据分析与知识发现, 2024, 8(3): 29-40. (WU Xinyu, LI Hanyu, ZHANG Zhixiong, et al. Reviewing research on semantic novelty in sic-tech literature [J]. Data Analysis and Knowledge Discovery, 2024, 8(3): 29-40.)
- [7] 索传军, 盖双双. 单篇学术论文的评价本质、问题及新视角分析[J]. 情报杂志, 2018, 37(6): 102-107. (SUO Chuanjun, GAI Shuangshuang. Analysis on the essence, problems and new perspective of individual academic paper evaluation [J]. Journal of Intelligence, 2018, 37(6): 102-107.)
- [8] 杨杰, 邓三鸿, 王昊. 科学研究的颠覆性创新测度——相对颠覆性指数[J]. 情报学报, 2023, 42(9): 1052-



1064. (YANG Jie, DENG Sanhong, WANG Hao. Measure of disruptive innovation in science: relative disruptive index (RDI) [J]. Journal of the China Society for Scientific and Technical Information, 2023, 42 (9): 1052-1064. )
- [ 9 ] WU Lingfei, WANG Dashun, EVANS J A. Large teams develop and small teams disrupt science and technology [J]. Nature, 2019, 566 (7744): 378.
- [ 10 ] JIANG Yuyan, LIU Xueli. A bibliometric analysis and disruptive innovation evaluation for the field of energy security [J]. Sustainability, 2023, 15 (2): 969.
- [ 11 ] MEYER C, NAKAMURA Y, RASOR B J, et al. Analysis of the innovation trend in cell-free synthetic biology [J]. LIFE-BASEL, 2021, 11 (6): 551.
- [ 12 ] 宋歌. 科研成果创新力指标 S 指数的设计与实证 [J]. 图书情报工作, 2016, 60 (5): 77-86, 124. (SONG Ge. Study on innovation power evaluation index of scientific research achievements: design and demonstration of S index [J]. Library and Information Service, 2016, 60 (5): 77-86, 124. )
- [ 13 ] 魏绪秋, 姜召昊, 常霞, 等. 基于引证意图的学术论文创新性评价研究 [J]. 情报理论与实践, 2023, 46 (9): 24-30, 46. (WEI Xuqiu, JIANG Zhaohao, CHANG Xia, et al. Research on the innovation evaluation of academic papers based on citing intention [J]. Information Studies: Theory & Application, 2023, 46 (9): 24-30, 46. )
- [ 14 ] 杨思洛, 江曼, 高强. 基于知识重组和变异的技术新颖性评估——以数字医疗技术为例 [J]. 数据分析与知识发现, 2023, 7 (12): 52-63. (YANG Siluo, JIANG Man, GAO Qiang. Technology novelty assessment based on knowledge reorganization and variation: case study of digital medicine [J]. Data Analysis and Knowledge Discovery, 2023, 7 (12): 52-63. )
- [ 15 ] YANG Jinqing, LU Wei, HU Jiming, et al. A novel emerging topic detection method: a knowledge ecology perspective [J]. Information Processing & Management, 2022, 59 (2): 102843.
- [ 16 ] 曹树金, 曹茹桦. 情报学论文创新性评价研究——LDA 和 SVM 融合方法的应用 [J]. 图书情报知识, 2022, 39 (4): 56-67. (CAO Shujin, CAO Ruyue. Evaluation of paper innovativeness in information science by a method integrated of LDA and SVM [J]. Documentation, Information & Knowledge, 2022, 39 (4): 56-67. )
- [ 17 ] MATSUMOTO K, SHIBAYAMA S, KANG B, et al. Introducing a novelty indicator for scientific research: validating the knowledge-based combinatorial approach [J]. Scientometrics, 2021, 126 (8): 6891-6915.
- [ 18 ] 汪雪锋, 于慧妍, 郑思佳, 等. 学术论文创新质量评价研究: 以多能干细胞技术为例 [J/OL]. 数据分析与知识发现: 1-21 [2024-03-21]. <http://kns.cnki.net/kcms/detail/10.1478.G2.20230608.1604.002.html>. (WANG Xuefeng, YU Huiyan, ZHENG Sijia, et al. Research on innovation quality evaluation of academic papers: a case study in the field of pluripotent stem cells [J/OL]. Data Analysis and Knowledge Discovery: 1-21 [2024-03-21]. <http://kns.cnki.net/kcms/detail/10.1478.G2.20230608.1604.002.html>. )
- [ 19 ] 侯剑华, 王东毅. 基于 SAO-ADV 模型的学术论文创新性的测度方法研究 [J]. 情报理论与实践, 2020, 43 (11): 129-136. (HOU Jianhua, WANG Dongyi. Research on innovation measurement method of academic papers based on SAO-ADV model [J]. Information Studies: Theory & Application, 2020, 43 (11): 129-136. )
- [ 20 ] 段庆锋, 陈红, 闫绪娴, 等. 基于知识结构突变的学科新兴主题识别研究 [J]. 情报学报, 2023, 42 (9): 1018-1028. (DUAN Qingfeng, CHEN Hong, YAN Xuxian, et al. Identifying emerging scientific topics by abrupt change of knowledge structure [J]. Journal of the China Society for Scientific and Technical Information, 2023, 42 (9): 1018-1028. )
- [ 21 ] 潘现伟, 崔雷. 基于文献网络节点属性的论文重要性评价模型的构建及评估 [J]. 情报理论与实践, 2023, 46 (9): 130-141. (PAN Xianwei, CUI Lei. Construction and assessment of paper importance evaluation model based on the node attributes of paper similarity network [J]. Information Studies: Theory & Application, 2023, 46 (9): 130-141. )
- [ 22 ] SHIBAYAMA S, WANG Jian. Measuring originality in science [J]. Scientometrics, 2020, 122 (1): 409-427.
- [ 23 ] AMPLAYO R K, HONG S, SONG Min. Network-based approach to detect novelty of scholarly literature [J]. Information Sciences, 2018, 422: 542-557.
- [ 24 ] 李楠. 知识进化视角的技术预见方法研究 [D]. 北京: 中国农业科学院, 2021. (LI Nan. Research on technology foresight method from the perspective of knowledge evolution [D]. Beijing: Chinese Academy of Agricultural Sciences, 2021. )
- [ 25 ] 沈阳. 一种基于关键词的创新度评价方法 [J]. 情报理论与实践, 2007 (1): 125-127. (SHEN Yang. A keyword-based innovation evaluation method [J]. Information Studies: Theory & Application, 2007 (1): 125-127. )
- [ 26 ] PESET F, GARZON-FARINOS F, GONZALEZ L M, et al. Survival analysis of author keywords: an application to the library and information sciences area [J]. Journal of the Association for Information Science and Technology, 2020, 71 (4): 462-473.
- [ 27 ] 党永杰, 明均仁. 节点度、节点强度在作者合作网络分析中的应用研究 [J]. 情报探索, 2018 (12): 15-20.

- (DANG Yongjie, MING Junren. Application of node degree and node intensity in the analysis on authors' collaboration network [J]. Information Research, 2018 (12): 15-20.)
- [28] 张玉, 魏华波. 基于 CRITIC 的多属性决策组合赋权方法 [J]. 统计与决策, 2012 (16): 75-77. (ZHANG Yu, WEI Huabo. Multi-attribute decision combination weighting method based on CRITIC [J]. Statistics & Decision, 2012 (16): 75-77.)
- [29] 曹艺文, 许海云, 武华维, 等. 基于引文曲线拟合的新兴技术主题的突破性预测: 以干细胞领域为例 [J]. 图书情报工作, 2020, 64 (5): 100-113. (CAO Yiwu, XU Haiyun, WU Huawei, et al. Study on radical innovation prediction to emerging technology topics based on citation curve fitting: taking the field of stem cells as an example [J]. Library and Information Service, 2020, 64 (5): 100-113.)
- [30] YU Hao, Wei Yiming, Tang Baojun, et al. Assessment on the research trend of low-carbon energy technology investment: a bibliometric analysis [J]. Applied Energy, 2016, 184: 960-970.
- [31] 戎军涛, 索传军, 周彦廷, 等. 基于创新知识元谱系的学术论文新颖性测度研究 [J]. 图书情报工作, 2024, 68 (1): 27-38. (RONG Juntao, SUO Chuanjun, ZHOU Yanting, et al. Novelty measurement of academic literature based on the innovative knowledge elements genealogy [J]. Library and Information Service, 2024, 68 (1): 27-38.)
- [32] 吴瑞丽. 数字人文视域下的非遗资源整合及保护机制 [J]. 图书馆学刊, 2018, 40 (10): 50-54. (WU Ruili. Integration and protection mechanism of non-heritage resources from the perspective of digital humanities [J]. Journal of Library Science, 2018, 40 (10): 50-54.)
- [33] 杨丽萍. 数字人文视域下的非物质文化遗产资源整合及保护 [J]. 文化产业, 2022 (14): 138-140. (YANG Liping. Integration and protection of intangible cultural heritage resources from the perspective of digital humanity [J]. Culture Industry, 2022 (14): 138-140.)
- [34] 孔凡晶, 王雪莲. 数字人文视域下哈佛大学图书馆知识管理研究 [J]. 四川图书馆学报, 2023 (3): 48-54. (KONG Fanjing, WANG Xuelian. Research on knowledge management of Harvard library from the perspective of digital humanities [J]. Journal of the Library Science Society of Sichuan, 2023 (3): 48-54.)
- 作者简介:** 常霞, 女, 1999 年生, 硕士生。研究方向: 信息计量, 科学评价与数据挖掘。魏绪秋 (通信作者, Email: weixuqiu@163.com), 男, 1990 年生, 副教授, 硕士生导师。研究方向: 信息计量, 科学评价与数据挖掘。张以迪, 女, 2001 年生, 硕士生。研究方向: 信息计量, 科学评价与数据挖掘。徐雨薇, 女, 1999 年生, 硕士生。研究方向: 信息计量, 科学评价与数据挖掘。
- 作者贡献声明:** 常霞, 论文初稿撰写。魏绪秋, 论文选题与修改。张以迪, 论文修改。徐雨薇, 论文修改。
- 录用日期:** 2024-05-27
- (上接第 129 页)
- [33] LIU Jiqun. Deconstructing search tasks in interactive information retrieval: a systematic review of task dimensions and predictors [J]. Information Processing and Management, 2021, 58 (3): 102522.
- [34] BORLUND P, SCHNEIDER J W. Reconsideration of the simulated work task situation: a context instrument for evaluation of information retrieval interaction [C] //Proceedings of the Third symposium on Information Interaction in Context. [S.l.]: ACM, 2010: 155-164.
- [35] KELLY D. Methods for evaluating interactive information retrieval systems with users [J]. Foundations and Trends in Information Retrieval, 2009, 3 (1/2): 1-224.
- [36] 李月琳, 肖雪, 仝晓云. 数字图书馆中人机交互维度与用户交互绩效的关系研究 [J]. 图书情报工作, 2014, 58 (2): 38-46, 120. (LI Yuelin, XIAO Xue, TONG Xiaoyun. Identifying influential users in micro-group using association rules [J]. Library and Information Service, 2014, 58 (2): 38-46, 120.)
- [37] LIU Jingjing, BELKIN N J. Multi-aspect information use task performance: the roles of topic knowledge, task structure, and task stage [J]. Proceedings of the American Society for Information Science and Technology, 2014, 51 (1): 1-10.
- 作者简介:** 章小童 (ORCID: 0000-0003-1875-8832), 男, 1991 年生, 讲师, 博士。研究方向: 信息行为, 交互信息检索, 健康信息学, 情报学理论与方法。李月琳 (ORCID: 0000-0002-1496-6741, 通信作者, Email: yuelinli@nankai.edu.cn), 女, 1970 年生, 博士, 教授。研究方向: 信息行为, 交互信息检索, 健康信息学, 数字图书馆与信息系统评估, 信息公开质量。辛然 (ORCID: 0000-0003-2866-9569), 1998 年生, 硕士生。研究方向: 心血管内科学。
- 作者贡献声明:** 章小童, 提出研究设计与分析方案, 数据收集与分析, 撰写论文初稿。李月琳, 完善研究设计, 论文终稿审定。辛然, 数据收集与分析。
- 录用日期:** 2024-05-27