doi:10.3772/j.issn.2095-915x.2018.03.012

文献相似性检测技术及其应用

蒋勇青　于洋

中国科学技术信息研究所　北京　100038

摘 要 学术不端行为屡禁不止，通过计算机技术自动发现或规避学术不端行为特别是文献抄袭和剽窃等需求日益迫切。文献相似性检测是发现文献抄袭剽窃的重要手段，而且可以帮助研究者规避文献抄袭和剽窃风险。自万方数据公司文献相似性检测系统正式提供服务以来，经过多次更新升级，相似文献监测效果显著，在中国学术界得到广泛关注和使用。本文对文献相似性检测系统工作原理进行介绍，结合万方数据公司文献相似检测系统应用情况分析了现有系统应用的有效性，并对现有系统的局限性进行深入分析，最后为文献相似性检测系统的应用发展提出了建议与方向。

关键词： 文献检测系统；相似性检测；学术不端；检测系统应用

中图分类号： G350, TP391

开放科学（资源服务）标识码（OSID）

Research on Literature Similarity Detection Technology and Its Applications

JIANG Yongqing　YU Yang

Institute of Scientific and Technical Information of China, Beijing 100038, China

Abstract Academic misconduct has continued for many years, and it has become increasingly urgent to automatically discover or circumvent academic misconduct such as plagiarism through computer technology. The similarity detection of documents is an important method to find plagiarism, and it can help

基金项目： 国家自然科学基金项目“基于文献的科研事件语义链接及其应用研究”（71503240）。

作者简介： 蒋勇青（1963–），高级工程师，研究方向：信息资源建设与服务、知识服务、科研管理、企业管理，E-mail：jiang@wanfangdata.com.cn；于洋（1984–），硕士，工程师，研究方向：知识服务，数字编辑，项目管理。

researchers to avoid the risk of plagiarism. Since the literature similarity detection system of the Wanfang Data Corporation has supplied services, after many updates and upgrades, the literature similarity detection system has achieved remarkable results and received extensive attention and use in Chinese academic circles. This paper introduces the working principle of the literature similarity detection system and analyzes the effectiveness of the existing system application based on the application of the literature similarity detection system of the Wanfang Data Corporation. It also deeply analyzes the limitations of the existing system. This paper also gives recommendations for the research and applications of the literature detection system in the future.

Keywords: Literature detection system; similarity detection; academic misconduct; applications of detection system

1 引言

科研诚信对科技发展起着重要作用，发现并惩罚抄袭剽窃行为对维护学术公平具有重要意义。随着计算机技术和互联网行业的飞速发展，数字化、电子化信息得以广泛传播与应用，但同时也为学术研究过程中学术抄袭、不合理引用等学术不端行为的发生提供了便利条件。为了维护健康公平的学术环境、促进学术创新与发展，如何有效探查并防范学术不端行为已引起国内学界的广泛关注。至今，我国有关机构和组织已针对学术不端行为制定和出台了各种规范和措施，如 2002 年 2 月 27 日教育部发布《关于加强学术道德建设的若干意见》[1]；国家科技部在 2007 年 1 月 1 日起实施的《国家科技计划实施中科研不端行为处理办法（试行）》中定义了 6 项科研不端行为 [2]；中国科学院 2011 年 11 月 9 日通过了《中国科学院院士科学道德自律准则》[3]；2016 年 9 月 1 日开始施行的《高等学校预防与处理学术不端行为办法》，将剽窃、抄袭、侵占他人学术成果及篡改他人研究成果列在学术不端行为的首要位置 [4]；中华医学会 2016 年 11 月 16 日发起《倡导学术诚信，抵制学术不端 —— 医学期刊从业者倡议书》[5]。

面对海量的学术成果以及大规模的学术文献产出，一些研究者从计算机技术应用角度入手，设计出用于探查学术不端行为的计算机软件，即文献相似性检测系统，希望依据系统提供的全面客观的相似性检测结果，规避杜绝学术不端行为，确保学术文献的真实性，进而提升整体学术与科技发展水平，有效贯彻落实在学术及科研领域的学术不端审查工作。近几年来，由于学术不端行为屡禁不止，市场需求不断扩大，文献相似性检测系统秉持客观、公正、精准、全面的原则，在检测算法、数据积累、用户体验等方面都进行了全面的升级，为各领域提供精准的文献检测，同时根据送检文献特征及所属领域特点提供便捷的检测信息统计、多级账号管理等增值性服务，为学术不端行为的发现与探查提供了有效的技术手段。

doi:10.3772/j.issn.2095-915x.2018.03.012

2 相关工作

文本相似性检测实际上是判断给定的两篇文档中是否存在相同文本，当两个文档中相同部分占比超过阈值时，则认为两篇文档具有相似性。文本相似性检测方法可分为：基于词频向量的文本相似性检测方法、基于关键词匹配的文本相似性检测方法、基于相似度估值的文本相似性检测方法和基于写作风格的文本相似性检测方法。

2.1 基于词频向量的文本相似性检测方法

基于词频向量的文本相似性检测方法将文档表示为词频组成的向量。对于指定相似监测文档集合，设定词典集合，根据词典中每个词出现的频率，对词典中不同字词赋予不同的权值，再通过利用相似度计算方法如欧氏距离或者余弦距离等计算文档之间的相似度。基于词频向量的文本相似性检测研究工作包括SCAM[6,7]、DSCAM[8]、CHECK[9]、CDSDG[10]和HFM[11]等。基于词频向量的文本相似性检测方法检测结果准确率低而且容易误判，因为基于词频向量的检测方法中每个词的权重相同，而在实际应用中，不同字、词和短语都具有不同的权重。如果词典中数目较多会造成文档模型的词频向量维度较高，造成文本相似度计算复杂度较高。

2.2 基于关键词匹配的文本相似性检测方法

基于关键词匹配的文本相似性检测方法从待检测文档中利用TF-IDF等方法选取关键词，并把关键词散列为哈希表中的数值形成文档的指纹。该类方法主要研究工作包括sif[12]、COPS[13]、KOALA[14]、Shingling[15]、YAP[16]和MDR[17]等系统。基于关键词匹配的文本相似性检测方法需要在分词处理后进行关键词选取，文本相似性检测质量依赖于分词质量和关键词的选取结果。

2.3 基于相似度估值的文本相似性检测方法

基于相似度估值的文本相似性检测方法采用降维[18]技术将文档向量或关键词集合转换为k个指纹的集合，并用指纹集合表征文档。当求解两篇文档相似度时，直接比对两篇文档对应指纹集合的相似度即可作为两篇文档之间的相似度。该方法中k越大，估值方差越，估计精度越高；k越小，估值的方差越大，估计精度就越低。基于相似度估的文本相似性检测研究工作包括：随机投影方法[19]、min-wise相似度估计算法[20]、sire_hash算法[21]和b位min-wise哈希算法[22]。基于相似度估值的文本相似性检测方法能够有效地解决相似度检测中时间复杂度偏高的问题。

2.4 基于写作风格的文本相似性检测方法

写作风格分析（Stylometric Analysis）用于分析和发现文本文件的作者信息[23]。基于写作风格的文本相似性检测不同于对文档内容本身的相似性检测，是对两篇文档写作风格维度的检测比对。该类方法检测的内容包括：词法、句法、结构、内容和特定的写作特点[24]。基于写作风格的文本相似性检测方法能够将不同来源甚至不同作者的文档挑选出来作为相似文档，对于文档抄袭和剽窃检测具有一定的参考价值。

上述方法中，基于词频向量的文本相似性检测方法、基于关键词匹配的文本相似性检测方法和基于相似度估值的文本相似性检测方法，主要基于文本内容用词和词义方面检查文档之间的相似性；而基于写作风格的文本相似性检测方法，则从文本作者的写作风格和习惯上鉴别文本之间的相似性。文本相似性检测需要从词语、句子[25]、段落甚至篇章结构[26]、写作风格多个方面去考虑。在实际利用中，需要根据文本相似性检测文档的规模和侧重点选择不同的文本相似检测方法，可以综合利用多种方法检测文档之间的相似性，甚至可以检测不同类型文献如论文和专利之间的相似度[27]以及网页之间的相似度[28]。

3 文献相似性检测系统工作原理

自 2008 年，各类文献相似性检测系统相继应运而生。各类文献相似性监测系统的工作原理大致相同，基本模块包括：输入模块、比对数据库模块、检测模块、检测报告模块和其他辅助功能模块。通过检测算法将待检测文献与比对库文献进行比对，从而获得相应的检测报告，此报告呈现了对送检文献检测后的结果，可作为学术评判的依据。图 1 显示了文献相似性监测系统的组成模块及相互关系。

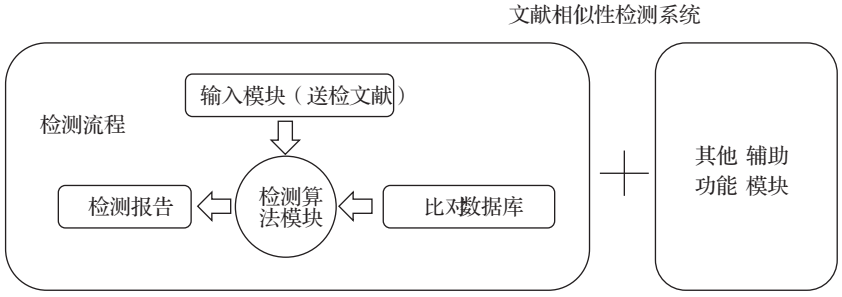

图 1 文献相似性检测系统的工作原理

3.1 主流检测算法

检测方法是搭建一个完善的、高性能、高质量文献相似性检测系统的核心关键。目前，国内实现对汉语文本检测的主要模型有向量空间模型、广义向量空间模型、语义索引模型等，基本采用数字指纹和词频统计两大类技术和方法[29]，并且大多根据检测的对象而对算法进行相应的参数、公式调整。以“万方数据文献相似性检测系统”[30]为例，其采用的是“句子级正交软聚类语义算法”，该算法基于空间向量模型，结合同义词、词性及语序的语义结构来实现，集词频统计检测算法和语义结构算法的优点于一体，从而实现高准确率、高稳定性的相似性检测服务。

3.2 比对资源情况

利用文献相似性检测系统实现学术文献的相似性检测并以此作为学术文献的质量评判依据，多种类型的海量比对资源则是整个系统的发展基础，主要原则包括：比对资源的认可度越高，权威性则越高；比对资源的文献类型覆盖越全，检测结果的精准度则越高。

doi:10.3772/j.issn.2095-915x.2018.03.012

各文献相似性检测系统收录的比对资源因其业务发展定位的不同而不同，以“万方数据文献相似性检测系统”和“知网学术不端文献检测系统”[31]为例，万方数据文献相似性检测系统比对库中学位论文及会议论文质量较高，而知网学术不端文献检测系统比对库收录的文献类型较多、相似监测覆盖度更广。随着互联网+及大数据应用逐步在学术领域的渗透，互联网资源的受重视程度越来越高，逐渐成为学术研究参考的重要数据源，这也是互联网资源被纳入相似性检测比对范围的主要原因。专利、标准、法律等行业性资源作为文献检测系统的比对资源的种类储备，更有利于开展行业针对性检测服务。

4 万方文献相似性检测系统应用现状

在具体使用文献相似性检测系统的过程中，人们从初始的尝试操作使用到之后对检测结果的理性认知，与逐步形成了对文献相似性检测系统深层次、多层面的了解和把握，具体表现为文献相似性检测系统的有效性认可、应用性选择、局限性分析、标准性界定等方面。

4.1 有效性认可

笔者使用关键词“学术不端检测系统”“相似性检测系统”查询了万方数据知识服务平台及知网服务平台已发表论文66篇，内容显示文献相似性检测系统的使用者基本认可通过文献相似性检测系统的使用，与以往人工审核学术成果相比较，检测更快捷有效，通过检测报告的阅读获取总相似比、参考文献相似比、单篇最大相似比、段落相似比等结果。李丹研究表明通过文献相似检测系统可以迅速处理文章，大大提高了编辑的工作效率，同时编辑利用文献检测性检测系统，在检测报告页面上可以直观了解文章的抄袭程度，直观地用四种颜色标示不同的结果，红色的为抄袭行为，黄色的为可能引用本人学位论文行为，绿色为参考文献引用情况，蓝色为可能引用本人已发表论文情况[32]。窦学俊、高巧林等使用已发表论文检测系统在2个月内完成了9565篇教师职称评审论文的检测并得到有效的检测结果，而且通过全文版检测报告可详细总览文献检测情况，同时打印一页式存档报告作为检测结果案底[33]。于晓舟等认为文献相似性检测系统的运用，对提高博士学位论文的质量起到一定的保障和促进作用[34]，首先是增强了师生质量意识，其次强化了论文撰写规范，同时明确了质量保证是重点。胡政平详细介绍文献相似性检测系统所具有的检测、对比分析、管理和自动生成检测报告四个主要功能的基础上，重点分析了系统对学术道德建设产生的积极影响，即预防和打击学术不端行为[35]。以上研究表明文献相似性检测系统在很大程度上实现的最基本的功能符合研制者初衷。

4.2 应用性选择

科技文献高效传播的同时，也为学术文献的抄袭剽窃、非法扩散等不道德行为提供了更多方便。在此背景下，文献相似性检测系统得以广泛应用于多个领域，成为遏制学术不端行为的有效工具。

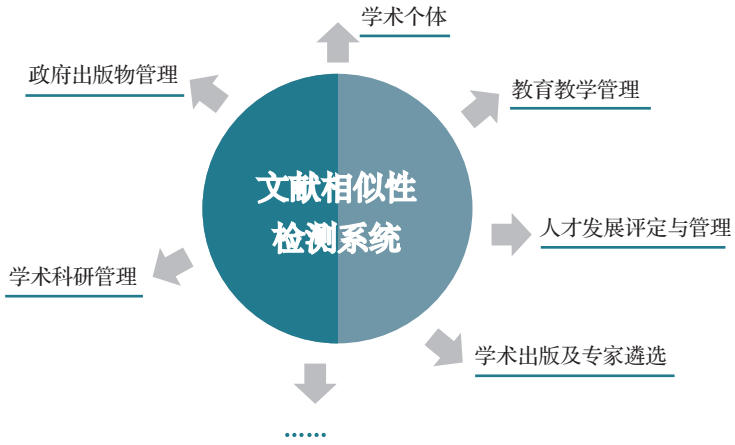

图 2 ADSDS 的应用领域

目前文献相似性检测系统主要应用于学术个体成果检测、教育教学管理、人事评定管理、学术出版管理、学术科研管理 [36]、政府出版物管理等多个领域，如图 2 所示。

（1）应用于教育教学管理领域：以“万方检测 - 硕博论文版”为例，该系统应用于高等院校对学位论文开题、数据筛选、论文答辩等各个阶段，为其提供较为科学的判定标准和详实的不端行为检测结果，助力学位论文质量的把控，规范学位论文管理，推进建立良好学风，提高人才培养质量。

（2）应用于人事评定管理领域 以“万方检测 - 职称论文版”为例，该系统能够根据职称论文特点精确比对范围并提供详实的检测结果，不仅提高论文评定效率，对人才发展的管理提供辅助工具。

（3）应用于学术出版管理领域，以“万方检测 - 学术预审版”为例，该系统可以查看学术文献不同版本间的关联性及差异性，可以调用学术出版机构自建的比对数据，获得更加精准的专有比对报告，从而作为审稿的判断依据，是编辑审稿的有利工具，同时可根据检测结果确定挖掘其潜在的遴选审稿专家的有效方法。

（4）应用于学术科研管理领域：符合学术科研管理场景（项目文档检测、专利申请审核等）与工作流程的学术文献检测系统，实现对各领域科研成果进行全程跟踪审核，辅助完成评估评价、成果转化等环节的统一规范管理。

（5）应用于政府出版物管理领域：以“万方检测 - 公文检测版”为例，该系统以政府出版物为基础比对数据源，有效区分未发布出版物与已发布出版物的重复性内容，区分核心精神指导内容与各地区因地制宜的差异性内容，帮助政府公务人员快速有效处理政府公文出版物管理相关工作。

4.3 局限性分析

目前文献相似性检测系统在比对数据容量、检测结果精准度等层面还存有一定的局限性，还不能完全取代人工核查，主要问题列举如下：

4.3.1 比对资源容量的局限性

使用万方数据文献相似性检测系统、学术不端文献检测系统等几个主流的检测系统，通过上千篇文献的检测发现，检测结果的不同大多由于比对数据的不同而造成的。譬如万方数

doi:10.3772/j.issn.2095-915x.2018.03.012

据文献相似性检测系统和学术不端文献检测系统均收录了期刊论文、学位论文、会议论文、中国专利数据，但同时收录的范围及数量都不尽相同，那么此种情况势必会影响检测结果。随着数据的融合与贡献，各检测系统需要在比对数据的来源、数量等层面进一步加强积累，确保检测结果的权威性。

4.3.2 检测结果精准度的局限性

文献相似性检测系统本身还存在一些问题，需要检测人员以高度的责任心对检测报告仔细解读，审慎出具检测结论。常见的问题包括：

（1）合理引用的误判，引用领导人讲话、权威著作、历史材料、法律法规、诊断疗效标准[37]、操作规程、专有名词等，被误判为段落抄袭或句子抄袭。

（2）检测技术问题，各检测系统使用的检测算法对文献的词句识别、段落结构识别、语义识别等方面都存在差异，这些检测技术的不同会造成同一篇论文在不同的检测系统进行检测时会得出不同的结果，并且有时差异较大[38]。

（3）文献格式检测误差，同一篇文献的不容格式（如 doc 格式与 pdf 格式）进行检测会得到明显不同的结果，这主要由两者文字识别率不同引起，譬如 pdf 格式的文献在进行文本抽取时，乱码率明显高于 doc 格式。

（4）需提高识别参考文献的水平，虽然在各文献相似性检测系统中均提供了“去除参考文献相似比”的指标，但有的参考文献未被识别，而是记录在了总相似比中。

4.4 标准性界定

在以防止抄袭，杜绝学术作假行为，净化学术市场，保障学术成果质量为目的的基础上，多个领域对学术作假行为做出了明确的规定，如自 2013 年 1 月 1 日起施行的中华人民共和国教育部令第 34 号《学位论文作假行为处理办法》[39] 已列出 5 种学位论文作假行为：（1）购买、出售学位论文或者组织学位论文买卖的；（2）由他人代写、为他人代写学位论文或者组织学位论文代写的；（3）剽窃他人作品和学术成果的；（4）伪造数据的；（5）有其他严重学位论文作假行为的。但是在比对资源收录范围要求、相似性的判定、文献组织结构内容送检等方面并没有一个统一的标准供检测者参考，主要的问题如下：

（1）在比对资源收录范围标准的界定上，大多文献相似性检测系统已经收录了学位论文、学术期刊论文、学术会议论文、互联网资源、中国专利、中国标准等，收入的容量及资源种类的范围是否需要统一的标准？

（2）文献相似性的判定标准的界定上，文献的总相似比最高不超过多少？通用术语、法律条款、标准条款等内容的引用比例最高不超过多少？

（3）在文献组织结构内容送检标准的界定上，不同类型的学术成果，文献组织结构不同，哪些学术成果的哪些内容需要检测？文献的题名、作者姓名、参考文献是否计入检测范围？

对于上述问题，笔者认为已有的法规或办法已经明确了学术不端行为，如果再出台更加详细的、量化标准，只能使检测者在总指导思想下视具体情况具体分析成为了在条条框框中生搬硬套，不能灵活处之。处理学术问题更多还是要遵循学术规律，采取学术办法[40]。因此，

我们建议文献相似性检测系统提供商在资源收录层面要以学术资源为主，讲究资源的质量，同时建议文献相似性检测系统使用者在评判送检文献的学术质量时进行综合性考察，最根本的是要找到学术不端的事实证据。

5 总结与展望

文献相似性检测对于建立科研诚信以及公平的学术环境有重要意义。目前国内文献相似性检测系统主要着眼于中文文本的相似性检测，对中文论文的相似性检测及抄袭排查等起到支撑作用，并广泛应用到多个领域。但是，现有的文献相似检测系统中，语义关系识别基本有停留在同义词等阶段，对于意义相同表达文字改写、跨语言剽窃、图片表格等多剽窃方式尚难以检测。

文献相似性检测系统在未来的发展过程中还需要在如下方面提升：（1）增加相似度检测范围，加强图书等类型数据的收录，同时加强多语种资源的收录；加强公式、图片、图表、代码等内容的相似性检测技术的研究与应用；（2）改进相似性检测算法，考虑各类型学术文献的组织结构特征，采用多层次多策略的方法，建立文本结构化识别处理模型与特征向量计算模型，提供自定义多维度的相似性检测服务；（3）深化文献内容理解，挖掘语义结构信息以及上下文的关联语境，利用学科领域特征的专有词典或词表，发掘献之间的隐含的相似关系，实现更深层次的内容相似识别。

参考文献

[1] 教育部 . 关于印发《关于加强学术道德建设的若干意见》的通知 [EB/OL]. (2002-02-27)[2018-1-10]. http://old.moe.gov.cn//publicfiles/business/htmlfiles/moe/moe_25/200407/943.html.

[2] 中华人民共和国科学技术部 .《国家科技计划实施中科研不端行为处理办法（试行）》. (2006-11-10) [2018-1-10]. http://www.most.gov.cn/kjbgz/200611/t20061109_37931.htm.

[3] 中国科学院 . 中国科学院院士科学道德自律准则 (2001 年 11 月 9 日学部主席团会议通过)[J]. 中国科学院院刊 , 2002, 17(1):6.

[4] 中华人民共和国教育部 . 中华人民共和国教育部令第 40 号《高等学校预防与处理学术不端行为办法》, 2016 年 9 月 1 日起施行 [EB/OL]. [2016-06-16]. http://www.moe.edu.cn/srcsite/A02/s5911/moe_621/201607/t20160718_272156.html.

[5] 中华医学会 . 中华医学会倡导学术诚信 抵制学术不端倡议书 [EB/OL]. [2018-1-10]. http://www.cma.org.cn/attachment/20161229/1482975418267.pdf.

[6] Shivakumar N, Garcia-Molina H. The SCAM Approach to Copy Detection in Digital Libraries[J]. 1995, 11(5).

[7] Shivakumar N, Garciamolina H. Building a scalable and accurate copy detection mechanism[C]. 1996:160-168.

[8] Garciamolina H, Gravano L, Shivakumar N. dSCAM: Finding Document Copies Across Multiple Databases[C]. International Conference on Parallel and Distributed Information Systems. IEEE, 1996:68-79.

[9] Si A, Hong V L, Lau R W H. CHECK:a document plagiarism detection system[C]. SAC '97: Proceedings of the 1997 ACM Symposium on Applied Computing. 1997:70-77.

[10] Song Q B, Yang X R, Shen J Y, et al. Detection algorithm for the illegal copying and distributing of digital goods[J]. Chinese Journal of Computers, 2002.

[11] Bao J P, Shen J Y, Liu H Y, et al. A fast document copy detection model[J]. Soft Computing, 2006, 10(1):41-46.

[12] Manber U. Finding Similar Files in a Large File System[C]. Usenix Winter Technical Conference.

doi:10.3772/j.issn.2095-915x.2018.03.012

1994:1--10.
[13] Brassil J T, Low S, Maxemchuk N F, et al. Electronic marking and identification techniques to discourage document copying[J]. Selected Areas in Communications IEEE Journal on, 1994, 13(8):1495-1504.
[14] Heintze N. Scalable Document Fingerprinting[C]. 1996 USENIX workshop on electronic commerce. 1996:6-9.
[15] Broder A. On the Resemblance and Containment of Documents[C]. Compression and Complexity of Sequences. Positano, Salerno, Italy, 1997:21.
[16] Wise M J. YAP3:improved detection of similarities in computer program and other texts[C]. Twenty-seventh Sigcse Technical Symposium on Computer Science Education. ACM, 1996:130-134.
[17] Zaslavsky A, Schmidt H. MatchDetectReveal: finding overlapping and similar digital documents[C]. Information Resources Management Association International Conference on Challenges of Information Technology Management in the Century. IGI Global, 2000:955-957.
[18] 陈涛, 谢阳群. 文本分类中的特征降维方法综述[J]. 情报学报 , 2005, 24(6): 690-695.
[19] Bingham E, Mannila H. Random projection in dimensionality reduction: applications to image and text data[C]. Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2001: 245-250.
[20] Broder A Z, Charikar M, Frieze A M, et al. Min-Wise Independent Permutations[J]. Journal of Computer & System Sciences, 2000, 60(3):630-659.
[21] Charikar M S. Similarity estimation techniques from rounding algorithms[C]. Proceedings of the 34th Annual ACM Symposium on Theory of computing. ACM, 2002: 380-388.
[22] Li P, König C. b-Bit minwise hashing[C]. Proceedings of the 19th International Conference on World Wide Web. ACM, 2010: 671-680.
[23] Bergsma S, Post M, Yarowsky D. Stylometric analysis of scientific articles[C]. Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2012: 327-337.
[24] Abbasi A, Chen H. Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace[J]. ACM Transactions on Information Systems (TOIS), 2008, 26(2): 7.
[25] Zhang J, Sun Y, Wang H, et al. Calculating statistical similarity between sentences[J]. Journal of Convergence Information Technology, 2011, 6(2):22-34.
[26] Zhang J, Sun Y, Bie R. Measuring Semantic-Based Structural Similarity in Multi-Relational Networks[M]. IGI Global, 2016.
[27] 曾文 , 徐红姣 , 李颖 , 等 . 基于 VSM 的科技期刊文献与专利文献的相似度计算方法研究 [J]. 情报工程 , 2016, 2(3):37-42.
[28] 李岩 , 徐硕 , 吴广印 , 等 . 一种采用 SpotSigs 算法的中文新闻网页相似性检测方法 [J]. 情报工程 , 2017, 3(3):116-124.
[29] 史彦军 , 滕弘飞 , 金博 . 抄袭论文识别研究与进展 [J]. 大连理工大学学报 , 2005, 45(1):50-57.
[30] 北京万方数据股份有限公司 . 万方数据知识服务平台 [EB/OL] .[2017-12-12]. http://www.wanfangdata.com.cn.
[31] 中国知网 . [EB/OL]. [2017-12-16]. http://www.cnki.net.
[32] 李丹 . 学术不端检测系统的理性使用 [J]. 学术界 , 2012(12):129-133.
[33] 窦学俊 , 高巧林 , 栾世杰 , 等 . 山东省部分高校论文学术不端检测分析及学术不端检测系统使用体会 [J]. 中国中医药图书情报杂志 , 2017, 41(6):18-22.
[34] 于晓舟 , 张俐 , 郭嘉 . 应用 TMLC 检测博士学位论文质量的成效分析 —— 以中国地质大学 (武汉) 为例 [J]. 安徽农业大学学报 (社会科学版), 2010, 19(2):130-133.
[35] 胡政平 . 学术不端文献检测与期刊质量的提升 [J]. 出版发行研究 , 2012(11):87-89.
[36] 赵晓兰 . 学术不端文献检测系统应用现状评析 [J]. 中国出版 , 2014(19):34-36.
[37] 杨晨晨 . 运用学术不端文献检测系统检测医学论文存在的问题及对策 [J]. 编辑学报 , 2014, 26(1).
[38] 李永莲 . 学术不端文献检测系统的检测盲区研究 [J]. 青岛职业技术学院学报 , 2014(5):78-80.
[39] 中华人民共和国教育部 . 中华人民共和国教育部令第 34 号《学位论文作假行为处理办法》自 2013 年 1 月 1 日起施行 .[EB/OL]. [2012-11-13]. [2018-1-20]. http://www.moe.edu.cn/srcsite/A02/s5911/moe_621/201211/t20121113_170437.html.
[40] 王宇 . 学术不端文献检测系统研究综述 [J]. 西南民族大学学报 (人文社科版), 2013, 34(4):236-240.