

基于著者共现的 CBM 机构名称规范研究^{*}

吴英杰 李军莲 孙海霞 王 蕾 陈 颖

(中国医学科学院医学信息研究所 北京 100020)

[摘要] 以中国生物医学文献数据库收录的期刊论文为数据源,从发文著者这个重要关联因素着手,开展基于共现分析的著者机构名称规范机制研究,提出基于著者共现的同一机构发现算法,经评测该算法推荐结果准确率达 85%,具有较好的可参考性。

[关键词] 机构规范;共现分析;中国生物医学文献数据库;著者共现

[中图分类号] R-056 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2019.12.011

Study on CBM Affiliation Name Normalization Based on Co-occurrence of Authors WU Yingjie, LI Junlian, SUN Haixia, WANG Lei, CHEN Ying, Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing 100020, China

[Abstract] Taking journal articles included in Chinese Biomedical Database (CBM) as the data source, the paper begins with the important associated factor of publication author to implement the normalization mechanism study on affiliation name of authors based on co-occurrence analysis, and puts forward the algorithm of the same institution discovery on the basis of author co-occurrence. Recommended results of the algorithm is up to 85% by evaluation and measurement, which has better reference value.

[Keywords] affiliation name normalization; co-occurrence analysis; China Biology Medicine disc (CBM); author co-occurrence

1 引言

学术论文是科研评价研究和实践的主要依据之一。精准统计机构发文与被引情况则是客观准确地开展机构科研评价的基础与关键。开展学术论文著

者机构规范控制研究,实现同一机构不同表达形式的汇聚,能够最大程度避免由于机构更名、合并、拆分和不同著者对同一机构名使用不同表达形式等原因影响机构论文查准率和查全率,从而优化科研机构学术评价^[1-2]。基于科技文献的机构名称规范任务一般包括两个核心环节:一是从著者单位著录项中识别出机构名称。在著者提交的论文及科技文献数据库中科研机构名称一般连同所在城市和邮编出现在著者单位著录项中。二是在同一机构不同名称表现形式间建立映射,实现同一机构名称不同表现形式的汇聚。由于数据库中著者单位著录具有一定的结构性和规范性,前者难度相对较低,相关研究更多聚焦于后者,以相似度聚类技术为基础来实现不同名称同一机构的自动汇聚规范,尽可能地减轻人工负担^[1]。

聚类是将物理或抽象对象集合按有关特性的相

[修回日期] 2019-09-27

[作者简介] 吴英杰,工程师;通讯作者:李军莲,副研究馆员。

[基金项目] 中国医学科学院医学与健康科技创新工程项目“生物医学科技信息支撑平台”(项目编号:2016-I2M-2-005);中国医学科学院医学信息研究所基本科研业务专项“基于共现分析的著者机构名称规范机制研究”(项目编号:2016RC33006)。

似程度进行分组的过程,目的是使同一簇中对象特性尽可能地相似,而不同簇对象之间的差异尽可能地增大^[3-4]。聚类效果与相似度量特征密切相关。在关于机构名称规范的众多研究中基于字符串、规则及二者相结合的相似度计算方法探讨较多^[5],已成为名称表达差异不大的同一机构识别的常用方法。中国生物医学文献数据库(China Biology Medicine disc, CBM)的机构名称规范早期即通过字符串相似度计算方法对大规模机构名称进行归一化处理,之后在长期实践中持续探讨优化规则与字符串相似度结合的混合方法,不断改进规范效果,为基于 CBM 的机构规范文档建设提供较好的技术支撑。然而由于上述方法主要基于机构名称表达的共性特征,无法有效解决表达形式差异较大的同一机构的识别与规范问题,即机构变革导致的名称变化、同一单位多个名称等,如无法识别“首都医科大学第六临床医学院”与“北京安贞医院”是同一机构。对此,有学者引入机构网页统一资源定位(Uniform Resource Locator, URL)特征,通过 URL 与机构名的共现情况来判断机构名称的相似度,取得了一定效果^[6-7]。受该研究启发,本文将从发文著者这个重要关联因素着手,开展基于著者共现的 CBM 机构名称规范机制研究,尝试从“机构-著者”维度探索因名称表达差异较大带来的同一机构漏归、错归问题,以减少机构规范过程中的人工干预,提升自动规范效果。

2 机构名称规范技术研究现状

在机构名称规范中,依据采用相似度量特征的不同可划分为基于字符串相似度方法、基于规则的相似度匹配方法、基于共现的关联统计方法以及综合考虑多因素的混合方法^[5,8-11]。(1)基于字符串相似度方法。基本思路是将机构名称字符串看作是字符序列,字符序列间相同字符越近表明这两个字符串越相似,二者指向同一机构实体的可能性越大。如 French JC 等^[11]先后采用 Hall-Dowling 编辑距离算法和 Jaccard 系数法进行天体物理数据系统(Astrophysics Data System, ADS)作者机构名称规

范文档半自动化构建研究; Jacob F 等^[12]采用 Levenshtein 编辑距离法对求职简历中的求职者单位名称进行匹配计算与规范; Jiang Y 等^[13]采用归一化压缩距离(Normalized Compression Distance, NCD)聚类算法实现同一机构多种名称的高效集成。(2)基于规则的方法。主要思想是根据机构名称构词特点建立一定的规则库,通过规则进行可能匹配的候选名称字符串识别及错误匹配过滤。如 Huang S 等^[14]、杨波等^[15]在利用 WOS 数据研究机构名称聚类中综合字面相似度、字长、字顺、子串、地区等特征信息构建识别可能匹配的机构名称对规则,然后基于机构对匹配频率进行错误匹配过滤。(3)基于共现的关联统计方法。主要是利用 Web 语料,通过计算不同机构名称字符串网络搜索结果中 URL 的共现情况来判定机构名称的相似度^[16-17]。如 Aumuller D 等^[17]基于 Google 和 Yahoo 搜索返回的 Top k 个 URL 的共现重叠情况来计算两个机构名称匹配程度,同时参考 TF-IDF 模型对共现 URL 的排序位置进行加权。(4)综合考虑多因素的混合方法。主要是结合规则、加权统计来克服单纯字面相似度匹配方法的不足。如 Jonnalagadda SR 等^[16]在开展 PubMed 数据库机构名称规范研究中,通过引入世界地区/邮编字典及机构-地区/邮编一致性规则来过滤错误匹配,提升相似度判断效果;贾君枝等^[17]在科研机构名称归一化研究中,通过构建机构特征词表和相关规则进行机构名称分级识别,再基于编辑距离算法、TF-IDF 及 K-means 算法实现机构名称归一,有效提高聚类的准确率和召回率。

3 CBM 机构规范文档建设现状

3.1 建设进展

机构规范文档建设主要解决同一机构因表达形式不一致、名称变更、隶属关系不清等带来的成果分散问题^[5]。基于 CBM 的机构规范工作始于 2010 年。经过近两年的数据分析和算法研究,从 2012 年起正式启动规范处理工作。目前机构规范工作组已完成千万篇文献所涉机构数据(超过 322 万条)的清洗,经拆分、去噪、去重、标识机构类型、所

在省区等处理，形成近 60 万条机构信息；通过优化规则与字符串相似度结合的混合方法对大规模机构名称进行归一化处理，形成各类规范机构 17 万条。其中医院类规范机构 8 万多条，高等院校类机构 3.9 万条，实验室机构 4 000 多条，其他机构 4.8 万条。这些数据已经成为 CBM 机构检索、分析等功能的重要支撑^[18-19]。

3.2 现存问题分析

尽管目前的规范文档已完成绝大部分机构的同义规范，但仍存在漏归、错归等情况，具体原因分析，见表 1。可以发现同一机构不同名称形式漏归、错归的主要原因是这些名称形式与同机构的其他名称相似度不高或极低，仅靠相似度计算很难被发现；而基于著者共现分析角度探索漏归、错归同一机构的发现问题，从理论上分析具有较好的可信度，将是对相似度同一机构发现方法的有效补充。

表 1 同一机构不同名称形式漏归、错归原因分析

漏归、错归常见原因	举例
同一机构的不同名称表达形式完全不同	“复旦大学附属中山医院”又称“上海市中山红十字医院”；“安徽省立医院”又称“中国科学技术大学附属第一医院”
三位一体的研究机构，因类型不同被分散	首都医科大学口腔医院（医院类机构）；首都医科大学口腔医学院（高校类机构）；首都医科大学口腔医学研究所（研究类机构）
医院与临床学院之间名称差异明显且机构类型不一致	首都医科大学附属北京安贞医院（医院类机构）；首都医科大学第六临床医学院（高校类机构）
名称变更造成机构名称间差异较大	“二炮总医院”变更为“火箭军总医院”；“辽宁医学院”变更为“锦州医科大学”
全称与简称之间相似度较低	国家卫生和计划生育委员会（简称“卫计委”）

4 基于著者共现的机构名称规范机制研究

4.1 发文著者机构特征分析

一般而言，发文著者的机构主要有两种情况：单机构和多机构。单机构指著者机构为某单一机构，如胡 XX - 中国医学科学院医学信息研究所，

胡 XX - 中国医学科学院医学情报研究所。当某著者所在机构包括多个机构名称时，这多个机构可能是某机构对外的不同名称，也可能确是多个机构但仅是存在某些关系而已，如孔 XX - 武汉大学人民医院 湖北省人民医院，赵 XX - 三峡大学第一临床医学院 宜昌市中心人民医院，程 XX - 蚌埠医学院药学院 安徽省生化药物工程技术研究中心。CBM 中机构著录主要包括机构实体和地址信息，二者之间用逗号分隔。其中机构实体部分主要包括机构名称、科室/学系等信息，机构地址部分包括机构所在省市和邮编信息。当某著者机构为多机构时，则在“机构（AD）”字段项著录多个机构的信息，多个机构名称间用“/”或空格进行分隔。如 AD 字段为“国家癌症中心“/”中国医学科学院、北京协和医学院肿瘤医院，北京 100021”、“首都医科大学附属北京安贞医院 首都医科大学第六临床医学院超声诊断科，北京 100029”，即描述了多个机构名称信息，见图 1。CBM 收录文献时遵照客观事实对发文著者及其所在机构进行全面揭示，也就是说 CBM 不仅描述著者机构的基本信息，还对著者与机构的对应关系进行规范揭示，为基于著者共现开展同一机构发现提供重要数据基础。

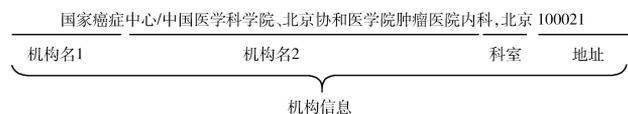


图 1 著者机构著录形式样例

4.2 方案设计

基于著者共现开展同一机构/相关机构发现研究主要是通过计算机构间发文著者的共现次数与重合度，实现一个机构不同表达形式的发现与推荐。其基本思想是：在一个地区，若两机构共有的发文著者越多则二者为同一机构/相关机构的可能性越大。具体方案，见图 2。首先利用 CBM 数据库的著者 - 机构关系构建机构 - 著者空间，在此基础上分地区计算各机构间的著者共现关联度，结合机构类型维度从高到低排序，著者共现关联度高的机构将被推荐为高相似度机构。方案的核心是基于机构的

发文著者量、不同机构间的共有发文著者量计算机构的著者共现关联度。因其主要测度机构发文著者集合间的包容度，不关注个体著者间的相关关系，也不受著者发文多少的影响，故采用相对简约的包容指数^[20]进行计算，具体计算公式为：

$$I_{ij} = C_{ij} / \min(C_i, C_j)$$

其中 C_{ij} 是文献集中机构对 (O_i 和 O_j) 共现的发文著者数； C_i 是机构 i (O_i) 在文献集中出现的发文著者数； C_j 是机构 (O_j) 在文献集中出现的著者数； $\min(C_i, C_j)$ 取 C_i, C_j 中的最小值。

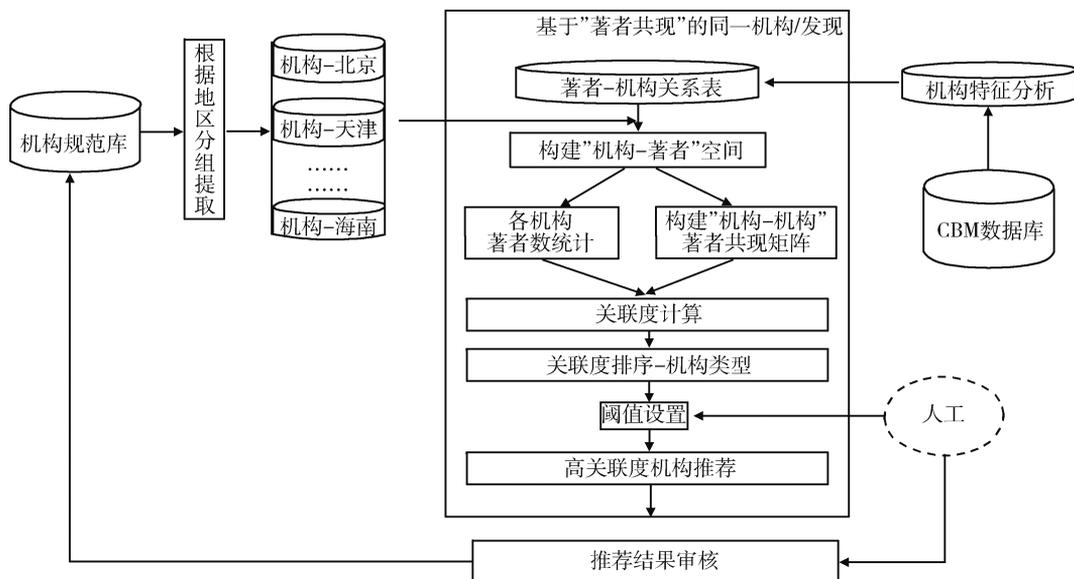


图 2 整体研究思路

4.3 处理流程

包括以下 8 个步骤：(1) 以机构规范库中各机构为处理对象，依据所在地区进行分组提取，生成地区机构子集。(2) 基于 CBM 数据库，结合发文机构所在地区、文献发表年等维度特征构建相应的地区著者-机构关系表。(3) 将著者按机构进行分组，构建“机构-著者空间”。(4) 统计各机构发文著者数。(5) 若某两个机构有一个相同著者，即表示该两个机构基于著者共现一次，依此遍历生成机构-机构著者共现矩阵。(6) 计算机构两两之间的“著者共现”关联度。(7) 结合机构类型对关联度按从高到低进行排序。(8) 结合人工评测初步划定输出阈值，对高关联度机构进行推荐。

5 效果评测

5.1 测试对象

初步遴选北京、天津、上海 3 个地区 82 家机构为

测试对象，其中北京 27 家、天津 35 家、上海 20 家。

5.2 统计数据源

CBM 2010 - 2016 年北京、天津、上海上述机构发表的所有文献数据 10.3 万条。

5.3 评测思路

基于提出的著者共现同一机构/相关机构发现算法，尝试在指定统计数据源中对各地区遴选医院的著者共现情况进行统计，根据共现频率计算机构间著者共现关联度，分析预设推荐阈值并对高关联度机构推荐结果进行输出。通过人工判定对推荐结果的准确性进行分析评价。

5.4 评测结果

经统计共 2 100 个机构对间存在著者共现情况，部分数据示例，见图 3。随机遴选 300 个机构共现对，对比分析阈值 <0.5 、 <0.4 、 <0.3 、 <0.2 及 0.1 的误判率，从希望较高准确性的角度考虑，各

段阈值内推荐结果的误判率，见图4。预设推荐阈值为 ≥ 0.4 ，共输出90个机构共现对。

ad_pre_et_a	ad_pre_et_b	C _{1j}	C _i	C _j	min(C _i , C _j)	I _{ij}
安徽医科大学解放军307临床学院	北京市心肺血管疾病研究所	1	30	1751	30	0.0333
安徽医科大学解放军307临床学院	解放军总医院	8	30	9650	30	0.2667
安徽医科大学解放军307临床学院	中国人民解放军总医院	3	30	4631	30	0.1
安徽医科大学解放军307临床学院	北京大学第一医院	3	30	4736	30	0.1
安徽医科大学解放军307临床学院	北京大学人民医院	2	30	4077	30	0.0667
安徽医科大学解放军307临床学院	首都医科大学附属北京胸科医院	2	30	811	30	0.0667
安徽医科大学解放军307临床学院	首都医科大学附属北京儿童医院	2	30	2063	30	0.0667
安徽医科大学解放军307临床学院	国家心血管病中心	2	30	1567	30	0.0667
安徽医科大学解放军307临床学院	空军总医院	1	30	2000	30	0.0333
安徽医科大学解放军307临床学院	北京大学第三医院	1	30	4077	30	0.0333
安徽医科大学解放军307临床学院	中国医学科学院阜外心血管病医院	1	30	804	30	0.0333
安徽医科大学解放军307临床学院	解放军307医院	12	30	541	30	0.4
安徽医科大学空军临床学院	北京市心肺血管疾病研究所	7	184	1751	184	0.038
安徽医科大学空军临床学院	北京大学肝病研究所	1	184	41	41	0.0244
安徽医科大学空军临床学院	解放军总医院	31	184	9650	184	0.1685
安徽医科大学空军临床学院	中国人民解放军总医院	18	184	4531	184	0.0978
安徽医科大学空军临床学院	北京大学第一医院	17	184	4786	184	0.0924
安徽医科大学空军临床学院	北京大学人民医院	15	184	4077	184	0.0815
安徽医科大学空军临床学院	北京大学第三医院	14	184	4077	184	0.0761
安徽医科大学空军临床学院	北京大学人民医院	9	184	1985	184	0.0489
安徽医科大学空军临床学院	首都医科大学附属北京儿童医院	8	184	2063	184	0.0435
安徽医科大学空军临床学院	解放军307医院	7	184	541	184	0.038
安徽医科大学空军临床学院	解放军总医院	5	184	1032	184	0.0272
安徽医科大学空军临床学院	国家心血管病中心	5	184	1567	184	0.0272
安徽医科大学空军临床学院	首都医科大学附属北京胸科医院	5	184	811	184	0.0272
安徽医科大学空军临床学院	北京大学第六医院	5	184	453	184	0.0272
安徽医科大学空军临床学院	清华大学第一附属医院	4	184	744	184	0.0217
安徽医科大学空军临床学院	北京安贞医院	3	184	336	184	0.0163
安徽医科大学空军临床学院	清华大学第一附属医院	1	184	89	89	0.0112
安徽医科大学空军临床学院	中国医学科学院阜外心血管病医院	1	184	804	184	0.0054
安徽医科大学空军临床学院	空军总医院	154	184	2000	184	0.837
北京安贞医院	解放军总医院	53	336	9650	336	0.1577
北京安贞医院	北京大学人民医院	36	336	4077	336	0.1071
北京安贞医院	北京大学第三医院	33	336	4077	336	0.0982
北京安贞医院	中国人民解放军总医院	32	336	4531	336	0.0952
北京安贞医院	北京大学第一医院	29	336	4786	336	0.0863
北京安贞医院	国家心血管病中心	21	336	1567	336	0.0625
北京安贞医院	北京大学第一临床医学院	1	336	19	19	0.0056
北京安贞医院	中国医学科学院阜外心血管病医院	18	336	804	336	0.0536
北京安贞医院	空军总医院	16	336	2000	336	0.0476
北京安贞医院	北京大学第三医院	15	336	1985	336	0.0446
北京安贞医院	首都医科大学附属北京儿童医院	14	336	2063	336	0.0417
北京安贞医院	解放军307医院	8	336	541	336	0.0238
北京安贞医院	火箭军总医院	2	336	89	89	0.0225
北京安贞医院	二炮总医院	1	336	45	45	0.0222
北京安贞医院	解放军第302医院	6	336	1082	336	0.0179
北京安贞医院	安徽医科大学空军临床学院	3	336	184	184	0.0163
北京安贞医院	清华大学第一附属医院	5	336	744	336	0.0149
北京安贞医院	北京市心肺血管病研究所	196	336	1751	336	0.5833

图3 著者共现关联度计算——北京地区部分数据示例

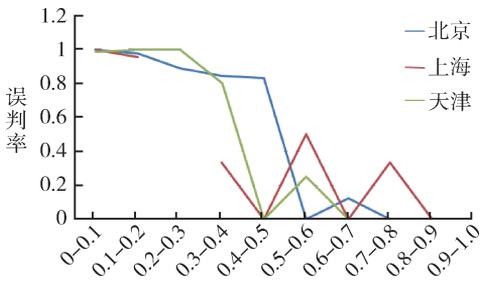


图4 各段阈值推荐结果误判率

机构i	机构j	C _{1i}	C _i	C _j	min(C _i , C _j)	E _{ij}	人工分析判断	
第二军医大学附属上海医院	第二军医大学附属上海医院	5	3181	5	5	0	0.9333	是
第二军医大学附属长海医院	长海医院	39	3181	49	49	0	0.1959	是
第二军医大学附属长征医院	上海交通大学附属第六人民医院	4	5	3457	5	0	0.8	否
第二军医大学附属长征医院	上海交通大学附属长征医院	2	5	1124	5	0	0.4	是
第二军医大学附属上海医院	第二军医大学附属长海医院	5	6	3181	6	0	0.8333	是
第二军医大学附属长征医院	第三军医大学新桥医院	6	6	2898	6	1	0.8	是
第三军医大学新桥医院	第三军医大学新桥医院	6	2898	6	6	0	1	是
复旦大学附属上海浦东五人民医院	上海市公共卫生临床中心	114	175	187	167	0.0826	是	
复旦大学附属中山医院	上海市公共卫生临床中心	120	178	547	178	0.0709	是	
复旦大学附属中山医院	复旦大学附属中山医院	3	7	20	7	0.4286	是	
复旦大学附属中山医院	复旦大学附属中山医院	3	20	7	7	0.4286	是	
解放军后勤学院附属部队	武警总医院附属部队	4	5	784	5	0.8	是	
解放军后勤学院附属部队	中国人民解放军第二五四医院	2	5	142	5	0.4	否	
北京安贞医院	北京安贞医院	14	18	697	18	0.1778	是	
北京安贞医院	北京市心肺血管病研究所	199	336	1751	336	0.5833	是	
北京大学第三医院	北京大学第三医院	11	16	4077	16	0.6875	是	
北京大学第三医院	北京大学第三临床医学院	11	4077	16	16	0.6875	是	
北京大学第一临床医学院	北京大学第一临床医学院	12	18	4786	18	0.6667	是	
首都医科大学第六临床医学院	首都医科大学第六临床医学院	1	2	1751	2	0.5	否	
首都医科大学第六临床医学院	首都医科大学第六临床医学院	1	2	4786	2	0.5	否	
首都医科大学附属北京儿童医院	首都医科大学附属北京儿童医院	126	2063	143	143	0.8511	是	
中国人民解放军第三〇二医院	解放军302医院	325	470	1082	470	0.6915	是	
中国人民解放军总医院	解放军总医院	3408	4531	9650	4531	0.1522	是	
中国人民解放军总医院	国家心血管病中心	1	4531	2	2	0.5	否	
中国医学科学院阜外心血管病医院	国家心血管病中心	419	804	1567	804	0.5211	是	

图5 机构著者共现关联度计算——专家判定部分数据示例

5.5 结果分析

人工测评结果，见表2。通过统计分析可以看出输出的90个机构共现对中，人工认为77个机构对是同一机构或高关联度机构，部分数据示例，见图5。推荐结果的准确率为85%，具有较好的可参考性。

表2 机构著者共现关联度计算人工测评结果

地区	共现机构对数量	人工审核正确数量	准确率(%)
北京	30	24	80
上海	32	28	87.5
天津	28	25	89.3
合计	90	77	85

6 结语

本文基于目标应用环境中中国生物医学文献数据库，从著者这个重要关联因素着手开展基于共现分析的机构名称规范机制研究，提出基于著者共现的同一机构/相关机构发现算法。经测试评估，该算法推荐结果具有较高的可信度，是对当前相似度同一机构发现方法的有益补充。后续将进一步优化算法，实现过程中的数据处理细节问题，尽快推进其在CBM机构规范文档建设中的实践应用。

参考文献

- 1 Khalid MA, Jijkoun V, Rijke MD. The Impact of Named Entity Normalization on Information Retrieval for Question Answering [EB/OL]. [2019-01-10]. https://link.springer.com/chapter/10.1007%2F978-3-540-78646-7_83.
- 2 李慧佳, 马建玲, 张秀秀, 等. 中文机构名称规范库建设的实践与分析——以“中科院机构名称规范库”建设为例 [J]. 图书与情报, 2016 (1): 133-139.
- 3 Jain AK, Dubes RC. Algorithms for Clustering Data. Prentice-hall Advanced Reference Series [M]. London: Prentice Hall College Div, 1988.
- 4 刘远超, 王晓龙, 徐志明, 等. 文本聚类综述 [J]. 中文信息学报, 2006, 20 (3): 55-62.
- 5 孙海霞, 李军莲, 吴英杰. 基于 K-means 的机构归一化研究 [J]. 医学信息学杂志, 2013, 34 (7): 41-44.
- 6 Bollegala D, Ishizuka M, Matsuo Y. Measuring Semantic Similarity between Words Using Web Search Engines [C]. Banff: Proceeding of the 14th International Conference on World Wide Web. 2007: 757-766.
- 7 Aumuller D, Rahm E. Web-based Affiliation Matching [C]. Potsdam: Proceeding of International Conference on Information Quality, 2009: 246-256.
- 8 Treeratpituk P, Giles CL. Disambiguating Authors in Academic Publications Using Random Forests [C]. USA: International Conference on Digital Libraries, 2009: 39-48.
- 9 邱均平, 董克. 作者共现网络的科学研究结构揭示能力比较研究 [J]. 中国图书馆学报, 2014, 40 (1): 15-24.
- 10 Abramo G, D'Angelo CA, Pugini F. The Measurement of Italian Universities' Research Productivity by a Non-parametric Bibliometric Methodology [J]. Scientometrics, 2008, 76 (2): 225-244.
- 11 French J C, Powell A L, Schulman E. Using Clustering Strategies for Creating Authority Files [J]. Journal of the American Society for Information Science, 2000, 51 (8): 774-786.
- 12 Jacob F, Javed F, Zhao M, et al. sCooL: a system for academic institution name normalization [EB/OL]. [2019-01-10]. http://xueshu.baidu.com/usercenter/paper/show?paperid=eceea4a5f6063289e1bf9e4533547dd0&site=xueshu_se.
- 13 Jiang Y, Zheng H T, Wang X, et al. Affiliation Disambiguation For Constructing Semantic Digital Libraries [J]. Journal of the American Society for Information Science and Technology, 2011, 62 (6): 1029-1041.
- 14 Huang S, Yang B, Yan S, et al. Institution Name Disambiguation for Research Assessment [J]. Scientometrics. 2014, 99 (3): 823-838.
- 15 杨波, 杨军威, 阎素兰. 基于规则的机构名称规范化研究 [J]. 现代图书情报技术, 2015, 31 (6): 57-63.
- 16 Jonnalagadda S R, Topham P. NEMO: extraction and normalization of organization names from PubMed affiliation string [J]. Journal of Biomedical Discovery and Collaboration, 2010, 5 (1): 50-75.
- 17 贾君枝, 曾建勋, 李捷佳, 等. 科研机构名称归一化实现 [J]. 图书情报工作, 2018, 62 (13): 103-110.
- 18 孙海霞, 李军莲. 知识服务驱动的 CBM 作者机构规范文档构建 [C]. 北京: 中国医学科学院/北京协和医学院医学信息研究所/图书馆 2014 年学术年会, 2015.
- 19 吴英杰, 孙海霞. CBM 数据库作者机构非规范著录数据自动检测研究 [J]. 医学信息学杂志, 2011, 32 (5): 38-40.
- 20 宋爽. 共现分析在文本知识挖掘中的应用研究 [D]. 南京: 南京理工大学, 2006.

关于《医学信息学杂志》启用

“科技期刊学术不端文献检测系统”的启事

为了提高编辑部对于学术不端文献的辨别能力, 端正学风, 维护作者权益, 《医学信息学杂志》已正式启用“科技期刊学术不端文献检测系统”, 对来稿进行逐篇检查。该系统以《中国学术文献网络出版总库》为全文比对数据库, 可检测抄袭与剽窃、伪造、篡改、不当署名、一稿多投等学术不端文献。如查出作者所投稿件存在上述学术不端行为, 本刊将立即做退稿处理并予以警告。希望广大作者在论文撰写中保持严谨、谨慎、端正的态度, 自觉抵制任何有损学术声誉的行为。

《医学信息学杂志》编辑部