

近五年国际医学信息学研究热点与前沿分析 ——基于22种高影响力外文期刊的文献计量研究

□ 秦方 刘硕 张士靖 / 华中科技大学同济医学院医药卫生管理学院 武汉 430030

摘要: 文章结合2010版JCR中Medical Informatics目录下的22种高影响力的外文期刊作为研究对象,以2007年到2011年为研究时限,并利用TDA、SPSS12.0、CiteSpaceII等统计分析工具,结合共现词聚类、突变术语检测等文献计量方法,分析了医学信息学的研究热点与前沿。认为信息技术应用与卫生信息分析方法是近五年来医学信息学的两大研究维度,且存在相互促进的关系;医学信息学在近五年间经历了从信息技术应用、卫生信息分析方法繁荣期到以新技术应用、新方法发明为主题的转型期,再到2011年的平稳期,今后的发展趋势可能是以服务卫生保健为目标,开发新的功能。

关键词: 医学信息学, 期刊, 研究热点, 研究前沿, 国际, 文献计量

DOI: 10.3772/j.issn.1673—2286.2012.09.009

1 前言

医学信息学(Medical Informatics)的概念最早起源于上世纪50年代,但是作为一门学科在20世纪70年代才被正式提出。近20年来,伴随计算机、网络、数据库等信息技术的发展,医学信息学的学科内涵日益丰富,研究领域日趋广泛,形成了大量有价值的研究成果,新的研究热点也不断出现。然而,医学信息学至今没有统一的概念,没有明确的研究范围,与其他学科存在广泛的交叉综合^[1],因此,通过追踪学科前沿,把握研究热点对促进该学科的研究、发展和教育具有重要的意义。

近年来,不少学者已经通过文献计量学方法对国际医学信息学的期刊进行了热点研究。如崔雷和陈东滨^[2]选取1990~2006年间的4种高

影响力医学信息学期刊进行引文聚类,得到医学信息学的4个研究重点;许丹、侯跃芳^[3]选取JCR中的7种核心医学信息学期刊,以期刊被数据库收录的年份至2010年为研究时限,利用可视化引文分析方法得出医学信息学的主要研究内容;Schuemie M. J等人^[4]运用N-grams方法对1993~2008年间Medline收录的16种医学信息学期刊进行聚类分析,也得出了医学信息学3大重点研究领域。但是,上述学者选取的研究对象相对偏少,研究的年限也相对较早。尽管如此,从其研究结果看医学信息学发展有着明显的动态性特点。

因此,为了进一步跟踪医学信息学的发展动态,更全面地了解医学信息学近年的研究热点,本文拟将2010版JCR收录的22种高影响力

期刊为研究对象,利用文献计量工具,结合共词聚类等文献计量学方法,分析医学信息学近五年的研究热点与前沿,以期厘清学科现阶段的主要研究内容与研究热点,为同行学者把握该学科未来研究方向,制订适应该学科发展的培养目标提供可资借鉴的参考和帮助。

2 数据来源

以2010版JCR中Medical Informatics学科目录下的22种高影响力期刊为研究对象(见表1),在SCI-E中进行检索,以表1中22种期刊名为检索词,用“or”连接;时间限定为2007~2011年,文献类型选择Article(期刊论文)、Proceedings Paper(会议论文)和Review(综述)三种,检索时间为

表1 医学信息学22种高影响力期刊(2010年)

序号	期刊名	出版国	2010 IF
1	Journal of Medical Internet Research	加拿大	4.663
2	Journal of The American Medical Informatics Association	美国	3.088
3	IEEE Engineering in Medicine and Biology Magazine	美国	2.844
4	Statistics in Medicine	英国	2.328
5	International Journal of Medical Informatics	爱尔兰	2.244
6	BMC Medical Informatics and Decision Making	英国	2.239
7	Medical Decision Making	美国	2.013
8	Medical & Biological Engineering & Computing	德国	1.791
9	Statistical Methods in Medical Research	英国	1.768
10	Journal of Biomedical Informatics	美国	1.724
11	IEEE Transactions on Information Technology in Biomedicine	美国	1.707
12	Artificial Intelligence in Medicine	荷兰	1.568
13	Methods of Information in Medicine	德国	1.472
14	International Journal of Technology Assessment in Health Care	英国	1.317
15	Computer Methods and Programs in Biomedicine	荷兰	1.238
16	Journal of Medical Systems	美国	1.064
17	Journal of Evaluation in Clinical Practice	英国	1.003
18	Informatics For Health & Social Care	英国	1.000
19	Cin-Computers Informatics Nursing	美国	0.957
20	Health Information Management Journal	澳大利亚	0.706
21	Journal of Cancer Education	美国	0.682
22	Biomedizinische Technik	德国	0.590

2012年2月29日,共检索到9417条文献。将这些文献的全记录以纯文本形式下载。

3 结果与分析

3.1 研究热点分析

笔者将9417篇文献的作者关键词、机标关键词和标题词(自然语言处理后)组合,用TDA进行数据清洗、合并,取频次Top50的术语生成共现矩阵,然后导入到SPSS12.0中进行系统聚类:类间距算法选

择离差平方和(Ward's Method),测度系数选取计数卡方(Counts: Chi-Squared Measure),得到聚类树图(见图1)。从图1中可以看出,50个高频术语被分成明显的2个大类,6个小类。

从图1中A、B两类的子类成员来看,A类中的术语多为网络、技术等信息技术常见词,而B类中多为试验、回归等统计学词汇。结合每类成员所代表的含义,可以把医学信息学研究热点分为以信息技术应用为代表的A类和以卫生信息分析方法为主要内容的B类。

3.1.1 A类: 信息技术应用

(1) A1: 现代信息技术对个人健康的影响

Cugelman Brian等人利用大规模随机对照试验(Randomized Controlled-Trials, RCT)来研究网络健康信息行为对公众健康习惯的改变^[5]。Meglic M等人对网络干预方法进行评价^[6],说明网络社区和专业健康网站逐渐受到患者和医护人员的青睐,我国研究者应该重视公众网络健康信息查询行为对健康习惯改善影响的研究。

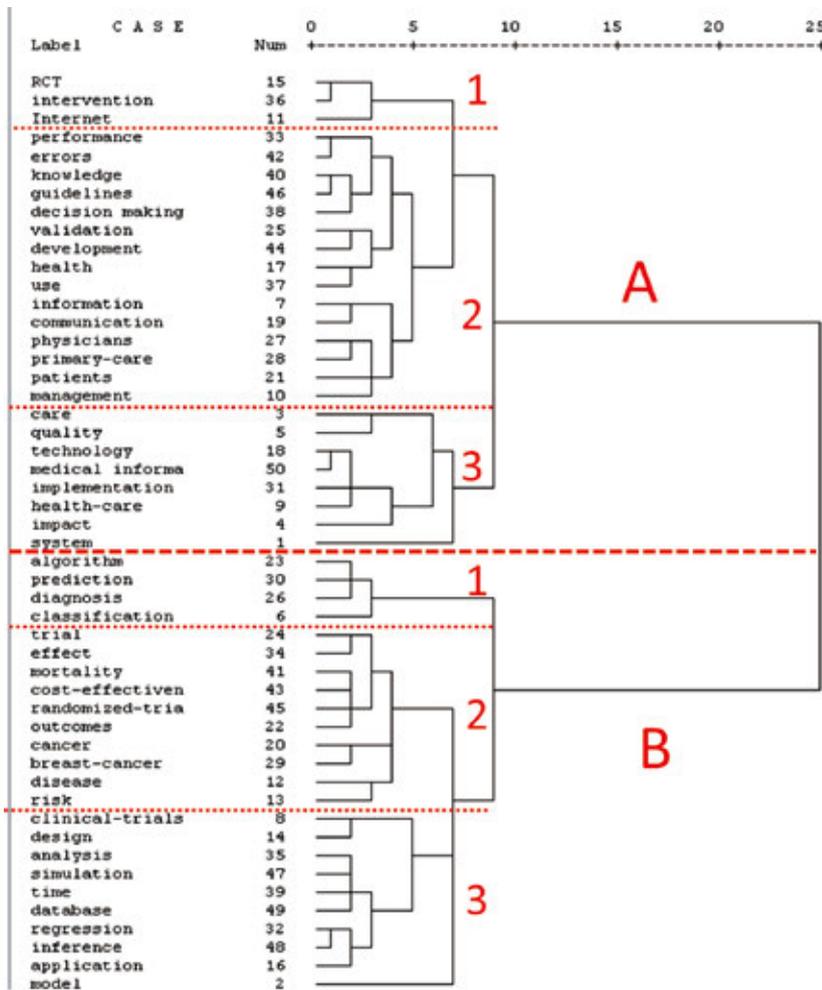


图1 近五年医学信息学高频关键词聚类树图

保健中发挥的作用进行了详细的介绍,为决策者如何评价卫生信息系统效益提供了方法上的支持,说明卫生信息系统在大规模应用之后,其效果评价开始受到研究人员的关注。

3.1.2 B类: 卫生信息分析方法

(1) B1: 算法在生物医学中的应用

Ellenius J和Groth T^[11]针对人工神经网络(Artificial Neural Network, ANN)在临床决策支持中不能很好地解释诊断指标的情况,提出了一个基于ANN算法的可视化诊断指标体系来解决这个问题,说明一个新的算法对解决一类生物医学难题会起到至关重要的作用。

(2) B2: 评估类分析及其应用

在卫生领域,类似于“一种新的治疗方法是否优于传统疗法”的问题层出不穷,评估性分析就是要解决这类问题,其中以成本效益分析(Cost-Effectiveness analysis)最为常见。Basch E^[12]等人评估了癌症病人在线自测化疗毒性网络平台的作用,指出网络平台自测不仅可以提高安全性,还能提升病人对护理的满意度,适合普及建设。

除此之外,Meta分析可以汇集同类研究的多个相互研究结果,进行定量分析来获得一个综合性结论。Covey J^[13]通过对大量已发表的论文和实验数据进行效应量Meta分析,采用三种不同指标对医患共同决策的治疗方法所产生的效果进行了评估。

(3) B3: 关系类分析及其应用

生物医学规律的发现过程往往先是观察到某种现象,然后再运用统计模型进行验证,确定变量之间的关系。Nikoloulopoulos A K和

(2) A2: 卫生信息系统的构建及应用

在A2中,卫生信息系统种类繁多,根据主要功能和应用对象可大致分为以下几种:

① 医院信息系统。其定义比较宽泛,有面向医院管理的门诊系统、临床管理系统,也有面向诊疗服务的医学图像信息系统、医嘱自动录入系统等。

② 决策支持系统。既有面向行政层面的决策支持,比如绩效考核,也有面向医护人员以及病患的决策支持,比如专家诊断系统。

③ 存储和处理病人信息的系统。电子病历(Electronic Medical

Records, EMR)和电子健康档案(Electronic Health Records, EHR)虽然在上世纪90年代就在欧美发达国家开展了深入研究,但至今仍然有很强的研究潜力,比如基于EHR的社区卫生和区域卫生信息平台建设^[7],基于自然语言处理的EMR文本信息挖掘等^[8]。

(3) A3: 对卫生信息系统效益的评估

Uslu AM和Stausberg J^[9]通过文献回顾分析了电子病历的成本和效益问题,并探讨了是否要使用以及如何使用电子病历才是值得的;

Rahimi B和Vimarlund V^[10]对如何评价基于IT技术的信息系统在卫生

Karlis D^[14]发明了多元逻辑固化模型来处理多元二分变量数据,克服了之前Copulas函数只能处理连续型变量的局限性。

关系类分析应用非常广泛,存在变量的地方就有其用武之地。Rohrer J E等人^[15]分析了患者授权的程度(以患者为中心,自我掌握健康的程度)与患者满意度之间的关系;除了研究患者的行为,医生的行为也在研究范围之内。Long A J等人^[16]就对医生在使用医嘱计算机录入系统前后对药品重复提醒的依赖程度进行了分析研究,说明医学模式的转变使越来越多的研究开始关注人的社会行为和疾病存在的关系。

3.2 研究前沿分析

陈超美^[17]认为,使用突然出现的主题术语要比使用出现频次最高的主题词更适合探测学科发展的新兴趋势和突然变化。在CiteSpaceII^[17]中,研究前沿是基于从题目、摘要、叙词和文献记录的标识符中提取出的突变专业术语(burst terms,又称涌现词)而确定的。本研究中,笔者选择节点类型为关键词,共探测到4101个突变专业术语,节点阈值设为每时段出现频次的前1%,之后将其中的高频术语和高突变术语导出,经过去重、排序等人工干预后制成前沿术语表(见表2),从中我们能够探测到近年来医学信息学的研究前沿和发展趋势。

表2显示,众多的主题概念在2007年就已经出现。以网络和数据库为代表的信息技术在医生诊疗、患者就医、预防保健、卫生管理、医学教育等诸多方面大规模应用,各种卫生信息系统也随之普及。决策

支持系统、绩效考核系统、专家诊疗系统等被广泛讨论。相应地,对这些系统的建设实施、方案制定、产出评估、成本效益分析等也成为热点。除了信息技术及其应用之外,卫生信息方法研究也是一大方向,RCT、Meta分析、循证医学、最大期望算法、回归模型等一系列的统计模型和方法被应用到疾病风险管理、治疗效果评估、健康行为改变等领域中,同时,怎样降低统计偏倚,如何根据目的、方法确定恰当的样本量等问题也被研究人员重视起来。另外,全科医学、初级保健和对中风的管理等成为健康管理的新方向。

2008年基本沿袭了2007年的研究方向,没有高突变术语出现,主题概念主要以卫生信息分析方法为主,生存分析开始广泛应用,对流行病学中的可信度、敏感度、分布、概率等问题的讨论也逐渐增多。生物信息学方面,基因表达的研究大量出现;护理信息学和三维图像重建等概念、技术被多次讨论。

2009年有多个高突变术语出现,而且主题概念与2008年有较大变化,数据挖掘和机器学习等情报学方法成为两个新研究点。卫生信息系统面临优化、整合的趋势,各地区的卫生技术评估也相继开展。卫生政策的影响成为医学信息学关注的问题。在健康管理层面,有关个人病史和家庭健康史的管理逐渐增多,心肌梗塞和高血压成为疾病管理的热点。

2010年和2011年属于稳定期,没有太多的突变术语出现。比较新的概念只有数据挖掘领域的支持向量机算法。此外,药物不良事件和电子健康档案重新得到重视。障碍和策略研究成为突变词,预示着传统医学信息学的发展遇到了新问

题,亟需新技术方法的出现来解决这些难题。

4 结论与启示

4.1 医学信息学研究范围不断扩大

高频共现词聚类分析得出近五年来医学信息学研究热点主要集中在信息技术应用与卫生信息分析方法两大维度。在应用维度下,研究人员对现代信息技术改变人们的健康行为的现象关注颇多,并作了大量研究;各种功能与用途的卫生信息系统正在大范围普及,同时针对它们的评估、测试与成本效益分析也成为研究的焦点。在方法维度下,以算法和模型为基础的分析方法不断推陈出新,以适应数量与日俱增、形式纷繁复杂的卫生信息数据;生物信息学算法、评估类分析方法和关系类分析方法是卫生信息分析的三大方向。

研究发现,应用与方法这两大维度不是彼此孤立,而是相互促进,交叉融合的:一方面,技术应用促进了研究方法的跟进,比如卫生信息系统大规模应用后,与之对应的评估方法大量出现;另一方面,新的方法为技术的开发、应用提供了理论保障,比如人工神经网络算法的出现,提升临床决策支持系统的可用性。应用的拓展与方法的创新是支撑医学信息学进步的两大重要因素。

4.2 研究前沿呈现动态性特点

通过利用CiteSpaceII的突变词探测功能,我们可以大概了解

表2 近五年医学信息学研究前沿术语表*

年代	术语	频次	年代	术语	频次
2007	care	保健	2008	depression	表达
	system	系统		survival analysis	生存分析
	Internet	互联网		nurses	护理
	model	模型		epidemiology	流行病学
	quality	质量		reliability	可靠性
	management	管理		sensitivity	敏感度
	risk	风险		reconstruction	重构
	design	设计		distributions	分布
	implementation	实施		probability	概论
	knowledge	知识		2009	health technology assessment (18.24)
	outcomes	产出	hospitals		医院
	performance	绩效	information-systems		信息系统
	physicians	内科医生	data mining		数据挖掘
	diagnosis	诊断	logistic-regression		逻辑回归
	education	教育	optimization		最优化
	randomized controlled-trials	随机对照试验	myocardial infarction		心肌梗塞
	primary-care	初级护理	health policy		卫生政策
	prediction	预测	hypertension		高血压
	cost-effectiveness	成本效益	machine learning (2.68)		机器学习
	evidence-based medicine	循证医学	history (15.73)	病史	
errors	过失	2010	adverse drug events	药物不良事件	
behavior	行为		prostate-cancer (2.91)	前列腺癌	
meta-analysis	元分析		support vector machines	支持向量机	
evaluation	评估		dementia (2.93)	痴呆	
bias (1.81)	偏倚	2011	barriers	障碍	
decision-making	决策		strategies	战略	
databases (3.35)	数据库		electronic health records	电子健康档案	
EM algorithm (2.44)	最大期望算法				
stroke (2.61)	中风				
regression models (2.78)	回归模型				
general practice (2.92)	全科医学				
sample size (2.56)	样本量				

到,从2007年~2008年这两年间,医学信息学的研究方向没有太大的变化,主要还是信息技术应用和

卫生信息分析方法两大维度。2007年已经有了关于两大维度的研究基础,后期是对这两大维度的再

深入研究,2008年卫生信息分析方法研究有了一些变化,2009年则是新技术的引入与老技术应用的

优化。而且医学信息学开始与卫生保健、健康管理、卫生政策有更多的交叉综合,2010和2011年医学信息学没有太大的概念突破与发展,甚至遇到一些瓶颈。总体来说,这五年医学信息学经历了三个时期:2007~2008年的繁荣发展期,2009年的转型期,和2010~2011年的平稳期。

上述动态性表明医学信息学的发展依赖于新方法技术的发明和新应用领域的开辟,尤其是需要后者。医学信息学最早应用于基础研究、临床实践、医院管理,现在扩展到区域卫生信息系统,国家卫生信息网络。笔者预测未来的应用方向可能更多趋向于新功能的实现,比如战略层的卫生决策支持系统开发、更加智能的健康服务网络等,而且其服务对象应该越来越倾向于普通大众。

4.3 启示

综上所述,针对我国医学信息学学科发展与教育现状,笔者提出以下三点建议:

(1) 学科设置要考虑该学科的研究热点

目前,医学信息学的研究热点主要集中在信息技术应用和卫生信息分析方法两大维度。信息技术应用尤其需要计算机科学的理论基础,强调技术突破;卫生信息分析方法以流行病学、卫生统计学、文献计量学等方法为基础,看重方法创新。而这两个方向的知识基础和理论框架存在很大差异。提示我们的医学信息学教育在设置课程体系时,应该关注这两个维度,尽量满足医学信息学的最新发展要求。特别是在研究生培养方向上要有意识地从事技术型和理论研究型两个维度进行培养目标的设计。同样,研究生导师也可以从这两个方面选择自己的研究方向。

(2) 关注细节,使研究结果更具有实际指导意义

从高频关键词热点文献分析可见,发达国家的医学信息学研究非常注重细节和对实际工作的指导意义。例如“大规模随机对照试验研究网络健康行为对公众健康习惯改变的影响”、“网络干预方法的评价”、“基于自然语言处理的EMR文本信息挖掘”、“基于IT技术的信息系统在卫生保健中的作用评价”、“通过Meta分析对医患共同

决策的治疗方法所产生的效果评估”以及“患者授权程度与患者满意度之间的关系”等等,所有的选题都基于一个很具体的研究主题,而且这些研究主题为我们未来在网络健康服务、IT技术应用、卫生信息系统效益分析、诊疗效果评价和患者满意度分析等方面如何去提供提供了很好思路和方法。

(3) 明确学科使命,促进该学科的国际化发展

由于历史原因,与西方发达国家不同,我国的医学信息学教育最早起源于医学图书情报^[18],因此,我们在培养目标、师资结构、课程设置、科学研究等方面与真正意义上的医学信息学差距较大,要使我国的医学信息学研究和教育与国际接轨,必须弄清医学信息学的学科内涵,明确其学科使命——提供更好的个体医疗和保健服务,更好的医疗和保健知识,更好的医疗和健康管理^[19]。我们的研究也证明,目前国际医学信息学的发展趋势正是在该使命的驱动下与卫生保健、健康管理、卫生政策有更多的交叉综合。因此,坚持跟踪国际发展前沿将有助于我们的研究更好地与国际接轨,促进该学科的国际化发展。

参考文献

- [1] Brittain JM, Norris AC. Delivery of health informatics education and training[J]. Health Libraries Review, 2000, 17(3):117-128.
- [2] 崔雷,陈东滨.国外医学信息学科研究热点的文献计量学分析[J].医学信息学杂志,2007(2):97-102.
- [3] 许丹,侯跃芳.国外医学信息学研究领域可视化引文分析[J].医学信息学杂志,2011,32(10):12-19.
- [4] SCHUEMIE M J, TALMO J L, MOORMAN P W, et al. Mapping the domain of medical informatics [J]. Methods Inf Med, 2009, 48(1): 76-83.
- [5] CUGELMAN B, THELWALL M. Online Interventions for Social Marketing Health Behavior Change Campaigns: A Meta-Analysis of Psychological Architectures and Adherence Factors [J]. Journal of Medical Internet Research, 2011, 13(1): e17.
- [6] MEGLIC M, FURLAN M, KUZMANIC M, et al. Feasibility of an eHealth Service to Support Collaborative Depression Care: Results of a Pilot Study [J]. Journal of Medical Internet Research, 2010, 12(5): e63.
- [7] LOPEZ D M, BLOBEL B G M E. A development framework for semantically interoperable health information systems [J]. International Journal of Medical Informatics, 2009, 78(2): 83-103.
- [8] DENNY J C, PETERSO J N F, CHOMA N N. Extracting timing and status descriptors for colonoscopy testing from electronic medical records [J]. Journal of the American Medical Informatics Association, 2010, 17(4): 383-388.
- [9] USLU A M, STAUSBERG J. Value of the electronic patient record: An analysis of the literature [J]. Journal of Biomedical Informatics, 2008, 41(4): 675-682.

- [10] RAHIMI B, VIMARLUND V. Methods to evaluate health information systems in healthcare settings: A literature review [J]. Journal of Medical Systems, 2007, 31(5): 397-432.
- [11] ELLENIUS J, GROTH T. Dynamic decision support graph - Visualization of ANN-generated diagnostic indications of pathological conditions developing over time [J]. Artificial Intelligence in Medicine, 2008, 42(3): 189-198.
- [12] BASCH E, ARTZ D, IASONOS A, et al. Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities [J]. Journal of the American Medical Informatics Association, 2007, 14(3): 264-268.
- [13] COVEY J. A meta-analysis of the effects of presenting treatment benefits in different formats [J]. Edical Decision Making, 2007, 27(5): 638-654.
- [14] NIKOLOULOPOULOS A K, KARLIS D. Multivariate logit copula model with an application to dental data [J]. Statistics in Medicine, 2008, 27(30): 633-640.
- [15] ROHRER J E, WILSHUSEN L, ADAMSON S C, et al. Patient-centredness, self-rated health, and patient empowerment: should providers spend more time communicating with their patients? [J]. Journal of Evaluation in Clinical Practice, 2008, 14(4): 548-551.
- [16] LONG A J, CHANG P, LI Y-C, et al. The use of a CPOE log for the analysis of physicians' behavior when responding to drug-duplication reminders [J]. International Journal of Medical Informatics, 2008, 77(8): 499-506.
- [17] CHEN C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature [J]. Journal of the American Society for Information Science and Technology, 2006, 57(3): 359-377.
- [18] 张美君,董建成.中美医学信息学教育的比较研究[J].中国高等医学教育,2005(2):93-94.
- [19] HAUX R. Medical informatics: past, present, future [J]. International of Medical Informatics, 2010, 79(9): 599-610.

作者简介

秦方 (1988-), 女, 在读硕士研究生, 研究方向医学信息资源检索与应用。E-mail: afangs@163.com

刘硕, 男, 本科生。

通讯作者: 张士靖 (1958-), 女, 教授, 硕士生导师, 研究方向为医学信息资源组织与管理、情报语言学、信息检索与评价、医学信息素养教育等。E-mail: zhangsj9999@163.com

Analysis to Research Hotspots and Fronts on Medical Informatics for the Nearly Past Five Years – Based on the Bibliometrical Study to 22 High-Impact International Journals

Qin Fang, Liu Shuo, Zhang Shijing / Department of Medical Informatics, Tongji Medical College of Huazhong University of Science and Technology, Wuhan, 430030

Abstract: The article analyzes research situation of international medical informatics for nearly past five years, and takes the subject research hotspots and fronts in order to provide the reference and help for the development, education of domestic medical informatics by retrieving 22 JCR2010 journals under the subject Medical Informatics in SCI-E from 2007 to 2011. Based on bibliometrics methods, co-occurrence term clustering and burst terms detect, the data are processed by statistical analysis tools, such as TDA, SPSS12.0, and CiteSpaceII. The research situation have been analyzed through research hotspots and fronts. The Conclusions is that the applications of information technology and health information analysis methods are two big dimensions of the Medical Informatics, and they promote each other. The research fronts of Medical Informatics in the five-year period from 2007 to 2011 have changed from the prosperity of information technology application, health information analysis method to the period which themes are new technologies' application and new methods' invention, and then to smooth period short for innovations; and the development trend in the future may be developing new technology's application for people's health-care.

Keywords: Medical informatics, Journals, Research hotspots, Research fronts, International, Bibliometrics

(收稿日期: 2012-08-21)

数图论坛促发展, 每期话题扣当前; 引领思潮话关键, 百尺竿头更耐看。

田胜立