

富知识发现型电子病历研发与应用进展

梁志伟¹ 罗云坚¹ 吕玉波¹ 李小华¹ 程翼宇² 刘保延³

(1. 广州中医药大学, 广州 510405; 2. 浙江大学, 杭州 310027;
3. 中国中医科学院, 北京 100700)

摘要:【目的】回顾电子病历有关结构化、标准与规范化、电子签名等方面研发文献, 论述具有知识发现潜力的电子病历技术发展状况, 预期适宜并可行的系统模型。【方法】通过 PubMed 和维普全文检索系统, 检索并分析电子病历结构化、标准与规范化研发应用的需求、动因、解决方案、实施效果等, 以自由结构化、标准与规范化、电子签名为共构基础, 预期具有丰富知识发现潜在价值的新一代电子病历系统研发模型。【结果】注意到 XML 技术的诞生引发了卫生保健及医疗信息存储格式、传输交换、共享利用的变革。以诊疗信息结构化、标准与规范化为基础的数据采集、存储交换、处理分析、输出利用, 以及医疗文档电子签名日益受到国内外卫生医药企事业重视。具有高维结构、表单化编辑界面的电子病历逐渐增多, 并催生出相应知识发现的应用软件。集电子签名和不限维数据结构于“即见所得”用户界面的富知识发现型电子病历将成为临床实践与科研一体化的知识发现载体。【意义】富知识发现型电子病历, 有利于从临床实践采集海量医疗信息, 有利于分析、总结、挖掘和发现其中的诊疗规律与知识, 进而推动临床科研一体化尤其是中医药临床研究与方法的现代化; 并在提高卫生医疗保健管理与服务质量方面具有广阔前景。

1 引言

信息技术的发展推动了医院信息化建设, 医院信息系统在行政、计财、物资、人力资源等方面业务取得了良好效益, 当前开始转向“以病人为中心”服务理念出发, 重视加强临床医疗信息化工作。管理信息和临床信息双线并重的新型医院信息管理模式悄然崛起。医院临床信息如何才能高效利用, 如何在保障诊疗记录来源的真实和患者隐私基础上, 加强临床信息规范与标准化来促进医疗“信息孤岛”的交流利用、提取加工信息发现新知和规律, 已逐渐成为当前我国医院信息系统(HIS)建设的新热点^[1,2]。

本文探讨分析了电子病历在结构化、标准与规范化、辅助决策、电子签名与加密等方面的研发与应用的现状, 旨在阐明具有丰富知识发现潜在价值的电子病历发展方向, 预期适宜并可行的研发模型。

2 方法

电子病历(Electronic Medical Records, EMR)是医院临床信息系统(Clinical Information System, CIS)中起重要作用的临床信息采集工具。为回顾电子病历结构化、规范化和标准化的、决策支持、数字签名与加密的研发进展,检索了国内的维普《中文科技期刊数据库》(CSTJ)1998-2006年和国外PubMed检索系统1998-2006年的文献,并参考国家已颁布的与医院信息管理相关法律法规、CIS行业学会与协会的学术动向和CIS产品市场情况等资料,对代表性文献或资料进行分析论述,着重探讨分析电子病历结构化、标准与规范化的研发应用动向、需求、解决方案、实施效果等方面内容,结合国内医院信息管理和技术的实际情况,通过优势整合,提出具有前瞻性、科学性、可行性、效益性、以“病人为中心”为服务理念,以“知识系统多维存储”为技术基础的新一代电子病历系统研发模型。

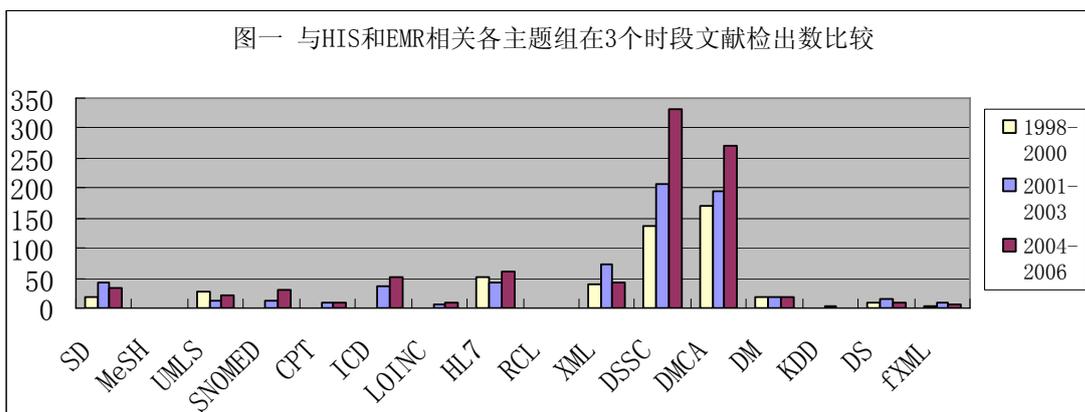
在进行有关标准检索时,先按主题分为若干个主题组,分别选择了部分目前国内外有所关注的标准,包括:Medical Subject Headings(MeSH)、Unified Medical Language System(UMLS)、Systematized Nomenclature of Medicine (SNOMED)、Current Procedural Terminology(CPT)、International Classification of Diseases(ICD)、Logical Observation Identifiers Names and Codes(LOINC)、Health Level 7(HL7)、Read Clinical Classification(RCC)、Extensible Markup Language(XML)等,并取后两者的词作为PubMed的文本词(Text Word)而其它在前面的词作为检索的MeSH主题词(MeSH Terms, 主题词),同时取主题词为标准化(Standization, SD)检索提及标准化的文献;在辅助决策方面,也按主题分为若干个主题组,中文取“决策支持”、英文取 Decision Support Systems, Clinical(DSSC)合并 Decision Making, Computer-Assisted(DMCA)作为MeSH主题词、Data Mining(DM)合并 Knowledge Discovery Databases(KDD)作为文本词。

3 结果

检索1998-2006年PubMed文献,并利用检出结果数据作图分析,先将过去9年的文献数按不同主题组列出条图,再对每一主题组分别以1998-2000年、2001-2003年、2004-2006年共三个时段加以比较列图(图一)。

从图中可见,DSSC及DMCA主题组检出文献数(检出数)最多,而且保持有较高的增长速率,第三段检出数分别达331条和271条。HL7、XML、ICD主题组在各段的检出数近50篇或以上,SD、UMLS、SNOMED、DM、DS主题组检出数接近10篇或以上,与SD主题组检出数较少并接近,而其余主题组MeSH、CPT、LOINC、RCL、KDD、fXML的检出数很少。从它们的时间段增幅来看,UMLS稳中有降、HL7、DM、DS、以及检出数很少的各组基本持平,XML、SD中段增幅大而后段也有所回落,

SNOMED、ICD 主题组检出数前段低而后段持续增幅（表一）。



表一 与 HIS 和 EMR 相关各主题组数量与变化分类及其分布情况

进展状况	主题组	备注
数量很多、增长快	DSSC、DMCA	***
数量较多、增长快	ICD	**
数量较多	HL7、XML	*
数量较少、增长快	SNOMED	*
增长快后回落	XML、SD	
数量较少	SNOMED、UMLS、DM、DS、SD	
增长慢或负增长	HL7、DM、DS；UMLS	
数量很少	MeSH、CPT、LOINC、RCL、KDD、fXML	

注：*号提示为：***大热类；**热类；*曾受关注类或值得关注类

4 讨论

4.1 XML 技术的诞生引导电子病历数据存储与交换及其信息利用的技术变革

卫生医疗数据长期以来对卫生事业决策不可置疑地发挥着重要作用。同时仍存在的一些不乐观因素，尽管业界有所共识并一直在寻找有效摆脱这些困境的良方。主要表现在：信息分散不利于信息的收集分析，数据管理系统众多且信息存储结构定维标识不一使数据可识别度差和易读性低，临床信息结构维度较低进一步更加剧临床信息统计分析不利性。这些不利因素一方面导致了整个卫生医疗资源利用不足甚至是投入的浪费，其影响包括卫生医疗信息化建设资源投入和社会卫生医疗基本保障资源分配等多个方面；另一方面可供统计分析和辅助决策的信息由于利用率不

高，难以为改善卫生医疗资源合理分配、提高医疗水平方面从技术上发挥应有辅助作用。计算机互联网络广泛普及为分散信息传输提供了极大的便利，同时也对解决信息的有效共享和交换利用难题提出了更为迫切的需求。

自XML于1996启动研究，至1998年2月10日W3C首次颁布XML Ver1.0，旋即引起业界的关注。国际标准化组织(ISO)的标准通用标记语言(Structured General Markup Language, SGML, XML是SGML子集)亦于当日声明Health Level 7(HL7)将被开发为支持XML的医疗标记语言^[3]。在欧洲德国Dudeck J同期指出应用XML比当时的卫生保健通讯标准交换格式能带给用户更巨大的灵便性和适用性^[4]。Marietti C紧接提出XML作为电子病历数据交换解决方案可能性，并认为XML是解决整个业界信息松散的关键^[5]。日本大约从1996年起研究医学标记语言(Medical Markup Language MML)，MML V2.21亦于1999年宣布采用XML作为元数据语言，并于2004年以HL7 CDA(Clinical Document Architecture)为基础完成了MML V3.0^[6]。我国刘氏于2001年从文献角度较早介绍以XML构建电子病历的方法^[7]。XML自问世后的头6年在HIS和ERM发展很快，近3年有所回落，同时有关HL7的研究相继遵循XML标准并平稳发展至3.0版本。

由于各国医疗制度、流程、病历内容与表达形式、以及相关业务仍存较大差异，近期内细分标准不易统一，因而出现除ICD外，标准文献总量偏小或增多不明显。我国要趁此把握好机遇，加强行政、财政、协会、学会的支持力度，发展丰富具有中国特色并引用兼容国外标准的国标，从战略上构筑国家技术壁垒，外可通而内可进并可守。建立中医药标准体系已被纳为国家科技部十一五科技规划重点项目。国家卫生部统计信息中心亦为此加强了标准化建设的国际合作^[1]。可以预见，随着HIS结构趋于以XML为格式进行存储和交换，海量临床信息有效利用、临床规律和知识发现、以及医疗卫生保健事业的决策支持，近期在国际区域内的医疗业务中更易于得以现实。

4.2 电子病历信息采集存储标准规范和高度结构化与文本书写自由化对立中趋于统一

2002年卫生部颁布了《医院信息系统基本功能规范》，2003年国家中医药管理局颁布了《中医医院信息化建设基本规范(试行)》，为医院信息采集，包括临床信息采集，提供了指导性的规范文件。大大改善了我国各医院信息管理系统在内容上更趋于完备性，有利于医疗信息在必要的时候能够汇总分类。国家标准《中医临床诊疗术语》(GB/T 16751-1995)、《中医临床诊疗术语(疾病部分、症候部分、治法部分)》(GB/T 16751.1-1997)和国家中医药管理局颁发的行业标准《中医病证诊断疗效标准》(ZY/T001-94)，促进了中医临床用语的规范和中医证候与疾病用词及其编码的统一。本世纪初始，为了更好地开展卫生决策管理的需要，国家中医药管理局会同中医院疾病监控信息中心试点推行了中医院病案信息汇总上报和广东省卫生厅推出了适用于广东省医疗

机构的病案首页并收集汇总信息。两者都对部分重要的临床信息用数据库加以较高维度的结构化，前都采用了 ICD-10 标准，而后者是采用 XML 标准进行交换汇总。

目前少部分医院已保存了数年的高维非自由结构数字化病历信息，而大多数医院采集和保存的是纸质病历信息、或维度较低甚至是平面文本文档病历信息。当前数据挖掘技术处理非结构化和文本信息的效果尚不理想，但这项内容已被列为数据挖掘技术未来研究方向和热点之一^[8]，吕氏等利用自选设计的平面文本分析的技术、建立数据挖掘模型和软件包，对低维度的半结构化的电子病进行了分析，发现被抽样病历中绝大部分的中医临床用词和术语均符合 GB/T 16751.1-1997，提示了这套规范对中医院临床实践的影响力，同时还可以通过临床的用语，结合运用 ICD 标准，发现临床的诊疗规律^[9]。更有不少学者认识到结构化信息存储的重要性，但在技术上却与自由结构录入法这个重要而实际的应用因素往往同时又是一对矛盾，并且长期试图寻找有效的解决方法。梁氏提出了电子病历的结构可分为平面文档无结构化、半结构化、高度结构化、自由结构化，以及难于实现的完全结构化几个层次^[10]。肖氏等开发了支持 XML 和即见所得和自由结构录入功能、具有高维结构电子病历^[11]。北京和广东地区一些单位也开始使用具有高度结构和即见所得界面的电子病历，并取得了初步的成效。

从有关 HIS 和 EMR 文献中，将 DSSC 合并 DMCA 为一组与 ICD 组曾作进一步的交集检索，检出数有 6 篇；另外也将 DSSC 合并 DMCA 为一组再与 XML 组作了同法处理，检出数有 13 篇，3 时段分别为 2、6、5 篇。提示国外有关 HIS 和 EMR 的研究中，涉及临床决策(包括临床诊断与治疗)或计算机辅助决策方面的研究共检出 1189 篇(国内相关中文文献仅有几篇)，比其他主题组检出数明显高出很多，然而涉及 ICD 和 XML 相对显得很少。这些方面的结合性研发工作虽然有所起步和进步，但显得与 ICD 和 XML 所历经过的快速发展不太相称，有待拓展更阔的研发空间。我国学者值得关注并可加强填补这些方面的空白。

利用 XML、临床信息标准与规范、电子签名技术，设计可不限维度结构化、具有自由文本录入的“即见所得”编辑界面、多层临床知识联想输入法的电子病历，将会逐步受到 HIS 界的关注与重视。从系统架构模型而言，这种电子病历的架构模型必须以病人信息为中心、具有数据结构维度高且不受限、信息规范标准、带有大量可供引用的临床知识库和术语库以用于构成病历显在或潜在临床信息、具有支持 XML 标准的数字签名与临床敏感信息安全保密功能、用户界面既可以表现为即见所得，又可以同时表现 XML 文档结构或树结构图，而在这种框架下的电子病历是一种富知识发现(Rich Knowledge for Discovery, RKD)型的电子病历(图二，略，参幻灯)。

卫生保健医疗决策管理需求拉动对富知识发现型电子病历产品的研发，而这两者都需要足够的临床信息标准化和数据的高维结构化。需求-产品-技术的互动将推动富知识发现电子病历的发

展，进而达到临床信息标准规范和高度结构化与文本书写自由化在对立中趋于统一。

4.3 富知识发现型电子病历是临床实践与科研一体化和中医药现代化的重要基础

利用临床信息进行知识发现和辅助决策时，必须在保障私隐权不受侵犯基础上，最大限度地满足跨区域、海量数据地开展医学研究。这种工具既可以采用回顾性研究和归纳法从临床信息挖掘诊疗规律，也可以按照前瞻性研究思路与要求、设计出随机分组方案及其病历报告表格(Case Report Form, CRF)，确定后可由第三方或客观无偏倚地从病案自动检索信息填充CRF，然后进行相应的统计分析。这种研究方法和工具所获取的数据自然朴实地来自于日常临床诊疗实践，科研数据的采集快捷方便、客观自然。可用于临床随机对照试验(Randomized Control Trails, RCT)样本估计、RCT模拟、发现潜在诊疗知识、总结诊疗规律。临床科研CRF对HIS的依赖开始受到重视^[12, 13]。

传统中医药学与现代西方医药学治疗病患的目标一致但所运用的理论与技术迥异。一是诊断与论治的思维基础理论不同；二是治疗手段的物质基础不同。现代中医药的发展应当吸纳各种科技成果。及时把握当今信息技术革命所带来的机遇，重视诊断依据合参理化诊断信息、诊断结果合诊中证西病、病证治疗合参中西手段（酌情单用中药及其方剂或合用中西药），既极大丰富和完善了现代中医药理论和实践体系，又符合了以病人为本的医学伦理观。

中医临床过程中宜引入多种诊疗信息指标作为潜在的可分析元素加以存储，然后根据研究主题，客观进行钻取分析处理，建立与病证相关的数学模型。这种在遵守医疗法规与医疗伦理条件下长期自然地生产呈海量级可供分析数据的存储方案，需要借助于集电子签名和不限维数据结构于“即见所得”用户界面的富知识发现型电子病历，以用签名和医疗法规保障数据的真实性，以数据加密技术保护病人私稳、以不限维数据结构保障中医诊疗思维演绎多样性、以即见所得保障直观易用性。如果能有效开展广域或跨域多中心合作采集中医药信息，共享共用，必将给中医药理论与实践带来新一轮飞跃。

5 意义与前景

集电子签名和不限维数据结构于“即见所得”用户界面的富知识发现型电子病历将成为国内外新的研发产，有利于从临床实践采集海量医疗信息，有利于分析、总结、挖掘和发现其中的诊疗规律与知识，进而推动临床科研一体化尤其是中医药临床研究与方法的现代化；并在提高医院与卫生管理和医疗卫生保健水平方面具有广阔前景。

参考文献：

[1] 雍传哲. 国家卫生部统计信息中心与英特尔公司签署电子病历标准化合作谅解备忘录[J]. 中华医学信息导报. 2006; 21(9):5.

[2] 刘晓辉 李小华 杨哲, 等. 集成临床数据仓库的建立与应用研究[J]. 医学信息(西安上半月). 2005; 18(5): 442-444.

[关键词] 集成 数据仓库技术 医疗信息系统 作用与意义 医院信息化 研究利用 诊疗信息 临床诊断 技术原理 辅助决策 [ISSN] 1006-1959

[分类号] R197.324 R221

[文摘] 本文探讨数据仓库技术在医院信息化应用中的作用与意义。研究利用数据仓库技术, 挖掘和整合医院各类医疗信息系统的诊疗信息, 建立集成临床数据仓库, 为临床诊断、治疗提供辅助决策的技术原理和方法。

[相关文献] 主题相关

[3] 张健飞. XML 实用培训教程[M]. 北京: 科学出版社. 2003 第一版. 2-8.

[4] Dudeck J. Aspects of implementing and harmonizing healthcare communication standards[J]. Int J Med Inform. 1998; 48(1-3): 163-71.

[5] Marietti C. Extensibility sensibility. Is XML the solution for electronic medical records data exchange? [J]. Healthc Inform.. 1998; 15(3): 23-6.

[6] Guo J, Takada A, Tanaka K, et al. The development of MML (Medical Markup Language) version 3.0 as a medical document exchange format for HL7 messages[J]. J Med Syst. 2004 Dec; 28(6): 523-33.

[7] 刘志文, 吴一民. 基于 XML 标准的电子病历实现技术[J]. 微型机与应用. 2001; 20(5): 37-39.

[8] 佚名. 数据挖掘资料汇编. [EB/OL]. 不敷出

<http://www.8sta.com/Article/base/database/200506/715.html>

[9] 吕玉波, 梁志伟, 杨志敏, 等. 从临床诊疗术语发现诊疗规律的方法学研究[J]. 广州中医药大学学报. 2006; 23(2): 179-182.

[10] 梁志伟, 刘保延, 吕玉波, 等. 临床实践科研一体化新技术——自由结构电子病历[R]. 第三届数字中医药学术研讨会. 北京 2005; 10

[11] 肖强, 吴伟斌, 陈联忠, 等. 自由结构录入法在电子病历系统中的应用[J]. 解放军医院管理杂志. 2005; 12(3): 222, 228.

[12] 梁志伟, 吕玉波, 刘保延, 等. 中西医结合治疗 SARS 多中心临床研究数据管理分析系统. 中国生物工程学报, 2007; 25(6): 683-687, 716.

[13] 董军, 刘志敏, 曹秀堂, 等. 小汤山医院 SARS 诊疗信息数据库建设[J]. 中国循证医学杂志, 2004, 4(3): 181-186, 209.