

新兴电子病历的结构化方法: OpenSDE

许陆飞, 雷健波

[摘要]介绍了电子病历结构化的重要意义和常用方法,系统论述了一种新兴的电子病历结构化方法,即 OpenSDE(Open Structured Data Entry)产生的意义、发展历程、原理和实现过程,讨论了 OpenSDE 的优缺点。

[关键词]OpenSDE; 电子病历; 结构化数据录入

[中图分类号]R197.323

[文献标志码]A

[文章编号]1671-3982(2012)01-0040-04

A new data entry method for structured electronic medical records: OpenSDE

XU Lu-fei, LEI Jian-bo

(Center of Medical Informatics, Beijing University, Beijing 100191, China)

Corresponding author: LEI Jian-bo

[Abstract]The importance and common methods for structured electronic medical records (EMR) were described, followed by a detail discussion on the significance, development, principle and implementation of OpenSDE, a new data entry method for structured EMR, and its advantages and disadvantages.

[Key words]OpenSDE; Electronic medical record; Structured data entry

1 电子病历及其结构化方法

1.1 电子病历及其优点

2003 年 ISO/TC 215(国际标准组织/负责卫生信息领域标准的技术委员会)将“电子病历”定义为^[1]:一个计算机可以处理,可安全存储和传输,能被多个授权用户访问,覆盖过去、现在和将来,与个体健康相关的信息库,具备独立于应用系统的标准化模型,目的是支持高效、连续、高质量的综合医疗保健。狭义上的电子病历(Electronic Medical Record, EMR)是指医疗机构(如医院)保存的个人健康档案资料的数字化记录^[2]。电子病历的应用给医疗卫生领域带来了巨大的变革和进步。

Biomedical Informatics 第 3 版对电子病历的优点进行了以下描述:EMR(电子病历)可以保存多个

副本,可以同时满足不同医生在不同场合同时浏览病历信息的需要,并可以根据研究人员的需要以不同的形式予以展示;EMR 可以记录最完整的病历信息,包括文本、视频、图片、音频等多媒体资料;EMR 记录格式标准统一,文字清楚,可以减少医师记录病历时的逻辑错误;EMR 可以支持即时查询和更新,整体或者部分数据可以为临床、科研重复利用^[3]。

1.2 电子病历结构化方法

电子病历的核心价值需要通过结构化的数据录入和存储体现。电子病历的结构化以及因为结构化提供的后续计算机处理是电子病历和传统纸质病历一个重要的本质区别。因为只有通过结构化的录入和存储,才能使计算机“读懂”电子病历,才能实现医疗过程中实时的临床决策支持^[4],提高医疗质量,体现电子病历的核心价值。也只有通过结构化的存储和查询机制,才能实现在语义层面的查询,数据才能更好地为科研所利用^[5]。例如,为了研究糖尿病的影响因素,需要从病历现病史或者既往史中筛选出患有糖尿病的病历。在自由文本形式的病历中,只能通过“糖尿病”这个关键词进行

[基金项目]国家自然科学基金(81171426):临床医生信息需求研究与“一键通”系统(一种基于临床现场的智能“临床决策支持”系统的研究和应用)。

[作者单位]北京大学医学信息学中心,北京 100191

[作者简介]许陆飞(1987-),男,河北抚宁人,在读硕士。

[通讯作者]雷健波,男,四川泸州人,北大医学部硕士生导师。E-mail: jblei@bjmu.edu.cn.

匹配,导致很多“无糖尿病”病历也被筛选进来。而在结构化的电子病历中,我们可以预先对“糖尿病”进行赋值,比如有=1,无=2,不知道=3。这样只需查询“糖尿病”=1的病历即可,提高了查询的效率和正确性。

电子病历结构化的方法按照时间的发展大致有自由文本、完全结构化、半结构化、自然语言处理^[6]4种。自由文本的优点是可根据需要随意录入,内容灵活丰富,录入时间短,方便自由,完全符合临床人员的习惯,还可改善临床人员之间的沟通;缺点是结构化程度最低,容易遗漏病历信息,没有录入校验机制,容易出现录入错误,病历各部分间难以保持一致,难以实现数据的二次利用。完全结构化的优点是病历结构完整,格式标准,数据能够实现结构化录入与存储,便于数据的检索查询以及各种决策支持;缺点是数据格式固定,限制了医师连贯的思维过程,不能适应不断变化的临床情况,容易丢失病人的特异性信息。半结构化的优点是兼顾了数据的结构化和表达的灵活性,确保病历录入的完整性和准确性,被标记的数据有助于进行信息的定位;缺点是结构化是基于预先定义的模板或逻辑概念集,当遇到复杂情况没有预先定义时,只能使用自由文本录入。自然语言处理的优点是临床人员录入自由文本,程序根据语义分析,自动结构化自由文本;其缺点是需要预先具备庞大而复杂的知识库和术语标准集以及自然语言处理算法,目前还难以完美地实现。

以下重点介绍的 OpenSDE 技术,按照其特点属于电子病历半结构化技术,也叫开放式的结构化数据录入技术^[7]。

2 OpenSDE

2.1 OpenSDE 的定义

OpenSDE 是英文“Open Structured Data Entry”的简称^[8],是用于病历结构化的一种方法。它在实现病历结构化的同时又能满足不同病历间的差异化表达,甚至能进行复杂病历的个性化定制。OpenSDE 是规定医疗数据以何种形式(组织形式和表现形式)录入和存储的理论原理,基于此原理可以进一步延伸开发出适合实际情况的程序模型。

2.2 OpenSDE 的目的

开发 OpenSDE 的目的就是为了弥补完全结构化方法带来的缺陷,实现数据结构化录入的同时又

兼顾表达的灵活性和多样性,适应各种复杂的临床实际情况,使临床人员在临床工作和回顾性、前瞻性的科学研究中能够方便地使用数据。从科研的角度来看,与自由文本数据相比较,结构化的数据对科研查询的意义更大,但是目前临床人员还是更倾向于使用自由文本这种形式进行数据录入。OpenSDE 在表现形式上必须减小与自由文本的差距,并且无缝整合到 EMR 系统中,才更容易被临床人员所接受。

2.3 OpenSDE 的历史

OpenSDE 理论是由荷兰鹿特丹大学医学中心(University Medical Center Rotterdam) MC(伊拉斯诺医学中心)的 Renske K. Los 等人提出的 2004-2005 年连续在 *International Journal of Medical Informatics* (国际医学信息学杂志), *Journal of the American Medical Informatics Association* (美国医学信息学会杂志)发表了 3 篇关于 OpenSDE 的论文,内容包括 OpenSDE 的原理、实现和限制因素等。随后,OpenSDE 的基础原理受到世界各地的医学信息学专业人员的关注,并根据各地的实际情况积极进行系统开发。

2.4 OpenSDE 的特点

OpenSDE 是以临床数据为研究对象,通过 OpenSDE 的一次性数据录入实现数据的可重复利用,具有以下特点。第一,OpenSDE 位于数据生命周期的前期。任何临床数据的生命周期都包括了数据的采集录入、展示、存储、整合加工和二次利用等几个阶段,OpenSDE 对于数据流的作用主要集中于之前的采集录入、展示、存储等几个阶段。第二,OpenSDE 主要用于描述性数据的结构化,即文本、数字、符号等数据的结构化。一些临床中出现的数据,如图片、视频等目前还无法使用此原理进行结构化表达。第三,OpenSDE 涉及数据语法表达阶段。临床数据结构化和标准化的过程共涉及数据的语法、语义、语用 3 个层次。语法层次主要负责数据组织形式的结构化,运用数据结构化理论搭建的域模型树状结构可以规定临床数据在短语或者句子中表达的先后顺序,规定数据在句子中所做的成分。

2.5 OpenSDE 的原理

2.5.1 域(Domain)和域模型(Domain Models)

根据维基百科解释,“域”指事物的范畴属性。在临床中,“域”更多地解释为能够明显区别于其他

专科的临床分支科学所包括的一切内容。域模型 (Domain Models) 是用于系统性描述域^[9] 在临床中包括域的医学术语集, 医疗流程和各种医疗规则。域模型规定了在 OpenSDE 中域的结构和内容的表达。例如, 临床小儿科是一个域, 其中包括小儿科在临床中所有的知识; 域模型则从组织数据录入的角度对小儿科域进行描述, 包括小儿科的诊疗流程, 小儿科涉及的所有医学术语以及各个医学术语间逻辑关系和限值等医疗规则。值得说明的是, OpenSDE 域模型中不应包括知识推论等逻辑内容, 如小儿肺炎导致并发症缺氧性脑病, 那么关于缺氧性脑病的详细知识描述不应包含在小儿科域模型中。OpenSDE 的通用性在于虽然不同的域模型间在内容上有所不同, 但在数据组织的结构和原理上是相同的。

在 OpenSDE 中, 域模型以逻辑概念树的形式表现出来。传统的医疗文档是典型的层次结构, 根据这一特征, 逻辑概念树有且仅有一个根节点, 根节点包括一个或多个叶节点。当叶节点以下不再包含节点时这个叶节点称为叶子节点, 一个叶节点可包括一个或多个叶子节点, 一个叶子节点有且仅有一个叶节点与之有隶属关系。从根节点到叶子节点有且仅有一条扩展路径。在节点上是预先定义的医学术语, 在 OpenSDE 中把节点上的医学术语定义为实体 (entity)。实体即视为一种逻辑存在, 有各自的属性, 如定义实体的“可见性”属性为 1=有 2=无 3=不知道。定义实体的编码可以帮助确定实体在逻辑概念树中的逻辑位置, 方便检索, 定义实体的有效值可以防止临床人员录入时发生错误。实体还可以根据特点分为核心实体和条件实体。核心实体是有具体语义的实体, 如“糖尿病”实体, “有”, “无”两种属性都能代表“糖尿病”实体具有一定的语义; 条件实体是有通用语义的实体, 如“数值”实体可以和“血小板”实体联合, 也可以和“血压”实体联合, 分别代表不同的语义。

可见, 逻辑概念树中的实体代表了域模型中的医学术语概念, 树的从根节点到叶子节点的路径代表了诊疗流程, 对于实体属性的定义代表了医疗规则, 从而在结构和内容上完成了对逻辑概念树的完整表达。

OpenSDE 的数据组织示意图, 见图 1。

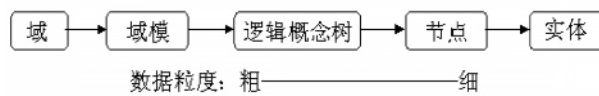


图 1 OpenSDE 的数据组织示意图

如果树的第一层根节点为“入院病历”, 第二层节点则包括“个人基本信息”、“主诉”、“现病史”、“既往史”、“体格检查”、“实验室检查”、“诊断”、“手术记录”等一系列与“入院病历”有隶属关系的信息。针对第二层的“手术记录”节点也可以再次展开第三层节点, 包括“术式名称”、“手术时间”、“手术记录”等一系列和“手术记录”有隶属关系的叶子节点, 叶子节点为病历逻辑概念树的最底层, 在此基础上进行数据录入。临床医师进行病历书写时, 通过遍历逻辑概念树, 针对特定的病历调用相关节点, 就形成了病历的行文逻辑结构。OpenSDE 建立的逻辑概念树见图 2。

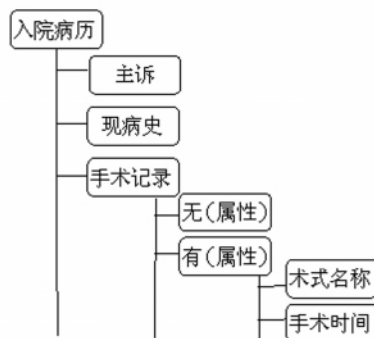


图 2 OpenSDE 建立的逻辑概念树

2.5.2 行模型 (Row Modeling)

OpenSDE 是从数据采集的角度开发的, 重点在于如何运用富有灵活性和表达性的方法支持临床数据录入。为了同时保证灵活性和通用性, OpenSDE 用行模型代替了传统的关系型二维表存储方式^[10]。

传统的存储方式, 建立了很多二维关系表, 依靠设立“主键”保持表之间的逻辑关系, 数据的语义被存储在二维表和表间的关系中。域模型中内容的改变将导致数据库结构的改变与软件程序和用户界面的改变。而在行模型中, 进行了一个列——行的转换, 把原来多张传统列模型表中的变量转换

为一张行模型表的变量进行存储,通过对行模型变量进行编码保持变量间的逻辑关系。转换前列模型表的行代表了一个病人的所有变量的记录,转换后行模型的行代表了一个变量的记录。转换前列模型的列代表了 N 个病人在一个变量上的记录,转换后行模型的列代表了一个病人在 N 个变量上的记录。当改变域模型节点,相当于改变了变量(也就是实体)的逻辑位置(编码),改变了每个实体的一系列属性值。新的数据记录将会被记录在新的节点位置,老的数据逻辑位置不变,因为它的属性

值记录没有变。OpenSDE 中通过行模型存储机制使域模型的改变对数据表、软件和数据结构没有影响。但明显的缺点是数据展现的格式和传统的统计分析软件需要的数据格式不同。

2.6 OpenSDE 实现过程

在软件系统中构建好逻辑概念树后,医师可以调用树中的节点组成病历文档,根据具体病历的不同遍历概念树选取特定的节点或者重复调用子树以实现病历的灵活表达。通过 OpenSDE 录入病历的过程见图 3。

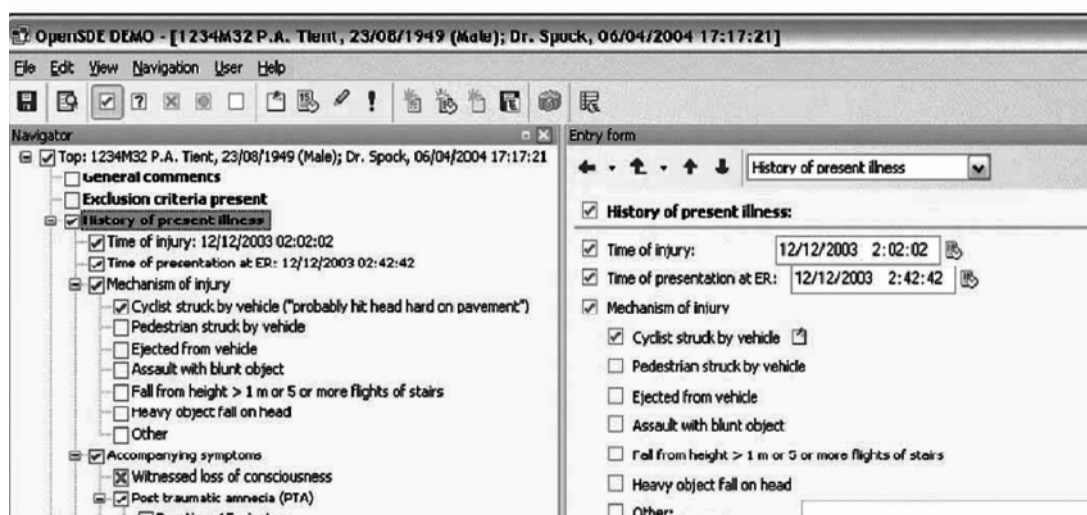


图 3 OpenSDE 录入病历的过程

从图 3 可见,屏幕左侧为根据域模型构建的逻辑概念树,包括了医学概念;右侧为调用节点组织的结构化病历文档,本例显示的为某病人的既往史。

基于 OpenSDE 原理,在软件程序实现中可进行多方面的拓展创新。OpenSDE 支持定制模板,针对特定临床专科中一些变异性较小的医疗文档,可以预先使用 OpenSDE 在电子病历系统中建立模板,需要使用时调用即可,大大节省了时间和减少了医生录入出现错误的概率^[11]。

3 OpenSDE 的优缺点

3.1 优点

首先,OpenSDE 克服了完全结构化方法带来的弊端,在实现数据结构化的同时又不失表达的灵活性。医师可以在一定程度上根据病历实际情况自定义病历文档结构。其次,基于病历结构的逻辑概念树符合医师的思维过程,医生从上至下、从根节点到叶

子节点遍历逻辑概念树的过程中可以帮助其回忆病人病程和相应诊断的流程,帮助其查漏补缺,使录入病历成为医师再学习的一种方式。第三,OpenSDE 具有良好的开源性。它是一种关于结构化的理论模型,不依赖于任何系统平台和开发平台,可根据实际要求进行系统开发。最后,利用 OpenSDE 可以根据各科室的具体情况,事先编辑一些适用于科室普遍情况的通用病历模板,既可实现在科室内的统一规范化,又可大大节省编辑病历的时间和提高效率。

3.2 缺点

首先,OpenSDE 基于行模型进行存储,在数据输出时与传统关系型数据库数据有所区别,不便于统计分析软件后续利用数据。近年来有学者开始对基于 OpenSDE 的数据在数据库中存储和分析进行研究。

(下转第 50 页)

研究的发展建立了以科学数据为基础的新型数据库;四是以文献库和科学数据集为基础,在数据之上开发的应用工具。系统分两期实施,第一期重点是文献库和知识库的建设,第二期重点是科学数据集等新型数据库的研制以及应用工具的开发,目前第二期正在进行中。

作为医学图书馆,在建设数字化图书馆和自主开发知识服务的过程中,一要立足于本图书馆的传统和优势,突出重点,形成自己的优势;二要根据国内外医学信息相关数据库的现状、部门自身的发展历史和现状、人才队伍的结构等,选择能在国内外产生影响或研究方向独特的研究领域形成特色数据库;三要在医学和信息学的发展呈现出不断分化和综合的趋势,新兴学科、交叉学科和综合学科不断涌现的情况下,使医学信息服务不断跟踪学科发展的前沿,寻找新的学科增长点,开发更多新的数据库产品。

【参考文献】

- [1] 孙卫,罗之兰,张蔚. 科学数据共享平台的数据管理研究[J]. 科学学与科学技术管理, 2005, 26(12): 32-36.
- [2] 李娜,张文举. 浅谈知识化医学图书馆[J]. 中华医学图书情

报杂志 2006, 15(3): 7-9.

- [3] 管建和,甘剑峰. 基于 Lucene 全文检索引擎的应用研究与实现[J]. 计算机工程与设计, 2007, 28(2): 489-491.
- [4] 张玉,张文举,李娜. 构建以知识服务和知识组织为基础的医药学知识库[J]. 医学信息学杂志 2010, 31(2): 26-29.
- [5] 张文举,李娜. 基于知识服务的医学知识服务系统研究[J]. 中华医学图书情报杂志 2007, 16(5): 1-5.
- [6] 李娜,张文举. 医药学知识咨询系统构建过程中相关概念的探讨[J]. 中华医学图书情报杂志 2006, 15(6): 67-69.
- [7] 张晓梅,杜云祥,雷春炳,等. 基于网络的生物医学信息导航系统的研发[J]. 中华医学图书情报杂志 2006, 15(6): 12-15.
- [8] 雷春炳,张晓梅,颜世刚,等. 西文生物医学期刊文献数据库(EMCC)建设[J]. 现代图书情报技术 2005, 21(8): 54-57.
- [9] Tyrkk E, Pelander A, Ojanper I. Differentiation of Structural Isomers in a Target Drug Database by LC/Q-TOFMS Using fragmentation Prediction [J]. Drug Test Anal (S1942-7611), 2010, 2(6): 259-270.
- [10] Riner A, Chan-Tack KM, Murray JS. Original Research: Intravenous Ribavirin - review of the FDA's Emergency Investigational New Drug Database (1997-2008) and literature Review [J]. Postgrad Med (S0032-5481), 2009, 121(3): 139-146.

[收稿日期: 2011-08-26]

[本文编辑: 杜云祥]

(上接第 43 页)

其次,域模型的构建完全基于已知的知识信息,虽然在一定程度上实现自定义表单可覆盖病历信息,但在临床中遇到复杂病历时,其内容难以完全用预先定义的域模型进行描述,只能添加自由文本类型的节点进行描述。这也与 OpenSDE 的初衷有所差异。

【参考文献】

- [1] 丁宝芬. 医学信息学[M]. 南京: 东南大学出版社 2009.
- [2] 伟斌,肖强,陈联忠,等. 电子病历系统的研究与开发[J]. 中华医院管理杂志 2004, 20(4): 204-206.
- [3] Shortliffe EH. Biomedical Informatics [M]. New York: Springer 2006.
- [4] Schnipper JL, Linder JA, Palchuk MB. "Smart forms" in an Electronic Medical Record: Documentation-based clinical decision Support to Improve disease management[J]. J Am Med Inf Assoc (S1067-5027) 2008, 15(4): 513-523.
- [5] 雷健波. 电子病历的核心价值与临床决策支持[J]. 中国数字

医学 2008, 3(3): 26-30.

- [6] 于一,廖睿,叶大田. 电子病历结构化方法概述[J]. 北京生物医学工程 2007, 26(1): 103-106.
- [7] Johnson S B, Bakken S, Daniel D *et al* . An electronic Health Record Based on Structured Narrative [J]. J Am Med Inf Assoc (S1067-5027) 2008, 15(1): 54-64.
- [8] Los R K, van Ginneken A M, van der Lei J. OpenSDE: A Strategy for Expressive and Flexible Structured Data Entry [J]. Int J Med Inf (S1386-5056) 2005, 74(6): 481-490.
- [9] Los R K, van Ginneken A M, Marcel de Wilde. OpenSDE: Row Modeling Applied to Generic Structured Data Entry [J]. J Am Med Inf Assoc (S1067-5027) 2004, 11(2): 162-165.
- [10] Los R K, van Ginneken A M, van der Lei J. Extracting data recorded with OpenSDE: Possibilities and Limitations [J]. Int J Med Inf (S1386-5056) 2005, 74(6): 473-480.
- [11] 肖强,吴伟斌,陈联忠,等. 电子病历系统的主要实现技术[J]. 中华医院管理杂志 2004, 20(4): 207-209.

[收稿日期: 2011-08-09]

[本文编辑: 王天津]