

• 医学信息研究 •

SNOMED CT 与 OpenEHR 的整合探究

李 斌

(上海生物信息技术研究中心 上海 200235)

〔摘要〕 详细介绍 SNOMED CT 与 OpenEHR 的发展历史、主要内容、主要架构,通过数据类型转换、原型中的临床术语与 SNOMED CT 的映射、候选术语集等方法进行 SNOMED CT 与 OpenEHR 的整合,从而使原型中的术语更加规范,也为互操作提供可能。

〔关键词〕 医学系统命名法——临床术语; OpenEHR; 映射; 自然语言处理

Discussion and Research on the Integration between SNOMED CT and OpenEHR Li Bin, Shanghai Center for Bio-informatics Technology, Shanghai 200235, China

〔Abstract〕 The paper concretely introduces the history, main contents, architecture of SNOMED CT and OpenEHR. The integration is realized through data type conversion, mapping clinical terms to SNOMED CT, set of candidate terms, so as to standardize the terms in the archetype and thus make it feasible for the interoperability.

〔Keywords〕 SNOMED CT; OpenEHR; Mapping; Nature language process

1 引言

电子健康档案 (Electronic Health Record, EHR) 若要通过安全、快速的方式被不同区域的研究者共享并使用,前提是必须有共同参考的数据标准。反之,具有可互操作性的信息系统将有利于改善医疗状况,减少错误和提高效率。数据的标准化需要正确地将临床数据映射到标准数据集中。医学系统命名法——临床术语 (Systematized Nomenclature of Medicine - Clinical Terms, SNOMED CT) 拥有完善的术语集,是目前国际上广泛使用的医学术语标准。将 EHR 中的临床数据映射到 SNOMED CT 中可以纠正拼写错误、归并同义语和还原缩略语,达到标准化,为互操作奠定基础。因此将 EHR 中的临床术语

映射到 SNOMED CT 中使其标准化是非常有必要的。

2 SNOMED CT 简介

2.1 发展历史

医学系统命名法——参考术语集 (Systematized Nomenclature of Medicine - Reference Terminology, SNOMED RT) 最早由美国病理学家协会 (College of American Pathologists, CAP) 开发使用。SNOMED RT 包含 10 万多个医学相关概念,用同义语 (Synonyms) 详细定义和描述。不同机构或个人收集的临床数据在集成和修复时可以参照 SNOMED RT 所制定的术语标准。

临床术语 (Clinical Terms, CT; 亦称 Read Codes Clinical Terms Version 3) 由英国国家卫生服务部 (National Health Service, NHS) 于 20 世纪 80 年代初开发。CT 包含 20 多万个医学相关概念,能

〔修回日期〕 2011-09-16

〔作者简介〕 李斌,博士,副研究员,发表论文 8 篇。

够帮助研究者将患者的就诊信息通过结构化的方式存储^[1]。

SNOMED RT 和 CT 的共同点:

- 组配式的概念体系;
- 以描述逻辑为基础;
- 通过概念间的关系对概念加以特化。

1999 年, CAP 的 SNOMED RT 与 NHS 的 CT 合并, 双方协议共享经验与资源。经过扩充和重组, 形成了 SNOMED CT。SNOMED CT 与其他的术语集 (如 ICD¹-x, LOINC² 和 OPCS-4³ 等) 之间还存在交叉映射。此外还支持 ANSI, DICOM, HL7 和 ISO 标准。2007 年 SNOMED CT 被国际卫生术语标准制定组织 (International Health Terminology Standards Development Organization, IHTSDO) 收购^[2]。

2.2 内容介绍

2.2.1 核心内容图示 (图 1)

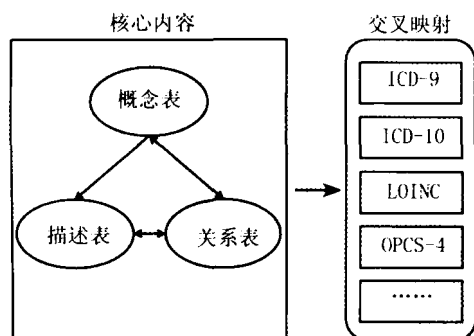


图 1 SNOMED CT 核心内容

2.2.2 概念表 收录 40 多万个准确定义的概念, 分类编入 19 个顶级概念轴中, 每个顶级概念轴又细分为包含多层子系统的树形结构。概念表的设计符合临床医师的诊疗流程, 也为数据管理和统计提供便利。

2.2.3 描述表 同一个医学概念可能对应多个专业术语。SNOMED CT 通过描述表来指定术语与概念的关系。描述表中收录了百万多个描述术语。一个特定的概念有首选的描述术语, 其他术语作为同义语存在。通过这种对概念和术语关系的编码化, 使拥有相同或相近概念的术语之间关系清晰, 从而

让临床概念表达更加灵活, 也便于计算机的数据挖掘和文本提取。

2.2.4 关系表 收录 140 多万组语义关联。这些语义关联既存在于同一概念轴内的概念之间, 也存在不同概念轴的概念之间。两个概念间通过一种特定关系的描述建立语义关联。SNOMED CT 通过强化概念间的语义关联并提供逻辑缜密、可直接通过电脑处理的医学概念, 从而保证了医学检索的准确性和有效性^[3]。

2.3 应用举例

- 电子病历;
- 计算机化医嘱录入;
- 急诊室表格记录;
- 癌症报告;
- 基因、蛋白质、代谢组学数据库;
- 临床决策支持系统;
- 临床试验管理系统;
- 临床用药管理系统;
- 实验室检查系统;
- 社区卫生、流行病学及公共卫生监测系统。

3 OpenEHR 简介

3.1 发展历史

OpenEHR 组织 (Open Electronic Health Record Foundation, OpenEHR Foundation) 是成立于 2003 年的非营利组织。综合了英国、澳大利亚及一些欧洲国家在电子健康档案方面的研究经验, 利用开源组件基于相关标准, 设计可交互操作的电子健康档案, 并通过临床示例去验证和完善。

OpenEHR 的主要目标有:

- 修正电子健康档案的规范;
- 改进和发布 OpenEHR 的信息架构和数据模型;
- 通过临床实践评价和确立 OpenEHR 原型;
- 实现互操作^[4]。

3.2 主要架构

3.2.1 概述 OpenEHR 的架构中包含一个名为

“本体的分隔 (Ontological Separation)” 的设计原则。在 OpenEHR 新近发布的 Release 1.0.2 中对该本体的结构描述, 见图 2。

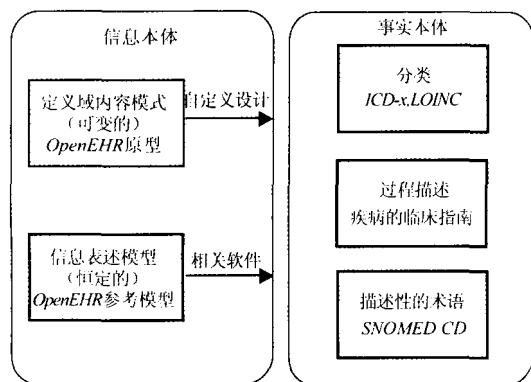


图2 OpenEHR Release 1.0.2 中的本体结构

3.2.2 信息本体 (Ontologies of Information) 包括电子健康档案的信息模型, 又分为: 定义域内容模型 (Domain Content Models), 包含了临床内容的规范化定义, 供使用者在 OpenEHR 提供的原型基础上, 根据临床需要自行设置, 因而是可变的; 信息表述模型 (Information Representation Models), 在 OpenEHR 中又被称为参考模型 (Reference Models), 是设置定义域内容模型的恒定基础, 可通过电子健康档案系统软件实现。

3.2.3 事实本体 (Ontologies of Reality) 包括: 分类, 如 ICD-9, ICD-10, LOINC; 过程描述, 如某一疾病的临床指南; 描述性的术语, 如 SNOMED CT^[5]。

4 SNOMED CT 与 OpenEHR 的整合

4.1 概述

通过共享的原型建立的 EHR 旨在方便不同健康组织之间实现数据交换。而原型的语义学 (Semantics) 需要“术语定义 (Term definition)”和“术语绑定 (Term binding)”来提供。

- 术语定义: 原型的节点由“名称”和“文本描述”赋予含义;
- 术语绑定: 原型中节点的定义需要参考外在

的术语集。

由此可见, 将 EHR 中的术语标准化具有重要意义。这不仅能使用术语使用更加精确规范, 还可以实现不同地域、不同机构之间的互操作^[6]。

4.2 数据类型转换

将原型中的临床术语映射到标准术语集之前, 可以通过 JAVA 程序语言, 将文本从构建 OpenEHR 原型的 ADL (Archetype Definition Language) 格式的语言转换成 XML (Extensible Markup Language) 格式的语言。

4.3 原型中的临床术语与 SNOMED CT 的映射

4.3.1 第 1 轮映射 主要是通过自然语言处理 (Nature Language Processing, NLP) 工具 (例如 EMT-P, Emergency Medical Text Processing) 对 XML 文本进行词汇加工 (Lexical Processing), 结合上下文内容检查拼写错误并剔除冗余信息。将所得结果在一体化医学语言系统 (Unified Medical Language System, UMLS) 中查找, 得到与之匹配的标准化词语。

4.3.2 第 2 轮映射 主要是对第 1 轮映射的结果进行词汇和语义的校对, 运用训练数据集 (Training Dataset) 对同义语和缩略语 (Abbreviations) 进行编辑。本轮映射可使用自然语言处理工具 GATE (General Architecture for Text Engineering) 中的“Morphological Analyzer”功能查找词根; “Word Net”功能查找同义语。

4.3.3 第 3 轮映射 将 SNOMED CT 中的数据导入 MySQL 数据库, 运行 SQL 查询并获取相关的信息。将前两轮映射并校对后的结果, 参照 MySQL 数据库进行过滤。对筛选结果进行过滤时需要制定相应的算法。Rahil Qamar 等^[7]使用的规则是: 如果输入的是子概念和父概念, 那么输出选择的是父概念。如:

```

"Filtering input = {Entire respiratory system, Respiratory
system structure} for 'respiratory'

Entire respiratory system is a Respiratory system structure
Output = {Respiratory system structure}"
  
```

如果输入的概念之间没有关联,那么输出选择的是这些彼此独立的概念。如:

" Filtering input = {Hydrogen ion concentration, Past history of, ph + , pH measurement arterial} for 'pH'

Hydrogen ion concentration (observable entity) is a Fluid observable

Past history of (context - dependent category) is a Context - dependent category

ph + (qualifier value) is a skin reaction grades

pH measurement arterial (procedure) is a pH measurement

Output = {Hydrogen ion concentration, Past history of, ph + , pH measurement arterial} "

4.4 候选术语集

通过 3 轮映射,筛选出符合研究者需要的标准化术语,组成候选术语集,为下一步构建临床模型提供帮助。候选术语集里包括的术语,有各自的数据结构、属性和数据元。技术简图,见图 3。

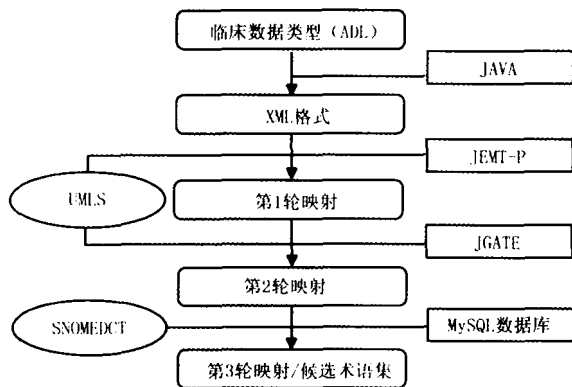


图 3 原型映射到 SNOMED CT 的技术简图

5 结语

由英国曼彻斯特大学开发的 MoST 系统 (Model Standardization Using Terminology System) 整合了 EMT - P 工具和 UMLS 系统,能帮助研究者快速高

效地完成映射的操作,但是最终的结果需要研究者进行人工筛选^[8]。目前关于 SNOMED CT 与 OpenEHR 的整合仅仅局限在几个原型中,随着计算机技术的进步和更多工具的应用,将会节省映射的时间和提高匹配的效率。此外具有临床医学背景同时又能熟练掌握计算机技术的研究者将在日后的开发中备受青睐。

参考文献

- 1 Coté RA, Robboy S. Progress in Medical Information Management: Systematized Nomenclature of Medicine (SNOMED) [J]. JAMA, 1980, (243): 756 - 762.
- 2 Price C, Spackman K. SNOMED Clinical Terms [J]. British Journal of Healthcare Computing & Information Management, 2000, 17 (3): 27 - 31.
- 3 SNOMED Clinical Terms User Guide, International Release July 2008 edition [EB/OL]. [2011 - 08 - 30]. <http://htg.his.uvic.ca/index.php?ContentFileID=57%5d>.
- 4 Dipak Kalra, Thomas Beale, Sam Heard. The Open EHR Foundation. Regional Health Economics and ICT Services [M]. Netherlands: IOS Press, 2005: 153 - 173.
- 5 Beale T, Heard S. OpenEHR Architecture: architecture overview. OpenEHR specification project release 1.0.2 [EB/OL]. [2011 - 08 - 30]. <http://www.openehr.org/releases/1.0.2/architecture/overview.pdf>.
- 6 Erik Sundvall, Rahil Qamar, Mikael Nyström, et al. Integration of Tools for Binding Archetypes to SNOMED CT [J]. BMC Medical Informatics and Decision Making, 2008, 8 (Suppl 1): S7.
- 7 Rahil Qamar, Alan Rector. MoST: a system to semantically map clinical model data to SNOMED - CT [C]. Semantic Mining Conference on SNOMED, 2006.
- 8 Rahil Qamara, Alan Rectora. Semantic Issues in Integrating Data from Different Models to Achieve Data Interoperability [C]. MEDINFO 2007, IOS Press, 2007: 674 - 677.