

基于本体的电子病历检索系统研究

赵 洋,李万龙,白杰英

(长春工业大学,吉林 长春 130012)

摘 要:适应循证医学成为电子病历的必然发展方向。但由于医学概念名称不规范或者缺乏一个统一的规范命名系统,导致一个概念可以有多种名称来表示。而且传统的查询方式无法进行语义扩展,存在查全率和查准率不高的问题。而由于本体是一种能在语义和知识层次上描述信息的概念模型建模工具,所以将本体引入电子病历中。文中提出了一种基于本体的电子病历检索系统,给出了一种映射算法实现概念的语义扩展,从而提高了检索的查全率和查准率,使得电子病历在医疗辅助决策中起到越来越重要的作用。

关键词:本体;电子病历;语义检索

中图分类号:TP391.3

文献标识码:A

文章编号:1673-629X(2010)03-0211-03

Electronic Medical Record Retrieval System Based on Ontology

ZHAO Yang, LI Wan-long, BAI Jie-ying

(Changchun University of Technology, Changchun 130012, China)

Abstract: To adapt evidence based medicine has become an inevitable development trend of electronic medical record. However, as non-standard of the medical concept or the lack of a unified nomenclature system norms, leading to a concept can be expressed by a variety of names. And traditional searching methods can not be semantic query expansion, recall rate and precision rate is not high. Therefore, ontology was introduced in the electronic medical record. This paper proposes an ontology-based electronic medical record retrieval system, presents a mapping algorithm to achieve expansion of the semantic of concept, to improve the recall rate and precision.

Key words: ontology; electronic medical record; semantic searching

0 引言

随着临床流行病学的发展,以此为基础的循证医学迅速发展起来。循证医学(Evidence Based Medicine, EBM)是遵循科学证据的临床医学。它提倡将临床医师个人的临床实践和经验与客观的科学研究证据结合起来,将最正确的诊断、最安全有效的治疗和最精确的预后估计服务于每位具体患者。而适应循证医学就成为重要临床资料——电子病历的必然发展方向。电子病历(Electronic Medical Record, EMR)是应用计算机技术实现的信息化病历^[1],EMR是基于一个特定系统的电子化病人记录,该系统提供用户访问完整准确的数据、警示、提示和临床决策支持系统的能力^[2]。这就要求电子病历能够体现规范系统化的治疗、准确详尽的病历记载、便捷的查询统计。

然而,由于医学概念名称不规范或者缺乏一个统一的规范命名系统,导致一个概念可以有很多种名称来表示,比如爱滋病和获得性免疫缺陷综合症等。而且,现在的检索查询关键词几乎没有任何语义,计算机不理解 Web 页面中词语表达的语义及其相互关系,从而导致查全率和查准率不高^[3]。自然语言中存在众多的同义词和近义词,传统的信息检索技术过分依赖用户的检索式,缺少语义分析和扩展能力,难以保证较好的查准率和查全率^[4]。因此检索的性能还是难以得到本质的提高。实现语义检索的关键在于对知识的表达和基于知识的推理^[5]。本体作为一种能在语义和知识层次上描述信息的概念模型建模工具,在信息检索,特别是基于知识的信息检索中得到广泛应用^[6]。本体技术的发展使上面的问题变得很容易解决。

本体(Ontology)是共享概念模型的明确的形式化规范说明^[7]。Perez 等人用分类组织法组织了 Ontology,归纳出 5 个基本的建模元语(Modeling Primitives)^[8]:类(class)或概念(concepts)、关系(relations)、函数(functions)、公理(axioms)和实例(instances)。本体提供对相关领域知识的共同理解,确定该领域内

收稿日期:2009-07-01;修回日期:2009-10-01

基金项目:吉林省教育厅科技计划项目(吉教科合字[2006]第75号)

作者简介:赵 洋(1982-),女,硕士研究生,研究方向为软件工程;李万龙,教授,研究方向为软件工程。

共同认可的词汇,并从不同层次的形式化模型上给出这些词汇和词汇间的相互关系的明确定义^[9]。本体能对知识共享和重用,故可以用来解决语义层次上信息的共享和交换。本体的描述语言主要有:SHOE^[10],XOL^[11],RDF,RDF-S^[12],OIL^[13],DAML,OWL^[14]等。

综上所述,将具有良好的概念层次结构和对逻辑推理的支持的本体引入到电子病历中,更够使其能更好地发挥作用。

1 基于本体的电子病历检索系统

1.1 电子病历检索系统体系结构

为了从专业的领域信息系统中全面地、准确地获取有效信息,文中利用领域本体来支持这一领域检索技术的实现。考虑到系统开发的高效性以及未来的可扩展能力,需要对其体系结构进行良好的定义,即既要有清晰的执行模块,又要有灵活的逻辑结构,这样才能为后续开发打下良好的基础。根据需要,把体系结构描述为三层模型,如图 1 所示。

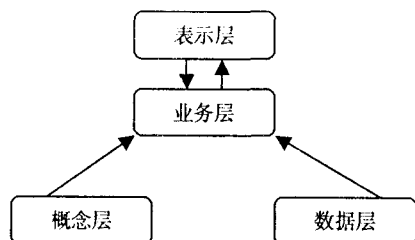


图 1 电子病历检索系统的体系结构

(1)概念层:指概念的构造,必须要明确本体的专业领域和范畴,而且要注意利用现有的本体。在构造时一般遵循 Gruber 提出的五条规则^[15]:

明确性和客观性:即 Ontology 应该用自然语言对所定义术语给出明确的、客观的语义定义;

完全性:即所给出的定义是完整的,完全能表达所描述术语的含义;

一致性:即由术语得出的推论与术语本身的含义是相容的,不会产生矛盾;

最大单调可扩展性:即向 Ontology 中添加通用或专用的术语时,不需要修改其已有的内容;

最小承诺:即对待建模对象给出尽可能少的约束。

(2)业务层:是本体系统的重要部分,主要是根据用户的查询要求,进行预处理,生成相应的查询条件,最后根据这些概念进行查询。在此层中,还会对最后返回的结果进行相应的处理。

(3)数据层:该层主要存放本体库和数据资源库,并且在数据源中间完成检索,返回检索结果。

1.2 电子病历检索系统详细框架

基于以上的系统结构,电子病历检索系统的详细框架如图 2 所示。

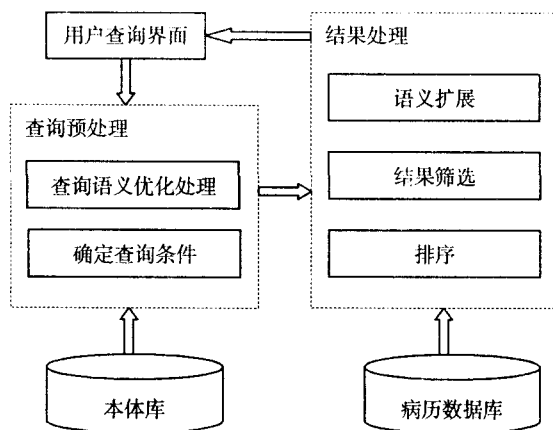


图 2 电子病历检索系统的详细框架

2 领域本体的构建

以肝炎为例,根据本体构建规则构造肝炎本体。在本体中可以描述的概念间的关系如下:(1)上下位关系,如:“肝炎”是“酒精性肝炎”的上位,“甲型肝炎”是“病毒性疾病”的下位。(2)整体部分关系,如:“肝脏疾病”是“消化系统疾病”的一部分。(3)同义词关系,如:“血清性肝炎”和“乙型肝炎”是同义词关系。另外在构建时还要注意对已有的本体库复用。

描述肝炎本体的部分代码如下:

```
<owl:Class rdf:ID="甲型肝炎">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="病毒性疾病"/>
  </rdfs:subClassOf>
</owl:Class>

<owl:Class rdf:ID="肝性脑病"/>
<owl:Class rdf:ID="酒精性肝炎">
  <rdfs:subClassOf>
    <owl:Class rdf:about="肝脏疾病"/>
  </rdfs:subClassOf>
</owl:Class>

<owl:ObjectProperty rdf:ID="同义词">
  <rdfs:range rdf:resource="血清性肝炎"/>
  <rdfs:domain rdf:resource="乙型肝炎"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID="病理改变">
  <rdfs:range>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <owl:Class rdf:about="急性肝炎"/>
```

```

<owl:Class rdf:about="慢性肝炎"/>
<owl:Class rdf:about="重型肝炎"/>
</owl:unionOf>
</owl:Class>
</rdfs:range>
<rdfs:domain rdf:resource="#乙型肝炎"/>
</owl:ObjectProperty>
<慢性肝炎 rdf:ID="慢性活动性肝炎"/>
<慢性肝炎 rdf:ID="慢性迁延性肝炎"/>
.....

```

3 查询过程及映射算法

根据系统的详细框架可知,查询过程是整个系统的主要部分,该模块是对用户查询条件进行处理并映射到本体库中得到扩展语义,在数据库中得到结果最终返回给用户的过程。

查询过程:

Step1:用户的初始查询进入到查询预处理模块;

Step2:在查询预处理模块中,首先进行查询语义优化处理,对初始查询进行剖析,最终返回查询中的核心语义词汇;

Step3:根据 Step2 中返回的核心语义词汇,确定查询条件。这样可以较好地保证用户的每一步查询请求对应到领域本体中的内容,从根本上确保查询用语的规范化,极大地提高了查询结果中信息的质量;

Step4:将查询条件映射到本体库中;

Step5:在结果处理模块中,根据上一层得到的概念的相关属性、关系和实例将概念进行语义扩展;

Step6:在病历数据库中进行查询,将所得到的结果进行筛选得到符合查询条件的结果,最后根据相似程度进行排序,返回结果。

在查询过程中会涉及到本体的映射,文中该处为一个映射算法,该算法的主要思想是本体 W_i ($0 < i < n$) 与本体库中的本体进行匹配,如果有完全匹配的就将匹配的 W_i 加入到结果集 $result$ 中,并将其子概念、父概念以及同义词等一放入结果集中。如果没有完全匹配的就查找 W_i 的父概念和子概念,具体算法如下:

本体的映射算法:

Input: 查询表达式 $Q = \{w_1, w_2, \dots, w_n\}$

Output: 结果集 $result$

Begin

$result = null$;

For ($i = 0$; $i < n$; $i++$) {

if(本体库中存在和 w_i 完全匹配的概念 w_j) {

$result = result +$ 相匹配的概念 w_j ;

确定概念 w_j 的关系和相关属性;

根据关系和属性得到 w_j 的相关联的概念,如子概念、父概念和同义词等;

$result = result +$ 相关概念集;

}

else

if(本体库中存在概念 w_i 的父概念集 $W - PARENT$) {

确定父概念的关系和相关属性;

根据关系和属性得到父概念的相关联的概念,如子父概念和同义词等;

$result = result + W - PARENT$ 以及相关概念集;

}

else

if(本体库中存在概念 w_i 的子概念 $W - CHILD$) {

确定子概念的关系和相关属性;

根据关系和属性得到子概念的相关联的概念,如子概念和同义词等;

$result = result + W - CHILD$ 以及相关概念集;

}

}

将得到的结果集 $result$ 根据相似度进行筛选;

Return $result$;

End

4 结束语

文中将本体引入了病历检索系统中,弥补了以往传统信息检索查全率和查准率不高的缺点,使得电子病历在医疗辅助决策中起到越来越重要的作用。但是本体研究是一项复杂的过程,还有很多工作有待进一步研究,如建立一个更为合理以及重用性更强的本体库,降低匹配算法的时间复杂度等。

参考文献:

- [1] 杨霜英,徐旭东,毛琦敏,等.电子病历系统在医院信息管理系统中的应用[J].医学研究生报,2006,19(3):258-260.
- [2] 满育红.电子病历的现状及应用[J].吉林医学,2007(1):139-141.
- [3] 蔡明,张体首.基于本体的搜索引擎研究[J].微计算机信息,2006,22(12-3):242-244.
- [4] 王存刚,王斌,姚文琳,等.基于 Ontology 的 Web 信息检索系统研究[J].计算机工程与设计,2008,29(24):6316-6319.

(下转第 217 页)

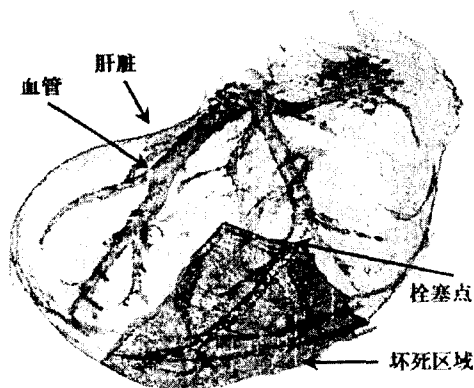


图5 带有肝实质的栓塞效果

在栓塞手术前无创地估算保留肝体积并选择合适的栓塞方案,尽可能彻底地切断肿瘤组织供血,同时确保栓塞手术后病人拥有足够维持生命的肝脏功能,是手术方案指定的重要原则。据临床研究,正常人肝脏可耐受70%组织损失而不影响生理需要,理论上栓塞后保留30%~40%的肝体积就能满足基本生命需要^[8]。根据栓塞仿真的结果,可以知道CT数据对应的肝实质点是否被栓塞。而在CT数据中,每个数据点代表一块体积,在实验中,每个体素为 $0.683 \times 0.683 \times 1.0$ 立方毫米的小长方体。模拟栓塞结果中,坏死肝脏的体积是195立方厘米,占肝脏总体积的17%。坏死区域内的肿瘤细胞将失去供血而“饿死”,肝脏损失17%的体积后不会影响病人生理需要。

4 结束语

文中提出一种肝脏血管栓塞治疗的计算机仿真方

法,用精确距离变换算法,把血管栓塞后的肝实质分为两部分:一部分为坏死区域,一部分为正常区域。利用该算法实验得到的效果图,可以清晰看到经栓塞后肝实质的坏死区域,从而可以为医生的栓塞治疗提供很好的参考评估。不过文中的理论基础是基于假定肝实质由距离最近的血管供血,而实际上肿瘤组织可能并非由单一血管供血,而是由多支血管供血的^[2],因此,在以后的研究中,文中的理论基础需要进一步深化。

参考文献:

- [1] 叶胜龙. 重视原发性肝癌的规范化治疗[J]. 中华肝脏病杂志, 2004, 12(8): 449-450.
- [2] 胡国栋. 肝癌的栓塞治疗及其超选择插管技术[J]. 生物医学工程与临床, 2001, 5(2): 118-119.
- [3] Rosenfeld A, Pfaltz J L. Sequential operations in digital picture processing[J]. Journal of ACM, 1966, 13(4): 471-494.
- [4] 刘相滨, 邹北骥, 王胜春. 一种新的完全欧氏距离变换算法[J]. 计算机工程与应用, 2005(13): 44-45.
- [5] Okabe N, Toriwaki J, Fukumura T. Paths and distance function on three-dimensional digitized pictures[J]. Pattern Recognition Lett., 1983, 1(4): 205-212.
- [6] Saito T, Toriwaki J I. New Algorithms for Euclidean Distance Transformation of An n-Dimensional Digitized Picture with Applications[J]. Pattern Recognition, 1994, 27(11): 1551-1565.
- [7] 程明, 黄晓阳, 王博亮. 医学图像中微细管道结构的提取[J]. 计算机工程与应用, 2007, 43(33): 221-223.
- [8] 黎一鸣, 吕凡, 白芝兰, 等. 国人正常肝脏体积计算公式的研究及意义[J]. 陕西医学杂志, 2002, 31(7): 634-636.

(上接第213页)

- [5] 丁晟春, 顾德访. Ontology 及其在信息检索中的应用研究[J]. 情报理论与实践, 2006, 29(1): 101-104.
- [6] 闫伟, 曹宝香, 张永红, 等. 基于本体的构件功能检索的设计与实现[J]. 计算机技术与发展, 2009, 19(5): 114-118.
- [7] Studer R, Benjamins V R, Fensel D. Knowledge Engineering, Principles and Methods[J]. Data and Knowledge Engineering, 1998, 25(1-2): 161-197.
- [8] Perez A G, Benjamins V R. Overview of knowledge sharing and reuse components: Ontologies and Problem-Solving Methods[C]//Proceeding of the IJCAI-99 workshop on Ontologies and Problem-solving Methods (KRR5). Stockholm, Sweden: [s. n.], 1999.
- [9] 邓志鸿, 唐世渭, 张铭, 等. Ontology 研究综述[J]. 北京大学学报: 自然科学版, 2002, 38(5): 730-737.
- [10] The SHOE Team. SHOE[EB/OL]. 2005. [http://www.cs.](http://www.cs.umd.edu/projects/plus/SHOE/)

[umd.edu/projects/plus/SHOE/](http://www.cs.umd.edu/projects/plus/SHOE/).

- [11] Karp P D, Chaudhri V K, Thomere J. XOL ontology exchange language[EB/OL]. 1999. <http://www.ai.sri.com/pkarp/xol/>.
- [12] W3C Working Group. RDF primer[EB/OL]. 2004-04-10. <http://www.w3.org/TR/rdf-primer/>.
- [13] Bechhofer S, Broekstra J, Decker S, et al. An informal description of standard OIL and instance OIL[EB/OL]. 2000. <http://www.ontoknowledge.org/oil/>.
- [14] McGuinness D L, van Harmelen F. OWL web ontology language overview[EB/OL]. 2004. <http://www.w3.org/TR/2004/REC-owl-features-20040210/>.
- [15] Gruber T R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing[J]. International Journal of Human-Computer Studies, 1995, 43: 907-928.