
本体在医疗信息检索中的应用

赵洋飞

(大连海事大学 信息科学与技术学院 辽宁 大连 116026)

摘 要 医疗信息涉及知识领域广泛,数据量庞大。由于用户的背景差异以及自然语言的二义性,导致医疗信息检索存在着一定的盲目性。由于本体具有良好的概念层次结构和对逻辑推理的支持,将本体引入到医疗信息检索中将大大提高医疗查询以及提高搜索医疗信息准确性。本文首先介绍了一些本体以及构建相关知识,进而阐述了本体在医疗信息检索中的相关应用,最后介绍了一个基于本体的医疗信息检索框架,该系统检索流程与传统信息检索系统相比,查全率和查准率均有一定提高,从而进一步说明本体应用在医疗信息检索中的重要性。

关键词 医疗信息; 本体; 医疗信息检索; 查准率

中图法分类号 TP351

Application of ontology in medical information retrieval

Zhao Yangfei

(College of Information Science and Technology, Dalian Maritime University, Liaoning Dalian,
116026, China)

Abstract Medical information covering a wide range of knowledge in the field, the huge amount of data. Due to differences in the user's background as well as the ambiguity of natural language, resulting in medical information retrieval there is a certain blindness. Because the body has a good support for the concept of hierarchy and logical reasoning, will be introduced into the body of medical information retrieval will greatly improve the health check and improve the accuracy of search for medical information. This paper describes some of the body and build knowledge, and then describes the ontology related applications in medical information retrieval, and finally introduces medical information retrieval framework based on ontology, the system retrieves the process compared to traditional information retrieval systems, recall and recall rates have certainly improved, thereby further illustrate the importance of the body in medical information retrieval.

Key words Medical information; Ontology; medical information retrieval; precision

1 引 言

随着互联网在全球化范围内的普及,网络数据正以前所未有的速度发展。从互联网上获取信息已经成为人们常用的方式。医疗信息是与人们息息相关的领域,患有不同程度疾病的人群和亚健康人群不可能频繁去医院,都希望通过网络得到相应的指导,但是目前的医疗网信息内容繁杂,结构混乱,不能给相关人员提供较好的服务。因此,提高医疗的信息检索服务至关重要的。

医疗信息的数据量与日俱增,用户的背景差异以及自然语言的二义性,导致医疗信息检索存在着一定的盲目性。本体^[1]作为一种能在语义和知识层次上描述信息系统的概念模型建模工具,通过概念间的关系来表达概念语义,从而能够捕获相关领域的知识,提供对该领域知识的共同理解,确定该领域内共同认可的词汇,并从不同层次的形式化模式上给出这些词汇和词汇间相互关系的明确定义。由于本体自身的优势,将语义本体论引入到医疗信息检索中,将大大的提高用户更高针对性的医疗查询以及提高搜索医疗信息准确性。

本体作为医疗信息检索的关键技术,其作用有三。第一,将医疗信息相关领域知识转化为医学领域本体;第二,本体进行结构化存储,提高信息检索效率;第三,基于本体进行医疗信息整合,整合医疗异构的数据源信息,为用户提供真正需要的数据。

2 相关概念和技术

2.1 本体

术语“本体论”源自哲学范畴,是与正在发生或存在的事物有关的研究领域。在哲学中,本体论是一种关于客观存在的理论。1998年 Studer 给出了本体较为精确的定义,认为本体是“共享概念模型的明确的形式化规范说明”。本体依照领域依赖程度可分为顶层本体、领域本体(Domain Ontology^[2])、任务本体、应用本体。领域本体描述的是特定领域中的概念及概念之间的关系,主要用来描述特定领域的知识(例如:医疗本体)。

定义^[3]: 领域本体可以用五元组表示:

$O = (C, R, Hc, rel, oA)$ 其中 C 表示概念标识符的集合; R 表示关系标识符的集合; $Hc \in C \times C$ 称作概念层次或分类; rel 表示概念间的关系,使概念不分类的联系起来; oA 表示本体公理集。

2.2 本体的构建

本体的构建是在领域专家的指导下建立相应的本体。常用的本体构建方法有^[4]: 企业建模法(Tove)、骨架法(Skeletal Methodology)、ETHONTOLOGY法、七步法等。本体构建工具主要有两大类:一类是基于AI的本体描述语言工具,一类是基于Web的本体描述语言工具。目前国内外众多本体研究机构的首选工具由斯坦福大学Stanford Medical Informatics开发的一个开放源码的本体编辑器Protégé^[5]。OWL本体构建的整个过程可以用图1描述。

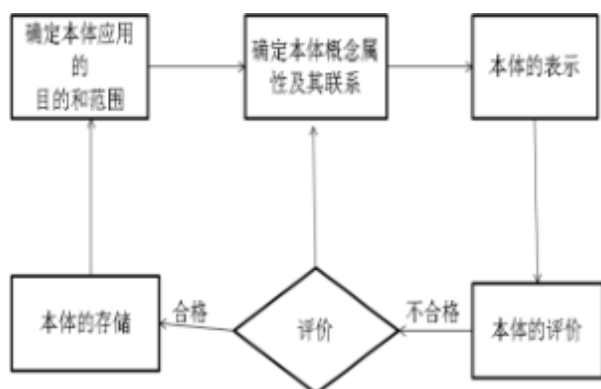


图 1 OWL 本体构建过程

2.3 信息检索

信息检索(IR, information retrieval)^[6]就是从大量收集的文档集合中,找到与给定的用户信息需求相关的恰当数目的文档子集。目前,信息检索技术可分为3类:全文检索、数据检索和知识检索。全文检索的特点是把用户的查询请求和全文中的每一个词进行比较,不考虑查询请求与文件语义上的匹配,这种方式可以保证查全率,查准率降低。数据检索的特点是查询要求和信息系统中的数据都遵循一定的格式,具有一定的结构,允许对特定的字段检索。知识检索强调的是基于知识的语义上的匹配,因此在查准率和查全率上有更好的保证。

3 本体在医疗信息检索中的应用

3.1 医疗本体

在医疗领域中构建医疗本体^[7]是为了有效地融合和关联疾病信息、药品信息、医生信息和医院信息,旨在帮助用户获取准确的医疗信息。图2为医疗本体模型。

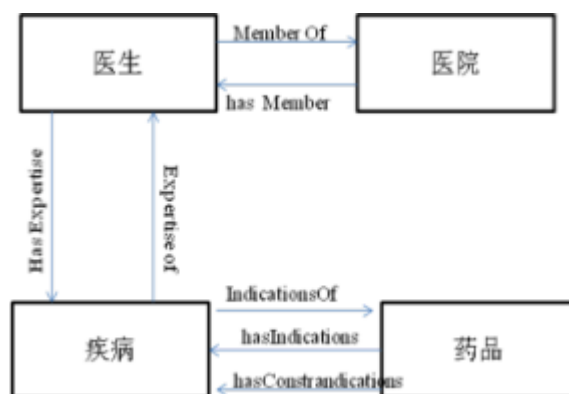


图 2 医疗本体模型

在构建的医疗本体为医疗系统间数据的交换提供了语义理解的基础但是各应用程序之间不能理解它们相互的语义表述,要解决语义异构,实现语义的互操作,主要方法就是将其分解成一系列的映射问题和集成问题,通过映射实现不同本体之间语义层面上的理解。在利用医疗本体进行检索的时候,首先提取用户查询语句中的概念(或实体),通过医疗本体进行语义分析,返回本体中用户所需要的知识;其次通过对用户查询的语义分析进行关键词扩展,挖掘潜在概念,实现更深层次的检索。

3.2 医疗本体信息整合

现代医疗系统依赖于对大量数据,信息的分析和处理,医疗信息系统可提供及时的和个性化的信息和知识服务。但由于各系统性质功能及服务对象的不同,造成医疗信息系统之间的资源沟通和信息交互存在着语义异构问题,这导致了医疗信息表示存储的异构多样性。为了解决医疗信息存在的语义异构问题,需要对本体的医疗信息整合^[8]。在基于本体的信息集成方法中本体被用作异构信息源语义的直接描述,可分为:单本体方法、多本体方法、混合方法。

3.3 基于本体的医疗信息检索过程

医疗本体在信息检索引擎中的应用的思路是利用医疗本体对信息的描述来检索知识库,从而提高检索的效率和精确度。其基本思想^[9]如下:(1)建立医学领域本体。(2)信息的收集、组织和存储:由信息收集器收集信息源中的数据,并参照已建立的本体,把收集来的数据按规定的格式存储在元数据库(关系数据库、知识库等)中。(3)查询处理:对用户检索界面获取的查询请求,查询转换器按照本体把查询请求转换成规定的格式,在本体的协助下从源数据库中匹配出符合条件的数据集合。(4)检索结果处理:检索的结果经过定制处理后,返回给用户。

4 基于本体的医疗信息检索框架

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,所以将本体引入医疗领域中。文中介绍了一种基于本体的医疗信息检索系统框架^[10],通过一种映射算法实现概念的语义扩展,从而提高了医疗检索的查全率和查准率。

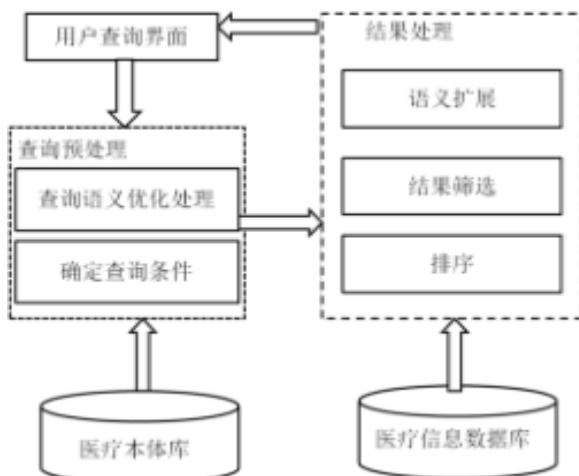


图 3 基于本体的医疗信息检索框架

图 3 中医疗检索模型各模块相互协作,检索过程包括以下几步。(1) 用户输入查询条件,进入到查询条件处理部分,查询条件处理部分包括语义化处理模块、查询条件确定模块:将语义化处理后的查询输入交给查询条件确定模块,然后把结果传递给查询转换模块。(2) 进入结果信息处理部分,包括三个模块。一是查询转换模块:按照所得信息对查询进行转换,使查询具有语义信息,在医疗信息数据库中进行查询,对不能确定的查询按照关键字查询处理;二是检索抽取:在查询结果中进行筛选;三是排序:采取文档相似度的方式进行显示,即按照语义相似度由高到低的顺序进行显示。

5 本体的医疗信息检索中的作用分析

通过上述模型分析,本体在医疗信息检索中的作用^[11]主要体现在以下方面:查询扩展:用户的检索表达式往往由一个或几个关键词构成,很难准确表达检索意图,一种有效地扩大查全率的方法。本体的作用可以说是组织“世界上的概念”,并且将它们关联到语言学上的表达。判断查询与文档的相关性,可以从识别文档相关的词汇作为出发点,来建立概念之间或其他代表这些概念或相关概念的词汇单元之间的关联。信息抽取:本体允许多重继承关系,即本体中每个概念能从不只一个更高层次域中得到属性和概念关系。如果本体与术语数据库有关,我们将获得与每个概念相关的术语的不同语言信息。因而能够提高信息抽取的正确率,提供准确简洁的信息源供检索查询。自动分类:对文档内容的正确分类通过缩小检索范围从而达到提高检准率的效果,可以看作是文档内容分析和表示的中间环节。语义形式化表示:在为信息检索构造文档工具的过程中,本体在特定的领域,通过语义形式化表示,提供了一种概念关系的清晰定义,使得能够得到一个逻辑的、一致的概念表达,能同时被人和计算机所理解,更深层地揭示文档内容。推理机制:利用本体的推理机制可以提高查询的语义能力。

6 结束语

基于本体的信息检索和信息存储,改善了搜索引擎和知识管理之间的界面。使得网络信息搜索智

能化程度更高。本文介绍了本体知识和医学专业搜索引擎在医疗领域搜索中存在的问题,将本体引入医疗信息搜索,并描述一个系统框架和实现的流程,分析了本体的医疗信息检索中的作用。为医学搜索引擎的设计和开发提供了良好的理论依据。但是,在未来的进一步研究中还有很多工作要做,相信未来本体在医疗信息检索领域会有广泛的应用前景。

参 考 文 献

- [1] 杜小勇,李曼,王珊 本体学习研究综述[J].软件学报,2006,17(9):1837-1847.
- [2] KuanDai,DongHan,HuiZhao,HangLi.FocusedWeb CrawlerBasedonDomainOntology[A].IEEEBeijingSection,China:2013:4.
- [3] 李明,王丽 基于本体的信息检索系统模型[J] .兰州理工大学学报, 2007, 33(2): 90-93.
- [4] 黄伟.本体构建与语义集成研究[D].东南大学硕士学位论文.2005
- [5] 基于本体论和数字图书馆的信息检索[J] .情报学报, 2003, 22(6): 648-652
- [6] 栾艳,丁二玉,骆斌等 基于 Ontology 的语义检索技术 [J]. 计 算 机 工 程 与 应 用 41(28):156-159.DOI:10.3321/j.issn:1002-8331.2005.28.048.
- [7] 杜海波. 本体在医疗信息检索系统中的应用研究[D].长春工业大学 2010.
- [8] 白杰英. 本体在医疗信息整合中的应用与研究 [D].长春工业大学,2010.
- [9] 张子振,李旭辉. 本体在市场信息检索中的应用研究 [J]. 广 东 石 油 化 工 学 院 学 报,2011,01:52-54
- [10] 孔田野,李万龙,张海鸥等.基于药品本体的信息检索系统研究 [J]. 河 北 科 技 大 学 学 报,2008,29(3):223-226.
- [11] 吴丹. 本体在信息检索中的作用及实例研究 [J].情报杂志,2006,06:72-75