

面向医疗行业的智能问答系统研究与实现

刘 芳, 于 斐

(扬州市职业大学 工商管理学院, 江苏 扬州 225009)

摘 要: 为了实现医疗信息化中的疾病与诊疗的智能问答系统, 本文提出一种基于人机交互的用户自然语言问题理解方法. 该方法在关键词库的基础上, 采用正向最大匹配算法对用户咨询进行词法分析, 进而对医疗领域的问题进行分类, 系统根据问题类别进行数据检索; 并利用模板匹配方法, 完成答案的提取, 以句子作为答案的基本单位. 最后, 在 JSP 和 Java 平台上实现了 B/S 模式的医疗领域智能问答系统, 验证了该方法的有效性.

关键词: 医疗行业; Q&A; 智能; 实现

中图分类号: TP315

文献标识码: A

文章编号: 1000-7180(2012)11-0095-04

Research and Development of Intelligent Q&A System for Medical Industry

LIU Fang, YU Fei

(School of Business and Management, Yangzhou Polytechnic College, Yangzhou 225009, China)

Abstract: In order to realize the intelligent question answering system for disease diagnosis and treatment in medical informationization, this article proposed a method for understanding the user's question described by nature language based on human-computer interaction. On the basis of the keyword library, the method use the maximum matching algorithm for lexical analysing of the user consulting. And then classify the issues of medical field, the system take the data retrieval according to the categories of issues. The system will also use the method of template matching for extracting the answer, with the sentence as the basic unit of answer. Finally, intelligent question answering system of medical field is realized based on B/S model on the platform of JSP and JAVA, it prove the effectiveness of this method.

Key words: medical industry; Q&A; intelligent; development

1 引言

医疗领域一直是信息化发展的重要方向, 医疗行业的信息化投资规模逐年增加. 医疗卫生事业关系国计民生, 而医疗信息化则是医疗卫生事业能否在信息时代更好地服务于患者、服务于社会的重要支撑, 正受政府部门及越来越多企业和组织的高度关注. 智能问答系统能够让患者家属、医生、护士等人群随时随地实时了解业内相关信息与最新动态, 自动化、信息化、高效化、集中化、智能化实时服务于社会^[1].

问答系统(Question Answering System, QA)是信息计算系统的一种高级形式^[2]. 它能用准确、简洁的自然语言回答用户用自然语言提出的问题. 其研究兴起的主要原因是人们对快速、准确地获取信息的需求. 问答系统是目前人工智能和自然语言处理领域中一个备受关注并具有广泛发展前景的研究方向^[3].

自动问答系统的历史可以追溯到 1950 年, 著名的英国数学家图灵(A. M. Turing)第一次提出了“机器智能”的概念, 并提出判断计算机是否具备智能的实验方法——“图灵测试”, 也就是通过自然语

言问答的方式,判断计算机是否具备人的智能^[4].“图灵测试”可以看作是问答系统的理想蓝图.60年代末到70年代,专家系统的出现使人工智能研究出现新的高潮^[5].80年代以来,人工智能已经进入了它发展高潮的黄金时代,而且其研究成果层出不穷,有很多项目获得了重大成果.特别是当前人工智能的一个重要子领域自然语言处理更得到了迅速的发展,目前在这个领域中也显示了技术的商业性应用.

本文研究的是医疗领域基于模板匹配的智能问答系统,系统主要包括两个模块:问题理解、模板匹配.在关键词库的基础上,采用正向最大匹配算法对用户咨询进行词法分析,进而对医疗领域的问题进行分类,系统根据问题类别进行数据检索;并利用模板匹配方法,完成答案的提取,以句子作为答案的基本单位.最后,在JSP和Java平台上实现了B/S模式的医疗领域智能问答系统,验证了该方法的有效性.

2 系统架构

面对智能问答系统的工作流程以及医疗行业的业务知识,结合用户的使用习惯,本系统可以解决以下问题:设计友好的人机交互界面,简单快速的操作;用户针对医疗卫生领域对象人进行自然语言提问;系统根据用户的提问返回自然语言答案;

系统主要功能包括:用户模块:注册、登录及身份验证;功能处理模块:问题理解、模板匹配.

智能问答系统主要由三部分构成,一部分是WEB浏览器端的人机交互界面,第二部分是位于WEB SERVER模块中的功能处理模块,第三部分是存储研究领域知识的知识库.其中第二部分是核心部分,主要负责处理用户的提问以及给用户反馈相关信息.其模型如图1所示.

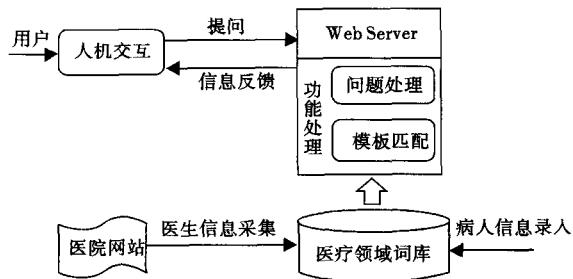


图1 系统模型图

3 用户问题的自然语言理解

问题理解有时也称为问题分析.在此过程中关

键就是对问题的内容进行深刻的理解.只有首先明确用户提出的问题所要表达的正确含义以后,才能够识别问题的疑问中心对其进行关键词抽取,用于模板匹配和答案生成,才能使用户得到满意的答案.

对于用户用自然语言提出的问题,问题理解模块对之进行基本的预处理,包括分词,提取关键词,最后将关键词的集合提交给模板匹配模块.问题理解流程如图2所示.

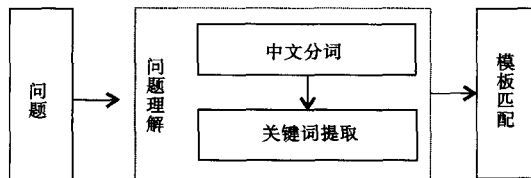


图2 问题理解流程

用户在提问时通常表现了一定的目的性,即知识需求,这种知识需求在使用自然语言进行查询时可以通过疑问句或者祈使句等形式体现出来.下面是用户可能对问答系统输出的不同的问题:

“YZH市第一人民医院 TLN 医生出诊时间是多少?”

“病人林海的病房号是多少住院时间是什么时候?”

“YZH市消化内科的医生有哪些?”

对于每种问句按上文的知识表现形式对知识需求的拆解来采集句模.表1中列出了本文研究所涉及的问题类型、所对应的疑问词举例、预期的答案类型,以及问题的举例.

表1 问题类型

问题类型	疑问词 举例	预期的答 案类型	问题举例
询问人	谁	人名	病人林海的主治 医生是谁?
询问地点	哪里	地点名词	病人林海的病房 在哪里?
询问时间	何时	时间名词	SZH 医生的出诊 时间是何时?

我们分析了常见的医疗系统的咨询问题集,以满足需求为基本要求,定义了300多个基本句型模板,常见的句型举例如下:

<模板1>::=<C>(的)<C.A>是多少?

例句:WXH 医生的会诊时间是多少?

林海病人的病房号是多少?

<模板2>::=<C.A>(的)[医生|病人]有哪些?

例句:消化内科的医生有哪些?

病房2-108的病人有哪些?

领域词库是中文分词的基础,词库的构建直接

影响分词的速度以及准确性.通过咨询业内人士,浏览多家医院网站,总结出本文研究涉及到的相关字典:

(1) 科室词库:消化内科、血液内科、儿科、神经科、皮肤科、中医科、妇产科、麻醉科、耳鼻喉科、超声诊断科、病理科等.

(2) 职称词库:主任医师、教授、主任、副主任、研究员.

(3) 疾病词库:神经系统疾病、消化系统疾病、循环系统疾病、呼吸系统疾病等.

(4) 传染病词库:甲类传染病、乙类传染病、丙类传染病.

在问题理解后,还需要根据理解的结果生成答案,这个过程,我们通过查询数据库完成.我们定义了一个问题模板查询数据库的 SQL 接口,完成这一功能.在本节,我们通过下面一组示例,展示问题模板的数据库接口设计和功能.

示例 1:YZH 市第一人民医院唐丽娜医生出诊时间是多少?

- NULL * * SQL * * <TLN><医生><是><# DEPARTMENT #><医生>,<# SPECIALISTTIME #>出诊,<YZH 市第一人民医院><的><联系电话><是><# CONSULTTEL #>,<欢迎就诊>.

- * * SQL: SELECT DEPARTMENT, SPECIALISTTIME, CONSULTTEL
FROM DOCTOR
WHERE HOSPITAL='YZH 市第一人民医院' AND DNAME='TLN'

- 智能回答:TLN 医生是妇产科医生,周四下午出诊.YZH 市第一人民医院的电话是 87906555,欢迎就诊.

示例 2:病人林海的病房在哪?住院时间是什么时候?

- NULL * * SQL * * <病人><林海><的><病房><在><# HOSPITAL #><# HOSPITALBED-NO #>,<住院时间><是><# PATIENTSTAY #>,<谢谢慰问>.

- * * SQL: SELECT HOSPITAL, HOSPITALBED-NO, PATIENTSTAY
FROM DOCTOR, PATIENT
WHERE DOCTOR. DNAME = PATIENT. RESPONSEPHYSICAN AND PNAME='林海'

- 智能回答:病人林海的病房在 YZH 市中医院 2-108,住院时间是 2011/1/30,谢谢慰问.

示例 3:病人林海的负责医生是谁?家庭电话

是多少?

- NULL * * SQL * * <病人><林海><的><负责医生是><# DNAME #>,<家庭电话是><# HOMETEL #>.

- * * SQL: SELECT HOMETEL, DNAME
FROM PATIENT, DOCTOR
WHERE PNAME='林海' AND DOCTOR. DID = PATIENT. RESPONSEPHYSICAN

- 智能回答:病人林海的负责医生是 WXH,家庭电话是 81234567.

示例 4:SB 医院 SZH 医生负责的病人有哪些?

- NULL * * SQL * * <SZH><医生><的><病人有><# PNAME #>

- * * SQL: SELECT PNAME
FROM DOCTOR, PATIENT
WHERE DOCTOR. DID = PATIENT. RESPONSEPHYSICAN AND DNAME='孙志浩'

- 智能回答:SZH 医生负责的病人有:HL XXY

示例 5:YZH 市消化内科的医生有哪些?

- NULL * * SQL * * <YZH 市><消化内科><医生有><# HOSPITAL #><# DNAME #>,<医生简介:><# INTRODUCTION #>.

- * * SQL: SELECT HOSPITAL, DNAME, INTRODUCTION
FROM DOCTOR
WHERE DEPARTMENT='消化内科'

- 智能回答:YZH 市消化内科医生有 YZH 市中医院 WXH,医生简介:美国留学 6 年,SS 医院 ZYP,医生简介:从业 15 年,资历丰富.

示例 6:林海病人是否出院?

- NULL * * SQL * * <LH><病人><# Discharge #>

- * * SQL: SELECT DISCHARGE
FROM PATIENT
WHERE PNAME='LH'

- 智能回答:LH 病人未出院.

4 智能问答的实现

智能问答模块是本系统的核心模块,该模块主要由问题理解和模板匹配两部分组成,以下主要介绍智能问答模块的核心代码.

数据流图(Data Flow Diagram)从数据传递和加工角度,以图形方式来表达系统的逻辑功能、数据

在系统内部的逻辑流向和逻辑变换过程,是结构化系统分析方法的主要表达工具及用于表示软件模型的一种图示方法.本系统智能问答模块的数据流图如图3所示.

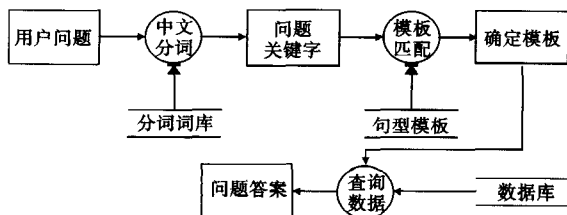


图3 智能问答数据流图

4.1 FMM 核心代码

```

public Vector<String> Segment(String sentence, Dictionary
dic) {
    int maxLength=dic.getMaxLength();
    int pos=0;
    int targetLength=maxLength;
    int restLength=sentence.length();
    Vector<String> seged=new Vector<String>();
    while (restLength> 0)
    {
        if (targetLength> restLength)
            targetLength=restLength;
        String tempStr= sentence.substring(pos, pos+target-
Length);
        if (dic.checkWord(tempStr) || targetLength==1)
        {
            seged.add(tempStr);
            pos+=targetLength;
            targetLength=maxLength;
            restLength=sentence.length()-pos;
        }
        else
            targetLength--;
    }
    return seged;
}
  
```

4.2 智能问答实例解析

本节以问题“YZH 市消化内科的医生有哪些?”当用户点击[提交问题]按钮之后,系统调用 Select-Answer. java Servlet 程序,对问题做如下处理:

Step1: 对问题“YZH 市消化内科的医生有哪些?”进行中文分词.

分词结果为:YZH 市|消化内科|的|医生|有|哪些? 为了表示方便将结果记为集合 R,即 $R=\{\text{YZH 市, 消化内科, 的, 医生, 有, 哪些?}\}$

Step2: 将集合 R 中的元素分别与句型模板的关键词匹配,其中“消化内科”与句型模板关键词“科”相匹配,确定

模板.

Step3: 执行 SQL 语句,“消化内科”是一个参数传递到 SQL 语句中作为查询条件查询数据:“SELECT HOSPITAL, DEPARTMENT, DNAME FROM DOCTOR WHERE DEPARTMENT LIKE ‘%’+param+‘%’”;

Step4: 根据查询结果,以自然语言的形式组织答案.答案模板如下:answer=param[0]+param[1]+“的医生有”+param[2]+“.”; 其中 param[0],param[1],param[2]分别依次为 SQL 语句中查询的三个字段 HOSPITAL、DEPARTMENT、DNAME 的值.即 answer=[HOSPITAL][DEPARTMENT]的医生有[DNAME].

Step5: 返回答案.

5 结束语

智能问答系统是目前 IT 行业研究的热门领域,是今后发展的必然趋势,对于医疗行业还没有相对成熟的智能问答系统问世.居民健康档案管理系统已经在国内多个城市展开研发,YZH 市居民健康档案管理系统即将投入使用.所以智能问答系统,可以作为居民健康档案管理系统的一个子系统便于居民对医疗信息的快捷获取.

本文采用的是关键词模板匹配算法,这种算法虽然提取关键词比较简单,但是受到分词算法词长的限制,要在两种算法之间权衡,举行模板的关键词提取也变得困难.

参考文献:

- [1] 吴晓珊 王茜.加强统计信息工作促进医院信息化管理[J].中国医院统计,2004,11(1):66.
- [2] 季永华,许华虎,沈敏,等.自动答疑系统的研究与实现[J].计算机工程与应用,2005,41(14):224-226.
- [3] Li SJ. Semantic computation in a chinese question answer system[J]. Computer. Science and Technology, 2002, 17(16): 933-939.
- [4] Burke R D, Hammond K J, Kulyukin. Question answering from frequently asked question files: experiences with the FAQ finder system P[J]. AI Magazine, 1997(18):57-66.
- [5] 钟义信.中国人工智能进展[M].北京:北京邮电大学出版社,2001:1129-1132.

作者简介:

刘 芳 女,(1979—),硕士,讲师.研究方向管理信息系统、企业建模及人机交互系统等.