

## 基于中医药知识图谱的智能问答与用药推荐系统

文/郑懿鸣 翟洁 胡晓龙 施轶劼 刘浩

## 摘要

将知识图谱与知识卡片相结合,使得用户既可以阅读文字以获取详细解释,又可以通过视觉的图形来感知知识本体之间的相关性,文字与图形相辅相成,配合得当。本智能问答系统是将无序的用户语料信息,进行科学有序的整理,通过CRF分词技术处理,提取自然语言关键词信息,并基于知识图谱基本原理获得反馈给用户的最终答案。作为用药的辅助推荐信息,以知识图谱和属性列表同时呈现。

【关键词】知识图谱 智能问答 中医药

## 1 引言

在现有的自然语言处理技术、知识图谱构建相关研究的基础上,结合各种方式的优势,同时利用尽可能丰富的数据源,专门针对中医

药的查询提出了较为适合的用于本系统的句式分析规则;同时,提供了知识图谱与知识卡片相结合的方式更清晰地展示一种中药实体的具体信息。

将知识图谱与知识卡片相结合,用户既可以阅读文字以获取详细解释,又可以通过视觉的图形来感知知识本体之间的相关性。文字与图形相辅相成,丰富了系统内涵,使得系统为用户提供的资源不再单调,提高了系统的感知性。

## 2 总体设计

## 2.1 设计重点

## 2.1.1 中医药知识库的构建

本系统采用自顶向下的策略,依次经过模式图定义、知识抽取、知识融合三个步骤完成基础知识库的构建。模式图包含知识库拥有的概念、概念的属性,以及概念之间的层次关系。知识抽取主要包括现有医学文献、医疗健康网站、中医药百科站点的知识抽取中医药相

关的实体、实体类型、同义词关系和“属性-值”关系。知识库的模式图常用中医药知识库主要包括中药材、中医证候、中医疾病等上层概念及概念的属性,构建模式图。基于双向LSTM-CRF模型的中医药实体识别方法。常用中医药知识库构建中,需要从专业中医药文献文本中识别药品相关实体。

采用基于特征模板的方法,利用大规模语料学习出标注模型,进而对语句进行标注,特征模板通常是人工定义的一些二值特征函数,挖掘命名实体内部以及上下文的构成特点。采用基于双向LSTM-CRF模型的中文实体识别方法来识别专业医学典籍文档中的中医药等实体。与传统LSTM不同,双向LSTM同时考虑了过去的特征和未来的特征,结果更具准确性。

## 2.1.2 基于自然语言处理技术的中医药智能问答

自然语言的处理通常采用自然语言技术(如命名实体识别等)对问题进行深层次的分割、理解。智能问答系统按照对数据的处理方

## &lt;&lt; 上接 133 页

PS软件也可以实现三维视觉效果,但是其设计空间非常狭隘,只可以通过色泽变化及阴影进行立体映射,而不能对立体图形的多维视角进行延伸,所以,需要借助3DMAX软件来补充视觉特征。

## 2.4.2 使用结构调整方案调整图像比例盲区

由于初始图片空间尺寸发生错误,导致产生图像比例盲区。这类设计问题,造成网页设计趋于单一,没有行成强烈的视觉效果。在网页设计时没有完全的黄金比例,大多都是根据设计师的以往经验形成,但是仍然存在明显的空间比例规律,特别是近年来移动终端的用户逐年增长,H5页面形成的视觉效果直接影响实际的网页流量。所以,针对网页主题、风格、比例的设计,必须综合考量H5页面与Web页面的对比效果,并且分析用户的浏览数据进行网页元素的补充调整方案。Web页面相关的文本信息在原则上可以扩充,但是由于H5页面元素设计要求集中。可以使用PS软件在设计时设置相应的网格比例分区,对主要网页元素设定空间比例,结合链接导向的重要性排序,对空间结构是否合理进行对比。网页设计中综合运用多类图形结构非常重要,比如,使用方形图像,对用户而言传递过来的信息比较严谨

正规,比较适用于学术研究网站或各类新闻网站的网页设计。圆形图像传达的圆润柔美的信息,能够带给用户愉悦的情感体验,比较适用于娱乐网站或购物网站的网页设计。三角线图像棱角尖锐,代表的着个性化、突出性的形象特点,多用于大多的各类个性化网站的设计。但是就设计本质而言各类图像结构的应用必须依附于网站的主旨,处理图像比例错误,才能够促使网页整体结构的主题更加突出,更加人性化。使网页设计内容更加丰富,页面效果更加突出。

## 2.4.3 对体验需求盲区进行个性化创新

网页浏览带有强烈的商业目的,设计评估的重要衡量标准是网页设计与用户体验需求是否契合。用户体验的需求愈来愈高,基于目前常用的网页设计软件已经无法满足用户的需求从而导致用户体验盲区的网页设计空白。所以,需要图像处理软件进一步的开发并且灵活的综合应用多种图像处理软件,才可以从真正意义上避免用户体验盲区的发生。力争人性化的网页设计,使设计元素更加丰富,更具有创意感,使体验效果更加舒适,在网页设计中综合运用多种图像处理技术,使网页设计别具一格。

## 3 结束语

在网页设计中合理的应用图像处理技术,能够使网页设计更加的科学合理更加的美观,并且精美的图像制作使网页更加的档次,能够带给浏览者更好的视觉体验,从而增加网页浏览量。因此,图像处理技术的合理应用在网页设计中具有重要意义。

## 参考文献

- [1] 丁黎明. 图像处理技术在计算机网页设计中的应用[J]. 信息与电脑(理论版), 2019(13): 3-4.
- [2] 包金海. 计算机网页设计中图像处理技术的应用[J]. 计算机产品与流通, 2019(07): 29.

## 作者简介

曹利(1976-),男,山西省大同市人。硕士学历。讲师。研究方向为公共计算机基础教学、计算机应用技术。

## 作者单位

晋中师范高等专科学校 山西省晋中市 030600

式可以分为多种, 尽管不同类型的问系统对于系统整体模块的分工和实现存在一定差异, 在一般情况下, 依据 Q&A 系统的数据流程处理过程中, 处理问题的框架都包括了问题理解、信息搜索和答案生成三个功能模块。

### 2.1.3 基于 KK 算法和节点生成算法的中医药知识图谱可视化

知识图谱是一种以“语义网络”为骨架而搭建起来的巨型网络知识系统。它能够捕获并展现领域概念之间的语义关系, 使得各种知识资源中零碎、松散的知识本体互相联系。将数据图形化是生成知识图谱的核心之一。作为一个可视化系统, 界面必须清晰易懂, 因此需要过滤部分次要信息, 提取主要信息, 并对结果随机排序。

KK 算法作为力引导图布局算法, 由一个双层循环构成。外层循环改变当前正在选择最佳位置的节点, 内层循环控制此节点可改变位置的次数。实则就是在微小移动一个节点的同时, 固定其他所有节点。KK 算法将迭代次数的限制致力于内循环上。控制所有节点的移动次数上限, 保证了即便无法达到最完美布局, 但整体效果依然呈现出平衡、稳定的效果。

### 2.2 设计难点

(1) 中文语言与其他的语言最大的不同的地方在中文的词汇之间没有空格使其分隔开, 这加大了语言处理算法对语句分割、语义理解的难度。

(2) 由于本系统是以中医药为核心, 必须格外注意源数据的权威性。同时在建立中医药知识库的构建过程中, 会存在词典需求特殊的问题, 需要进行特别的词性标注。

(3) 在带有歧义词语的句子中, 经常会出现非期望的结果。在对用户输入的自然语言进行处理时, 需要通过调用 HanLP 的相应方法来实现对分词进行整合。但是, 在带有歧义词语的句子中, 仍会不可避免地出现非期望的结果。

(4) 可视化的知识图谱布局平衡问题。需要使用相应算法来实现知识图谱布局的美观和易读性。否则图谱节点连线会出现互相交叉的现象, 降低图谱的阅读性。

## 3 功能实现

### 3.1 分词实现

基于对用户自然语言处理接受的需要, 必须对用户输入的自然语言进行处理。处理的方法采用了 CRF++ 技术对用户输入进行基本分割, 并根据基本名词短语识别 (BaseNP)

的 4tag 模式。通过对已经标注好的每个单字, 根据它的 BaseNP 标记进行整合。这一部分通过调用 HanLP 的相应方法来实现。经过多次实践发现, CRF 对于陌生的词汇具有较高的识别率, 但也不可避免地出现少数非期望的结果, 特别是在带有歧义词语的句子中。此时需要调用其他的分词器来进行校正。

### 3.2 词典实现

通过 CRF 获得分词结果之后, 需要对结果进行整理, 并标注词性。采用了人民日报 2014 年词性标注库进行词性标注。针对该系统的特殊需求, 新增了中药、疾病、信号词、疑问词等词性的特殊词性标注, 方便了该系统通过此行来过滤关键词、信息词的提取。

由于用户时常不能准确定位所需检索的问题的关键词信息, 所以有必要进行同义词匹配的处理。本系统所采用的词典是根据哈工大信息检索研究中心的《同义词词林 (扩展版)》而修订的同义词词典。其为原词典的简化版, 并且加入了很多专用于中药特性的词语, 以更好地进行同义词匹配。

### 3.3 用药推荐

当用户输入的词语是疾病或者是症状类型词时, 系统将对对应的症状与数据库进行匹配, 检索出与之相关的中医药品类, 按照相关程度给出适合该症状或疾病的中医药列表作为用药的辅助依据。

## 4 小结

本系统实现了依照用户输入的书面或者口语化语言, 进行汉语语言标注; 根据信号词进行句式模板匹配, 根据一定的匹配原则, 产生相应的查询语句, 在已有的中医药知识库中检索需要提供给用户的信息, 并通过知识卡片、可视化图谱、用药推荐的方式呈现出来。

(通讯作者、指导教师: 翟洁)

## 参考文献

- [1] 田久乐, 赵蔚. 基于同义词词林的词语相似度计算方法 [J]. 吉林大学学报信息科学版, 2010, 28 (6): 602-608.
- [2] 贾李蓉, 刘静, 于彤等. 中医药知识图谱的构建 [J]. 医学信息学杂志, 2015, 36 (8), 51-59.
- [3] 王秀萍. 基于中文分词及知识本体技术的课程智能问答系统的研究 [D]. 长沙: 湖南大学图书馆, 2015.
- [4] 秦小波. 设计模式之禅 [M]. 北京: 机械

工业出版社, 2014.

- [5] 刘凯, 周雪忠, 于剑等. 基于条件随机场的中医临床病历命名实体抽取 [J]. 计算机工程, 2014 (9): 312-316.
- [6] 李爽, 陈丽. 国内外网上智能答疑系统比较研究 [J]. 中国电化教育, 2003, 192: 80-83.
- [7] 徐永东, 权光日, 王亚东. 基于 HL7 的电子病历关键信息抽取技术研究 [J]. 哈尔滨工业大学学报, 2011, 43 (11): 89-94.
- [8] 陈扬. 基于命名实体识别的学科智能答疑模型研究 [D]. 东北师范大学, 2010.
- [9] 邢超. 智能问答系统的设计与实现 [D]. 北京: 北京交通大学图书馆, 2015.
- [10] 陈宏朝, 李飞, 朱新华等. 基于路径与深度的同义词词林词语相似度计算 [J]. 中文信息学报, 2016, 30 (5): 80-88.
- [11] 阮彤, 孙程琳, 王昊奋等. 中医药知识图谱构建与应用 [J]. 医学信息学杂志, 2016, 37 (4): 8-13.
- [12] 龚凡, 王梦婕, 阮彤等. 电子病历文本症状自动识别方法 [J]. 医学信息学杂志, 2016, 37 (7): 7-14.
- [13] 叶枫, 陈莺莺, 周根贵等. 电子病历中命名实体的智能识别 [J]. 中国生物医学工程学报, 2011, 30 (2): 256-262.
- [14] 胡芳槐. 基于多种数据源的中文知识图谱构建方法研究 [D]. 华东理工大学, 2015.
- [15] 瓦尔兹, 朱一凡, 赵鑫虎. 自然语言问答系统 [J]. 新疆社科信息, 2013 (3): 10-14.
- [16] 陈飞, 刘奕群, 魏超等. 基于条件随机场方法的开放领域新词发现 [J]. 软件学报, 2013 (5): 1051-1060.
- [17] 谢天罡. 基于 Web 的自然语言问答系统的研究 [D]. 东南大学, 2001.
- [18] 王博. 基于本体的智能化问答系统研究 [D]. 江西师范大学, 2009.
- [19] HONG MingCai, ZHANG Kuo, TANG Jie, 等. A Chinese Part-of-speech Tagging Approach Using Conditional Random Fields 基于条件随机场 (CRFs) 的中文词性标注方法 [J]. 计算机科学, 2006.

## 作者简介

郑懿鸣 (1997-), 男, 河南省焦作市人。大学本科在读。计算机科学与技术专业。

## 作者单位

华东理工大学 上海市 200237