



开放科学
(资源服务)
标识码
(OSID)

专利计量情报问答系统研究与实践

吕璐成^{1,2} 韩涛^{1,2} 王燕鹏¹

1. 中国科学院文献情报中心 北京 100190;
2. 中国科学院大学经济与管理学院图书情报与档案管理系 北京 100190

摘要: 面向科技情报机构对于高效、实时、精准、智能科技情报服务方式和产品的需求, 针对专利计量分析服务, 开展专利计量情报问答系统的研究和实践。研究对专利计量情报问答系统的概念进行了界定, 基于自然语言理解和自然语言生成技术, 提出了以问句理解模块和答句生成模块为核心的专利计量情报问答系统实现技术方案。研究设计并实现了人工智能领域专利计量情报问答系统, 能够应对同一问题的多种问法给出正确答案, 对科技情报机构提升科技情报服务质量和用户体验具有积极意义。

关键词: 问答系统; 智能情报服务; 任务型对话系统; 专利计量

中图分类号: G306, G350.7

Research and Application of Intelligence Question-answering System on Patent Bibliometrics

LYU Lucheng^{1,2} HAN Tao^{1,2} WANG Yanpeng¹

1. Development of Documentation and Information Centre, Chinese Academy of Sciences, Beijing 100190, China;
2. School of economics and management, Chinese Academy of Sciences, Beijing 100190, China

Abstract: Aiming at the demand of high-efficiency, real-time, accurate and intelligent S&T information service mode and products from S&T intelligence agencies, the research and application of intelligent Q&A system on patent bibliometrics is carried out for the service of patent bibliometrics. The research first defines the intelligence question-answering system on

基金项目: 中国科学院文献情报新增能力建设专项“情报研究数据中心及服务体系建设”(院 1680)和中国科学院青年创新促进会(2016157)。

作者简介: 吕璐成(1989-), 博士研究生, 助理研究员, 研究方向: 专利文本挖掘、智能情报方法与技术, E-mail: lucheng918@126.com; 韩涛(1980-), 博士, 研究员, 研究方向: 智能情报方法与技术、科学计量与复杂性科学; 王燕鹏(1992-), 硕士, 助理研究员, 研究方向: 文献计量与文本挖掘。

patent bibliometrics. And then based on natural language understanding and natural language generation technology, a technical proposal of the intelligent Q&A system on patent bibliometrics with question understanding module and answer generation module as the core is proposed. Finally, we design and implementation the intelligent Q&A system on patent bibliometrics in the field of artificial intelligence. The system can answer various expression forms of the same question and give the correct answers. It is of positive significance to improve the service quality and user experience of scientific and technological intelligence.

Keywords: Question-answering system; intelligent intelligence service; task-oriented dialogue system; patent bibliometrics

引言

随着移动互联网的普及,信息过载问题愈发严重。由于问答系统能够理解用户通过自然语言表达的问题,并帮助用户完成特定任务或提供相对准确的回答,因此一直受到社会各界的关注和重视。近年来由于机器学习技术的不断应用和发展,问答系统研究热度持续升温,不断有产品成果产出并落地应用,典型代表如苹果 Siri、微软小冰、百度度秘、贤二机器僧等问答系统。目前,这些问答系统已经初步实现了如查询天气、预定航班、购买产品等任务型功能以及闲聊谈心等社交型功能^[1]。

在这样的时代背景下,科技情报工作也步入了以智能服务为标志的情报 3.0 时代^[2],使用机器人处理情报人员繁重的机械劳动、加快用户需求响应的观念受到了研究人员的广泛认可^[3-6]。但是在情报 3.0 时代,科技情报服务工作同样面临信息过载的问题^[7],如何及时地、有针对性地为用户提供有序、增值、准确的信息仍是一个待解决的问题。由于问答系统能够基于人工智能算法以及加工的专业知识为用户提供具有及时性、针对性、智能性和准确性的结果,因此,问答系统为智能化科技情报服务模式的探索提供了机遇和突破口。专利计量分析是科技情报工作的重要组成部分,包括专利申请量、

授权量、专利有效性、专利权人、技术发明人和 IPC 分类号等专利相关字段的统计分析,能够反映技术布局趋势、挖掘技术优势机构以及分析技术发展现状^[8,9]。对于创新主体、科研机构和企业的人员而言,能够实时、准确地获取技术研发的最新动态、竞争对手的专利布局动态,对其把握技术发展趋势、调整研发方向具有重要意义。

基于上述分析,结合情报 3.0 的工作思路,本研究提出开展专利计量情报问答系统的研究与实践。区别于一般自然语言问答系统,专利情报问答具有基于统计分析呈现答句的特点,即问答反馈的答句需要基于计量规则对专利文献数据进行统计分析得到,这个过程需要专利情报人员介入提供计量规则。因此,本研究界定专利计量情报问答系统的概念为:针对专利计量分析这一科技情报服务场景,基于专利文献数据,利用专利情报人员凝练的计量统计规则,通过智能技术和数据挖掘算法,实时地、精准地、智能地为科技情报用户提供专利计量分析服务的问答系统。

1 国内外研究现状

目前,国内外已有科技情报问答系统的相关研究和实践,按照服务对象和服务功能的不

同,大致可以将其分为三类。

2.1 面向图书馆信息服务的科技情报问答系统

问答系统技术在图书馆文献和信息服务中已开展了相关的实践,主要集中在虚拟馆员、参考咨询、信息检索、个性化推荐服务等方面。

在国外,德国汉堡大学等一些欧洲图书馆最先开始研究问题系统在图书馆的应用,之后其他图书馆也开展了进一步的实践。Allison D^[10]介绍了内布拉斯加大学林肯分校的一个聊天机器人实验项目,该聊天机器人通过挖掘图书馆网站和聊天日志记录创建数据库,并且通过分析日志不断改善机器人的聊天;Zhang S, Gu J 和 Yang Y^[11]提出基于智能聊天机器人构建个性化推荐系统,通过分析用户注册信息和使用历史建立可动态调整的初始用户模型,聊天结束时可以向用户提供个性化推荐服务。

在国内,清华大学图书馆^[12]开发了智能聊天机器人“小图”,在对图书馆资源、可信第三方资源、优质网络资源加以组织、整理、整合和利用基础上,利用成熟的开源软件进行开发而成,包含了“逗闷子”、问题咨询、图书查询、教学训练等功能,旨在探索在线参考服务新模式;南京大学图书馆^[13]开发了微信智能问答系统,该系统可对图书馆电话、地址、开放时间等常识进行关键词回复,同时系统利用图书馆已有的FAQ常见问题系统和中文分词技术来实现对于读者的智能回复功能,经测试,该系统能够较好的解决读者的常见咨询;上海闵行区图书馆^[14]推出了“小敏”实时虚拟咨询微信机器人,可以协助图书馆工作人员在实时咨询中完成90%以上的问答,较大程度减轻了

馆员的工作量,工作效率得到了极大的提升,同时读者可以24小时随时随地咨询。此外,也有较多围绕图书馆智能化咨询问答系统的实践,如刘宝瑞等^[15]提出的基于Deep QA的图书馆数字参考咨询问答系统,来云^[16]提出的图书馆智能化咨询问答机器人系统,沈奎林等^[13]基于微信构建的图书馆智能问答系统,胡明明^[17]等基于分词技术提出虚拟参考咨询系统的智能优化解决方案。

2.2 面向专业领域知识服务的科技情报问答系统

随着科学技术的跨越式发展,各垂直科技领域的专业信息和知识呈现井喷式增长。问答系统的发展将帮助科研和产业人员、社会公众以自然语言这种低能力壁垒的检索方式,快速获取精炼的高价值专业知识和情报,目前在多领域专业知识服务中已有相关的研究和实践,主要包括生命健康、航空航天、物理学、农业科学、金融业等领域或行业。

在国外,Abdi A^[18]提出了一种基于物理领域本体的问答系统QAPD,该系统集成了自然语言处理、本体和信息检索技术,可辅助用户以自然语言的方式获取物理领域的专业知识;Lee HG^[19]利用生物医学文档多排序模型开发出了一套智能问答系统,可以基于用户的自然语言查询,从大规模医学文档中返回相关的片段;Devi M^[20]提出了一种基于农业领域本体的问答系统ADANS,该系统采用了自然语言处理技术和语义网技术,用于回答用户的自然语言问题。此外,John AK^[21]提出了基于深度学习算法的法律领域对话系统;Okuda T^[22]开发了金融服

务解决方案 Finplex 企业级聊天机器人, 该聊天机器人得到了索尼公司的引进。

在国内, 马晨浩^[23]提出了基于甲状腺知识图谱的自动问答系统, 可有效地回答病人在甲状腺类疾病方面的用药以及检查等方面的问题, 节约病人问诊时间的同时, 医生可以使用该系统对患者以及处方等信息进行相关查询, 更加快速便捷, 节约了医生的时间; 梁敬东^[24]开发了水稻 FAQ 问答系统, 该系统构建一个基于 word2vec 和 LSTM 神经网络, 包括输入层、嵌入层、LSTM 层、全连接层和输出层的句子相似度模型, 能够准确地匹配用户问题和水稻 FAQ 中的问题, 更好的帮助农户解决水稻生产中遇到的问题。此外, 杨芳权^[25]开发了基于包装产业大数据知识图谱的智能问答系统; 陈程^[26]提出了基于中医药知识图谱的智能问答系统。

2.3 面向综合领域知识服务的科技情报问答系统

面对日益复杂和快速变化的用户需求, 以及科技、社会重大问题的多学科交叉性, 单一领域的问答系统所能发挥的科技情报服务效果是有限的, 因此必须加强跨界数据和资源的整合、服务融合和优势互补, 构建能够应对复杂、综合问题的问答系统。面向综合领域知识服务的问答系统已有相关的研究和实践。

在国外, 最具有代表性的是 IBM 公司^[27]开发的 Waston 系统, 该系统采用了 IBM Deep QA 技术和 Apache UIMA 框架, 其数据来源包括了非结构化文本 (如百科全书和新闻文档等) 和 DBpedia、Wordnet 等结构化知识库。Waston 系统曾在智力比赛中战胜人类选手获

得冠军, 同时, 该系统目前已被生命健康行业引进, 实现商业化应用。此外, 面向综合领域知识服务的科技情报问答系统还包括了 MIT 人工智能实验室^[28]开发的 Start 问答系统、沃尔夫勒姆研究公司^[29]开发的 Wolfram Alpha 问答系统。

相较国外, 国内在面向综合领域知识服务的科技情报问答系统方面仍处于研究和探索阶段, 成果大多体现在学术论文方面, 而在辅助科技决策和实际商业应用方面仍有很长的一段路要走。

综上所述, 在人工智能、机器人等技术快速发展的时代背景下, 国内外各机构已经开展了一系列科技情报智能服务的探索, 但是针对专利计量分析的科技情报问答服务的研究和实践还处于空白。因此, 面对科技情报机构对于智能科技情报服务的新方式和新产品的迫切需求, 本研究开展专利计量情报问答系统的研究和实践。

3 技术方案

从功能角度看, 对话系统可分为两类^[30], 一类是任务导向型 (task-oriented) 对话系统, 能够实现如查询天气、预定航班、购买产品等任务; 另一类是非任务导向型 (non-task-oriented) 对话系统 (也称聊天机器人), 能实现闲聊谈心、社交等对话需求。由于专利计量分析服务面向具体的用户需求场景, 能够解决特定的用户问题, 属于典型的“任务导向”, 因此专利计量情报问答系统属于任务型对话系统的范畴。本研究基于任务型对话系统设计的流程^[31], 提出

专利计量情报问答系统的技术方案。

专利计量情报问答系统的整体实现流程如图1所示。首先问答系统接收用户输入的问候，

第二步，系统对问句进行理解，第三步，根据理解的信息生成答句，最后，将答句返回给用户，完成问题的解答。

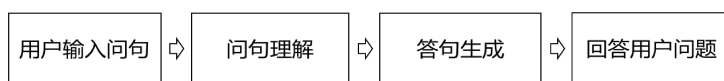


图1 问答流程

在上述流程中，核心的任务流程是问句理解和答句生成，分别对应系统架构中的问句理解模块和答句生成模块，如图2所示。其中问句理解模块具体包括问句分类器、意图检测器和语义槽填充器；答句生成模块具体包括答案搜索器和模板填充器。

问句理解模块分为问题领域识别、用户意图检测、语义槽填充等三个子任务^[32]。本研究问句理解过程的问题领域、用户意图和语义槽示例如图3所示。

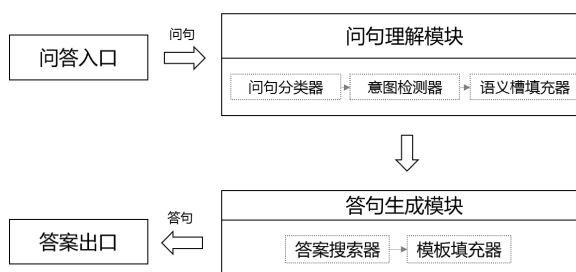


图2 专利计量情报问答系统的技术架构图

3.1 问句理解模块

问句理解模块是问答系统对问句进行语义理解的部分。在本专利计量情报问答系统中，

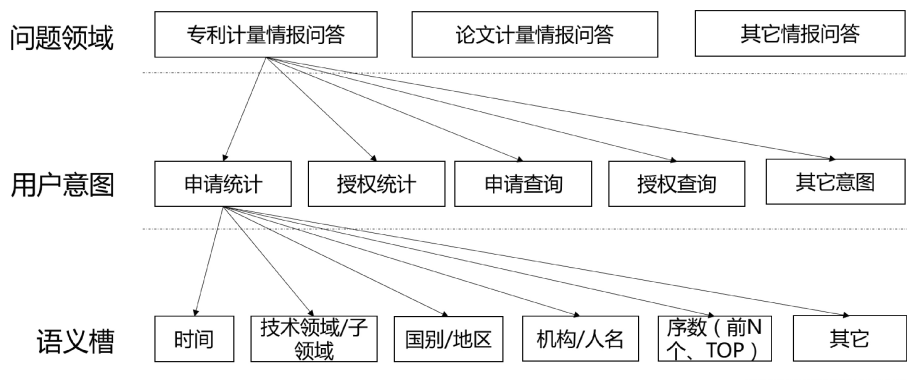


图3 专利计量情报问答系统问句理解过程

问题领域识别是指识别科技情报用户的问题所属的领域，如专利计量情报问答服务、论文计量情报问答服务或者其他情报问答服务。

用户意图检测是在明确用户的问题领域后，再进一步明确用户所问问题的具体意图，如明确用户问的是专利计量情报问题，则进一步判断用户的问题是属于专利申请统计、授权统计、

申请查询、授权查询还是其他意图。

语义槽填充的主要目标是在已知特定领域或意图的语义框架的前提下，从输入语句中抽取出该语义框架中预先定义好的语义槽的值^[31]，例如时间、技术领域/子领域、国别/地区、机构/人名、序数（前N个、TOP）等。

问句理解过程的具体流程如图 4 所示。首先问答系统调用问句分类器对于问句的领域进行识别，然后调用意图检测器对用户的意图进行检测，最后调用语义槽填充器进行语义槽的填充。以问句“2017 年中国在语音合成方面的

专利申请量有多少？”为例，问答系统将首先理解该问句的领域为专利计量情报问答，然后检测用户意图为申请统计，最后识别用户问句中的“时间”、“国别/地区”、“技术领域/子领域”三个语义槽。



图 4 问句理解实现流程及案例

3.1.1 问句分类器

问句分类器，用于实现问题领域识别。由于科技情报服务问句特征较为明显，因此问句分类器采用识别关键词特征的方法来实现。如问句出现“专利”、“patent”等关键词则识别为专利计量情报问答领域。

3.1.2 意图检测器

意图检测器用于检测用户问题的意图。本研究对专利计量问答服务的可能场景进行了梳理，共定义了 44 类意图，表 1 给出了部分意图的示例。对于每一个意图，分别定义了语义框架，即语义槽。

由于意图数量可控，因此意图检测器采用的是基于正则表达式匹配的意图检测方法。如果通过正则表达式匹配到多个意图，那么将继续向后判断，判断问句中的语义要素是否符合该意图的语义框架，直至确定最后意图；如果

最终未匹配到符合用户问句的语义框架，则返回用户“暂无相关答案”等反馈信息。

表 1 专利计量问答用户意图示例

意图序号	意图名称
1	专利申请统计
2	专利授权统计
3	专利申请量增长最快的年份计算
4	专利申请量top N来源国
5	最新申请的专利名称
6	三方专利统计

3.1.3 语义槽填充器

语义槽填充器是在明确用户问题领域和意图之后，对问句中的语义实体进行识别，并将其转换为问答系统能够识别的规范形式。本研究中，专利计量问答意图涉及的语义槽具体参见表 2。

表2 专利计量情报问答语义框架

语义槽	序列标注标签	实例
时间	TIM	➢2017年, 2016年… ➢今年, 去年… ➢近三年, 近五年… ➢90年代, 80年代…
技术领域/子领域	FIE	➢人工智能领域 ➢语音合成方面 ➢图像处理方面 ➢纳米领域
国别/地区	CON	➢中国, 美国, 日本 ➢北京, 山东 ➢海淀区, 朝阳区
机构	ORG	➢中国科学院, 清华大学
人名	PER	➢周志华, 唐杰
序数	ORD	➢前N个, TOP3 ➢最新、最早、最多、最少

语义槽填充是决定任务型对话系统问答效果的关键步骤^[31]。本研究中, 语义槽填充任务采用基于条件随机场 (CRF) 的序列标注模型实现^[33], 标注方案采用 IOB² 序列标注方法^[34]。IOB² 标签方案是指对于每一个文本块, 其开头字符标记位 B, 而中间和结尾字符都标记为 I, 除此之外的其他字符标记为 O。

以问句“2017年, 南京大学的周志华在机器学习领域的专利申请量是多少?”为例, 涉及的语义槽包括时间、技术领域/子领域、机构、人名四种。语义槽填充结果具体参见表3。

表3 序列标注示例

字符	2	0	1	7	年	,	南	京	大	学	的	周	志	华	在	机
标注	B-TIM	I-TIM	I-TIM	I-TIM	I-TEM	O	B-ORG	I-ORG	I-ORG	I-ORG	O	B-PER	I-PER	I-PER	O	B-FIE
	器	学	习	领	域	的	专	利	申	请	量	是	多	少	?	
	I-FIE	I-FIE	I-FIE	O	O	O	O	O	O	O	O	O	O	O	O	O

通过语义槽填充器识别的语义信息在多数情况下需要经过转换处理才能转换为问答系统识别的数据形式。具体地, 对于“时间”语义信息, 如“今年”、“去年”这类代词需根据查询产生时的系统时间来转化为相应的公元年份; 对于“国别/地区”语义信息, 需要通过国名别名映射规则库来进行转换; 对于“技术领域/子领域”语义信息, 通过构建的“领域-子领域对应规则库”进行识别转换; 对于“机构”和“人名”语义信息, 如机构名采用简写 (例如“哈尔滨工业大学”被简写成“哈工大”), 外国人名有可能使用不同音译, 或作不同的简写缩写 (例如“Albert Einstein”被写成其音译“爱因斯坦”), 需要通过机构和人名清洗规则库来清洗转换;

对于“序数”语义信息, 其形式相对固定, 一般形如“前3个”、“最多的五个”等, 因此采用正则表达式来进行识别, 并且转化为阿拉伯数字标准形式。

3.2 答句生成模块

答句生成模块是在问句理解模块对用户问句进行语义理解之后, 搜索答案要素自动生成答句的模块。

答句生成模块的具体流程如图5所示。根据问句理解模块理解到的问题领域、问题意图、语义槽标签信息, 调用答案搜索器生成知识库查询语言, 查询得到答案数据, 然后调用模板填充器从答句模板库中获取答句模板组装生成问题答句。

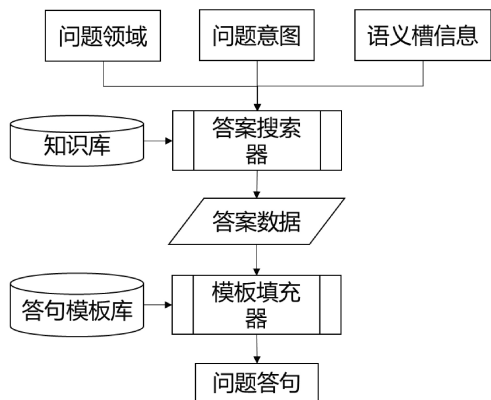


图 5 答句生成模块具体流程

3.2.1 答案搜索器

答案搜索器是搜寻问句答案数据的模块，通过问句理解模块识别到的问题领域、意图以及语义槽信息，生成数据库查询语句，从知识库中搜索问句的答案数据。

为了实现高效率的答案检索，本系统预先将系统内各个问题的可能答案全都预先生成了一遍，将其存入一个“关键词 - 答案对知识库”里，表 4 给出了数据库记录部分示例。

表 4 关键词 - 答案对知识库示例

问题类型	优先权年	授权年	申请年	来源国	流向国	领域	子领域	申请人	发明人	topN数目	答案
2			2016	中国		语音处理					11799
16			2009			语音处理		韩西杰			1

例如对于问句“2016 年，中国在语音处理领域的专利授权量是多少？”，答案搜索器会通过问句分类器和意图检测器识别其问句意图为“专利授权统计”，然后根据语义槽信息 { 申请年: 2016 年, 来源国: 中国, 领域: 语音处理 } 构造 SQL 查询语句从关键词 - 答案对知识库中查询得到结果为“11799 件”。

3.2.2 模板填充器

模板填充器是组合答案数据和答句模板形成问题答句的模块。答句模板是根据系统内定义的各种问句类型而手工录入的，本研究针对专利计量问答服务种的 44 种意图分别组织了答句模板，用于答句生成。答句模板中用符号“【”和“】”来标注出需要用答案数据进行填空的

区域。每类问句可能会有多个不同的答句模板，在最终生成时将随机挑选其中的一个使用。

表 5 给出了模板填充器的工作示例，表中“模板”行展示了答句类型“2”对应的一条答句模板。例如对于问句“2016 年，中国在语音处理领域的专利授权量是多少？”，通过问句分类器和意图检测器提取的问句类型“2”定位其中一条答句模板“【申请年】年，【来源国】在【领域】领域的专利授权量是【专利授权量】件”，然后结合意图检测器识别出的“申请年”、“来源国”、“领域”信息，再组合答案搜索器搜索出的答案“11799”，就组装得到了最终的答句，即返回结果为“2016 年，中国在语音处理领域的专利授权量是 11799 件。”

表 5 模板填充器工作示例

答句类型	2
模板	【申请年】年，【来源国】在【领域】领域的专利授权量是【专利授权量】件
答案数据	{ “问题类型” => “2”，“申请年” => “2016”，“来源国” => “中国”，“领域” => “语音处理”，“专利授权量” => “11799” }
输出答句	2016 年，中国在语音处理领域的专利授权量是 11799 件

4 研究实践结果

本研究基于以上技术方案,针对中国人工智能专利测试数据集,进行专利计量情报问答系统的实践。

在语义槽填充器环节,研究选取 150000 个句子作为序列标注模型训练过程的数据集,采用 Python-crfsuite 进行模型训练,序列特征选取上下文特征,训练算法为 L-BFGS,采用 Elastic Net (L1+L2) 进行正则化, L1 惩罚系数 1.0, L2 惩罚系数 0.001。训练迭代次数为 100 轮。表 6 展示了在测试集上的语义槽要素识别结果。从中可以看出,模型具有较高的准确率和召回率,能够用于问答系统的语义槽识别。

表 6 序列标注模型识别语义槽效果

语义槽类型	语义槽识别结果		
	准确率%	召回率%	F1%
TIM-时间	99.28	99.19	99.23
ORG-机构	85.51	84.08	84.79
CON-国别/地区	88.08	93.45	85.70
FIE-技术领域/子领域	99.81	100.00	99.91
PER-人名	93.49	90.39	91.91
ORD-序数	96.60	94.54	95.56

最终,研究实现了专利计量情报问答系统的 DEMO (包括微信公众号版本和网页在线版本)。如图 6 所示。

在这个示例中,问答系统对于用户关于“中国在 2016 年语音处理领域的专利授权量”的问题的多种问法都能有效理解并进行回答,能够保证服务的效果和质量。基于该成果,研发人员可以在零碎时间快速了解本领域专利技术布

局动态,把握技术发展趋势。

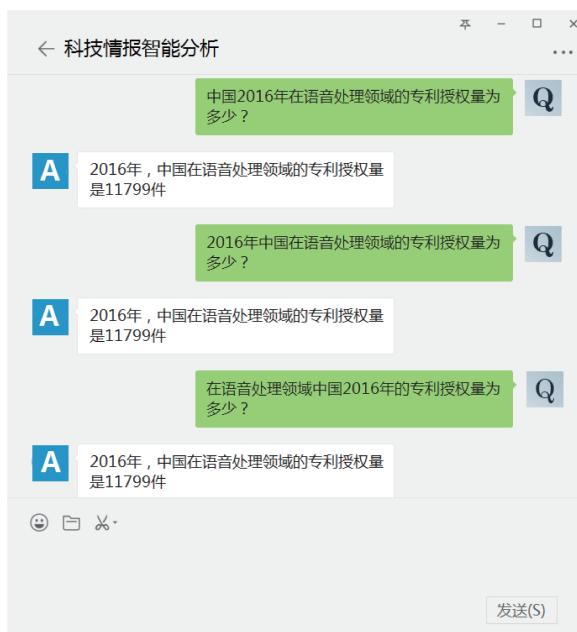


图 6 专利计量情报问答系统 DEMO

5 总结与展望

科技情报服务的重要任务是服务创新,虽然机器不能完全替代人类的创造力,但是人工智能技术的快速发展为机器激发人类创造力提供了可能^[35]。智能问答系统作为一种新型的人机交互形式,利用人工智能技术能够有效提升科技情报机构的情报协同服务能力^[36],为服务创新提供了新的方法和思路。

本研究面向科技情报机构对于科技情报服务的新方式和新产品的迫切需求,研究并实现了专利计量情报问答系统,该系统能够理解用户问题,针对同一问题的不同问法给出正确答案。

该系统不但能够加快情报机构、情报人员的服务效率,还能够在服务过程中不断“学习”进而改善服务质量,对于提升科技情报服务质

量和用户体验具有重要作用和意义。同时实现该系统的技术方案可以在其他文献数据（如论文数据和基金项目数据等）进行运用，进而扩展出其他文献类型的科技情报问答系统。

但是，本研究仍有一些问题有待在进一步工作中优化，包括：（1）专利计量分析的语义框架有待进一步完善，以保证系统的鲁棒性和健壮性；（2）目前的科技情报问答系统的答句生成还比较依赖于情报人员梳理的模板，随着系统上线使用、相关数据的不断积累，可以继续开展基于深度学习的答句生成方法研究与实践。

参考文献

- [1] 冯升. 聊天机器人问答系统现状与发展[J]. 机器人与应用, 2016(4):34-36.
- [2] 吴晨生, 李辉, 付宏, 等. 情报服务迈向3.0时代[J]. 情报理论与实践, 2015, 38(9):1-7.
- [3] Yeh P F. 情报分析中使用机器人(上)[J]. 情报理论与实践, 2017, 40(9):145.
- [4] Yeh P F. The case for using robots in intelligence analysis[J]. Studies in Intelligence, 2015, 59(4):1-8.
- [5] 傅平, 邹小筑, 吴丹, 等. 回顾与展望: 人工智能在图书馆的应用[J]. 图书情报知识, 2018(2):50-60.
- [6] 刘如, 张惠娜, 杜丽萍, 等. 基于情报3.0工作思路的自动简报系统设计与实现[J]. 情报学报, 2018, 37(2):172-182.
- [7] 时艳琴, 陈雪飞, 谢威, 等. 情报3.0时代情报的特征、任务与工具[J]. 情报杂志, 2017, 36(10):1-6.
- [8] 文庭孝, 杨忠, 刘璇. 基于专利计量分析的湖南省专利战略研究[J]. 情报理论与实践, 2012, 35(1):58-64.
- [9] 栾春娟, 侯海燕. 全球纳米技术领域专利计量分析[J]. 科技与经济, 2008, 21(4):38-40.
- [10] Allison D A. Chatbots in the library: is it time?[J]. Library Hi Tech, 2012, 30(1): 95-107.
- [11] Zhang S, Gu J, Yang Y. Research on Personalized Recommendation System based on Intelligent Chat Robot[EB/OL]. [2018-07-03]. <http://www.ica.stc.sh.cn/picture/article/176/b1/e0/dbc385694bf3b9c538a38c57347c/ac9d5a95-67d9-41e0-942e-a76e2aaf679f.pdf>.
- [12] 姚飞, 张成昱, 陈武. 清华智能聊天机器人“小图”的移动应用[J]. 数据分析与知识发现, 2014, 30(z1):120-126.
- [13] 沈奎林, 邵波, 赵华. 利用微信构建图书馆智能问答系统[J]. 图书馆学研究, 2015(8):75-80.
- [14] 闵腾超, MinTengchao. 微信机器人在图书馆实时虚拟咨询服务中的应用研究——以上海闵行区图书馆为例[J]. 新世纪图书馆, 2015(8):54-57.
- [15] 刘宝瑞, 郭宏娇. 基于Deep QA的图书馆数字参考咨询问答系统研究[J]. 情报科学, 2017(4):103-108.
- [16] 来云. 图书馆智能化咨询问答机器人系统设计与语料技术研究[J]. 现代情报, 2017, 37(11):121-124.
- [17] 胡朝明. 基于分词技术提升虚拟参考咨询系统智能服务[J]. 图书情报工作, 2012, 56(9):110-113.
- [18] Abdi A, Idris N, Ahmad Z. QAPD: an ontology-based question answering system in the physics domain[J]. Soft Computing, 2018, 22(1):213-230.
- [19] Lee H G, Kim M, Kim H. Implementation of a High-Performance Answer Snippet Retrieval System Based on Multiple Ranking Models for Biomedical Documents[J]. Advanced Science Letters, 2017, 23(10):9486-9490.
- [20] Devi M, Dua M. ADANS: An agriculture domain question answering system using ontologies[C]. International Conference on Computing, Communication and Automation. 2017:122-127.
- [21] John A K, Caro L D, Robaldo L, et al. Legalbot: A Deep Learning-Based Conversational Agent in the Legal Domain[C]. International Conference on Applications of Natural Language to Information Systems. 2017:267-273.
- [22] Okuda T, Shoda S. AI-based chatbot service for

- financial industry[J]. Fujitsu Scientific & Technical Journal, 2018, 54(2):4-8.
- [23] 马晨浩. 基于甲状腺知识图谱的自动问答系统的设计与实现[J]. 智能计算机与应用, 2018(3):102-107.
- [24] 梁敬东, 崔丙剑, 姜海燕, 沈毅, 谢元澄. 基于 word2vec 和 LSTM 的句子相似度计算及其在水稻 FAQ 问答系统中的应用[J/OL]. 南京农业大学学报, 2018:1-9 [2018-07-03]. <http://kns.cnki.net/kcms/detail/32.1148.S.20180518.1001.004.html>.
- [25] 杨芳. 基于包装产业大数据知识图谱的智能问答系统设计[J]. 现代电子技术, 2018, 41(4):143-146.
- [26] 陈程, 翟洁, 秦锦玉, 等. 基于中医药知识图谱的智能问答技术研究[J]. 中国新通信, 2018, 20(2):204-207.
- [27] IBM. IBM Watson[EB/OL]. [2018-07-03]. <https://www.ibm.com/watson/index.html>.
- [28] MIT. The START Natural Language Question Answering System [EB/OL]. [2018-07-03]. <http://start.csail.mit.edu/index.php>.
- [29] Wolfram Research. wolfram alpha - Wolfram|Alpha: Computational Intelligence[EB/OL]. [2018-07-03]. <http://www.wolframalpha.com/>.
- [30] Chen H, Liu X, Yin D, et al. A Survey on Dialogue Systems: Recent Advances and New Frontiers[C]. Acm Sigkdd Explorations Newsletter, 2017, 19(2).
- [31] 车万翔, 张宇. 任务型与问答型对话系统中的语言理解技术[J]. 中国计算机学会通讯, 2017, 13(9):10-14.
- [32] Tur G, De Mori R. Spoken language understanding: Systems for extracting semantic information from speech[M]. John Wiley & Sons, 2011.
- [33] 张祝玉, 任飞亮, 朱靖波. 基于条件随机场的中文命名实体识别特征比较研究[C]. 全国信息检索与内容安全学术会议. 2008.
- [34] Ramshaw L A. Text chunking using transformation-based learning[C]. Proc. Workshop on Very Large Corpora. 1995.
- [35] Christopher Steiner. Can Creativity be Automated? MIT Technology Review[EB/OL]. [2012-07-27]. <http://www.technologyreview.com/news/428437/can-creativity-be-automated>.
- [36] 李辉, 张惠娜, 侯元元, 等. 情报 3.0 时代科技情报服务能力研究——基于工程技术视角的服务能力四层结构模型[J]. 情报理论与实践, 2017, 40(3):1-4.