

基于知识图谱的知识导航服务研究

朱 鹏

(云南师范大学 信息学院, 云南 昆明 650500)

摘 要: 随着知识生产速度和规模的增加, 传统的基于“字符串”的检索技术在正确性和扩展性方面都存在着严重的不足, 且容易造成知识迷航。文章讨论了基于知识图谱的知识导航系统模型, 把人的思维模式应用到知识导航中, 降低使用者的知识检索成本, 同时提升知识导航的服务质量。

关键词: 知识图谱; 知识导航; 知识服务

在这信息爆炸却大多无用的世界里, 清晰的见解就成了一种力量。知识导航的方式因知识管理技术的发展而改进, 传统的信息检索技术主要有关键字匹配、目录分类和概念搜索, 都是以“字符串”匹配的方式去检索信息, 以提供最接近的结果列表^[1]。Google于2012年5月推出了知识图谱(Knowledge Graph, KG)技术, 增强其搜索引擎的搜索结果, 标志着大规模知识成功应用于互联网信息的语义搜索。知识图谱中表示互联网信息的形式更接近人类的认知形式, 增强了人们对海量的互联网信息的组织、管理和理解能力。知识图谱给互联网信息的语义搜索带来了活力, 同时也为提升知识导航服务的质量和效率提供新的可能。

1 知识导航服务的内涵

知识导航起源于知识管理, 可根据知识与检索主题间的语义相关程度为用户呈现结构清晰的知识体系。垂直的知识分类, 实现基于学科本体, 面向主题的相关关系、等级关系和等同关系术语的展现^[2]。为适应社会发展以及人们的需要, 知识导航服务的发展在诸多方面都呈现出以下新的特征: 服务范围全球化、服务对象具体化、服务方式智慧化、服务内容深层化。

2 基于知识图谱的知识导航服务系统

知识图谱揭示了实体之间的关系, 实现了对现实世界的事物及其相互关系进行形式化的描述^[3]。知识图谱的一般表示为 $G=(E, R, S)$ 。其中 $E=\{e_1, e_2, \dots, e_{|E|}\}$ 表示了客观世界中的具体事物, 是知识库中的实体集合, 共有 $|E|$ 个不同实体; $R=\{r_1, r_2, \dots, r_{|R|}\}$ 表示知识库中的关系集合, 描述了概念、实体、事件之间客观存在的联系, 共有 $|R|$ 种不同关系; $S \subseteq E \times R \times E$ 表示知识库中的三元组集合。三元组的基本形式主要包括实体1(E_1)、关系(R)、实体2(E_2)和概念(C)、属性(P)、属性值(V)等, 知识图谱中的基本元素是实体, 不同的实体间有不同的关系。概念是指人们在认识世界过程中形成的对客观事物的概念化表示^[4], 如交通工具、公司企业、政府机关等。

基于知识图谱的知识导航系统的技术架构如图1所示, 自低向上可以分为数据层、知识图谱构建层和知识导航应用层3个部分。与传统技术相比, 应用知识图谱技术构建的知识库质量更高, 知识之间的语义关系更丰富。基于知识图谱的知识导航可以对用户的检索内容进行语义推理, 精准分析用户知识需求, 快速地帮助用户找到所需的知识。

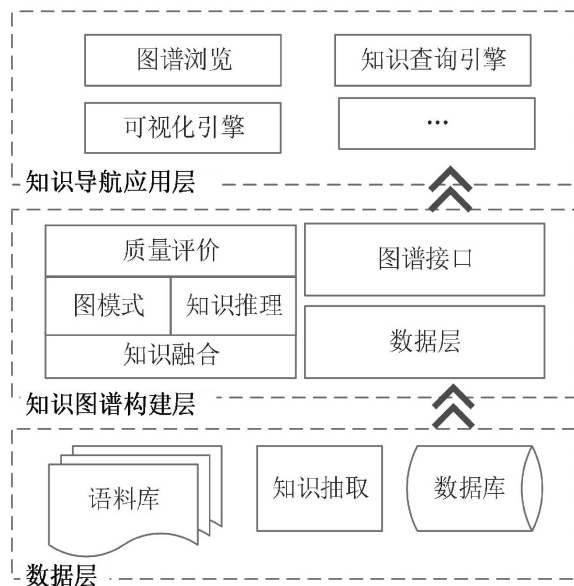


图1 基于KG的知识导航服务技术框架

2.1 数据层

数据层主要任务有语料库的存储、知识抽取和知识存储。语料库中有选择地从互联网上采集了大量信息和数据, 并存储到本地; 知识抽取则是采用自然语言处理技术, 从语料库中有选择地抽取知识三元组并存储到数据库中; 数据库采用非关系型数据库(NoSQL)存储数据, 向上层提供统一接口并完成一系列的数据操作任务。

2.2 知识图谱构建层

构建层负责知识存储库的构建和管理。其中图模式一般

表示形式为本体^[5],在图模式的约束下进行知识融合、知识推理,融合和推理后产生的新知识通过质量评价后存储到知识库中。知识图谱构建层中向上提供的抽象接口封装了具体数据操作的内容,知识导航程序通过接口进行图的数据访问操作,而保持底层数据存储透明。

2.3 知识导航应用层

知识导航应用层负责导航服务的应用实现。所以知识查询引擎通过图接口中的抽象概念来构建计算和查询过程,图算法、查询语言端点、知识点搜索也在该层实现。查询结果集返回给可视化引擎进行图谱渲染,然后在图谱浏览界面显示给用户。知识导航服务平台如图2所示。

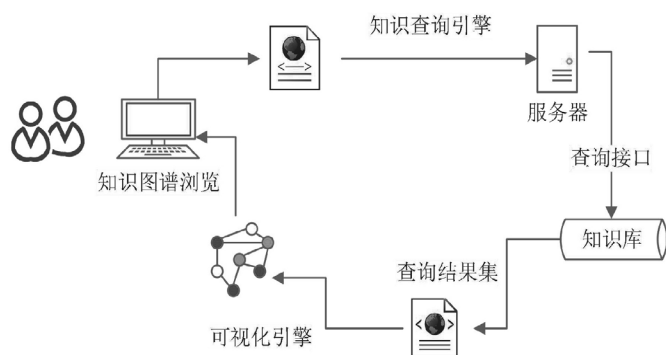


图2 知识导航服务平台

对知识图谱来说,我们的搜索对象是一个实在的事物(Things),而不是单纯的字符串(String)^[6]。知识查询引擎可以提供特定查询模式对知识进行高效的查询。例如,斯坦福大学的Protégé^[7],Apache开源项目Jena^[8]以及Google的

Cayley^[9]。查询引擎采用语义分析技术分析用户输入,理解用户意图和需求,查询并向用户提供正确的信息。图谱可视化渲染引擎负责创建可视化的图谱结构的知识导航界面。当用户点击某个知识点后,与其相关的知识将会被列出,这时用户可以从这些知识中找出自己所需求的知识,点击相关链接获得更加详细的资源。应用可视化的图结构展示,用户可高效地获得核心与边缘知识、上下位知识、综述和摘要知识等。

3 结语

知识图谱技术的潜力很大,其研究了如何从互联网中获取知识,以知识驱动的语言理解以及可视化的图结构展示知识关系等。把知识图谱技术应用到知识导航系统中,弥补了传统技术方案的不足。知识导航系统通过知识图谱来对知识内容进行组织,提供知识结构的可视化导航。应用知识图谱进行知识组织与管理,实现语义的表达和识别,并通过逻辑推理机制,根据用户的思维方式实现知识导航,把人的思维模式应用到知识导航中,更好地体现人机交互,同时借助可视化的导航界面帮助用户更全面地理解知识内容。在未来的工作中,还可以建立用户的知识状态模型,构建用户特征分析模块,以便从个性化和专业性等角度来构建知识图谱,为向用户提供个性化导航服务提供依据。采用知识推送策略,根据用户的信息需求和兴趣爱好,将知识推送给使用者,这样可以有效提高知识的利用率,降低使用者的检索成本。在进行知识推送时,要十分慎重,既不遗漏重要的知识,又不能使使用者被无用的或者低效的知识所干扰。

[参考文献]

- [1] HARARI Y N. 21 Lessons for the 21st Century[M]. New York: Spiegel & Grau, 2018.
- [2] 盛东方,孙建军.基于语义搜索引擎的学科知识服务研究——以GoPubMed为例[J].图书情报知识, 2015(4): 113-120.
- [3] 徐增林,盛泳潘,贺丽荣,等.知识图谱技术综述[J].电子科技大学学报, 2016(4): 589-606.
- [4] 李涓子,侯磊.知识图谱研究综述[J].山西大学学报(自然科学版), 2017(3): 454-459.
- [5] 杜小勇,李曼,王珊.本体学习研究综述[J].软件学报, 2006(9): 1837-1847.
- [6] SINGHA A. Introducing the Knowledge Graph: things, not strings[EB/OL]. (2018-05-16) [2018-09-20]. <https://www.blog.google/products/search/introducing-knowledge-graph-things-not/>.
- [7] RESEARCH SCFB. Protégé[EB/OL]. (2017-05-23) [2018-09-20]. <https://protege.stanford.edu>.
- [8] FOUNDATION AS. Apache Jena[EB/OL]. (2018-06-07) [2018-09-20]. <https://jena.apache.org>.
- [9] CAYLEY GRAPH. Cayley[EB/OL]. (2018-01-06) [2018-09-20]. <https://github.com/cayleygraph/cayley>.

Research on knowledge navigation service based on knowledge graph

Zhu Peng

(School of Information, Yunnan Normal University, Kunming 650500, China)

Abstract: With the increase of the speed and scale of knowledge production, the traditional “string”-based retrieval technology has serious shortcomings in terms of correctness and scalability, and it’s easy to get lost in knowledge. This paper discusses the knowledge navigation system based on knowledge graph to reduces the cost of knowledge retrieval and improves the service quality.

Key words: knowledge graph; knowledge navigation; knowledge service