

基于知识图谱实现的南理工校园智能地图



王佳渊, 周倩如*

南京理工大学计算机技术与工程学院, 江苏南京 210094

摘要: 为了实现带有语义信息的智能地图查询服务, 实现多维度、多来源、多平台综合宏观地理信息查询, 本文利用知识图谱技术, 将不同来源、以不同格式存储于不同数据库的数据整合入南理工校园地图知识库, 构造了集地理信息、声音、气味、花朵、美食等服务为一体的智能地图信息服务。本文还在 IEEE 上层本体标准 SUMO 和地理信息本体 wgs84 的基础上构造了智能地图本体, 并用本体驱动开发了智能地图服务, 最后, 提供了开放的地图 SPARQL 前端查询接口, 和查询结果的地图显示界面。验证阶段, 本文利用真实南理工校园地图和声音、气味、花朵、美食数据, 构建了真实的智能地图知识库, 经真实知识库的 SPARQL 查询检测, 准确率达到 100%, 并且 SPARQL 查询平均每条查询耗时 5 毫秒, 可满足生产生活需要。

关键词: 知识图谱; 本体; 地理信息服务; SPARQL

DOI: [10.57237/j.cst.2023.03.004](https://doi.org/10.57237/j.cst.2023.03.004)

Ontology Driven Semantic Campus Map Application for NJUST

Jiayuan Wang, Qianru Zhou*

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Abstract: Intelligence GIS service is of vital importance in industry and everyday life. Semantic Web technologies, such as ontology, is the most important tool in accomplish an intelligence GIS services, capable of deal with the heterogeneity of data from different sources and interoperability between these sources. This paper proposes a prototype system that can leverage from existing ontology engineering technologies, build a prototype intelligent GIS service for the campus of Nanjing University of Science and Technology. A formal ontology is built, an open SPARQL endpoint is provided as a web service. Evaluation experiments are carried out using real-life data from campus geo locations, smell, sound, food, and blossom. It is evident from the evaluation results that the accuracy of the queries are 100%, average time overhead is less than 5 ms/query.

Keywords: Knowledge Graph; Ontology; GIS; SPARQL

*通信作者: 周倩如, zhouqianru@njust.edu.cn

1 前言

物质世界中所有的事物都具备时空位置属性,互联网中百分之八十的信息都与地理位置信息有关[1, 3-8]。在很多应用场景中,地理信息是改善系统对用户请求的响应关键因素,地理知识库和地理空间推理也可能在应急响应或交通规划中发挥关键作用,将不同类型数据集的数据结合起来对于从系统中获得大量信息来说是非常有用的。例如,当组织一次旅游时,用户可以将文化遗址和交通工具的信息与酒店住宿结合起来考虑它们之间是否接近。在紧急响应情况下,可以通过结合数据来获得一个有用并且及时的响应。在例如地震这种紧急情况下,用户应该能够查询数据库来查找集合营地、医院、救援地点、直升机降落地区和无家可归的人所搭建的帐篷的位置信息。但是目前的互联网依然是以超文本链接 HTML 为基础的,也就是说必须要人来处理网页的内容,因为只有人才能看得懂网页的内容,计算机不能直接处理网页的具体内容。下一代互联网,语义网,是以链接数据 Linked Data 为基础的,通过本体、元数据等知识工程技术,让计算机不仅可以获取网页链接,而且可以直接读懂每个网页上的内容,并且基于这些内容进行查询,进行人机之间、计算机与计算机之间的交流和沟通,将整个互联网变成一个庞大的数据库,可以随意查询、获取知识[10-13]。

本体,是起源于哲学领域的一个概念,意思是“事物的本质”[10-16]。上个世纪中后叶,哲学家、计算机学家和心理学家在创立人工智能这个学科的时候,把本体的概念引入到计算机领域,用来表示一个领域的本质,因为在编程时,也就是与计算机对话时,需要将人类的世界观传递给计算机,包括这个世界的事物、以及事物之间运转的规则等,而且要用计算机能懂的语言,这就是本体。具体来说,本体是一个领域的知识表达,它是一种协议、一个标准,一个词汇表。语义网的实现,是建立在本体的基础之上的,因为要将整个互联网变成一个知识库,首先就要用一个统一、标准的语言,来描述整个互联网的网页的知识。建立整个互联网的本体是一件非常繁杂的任务,这相当于将整个世界的知识建立一个模型,美国的人工智能公司 Cycorp 花了 35 年时间,动用了 3500 名科研人员,终于于 2020 年完成了一个人类所有知识领域的本体并建立知识图谱 Cyc [17-19],此外开源的众筹的人类百科知识图谱还有 Geonames [20]和 LinkedOpenData [21]、WikiData [22]、DBpedia [23]等。

在众多有关语义网技术收集和发布地理信息的项目中, Linked Open Street Map (LOSM) 是目前为止最全面,最成熟的一个[8, 9]。它是以开源、众筹的形式持续收集和更新地图数据。它的数据库每 15 分钟更新一次,截至今天,它包含了由 2,445,598 个用户贡献的 5027,330,590 个 GPS 点。利用语义网技术, LOSM 可以将 Open Street Map (OSM) 数据与其他来源的数据进行动态集成, 与其他的语义网地理信息项目(如 LinkedGeoData5)不同的是, LOSM 不依赖定期转储数据,而是永远公开最新的数据。虽然 LOSM 提供的 GIS 数据本身包含部分语义数据,但其语义数据并不丰富,无法覆盖很多地理实体,为了实现多维度的知识图谱服务, Daniele Quercia、Rossano Schifanella 和 Luca Maria Aiello 三人团队的 GoodCityLife 项目以城市生活为背景,制作了一系列有趣并令人深思的地图。他们使用社交媒体数据有效、廉价地大规模跟踪城市气味、声音等信息,并将这些信息与街道关联起来,生成了 Smelly Maps、Chatty Maps 等特殊的城市地图[1]。本文中我们尝试实现了基于语义网技术的集气味、声音、美食等为一体的智能地图服务,并在南京理工大学校园地图上进行了测试和验证。

2 本体

本体的构建主要有以下几种方式: 1)自底向上, 2)自顶向下, 3)自底向上或自顶向下相交替使用, 4)用形式概念分析 (Formal Concept Analysis, FCA) 从数据中半自动的提取本体[28]。在现在的人工智能应用中,实际上并非采用传统的自底向上或自顶向下方式构建本体,甚至在软件开发的前期根本没有考虑本体的构建,而是将软件开发与构建知识图谱环节相互交错,让语义化数据“野蛮生长”,然后在语义数据量足够丰富的情况下,再以自底向上的方式构建本体,用这种方法开发的软件潜在有不少问题,比如结构的紊乱、难以理解、可扩展性差等。因此,必须要将人工智能应用的开发建立在本体的构建的基础之上。

为了更好的构建应用所需的本体,本文复用了已有的成熟的本体, Suggested Upper Merged Ontology (SUMO) 和 wgs84 Ontology。SUMO 是 IEEE 标准的上层本体,描述了世界最抽象、最上层的概念,比如它把世间万物分类为两类“物质实体 Physical Entity”和“抽

象概念 Abstract Concept”，本文复用了这两个概念。
wgs84 是由开放地理空间协会 OGC 设计并发布[2]，是应用最为广泛的描述地理位置和相关信息的本体。本文

中本体的构建工具选择了 Protégé，这是一个免费的开源工具，用于构建、编辑、和形式化验证本体[24-30]。

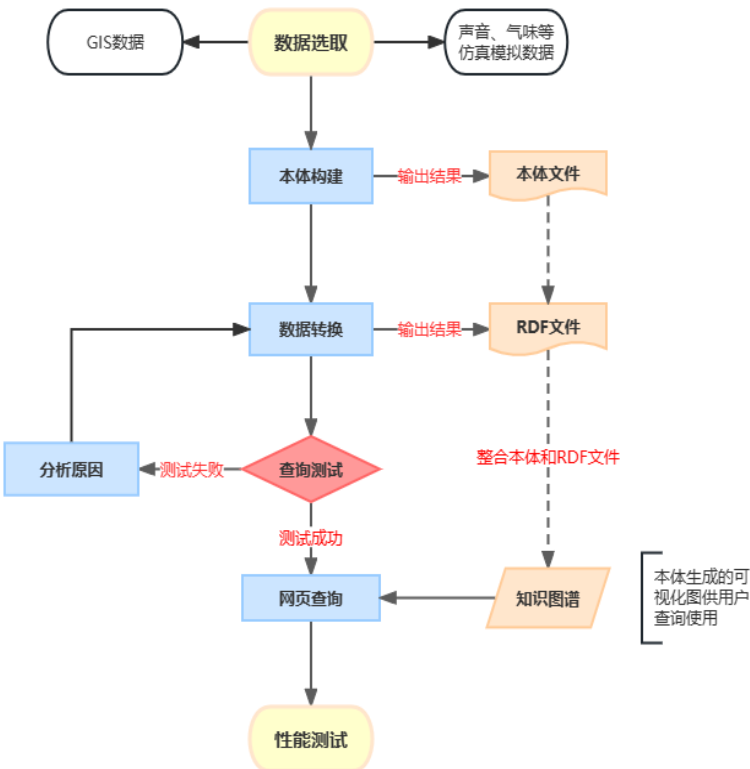


图 1 南理工智能地图服务开发流程

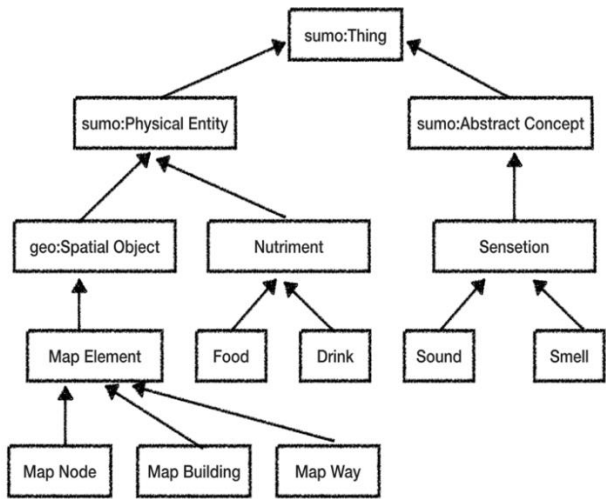
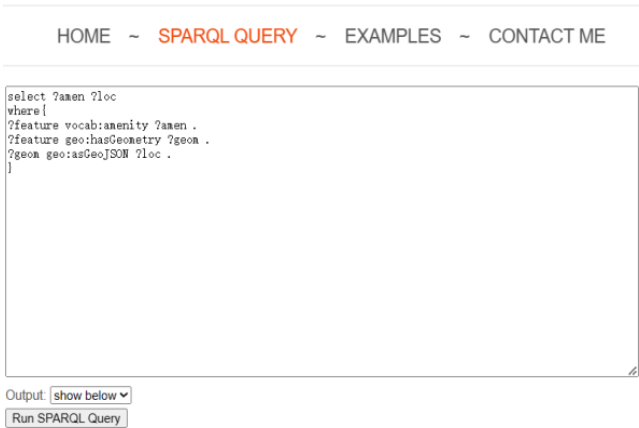


图 2 南理工智能地图服务使用的本体

本体中使用到的前缀如下表所示：
: <http://example.com/> .
geo: <http://www.opengis.net/ont/geosparql#> .
sumo: < http://www.ontologyportal.org/SUMO.owl#> .
xsd: <http://www.w3.org/2001/XMLSchema#> .

NJUST Map Designed For You



If you choose to output on this page, the answer will be displayed below.

图 3 南理工智能校园地图 SPARQL 查询前端界面

除了地理元素，如“位置点 Map Node”、“建筑物 Map Building”和“路径 Map Way”以外，本文地图中还包括的元素有“声音 Sound”、“气味 Smell”、“食物 Food”和“饮料 Drink”。其中“食物 Food”和“饮料 Drink”属于“Nutriment”，是“物质实体 Physical Entity”的一种，而“声音 Sound”和“气味 Smell”属于“感知 Sensation”，是“抽象概念 Abstract Concept”的一种。

使用这个本体，就可以描述如下问题：

“南理工占地面积最大的建筑是哪些？”

“南理工校园里哪里可以闻到花香/桂花香/咖啡的香气/烤面包的香气……？”

“南理工校园里哪条路自行车的噪音最少？”

“如果想要赏花，同时又能闻到花香，又不想听到自行车的噪音，应该选哪条路？”

“南理工校园里哪条路既可以赏花，又可以闻到咖啡的香味，同时又没有噪音？”

……

3 实验与验证

根据设计的本体，本文利用真实的地理数据和仿真的附加数据，构建了南理工校园智能地图的知识库，而

且搭建了 SPARQL 前端网站，实现了南理工的智能校园地图查询接口。

3.1 实验环境

具体实现工具和环境如下。

RDFLib: 本课题使用 RDFLib 工具来提取解析现有的数据，生成 RDF 形式化存储的知识图谱。

GeoPandas、Shapely、和 Matplotlib: 本文选取了 GeoPandas 包和 Shapely 包对查询结果可视化。GeoPandas 是一个开源项目，可以更轻松地使用 python 处理地理空间数据。它扩展了 pandas 使用的数据类型，允许对几何类型进行空间操作。几何运算是通过形状运算完成的。GeoPandas 进一步依赖 fiona 进行文件访问，并依赖 matplotlib 进行绘图。

Django: 选取了成熟的 Web 开发框架 Django 搭建网页服务器，页面存放于模板 template 文件夹中。在 vscode 本地终端执行 Django 的 runserver 命令启动服务器。

3.2 实验验证

本地图服务的前端 SPARQL 查询界面如图 3 所示。

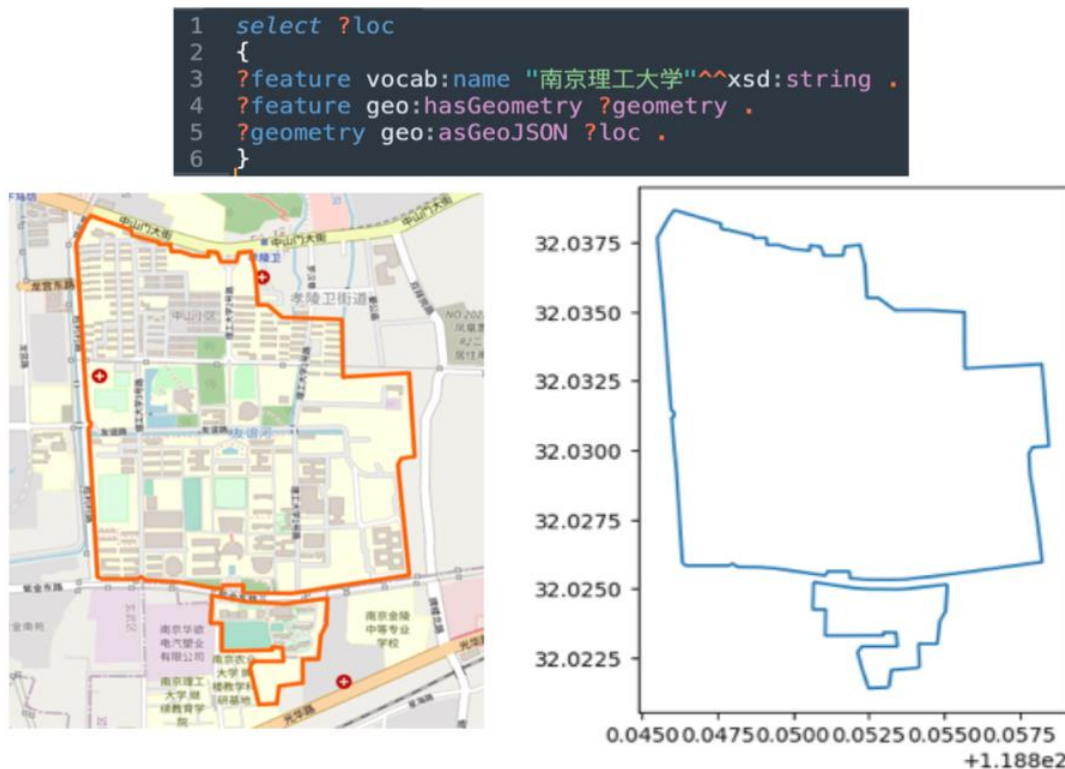


图 1 查询南理工校园轮廓的 SPARQL 语句和查询结果

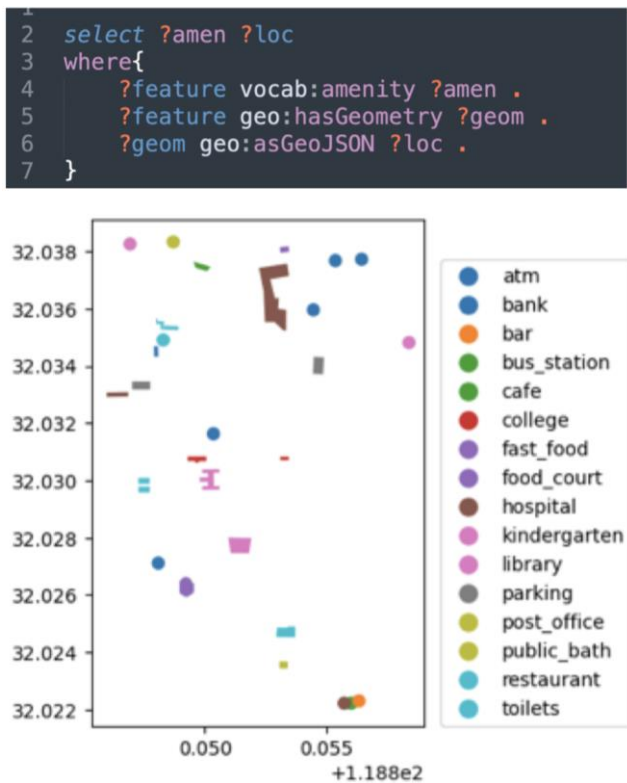


图5 南理工校园基础设施查询 SPARQL 语句和查询结果

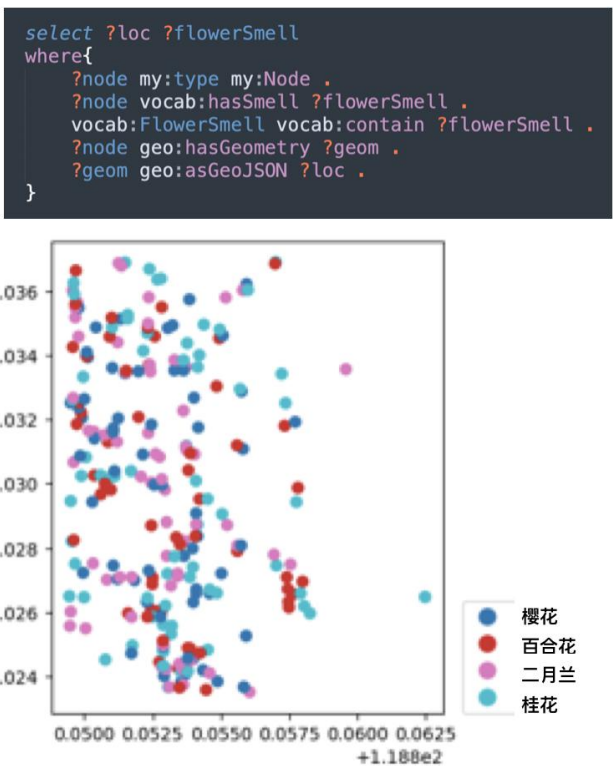


图7 南理工校园的能闻到花香的地方查询 Sparql 语句和查询结果

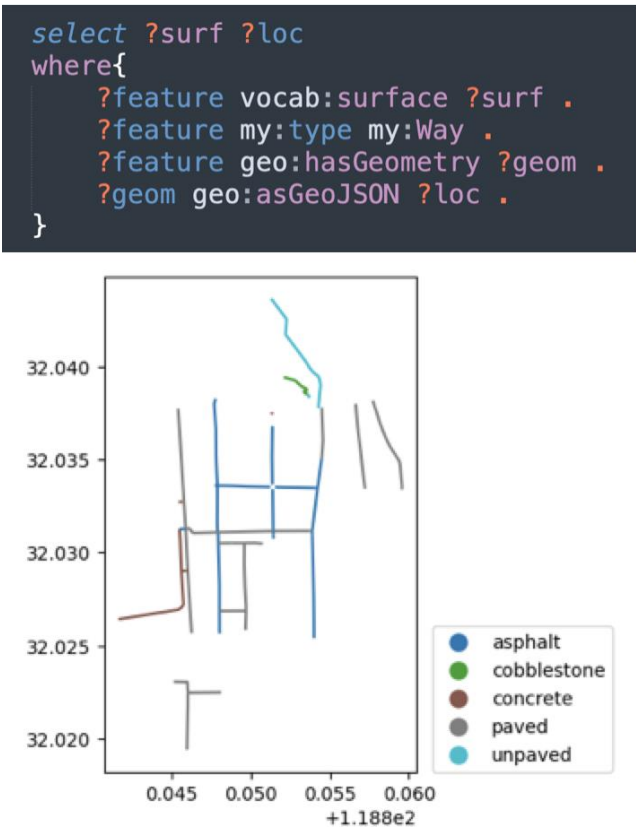


图6 南理工校园道路查询 SPARQL 语句和查询结果

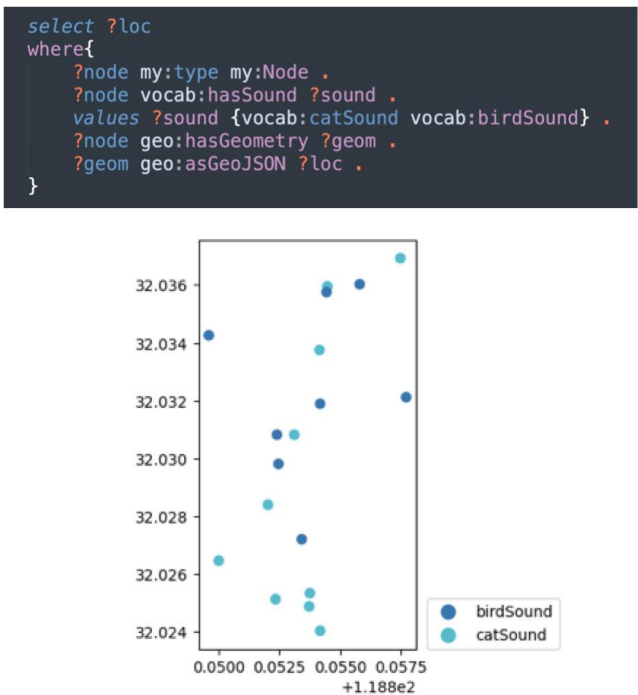


图8 南理工校园内能听到猫叫或鸟叫的地方查询 Sparql 语句和查询结果

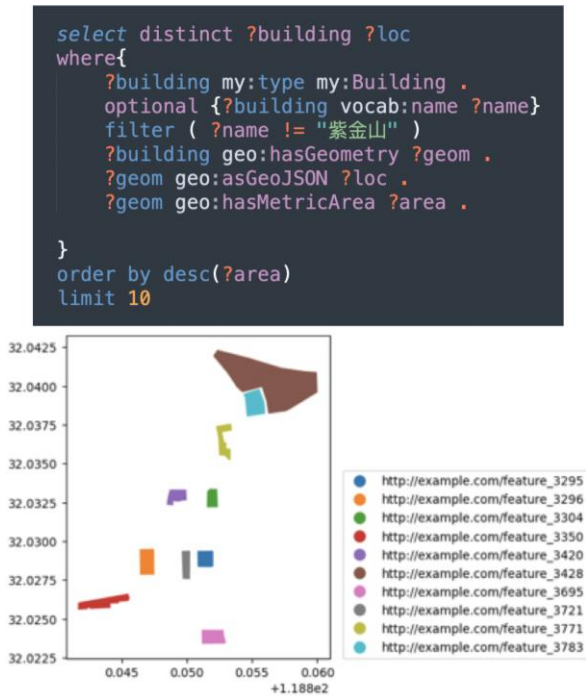


图 9 南理工校园内占地面积最大的 15 个建筑查询 SPARQL 语句和查询结果

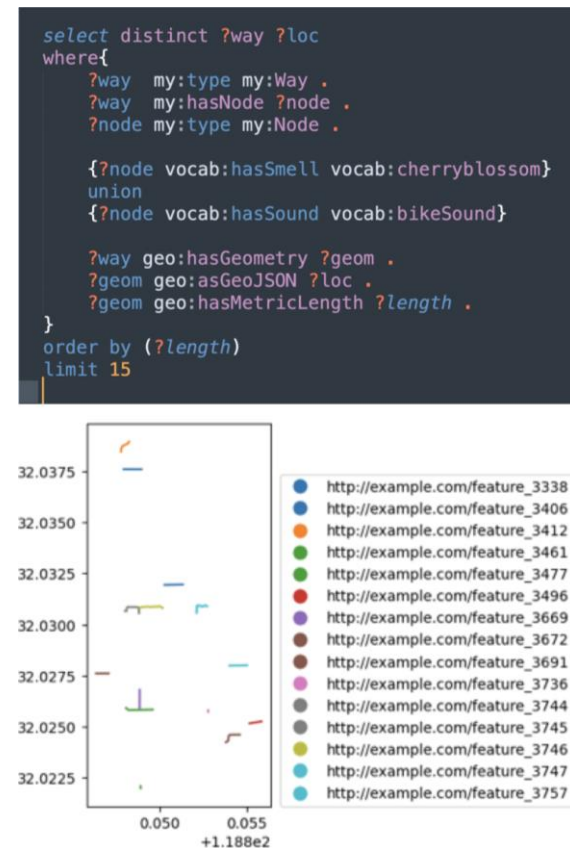


图 10 南理工校园内骑行赏花最佳路线，暨能闻到花香且听到自行车噪声最短的 15 条路径，查询 SPARQL 语句和查询结果

查询结果如图 4-10 所示，详细阐述如下。图 4 查询了“学校的轮廓是什么？”，查询结果如图所示，与真实校园轮廓是相符合的。图 5 查询了“南理工的基础设施都有哪些，分别在哪里？”查询结果如图所示，包括了咖啡厅、酒吧、车站、医院等地方，与事实相符合。图 6 回答了“南理工的主干道有哪些，分别在哪里？”，并且在地图上标注了出来。

图 7 回答了“南理工校园哪里能看到花，分别是什么花？”，并在地图上标注出了主要的花，比如二月兰、桂花、樱花和百合花的种植地点。图 8 回答了查询“南理工校园哪里猫咪和鸟出没的比较多？也就是能经常听到鸟叫和猫叫的地方？”并在地图上标注出来。

图 9 回答了查询“南理工校园占地面积最大的 15 个建筑是哪些，分别在哪里？”查询所有的建筑物及其占地面积，并进行递减排序，得到排名前 15 位占地面积相对较大的一些建筑物。查询语句里排除了“紫金山”，因为地图里包含了紫金山的数据。图 10 查询了最佳赏花路线，这里最佳赏花路线的定义是：“能看到花，并且自行车最少，最清静的路线，也就是能听到自行车噪声最短的路线”。如图所示，该服务可以迅速、准确的标注出符合条件的路线。

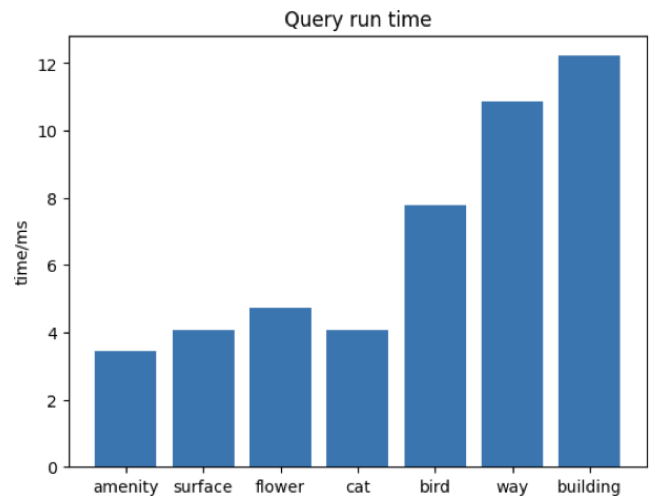


图 11 图 4 至 10 中示例 SPARQL 查询语句所用的时间损耗

以上示例 SPARQL 查询语句所用的时间损耗如图 11 所示，其中“amenity”对应的是图 5 的基础设施查询，“surface”对应的是图 4 中的轮廓查询，“flower”对应的是图 7 中的花香查询，“cat”和“bird”对应的是图 8 中的猫叫和鸟叫的查询，“way”对应的是图 10 中的最佳赏花路线的查询，“building”对应的是图 9 中的占地面积最大的建筑查询。由图中可见，绝大多数查询耗时都不

超过 10 毫秒（除了“way”和“building”），耗时最长的“building”也只是耗时 12 毫秒。大部分查询（“amenity”、“surface”、“flower”和“cat”）只需要 4 毫秒。因此可知，本文的南理工校园智能地图服务效率很高，可满足日常查询需求。

4 结论

本文利用语义网技术，整合了跨平台、跨数据库、跨格式的数据，比如声音、气味、美食、生活设施等数据，构造了南理工智能校园地图服务，并提供了前端 SPARQL 查询端口，向大众开放。本服务可以提供语义丰富、类似自然语言的智能查询，并在实际地图上进行了测试，结果符合预期，准确率 100%，时间损耗在 10 毫秒以内，可达到在实际生产生活中应用的需求。

参考文献

- [1] Aiello Luca Maria, Schifanella Rossano, Quercia Daniele, Aletta Francesco. Chatty maps: constructing sound maps of urban areas from social media data. [J]. Royal Society open science, vol. 3, no. 3, 2016.
- [2] Battle, R., and Kolas, D., Geosparql: enabling a geospatial semantic web. Semantic Web Journal, vol. 3, no. 4, pp. 355-370, 2011.
- [3] Neis, Pascal, and Dennis Zielstra. "Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap." Future internet, vol. 6, no. 1, pp. 76-106, 2014.
- [4] Kounadi, O. Assessing the quality of OpenStreetMap data. Msc thesis, geographical information science, University College of London Department of Civil, *Environmental And Geomatic Engineering*, 19, 2009. Berners-Lee T. Semantic web road map [J]. 1998.
- [5] Bennett J. OpenStreetMap [M]. Packt Publishing Ltd, 2010.
- [6] Grinberger A. Yair, Minghini Marco, Juhász Levente, Yeboah Godwin, Mooney Peter. OSM Science—The Academic Study of the OpenStreetMap Project, Data, Contributors, Community, and Applications [J]. ISPRS International Journal of Geo-Information, vol. 11, no. 4, 2022.
- [7] H. Mordechai, and P. Weber. "Openstreetmap: User-generated street maps." IEEE Pervasive computing vol. 7, no. 4, pp. 12-18, 2008.
- [8] Tempelmeier, N., & Demidova, E. (2021). Linking OpenStreetMap with knowledge graphs—Link discovery for schema-agnostic volunteered geographic information. *Future Generation Computer Systems*, vol. 116, pp. 349-364.
- [9] Tempelmeier Nicolas, Simon Gottschalk, and Elena Demidova. "GeoVectors: A Linked Open Corpus of OpenStreetMap Embeddings on World Scale." In *Proc. ACM CIKM*, 2021.
- [10] Kang J., Aletta F., Gjestland T. T., et al. Ten questions on the soundscapes of the built environment [J]. *Building and environment*, 2016, vol. 108, pp. 284-294.
- [11] Berners Lee Tim, Hendler James, Lassila Ora. The Semantic Web [J]. *Scientific American*, vol. 284, no. 5, 2001.
- [12] Hitzler Pascal. A Review of the Semantic Web Field Tracing the triumphs and challenges of two decades of Semantic Web research and applications [J]. *COMMUNICATIONS OF THE ACM*, vol. 64, no. 2, 2021.
- [13] Nigel Shadbolt, Tim Berners-Lee, Wendy Hall. The Semantic Web Revisited. [J]. *IEEE Intelligent Systems*, vol. 21, no. 3, 2006.
- [14] 朱木易洁, 鲍秉坤, 徐常胜. 知识图谱发展与构建的研究进展 [J]. *南京信息工程大学学报 (自然科学版)*, vol. 9, no. 06, pp. 575-582, 2017.
- [15] Yuhan Sun, Mohamed Sarwat. A generic database indexing framework for large-scale geographic knowledge graphs [P]. *Advances in Geographic Information Systems*, 2018.
- [16] 罗强, 胡中南, 王秋妹, 石伟伟, 贾玥. GIS 领域知识图谱进展研究 [J]. *测绘地理信*, vol. 48, no. 01, pp. 60-67, 2023.
- [17] Lenat, Douglas B. "CYC: A large-scale investment in knowledge infrastructure." *Communications of the ACM* 38.11 (1995): 33-38.
- [18] Lenat, D. B., Prakash, M., & Shepherd, M. (1985). CYC: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks. *AI magazine*, 6 (4), 65-65.
- [19] Lenat, Douglas B., et al. "Cyc: toward programs with common sense." *Communications of the ACM* 33.8 (1990): 30-49.
- [20] Bond, F., & Bond, A., GeoNames wordnet (GeoWN): extracting wordnets from GeoNames. In *Proc. 10th Global Wordnet Conference*, July 2019, pp. 387-393.
- [21] Bauer, F., and Kaltenböck, M., Linked open data: The essentials. Edition mono/monochrom, Vienna, vol. 710, no. 21, 2011.
- [22] Vrandečić, Denny, and Markus Krötzsch. "Wikidata: a free collaborative knowledgebase." *Communications of the ACM*, vol. 57, no. 10, pp. 78-85, 2014.

- [23] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z., DBpedia: A nucleus for a web of open data. In international semantic web conference, Nov. 2007, Berlin, Heidelberg, pp. 722-735.
- [24] Tang J., Long Y., Measuring visual quality of street space and its temporal variation: Methodology and its application in the Hutong area in Beijing [J]. Landscape and Urban Planning, pp. 191-103, 2019.
- [25] Noy, N. F., Crub ézy, M., Fergerson, R. W., Knublauch, H., Tu, S. W., Vendetti, J., & Musen, M. A. Prot égé2000: an open-source ontology-development and knowledge-acquisition environment. In Proc. AMIA, Jan. 2003, pp. 953-953.
- [26] 许鑫, 杨佳颖. 国外语义网研究现状与动向——基于 2002—2018 年 ISWC 会议 [J]. 情报学报, vol. 39, no. 07, pp. 761-776, 2020.
- [27] You, Lin. Towards a research agenda for knowledge engineering of virtual geographical environments [J]. Annals of GIS, vol. 22, no. 3, 2016.
- [28] 章勇, 吕俊白. 基于 Protege 的本体建模研究综述 [J]. 福建电脑, vol. 27, no. 01, pp. 43-45, 2011.
- [29] Song Dan, Ye Xin, Wu Wenrong, Zhang Zhijing, Saren Qimuge, Qian Jiahui, Zhu Dongsheng. Ontology-based assembly knowledge representation and process file generation [J]. MATEC Web of Conferences, pp. 355, 2022.
- [30] 敬国伟, 黄大池. 基于 ECharts 的数据可视化研究 [J]. 西部广播电视, vol. 43, no. 20, pp. 227-230+234, 2022.