

齐白石知识图谱构建方法与应用形式研究

文 / 石倩雯

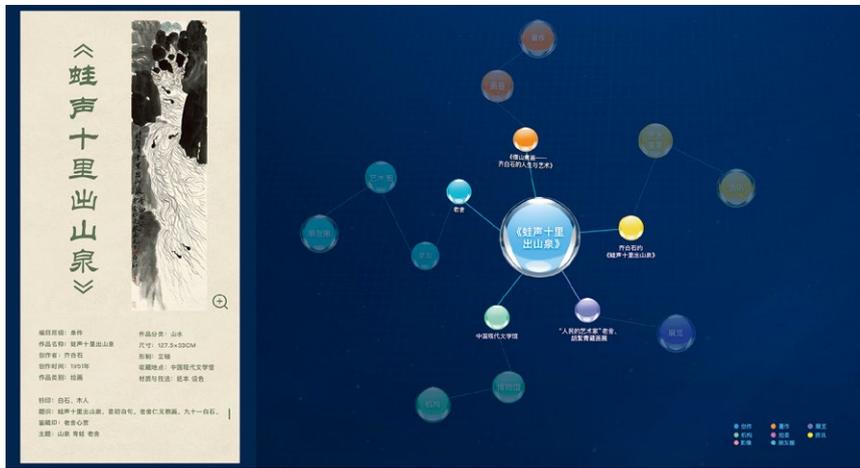
一、研究缘起

知识图谱是一种基于图的数据结构，由节点（Point）和边（Edge）组成，每个节点表示一个“实体”，每条边为实体之间的“关系”，知识图谱本质上是语义网络。实体指的是现实世界中的事物，譬如人物、地名、机构、书籍等；关系则用来表达不同实体之间的某种联系。知识图谱不仅可以为互联网中的信息表达成更接近人类认知世界的形式，而且提供了一种更好的组织、管理和利用海量信息的方式，其发展得益于多个研究领域的成果，是知识库自然语言处理、语义网技术、机器学习、数据挖掘等众多知识领域交叉融合的产物¹。

文化遗产是人类在社会历史发展过程中创造的具有历史、艺术、科学等价值的文化财富，深入研究和挖掘文化遗产，有利于阐发文化精髓，保护与传承文化遗产，推动中外文化交流互鉴²。在文化遗产领域，知识图谱目前在国内外已经得到了广泛的应用，国内典型的知识图谱应用案例如上海图书馆的中国家谱知识服务平台、盛宣怀档案知识库，上海博物馆的董其昌数字人文项目等。国外自 2010 年起，大英博物馆、美国弗利耳美术馆、荷兰国立博物馆等众多大量的国际博物馆均已开展知识图谱相关的项目与应用，例如大英博物馆关联开放数据项目、俄罗斯文化遗迹云平台、荷兰国立博物馆知识图谱项目、Europeana 等。其中，Europeana 是由欧盟各国的图书馆、档案馆以及博物馆联合建设的大型数字博物馆，2012 年 10 月，Europeana 数据集的 2000 万条文字、图像、视频和声音的预算内数据被转换为关联数据，并可以从网址获得，这些数据给欧洲的文物、文化遗产提供了新的共同接口，构建了文化资源不同概念之间的关联关系的语义知识本体模型³。

齐白石作为 20 世纪中国最负盛名、最具创造力和影响力的中国画大师之一，与之相关的作品、手稿、图书等相关内容数量也十分庞大。中国艺术品数据库（以下简称中艺库）收录齐白石数据 73, 087 条，其中创作 70, 339，资讯 1, 180，著作 374，影像 1, 079，展览 47，拍卖 68。可看出，虽然数据库中数据量较为可观，但是彼此间缺乏有效的关联，暂未具备较强知识性和传播性的功能。因此，为更好地展示数据之间的关联，形成较为优质的可视化展示方式。2021 年，雅昌艺术数据中心联合北京画院，利用知识图谱构建技术将齐白石相关资料进行数据挖掘、建立关联关系，构建齐白石知识图谱模型，在数字大屏中已有阶段性的研究成果。本文将重点阐述齐白石知识图谱构建方法与应用形式。

1: 陈涛, 刘炜, 单蓉蓉, 朱庆华. 知识图谱在数字人文中的应用研究 [J]. 中国图书馆学报, 2019, 45(06): 34-49.
2: 李章超, 何琳. 文化遗产语义组织研究进展 [J]. 图书情报工作, 2020, 64(07): 4-12.
3: 欧洲数字博物馆. [EB/OL]. [2021-9-22]. <https://www.europeana.eu/en>
4: 刘峤, 李杨, 段宏, 刘瑶, 秦志光. 知识图谱构建技术综述 [J]. 计算机研究与发展, 2016, 53(03): 582-600.



图三、《蛙声十里出山泉》示例



图四、齐白石年表之大事记



图五、齐白石五出五归

(二) 足迹

齐白石的五次出游，美术史称之为“五出五归”，对其在人际交往、思想境界以及绘画创作等方面均得到相应的提高。为此，本研究梳理了五出五归之概述、背景、意义以及与“衰年变法”间的联系，并且将五次出行的路线与作品也作了关联展示，如图五所示。

(三) 关联关系

由于齐白石的知识图谱概要模型与框架可作延伸性的知识拓展，可增加本体、实体以及实体对应的元数据，而不断随着数据的增加与迭代，数据量将十分庞大，二维平面已经无法满足展示的需求。因此，雅昌艺术数据中心选定以三维形式展现，如图六和图七所示。

三、齐白石知识图谱在展示大屏中的应用

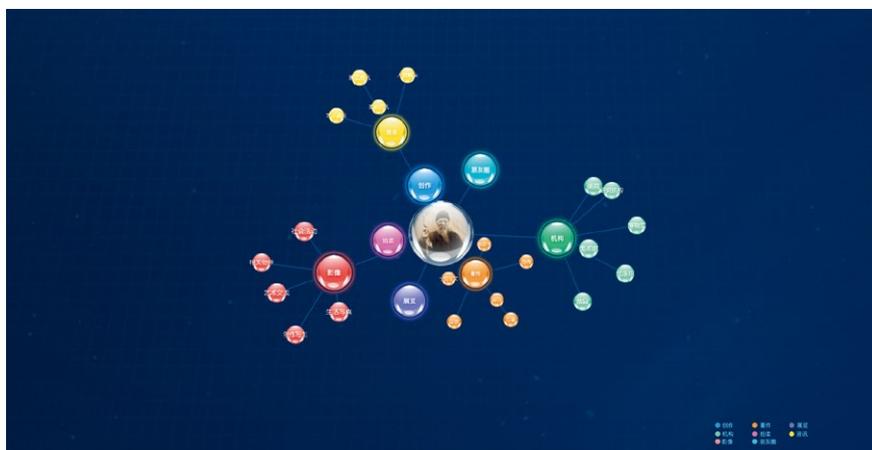
将知识图谱与数字大屏相结合是当下将数字人文技术应用于数字图书馆、博物馆等文化机构发展的必然趋势。齐白石主题知识图谱展示系统是基于北京画院馆藏资源以及雅昌中国艺术品数据库中结构化的元数据为主，并在此基础之上作了数据清洗，借助大数据、知识图谱等现代信息技术，集文化资源与展示于一体，既为用户提供线上专题知识服务，同时支持人文研究。展示分为齐白石年表、足迹、关联关系。三个单元相互独立，又有所关联，尽可能通过可视化的方式为用户提供更为友好的交互体验。

(一) 年表

齐白石年表版块主要梳理1864年(甲子)—1957年(丁酉)这段时间其所有事迹，以及每年所创作的重要作品。时间轴按照年份为单位，从中可读取不同时间的生平事迹示例数据，为用户在齐白石人物大事记方面有较为直观的了解。此外，在年表中部，通过大数据计算，可以统计出每年的书法、绘画、印章的创作数量，并形成曲线图，使得观众可较为清晰的看出创作高峰期，如图四所示。



图六、齐白石关联关系首页



图七、齐白石关联关系次页

四、结语

本研究提出了齐白石知识图谱的构建方法，基于知识图谱模型研究数字展示设计与应用。首先建立本体概念模型，其次采用数据关联三元组方法构建齐白石知识图谱，并利用计算机技术开发了齐白石数字展示服务。

齐白石知识图谱的构建与应用是北京画院与雅昌艺术数据中心对文化遗产领域知识图谱构建与应用的一次初步尝试，目前数据量较小、复杂度不高、构建自动化程度较低。未来无论是知识图谱的构建工作，还是将其服务于更多专题知识等许多方面都需要进一步深入研究，主要包括：（1）获取更多的知识数据源，建立更加完善、权威、多样的文化资源；（2）运用更加智能化的知识抽取、知识融合与知识推理等技术，实现自动化程度更高的知识图谱建模与构建；（3）进一步开发知识问答、智能推荐、知识检索等功能，实现更加丰富的知识应用。

雅昌 艺术数据

文化和科技交融，共同推动万物数字化的未来！

遵循国家“数字产业化”和“产业数字化”的指导方针，依托雅昌丰富的数据资源和艺术沉淀，通过领先行业的数据采集编辑能力、专业的数据治理能力、前沿的IT技术研发能力、多元的场景数据展示应用策划制作能力，为文博机构、艺术机构、艺术院校的智慧保护、智慧管理和智慧服务提供全流程的综合解决方案，为大众提供丰富的艺术教育内容、优质便捷的应用产品体验。

用科技助推文物保护，让文化资源永久保存、可持续再现。

用科技赋能博物馆智慧管理，让博物馆成为“智慧生命体”。

用科技带给观众永不落幕的展览，优化未来博物馆展示与路径。

用科技打造数字知识内容，让知识与人文触手可及。



雅昌艺术数据
官方二维码