

# 基于关联数据的科学家学术谱系构建——以钱伟长学术谱系为例\*

张昊

(上海大学文化遗产与信息管理学院 上海 200444)

〔摘 要〕构建科学家的学术谱系对于纪念科学家及其学术后代的学术荣誉、学术成就，保存属于科学家学术群体乃至学术共同体的社会记忆尤为重要。本文提出了基于关联数据的学术谱系构建路径，并将其划分为四个层级；其次以钱伟长的学术谱系为例，通过梳理钱伟长学术谱系相关信息，设计学术谱系本体，通过实体 RDF 化、本体语义映射、实体关联与关联数据发布等关键步骤完成钱伟长学术谱系构建，最后从知识关联、知识发现两个角度进行应用展望。基于关联数据构建科学家学术谱系能有效避免以往构建的弊端，丰富学术谱系的知识内容，对于理解科学家的教育思想及学术传承具有重要的实践意义。

〔关键词〕学术谱系 关联数据 本体 钱伟长

〔中图法分类号〕G250

## 0 引言

2019年6月11日，中共中央办公厅、国务院办公厅印发的《关于进一步弘扬科学家精神加强作风和学风建设的意见》中明确指出要大力弘扬科学家甘为人梯、奖掖后学的育人精神<sup>[1]</sup>。科学家学术谱系是传承科学家育人精神的载体，是以师徒关系的历代传承为连接进行知识转移的系统<sup>[2]</sup>。除了对科学家精神的传承，结合当前科技发展的实际情况，研究学术谱系还能可以科学地预测该学科未来的发展重点和突破方向，为学科的未来规划提供了理论依据和方向指导。利用科学家学术谱系进行相关领域的科学史研究<sup>[3,4]</sup>、学术谱系的测度评价<sup>[5,6]</sup>以及创新应用<sup>[7]</sup>等受到人们普遍关注，但是鲜见运用系统的工具和科学的方法构建科学家学术谱系及其实证研究。钱伟长被誉为中国近代力学之父和应用数学之父，目前学界对钱伟长的研究主要聚焦于钱伟长的教育思想<sup>[8]</sup>、爱国精神<sup>[9]</sup>和学术成就<sup>[10]</sup>上，但是缺乏钱伟长学术谱系的研究。本文在以钱伟长的学术谱系为例，提出基于关联数据的科学家学术谱系构建路径，以增进人们对科学家育人精神的感悟，为加强我国学界作风和学风建设提供参考。

## 1 基于关联数据构建科学家学术谱系的可行性分析

关联数据是国际互联网协会(W3C)推荐的一种规范，用来发布和连接各类数据信息和知识，它采用 RDF 数据模型利用 URI（统一资源标识符）命名数据实体来发布和部署实例数据和类数据，从而可以通过 HTTP 协议揭示并获取这些数据，同时强调数据的相互关联、

---

\*本文系国家社科基金重点项目“数字时代的开放科学政策、实施路径与评价研究”（项目编号：22ATQ005）研究成果之一。

相互联系和有益于人机理解的语境信息<sup>[11]</sup>。关联数据技术的应用已经相对成熟，基于关联数据的科学家学术谱系构建路径具有较高的可行性。

### 1.1 可实现数据互联和集成

关联数据技术在实现数据互联和集成方面的显著优势为构建科学家学术谱系提供了可行的技术保障。关联数据采用统一资源标识符（URI）和资源描述框架（RDF），保证不同数据源中的信息能够被唯一标识并且结构化表示。不同的 URI 使每个科学家、出版物或研究机构都可以获得一个全球唯一的地址，而 RDF 则提供了描述这些实体及其相互关系的语义信息。无论数据最初来源于何处，不同的数据库、期刊或者档案，它们都可以被有效地链接和整合在一起，而这正满足了科学家学术谱系构建的需要。学术谱系需要整合包括个人公开信息、论文、专利、报告等大量分散的数据，关联数据使这些信息能够跨越不同的数据库进行整合，从而形成连贯且全面的学术网络。

### 1.2 可提供统一的语义框架和数据整合过程的一致性

语义标准化和数据一致性是构建科学家学术谱系过程中的核心要求。关联数据通过采用 RDF 数据规范，为数据的描述和查询提供了统一的语义框架，同时要求准确定义和连接各种实体（如科学家、研究机构、学术论文等）及其相互关系，使不同数据源中的同一实体和关系得到统一识别和解释，进而保证数据整合过程的一致性。例如一个科学家可能在不同的数据集中以不同的名字或标识符出现，关联数据将这些不同的表达链接到同一实体上，满足信息归一化的要求。此外，标准化的查询语言如 SPARQL 允许研究人员通过精确的语义查询，高效地检索和分析学术谱系数据。诸多特性不仅支持了科学家学术谱系中复杂关系的准确映射，也提高了数据处理的效率和可靠性，为维护和更新科学家学术谱系数据库的长期可行性创造条件。

### 1.3 可提高学术谱系的动态性和扩展性

科学家实时变动的个人信息和不断涌现的学术成果对学术谱系提出了动态扩展的要求。科学研究是一个不断进步和变化的过程，新的研究成果和合作关系不断涌现，学者的师承链接关系和合作网络也随之发生变化，这就要求科学家学术谱系能够适时动态更新与扩展以准确地记录和展示不断变动的成员信息。当新的研究成果或学者信息出现时，关联数据支持数据的动态更新的特性，有直接添加数据到现有的数据集，而无需重新构建整个谱系的应用前景：一方面保证科学家学术谱系能够持续地更新和维护，使之成为一个活生生、反映当前科学研究状况的资源，另一方面允许在现有架构基础上添加新的数据类型和关系，使构建的科学家学术谱系不仅能够持续地反映学术界的最新进展，还能适应学术研究领域的演变和扩展。

## 2 基于关联数据的科学家学术谱系的构建

基于关联数据的实践对技术依赖较高，需要应用多种技术手段。主流的关联数据应用

是在 JavaWeb 三层架构的基础上增添了用于领域知识组织的本体层，基于此本文提出了一种新的学术谱系构建路径试图厘清谱系成员背后的概念与学术关系，见图 1。

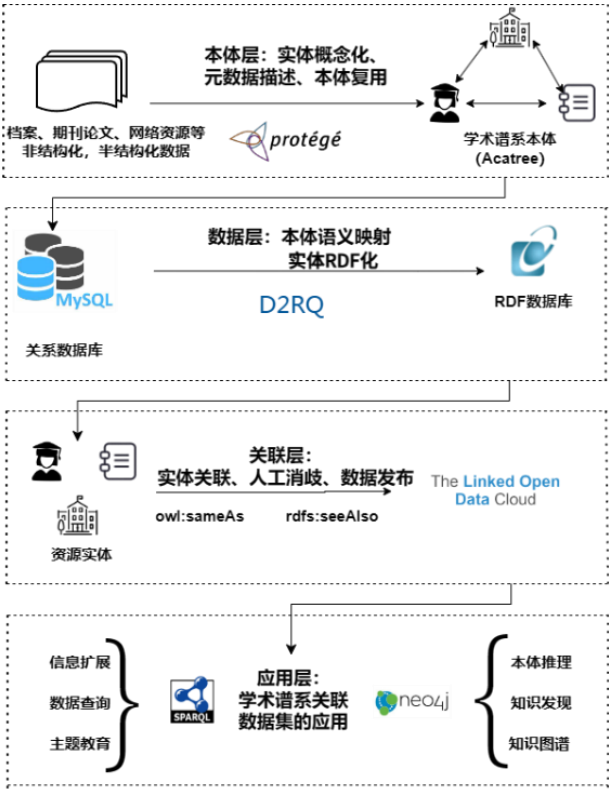


图 1 基于关联数据的学术谱系构建路径

最顶层是本体层，本体是以共享概念模型呈现的明确的形式化规范说明，用于描述和组织特定领域的信息和知识。本文对学术谱系本体定义为学术谱系内所含的信息和知识经组织、抽象与描述后形成的明确、可共享的结构化概念模型。第二层是数据层，其数据来源于档案、网络资源、期刊论文等的非结构化和半结构化的数据。数据层将经数据规整后存储在关系数据库中的数据进行本体语义映射以实现学术谱系关联数据的形成与组织，后专门存储在 RDF 数据库中。第三层是关联层，该层是实现关联数据的关键，要求尽可能多的将学术谱系关联数据集中的实体与外部资源库的资源相匹配，以实现学术谱系数据与更广泛的知识网络相连接，为学术研究和教育领域提供了更多的资源和视野。第四层是应用层，应用层能提供用户视图或操作平台，是为用户多样化利用学术谱系关联数据集的技术端口，同时应用层也是用户需求的满足层，应用层的技术选取应根据实际需求和工具的应用倾向性采取适合的技术手段来满足用户的需求。本文将依据以上层级的划分，构建学术谱系本体模型，获取谱系成员数据，实现谱系实体关联与谱系数据发布，以钱伟长学术谱系为例完成基于关联数据的科学家学术谱系的构建。

### 2.1 学术谱系本体模型设计

本体凭借其优秀的知识组织功能助力其在知识图谱、数字图书馆、语义网、自然语言

处理等领域的广泛应用。本文在参照文献<sup>[12-14]</sup>构建的相关本体模型的基础上，使用本体构建工具 Protégé 构建学术谱系本体模型，试图揭示和描述学术谱系内的显性和隐性的概念与关系，有效地组织起谱系领域知识，为后续工作奠定基础。学术谱系本体模型见图 2。

2.1.1 本体模型核心类的设计

前人学术谱系的研究往往将谱系成员作为学术谱系的唯一类别，依据家谱构建“学谱”，却忽视了学术谱系的关联性所能反映的学术价值，束缚学术谱系在跨学科领域中的交叉应用。鉴于此，本文尝试构建 Person（成员类），Academic\_Achievement（学术成果类）和 Organization（机构类）三个核心类以求能较为全面地概括学术谱系的众多概念和关系。成员类是学术谱系本体中最重要的类目，可以说所有有关人的实例的构建都直接或间接依赖于成员类。学术成果是通过各种科研活动，包括实验与观察、调研、综合分析、技术开发和生产效果评估等所进行的脑力及体力劳动成果，并力求涵盖尽科学家所有可能的学术成果。机构类通常是指符合国务院学位委员会要求的具有硕士、博士学位（或同级）授权点的高校或科研院所。

2.1.2 本体模型属性的设计

统计学术谱系本体模型对象属性 10 个，数据属性 29 个，示例见表 1。对象属性类似于语法学中的谓语，体现主语和宾语关系。其中不乏有逆属性，如 dcterms:hasPublisher 和 saws:hasWritten 有相反的主语和宾语。数据属性用于将实体与具体的数据值相关联，通常用于表示实体的某些性质、特征或属性。数据属性与对象属性不同，对象属性用于描述实体之间的关系，而数据属性用于描述实体的属性值。值的类型包括日期、文本、字符串、布尔值等。构建本体命名空间为<http://acatree.cn/>，缩写为“aca:”，资源的 URI 命名格式为“命名空间/类名#实体代码”，如<http://acatree.cn/Person#G0>表示代号为 G0 的实体，即钱伟长。鉴于我国研究生教育发展的特殊性，在创建实体时，特定时期产生的指导和被指导关系需经多方鉴别。

表 1 学术谱系本体模型属性示例

属性类别	属性名称	描述	示例
对象属性	“hasAuthor”	学术成果的作者	“学位论文 A” hasAuthor “作者 A”
	“publisherOf”	机构/学校发布的成果	“机构 A” publisherOf “论文 A”
	“supervisedBy”	论文由某人指导	“论文 A” supervisedBy “指导老师 A”
	“supervisorOf”	某人担任论文的指导者	“指导老师 A” supervisorOf “论文 A”
	“graduatedFrom”	某人从某高校毕业	“毕业生 A” graduatedFrom “学校 A”
数据属性	“person_id”	赋予个人 ID	“G2_5_3”
	“thesis_id”	赋予论文 ID	“T2_5_3-M”
	“ORCID”	ORCID 标识符	“0000-0002-1234-5678”

“location”	机构所在位置	“上海”
“email”	电子邮件地址	“user@example.com”

2.1.3 本体复用

本体复用是指在构建本体时，利用现有的本体、词汇或概念来构建新的本体，避免重复劳动并促进本体的可重用性和互操作性。本体复用是本体工程中的一项重要原则，可以加速本体的创建过程，提高本体的质量，并促进不同本体之间的集成。本文复用了 FOAF, DCterms, RDFS, BIBO, VIVO 等十余个通用本体和元数据集，如 foaf:person, dcterms:publisher, bibo:doi，在保证学术谱系本体的共享性的同时，还能实现对领域知识的标准化<sup>[15]</sup>。

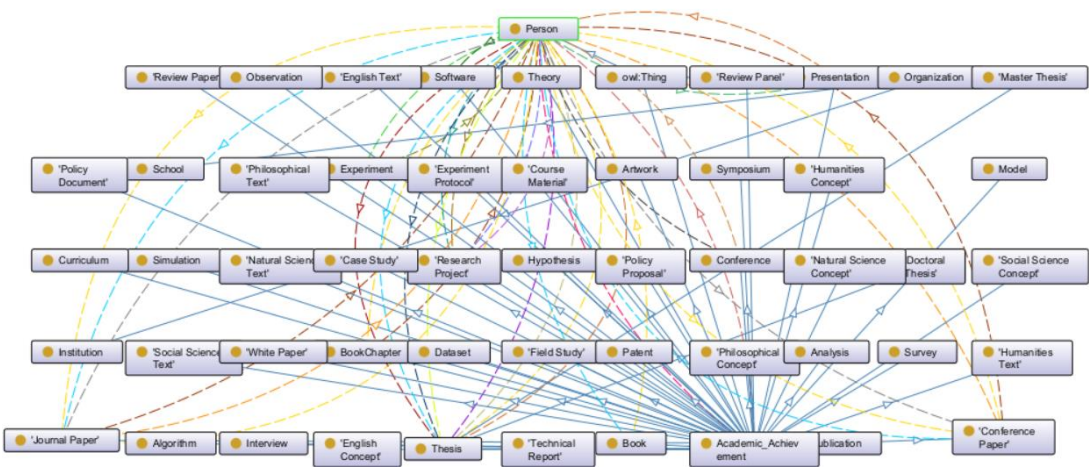


图 2 学术谱系本体模型

2.2 实例——钱伟长学术谱系的构建

2.2.1 钱伟长学术谱系数据的获取与处理

钱伟长的学术谱系已扩展至第四代。钱伟长（G0）至其指导的研究生（G1）信息源自上海大学档案馆“钱伟长全宗”，辅以纪念馆、专著<sup>[16]</sup>、博客<sup>[17]</sup>和邮件询问等，第二至第四代成员信息通过数据库检索先辈姓名、导师信息，并结合多个学术和官方资源进行验证和比对，如 CNKI、万方数据库、国家图书馆博士论文库等，以确保信息的一致性和准确性。虽然存在部分数据未公开，但不妨碍本文对已获得的非结构化或半结构化文本信息和结构化的学术论文元数据进行预处理。对前者实体识别与抽取，辅之以人工重检获取成员间的结构化的实体数据并保在 MySQL 数据库中，对后者则编写程序脚本将其转化为关系型数据进行存储。

谱系涉及时间跨度为 1952-2022 年，自钱伟长指导第一个研究生陈至达毕业起至 2022 第四代研究生毕业论文完成终。其中 G1 代 58 人，G2 代 448 人，G3 代 1865 人，G4 代 7098 人。钱伟长的教育生涯中培养了很多优秀的学生，不乏有中国科学院学部委员、中国工程院院士郑哲敏，也有长期从事力学教育与固体力学，创建西北地区唯一的理科力学系的叶

开沅。钱伟长学术谱系延续了钱伟长的治学风格，成员中钱伟长、郑哲敏、陈至达、何满潮、胡海昌、周又和、郑晓静等 7 人被评为院士。可见，钱伟长学术谱系枝繁叶茂，发展态势良好。

2.2.2 学术谱系的实体 RDF 化

将 Mysql 数据库中构建的关系表依据本体实现从关系数据库到 RDF 数据库的语义映射（实体 RDF 化），这一过程本质上是 将关系表中的二元组映射成“资源-属性-属性值”的三元组形式，是将关系表结构转换成更善于处理复杂关系、更富有语义信息的关联数据<sup>[18]</sup>的过程。

D2RQ 是使数据真正走向关联的重要工具，其中 d2rq-mapping 文件是实体 RDF 化的关键组件。它能实现的映射类型有数据库表向本体类的映射，关系表列向本体属性的映射，关系表的主键向本体实体的映射和关系表外键向本体关系的映射。通过合理配置映射文件，D2RQ 能够根据映射规则将关系数据库中的数据转化为 RDF 数据模型。映射过程要求原关系数据库表中必须包含主键，作为三元组的“主语”。以 person 表和 person\_to\_thesis 表为例，映射关系见图 3，谱系数据 RDF 化效果示例见表 2。

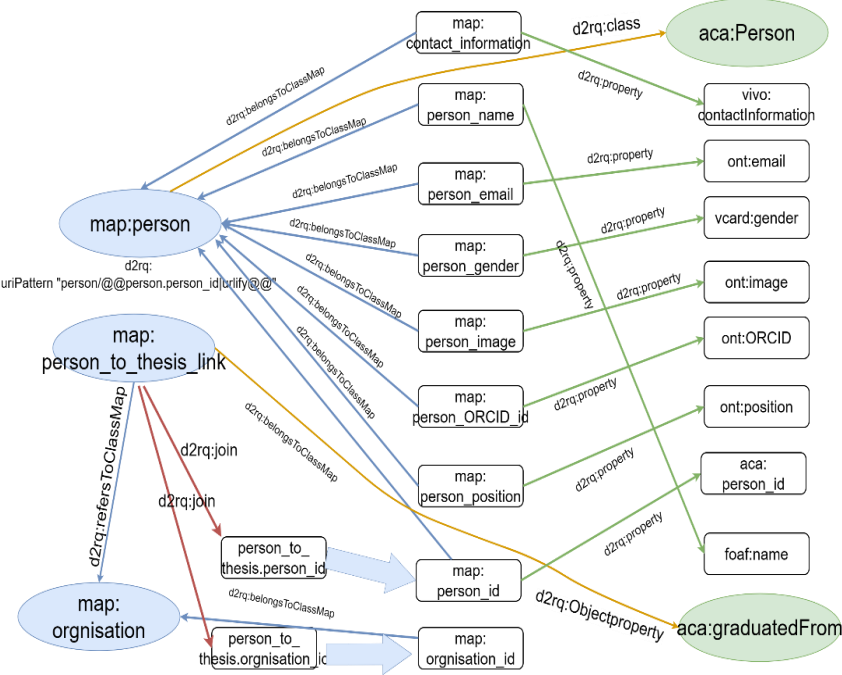


图 3 person 表与 person\_to\_thesis 表本体语义映射过程

表 2 谱系原始数据经 RDF 化示例

---

```

@prefix acatree: <http://www.acatree.cn#> .
acatree:thesis/T2_7_13-M acatree:citation_count "0" ;
    acatree:number_of_volumes "9" ;
    acatree:degree_of_thesis "硕士" ;
    acatree:publication_DateTime "2002/1/3" ;
    acatree:number_of_pages "" ;
    acatree:URI "" ;
    acatree:thesis_id "T2_7_13-M" ;
    acatree:language "中文" ;
    acatree:title "气动声学计算中振制短波的方法及非均匀网格上数值格格式的研究" ;
    acatree:DOI "" .

```

---

### 2.2.3 钱伟长学术谱系的实体关联

将数据关联到其他开放 RDF 数据集与词汇表是丰富关联数据语义的关键步骤<sup>[19]</sup>，本文选取通用关联数据集 DBpedia、CN-DBpedia 实现学术谱系实体与外部资源的关联。DBpedia 是一个跨多语言的知识库，它由维基百科信息中提取出来的结构化的内容构建而成。与 DBpedia 类似，CN-DBpedia 是由复旦大学知识工场实验室研发并维护的大规模通用领域结构化百科，其前身是复旦 GDM 中文知识图谱，是国内最早推出的也是目前最大规模的开放百科中文知识图谱，涵盖数千万实体和数亿级的关系<sup>[20]</sup>。

关联 DBpedia 与 CN-DBpedia 资源的具体方法不同。DBpedia 语言以英语为主，因此在关联时需要先通过维基百科将中文实体转换为英文实体，再通过 DBpedia 的 SPARQL Endpoint 使用查询语句关联对应的实体。结果中出现冗余信息（如上海大学站、上海大学悉尼工商学院）时，需要人工手动消歧。CN-DBpedia 支持中文检索，用户在检索页面输入实体中文名称出现结果后，将结果页面 URI 与 RDF 数据库中实体通过 owl:sameAs 或 rdfs:seeAlso 关联实现。实体关联阶段是钱伟长学术谱系关联数据集的最终构建步骤，本文将关联数据集以 N-Triples 的格式存储在开源软件 OpenLink Virtuoso 中。

### 2.2.4 关联数据集的发布

学术谱系关联数据的发布是将生成的富含链接的 RDF 文档发布到网络上实现公开或分享。由于学术谱系关联数据集有更新速度快且具有持续性的特点，一般关联数据发布工具很难满足数据迭代需求，故本文采用 D2RQ 的 d2r-server 工具发布钱伟长学术谱系关联数据集。d2r-server 有三个发布功能：一是提供关联数据集的 HTML 页面浏览服务，可以将实体链接指定数据以实现关联浏览目的；二是提供 RDF 视图，在支持语义网的浏览器中可以直接浏览 RDF 三元组关系图，实现原生数据的可视化；三是提供 SPARQL Endpoint 端口，用户可以使用 SPARQL 语句实现关联数据的查询服务。

d2r-server 需要从关系数据库经本体语义映射生成的映射文件中读取关联数据集，在本地建立 Web 界面提供数据集的浏览和使用服务。运行成功后，点击 Home 处的三个关系表



名称即可呈现出所属内容。使用 `d2r-server` 浏览钱伟长学术谱系关联数据集见图 4。示例实现了一个 `Thesis` 实体与其关联作者、关联出版方、关联实体类型的连接，以及部分论文元数据描述的展示。

property	hasValue	isValueOf
abstract	“湍流是自然界中最普遍存在的流动现象。它与人类生存、国防建设以及基础学科中的许多领域都有十分密切的关系。湍流研究的重要意义是不言而喻的。然而由于湍流的复杂性，虽然经过一百多年的研究，它至今还是经典物理领域中少数几个尚未解决的难题之一。本文从模式理论和拟序结构模型两方面着手对湍流问题进行了研究。首先，本文基于对大小涡的不同特性的考虑，将湍流脉动分成两个部分，即与大涡旋运动相关部分和小涡旋运动相关部分，并假定：一、小涡旋是准定常并接近均匀各向同性的，其相关的尺度由Taylor微尺度给出，并且其二阶、三阶关联项的表达式可以采用泊松方程用流式子；二、大涡旋是各向异性的，其尺度向很长，大涡旋雷诺应力不能用某一点上的局部物理量表示出来，大涡旋相关尺度由大涡旋动能和湍流耗散率表示。在引入雷诺平均以及小涡旋平均两种平均后导出了一个新的大小涡旋分考虑的双尺度湍流模式。该模式中，小涡旋应力方程采用 $q_{\tau}$ - $\epsilon$ 二方程封闭，而大涡旋应力方程采用完全的二阶矩封闭。”	-
academic_discipline	“固体力学”	-
citation_count	“1”	-
degree_of_thesis	“博士”	-
hasAuthor	db:person/G2%5F7%5F3	-
hasPublisher	db:organisation/10280	-
keywords	“输运；模式理论；湍流”	-
language	“中文”	-
number_of_volumes	“6”	-
place_publication	“上海大学力学与工程科学学院”	-
publication_DateTime	“1998/1/1”	-
thesis_id	“T2 7 3-D”	-
title	“双尺度湍流及标量输运模式理论和拟序结构的输运机理研究”	-
uri	https://d.wanfangdata.com.cn/thesis/ChiJuaGVzaNXOZdMjAyMzAwMTS1XhMDAwMDAyNz4MjMjSgghkn1z2hN3A30D%3D”	-
volume	“目录第一章前言1.1湍流研究简史1.2湍流模式及标量输运模式理论1.3湍流拟序结构研究及意义1.4本文的工作第二章大小涡旋分考虑的双尺度湍流理论2.1引言2.2大小涡旋分考虑时梯度的应力输运方程2.3小涡旋力输运方程的演化2.4大涡旋力输运方程的演化2.5关于湍能输运率的讨论2.6小结第三章湍流被标量输运的复数模型3.1引言3.2大小涡旋分考虑的精确标量输运方程的导出3.3小涡旋标量输运的显式代数模型3.4大涡旋输运方程的简化及扩散系数模型的建立3.5标量输运现象的物理解释3.6模型参数的确定及大涡旋输运模型的检验3.7小结第四章湍流拟序结构的模型及传热计算4.1引言4.2拟序结构流动模式和数学模型4.3基于湍流拟序结构模型的传热计算4.4计算结果讨论4.5小结第五章结论与展望主要参考文献致谢”	-
rdf_type	Thesis	-

图 4 d2r-server 关联数据集浏览页面

### 3 科学家学术谱系的应用展望

钱伟长学术谱系未来应用依托于对谱系数据的深度挖掘，不仅局限于对过往数据的整理和分析，更应关注如何借助现代技术手段，实现学术信息的动态关联与直观展示。可视化技术揭示学术网络内蕴藏的丰富知识和深层次联系，将有助于学术界对钱伟长及其学术影响的全面理解，也促进了学术信息的有效传播和利用。而且未来应用将推动学术谱系的探索不再是静态的历史回顾，而是转变为一个动态的、互动的、能够持续发展的学术资源网络，为学术领域带来新的视角和启发。

### 3.1 实现学术谱系领域知识关联及其可视化

对学术谱系关联数据的应用，本文并未采用复杂的数据分析与挖掘技巧，而是试图利用图技术将学术谱系关联数据可视化，欲从视觉上清晰地展示学术领域内的知识结构和关联关系，尝试弥补传统学术谱系研究中数据呈现单一、测度研究偏向的不足的同时，更为学术界和大众提供了一种清晰、直观的可视化途径，以增强学术谱系的互动性、趣味性，推动对学科传承和发展的深入理解。

本文采用 Neo4j 及其支持语义网和 RDF 的插件 neosemantics 实现钱伟长学术谱系的知识关联。Neo4j 是一种高性能的图数据库管理系统，它专门用于存储、查询和分析图结构的数据。关联数据是以 RDF 表示的语义数据，具有描述丰富、关系复杂的特点，neosemantics 插件使 Neo4j 这一图数据库能够更方便地存储、查询和分析 RDF 数据，为构建知识图谱和进行语义关联查询和可视化分析提供了便利。要使用 neosemantics，首先需要将插件添加到 Neo4j 数据库中，然后通过 Cypher 查询语言导入 RDF 数据。在导入过程中，neosemantics



会将 RDF 三元组映射到 Neo4j 的图形结构中，将主体、谓词和宾语分别映射为图数据库中的节点和关系。

图 5 是钱伟长学术谱系关联数据集的可视化成果。应用 Neo4j 部分为底层的节点关系关联展示：红色节点表示学位论文，黄色节点表示机构，绿色节点表示谱系成员，三者之间关系通过本体定义的关系连接，如“双尺度湍流及标量输运理论和拟序结构的输运机理研究”aca:has Author “卢志明”。为清晰地展示关系，笔者对图中部分节点做了隐藏处理。右侧的节点属性部分展示了学术谱系成员“卢志明”的相关公开信息，包括 ORCID，职位，邮箱等。

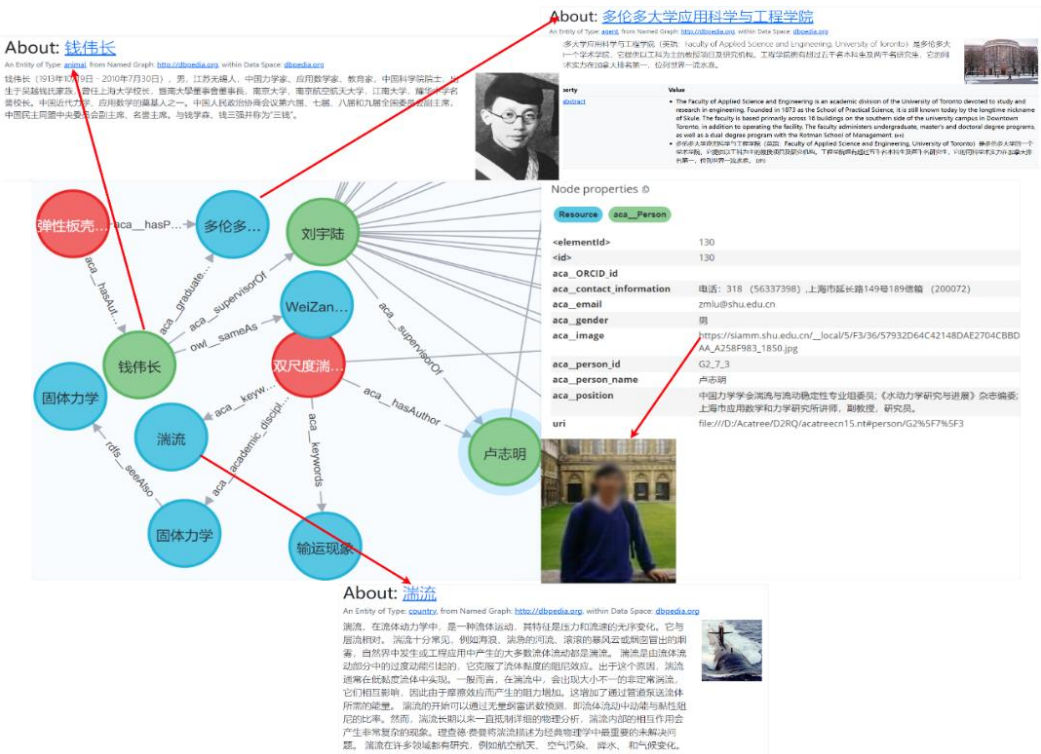


图 5 钱伟长学术谱系关联数据集的可视化

### 3.2 实现钱伟长学术谱系领域的知识发现

知识发现通常能够通过一些严谨的推理规则实现。本体推理是指利用本体模型中定义的概念（类）、实例（个体）和关系（属性），通过逻辑推理发现隐含的知识和关系的过程。它依赖于本体中定义的规则和逻辑，揭示数据之间未明显表达的联系，增强数据的语义理解和知识发现。如：“如果一个人是某篇论文的作者，并且这篇论文由某个机构出版，则可以推理出这个人与该机构存在学术关联”的本体推理的逻辑表达为“**hasAuthor(Thesis, Person)<sup>^</sup>publisherOf(Thesis, Institution) -> hasAcademicLink(Person, Institution)**”；再如推理学术谱系成员内部的相识（Knows）关系：“如果两个人（Person1 和 Person2）的论文（Thesis1 和 Thesis2）分别发表在不超过三年的时间差，并且这些论文都由同一个组织发表，同一位导师指导，那么可以推理出这两个人相识（Knows）”，根据本体推理规则构建的

Cypher 式为表 3，该关系确立实现了钱伟长学术谱系成员间小团体和关键人物的发现。

表 3 “Knows”推理的 Cypher 语句

<pre>MATCH   (a:aca__Thesis)-[:aca_hasAuthor]-&gt;(author1),   (b:Thesis)-[:aca_hasAuthor]-&gt;(author2),   (a)-[:aca_hasPublisher]-&gt;(publisher),   (b)-[:aca_hasPublisher]-&gt;(publisher) WHERE   a &lt;&gt; b AND   abs(apoc.date.parse(a.aca_publication_DateTime,'s', 'yyyy/MM/dd')   - apoc.date.parse(b.aca_publication_DateTime, 's', 'yyyy/MM/dd'))  &lt;= 94608000 WITH   a, b, author1, author2 WHERE   id(a) &lt; id(b) OPTIONAL MATCH   (author1)-[:aca__supervisorOf]-(author2) WHERE   author1 &lt;&gt; author2 CREATE   (author1)-[:Knows]-&gt;(author2) RETURN DISTINCT   author1.aca__person_name as author1_name,   author2.aca__person_name as author2_name;</pre>
---

4 结语

前事不忘，后事之师。科学研究本就是一个不断传承、迭代与更新的过程，一流的科学传统能够为在此传统下工作的科学家提供一流的学科问题域、研究思路和方法，也为之提供一流的科学观和价值观<sup>[21]</sup>。而科学传统延续的载体便是科学家的学术谱系。较于以往研究，本文创造性地提出并实证了一种基于关联数据的科学家学术谱系构建路径。未来，凭借丰富的语义信息和错综复杂的关系网络，学术谱系关联数据可能为学术推荐系统注入生命力，系统通过分析学者之间的合作关系和研究方向的相似性，可为用户提供个性化、针对性的学术建议，提高学术资源的利用效率。此外，关联数据的应用还可拓展到其他知识服务功能，如可以构建更为细致、全面的学术社交网络，为学术交流和合作提供更多可能性。另外，关联数据还能科研决策提供依据，更好地制定研究方向、科研政策和资源投入策略。

不过，本文构建的钱伟长学术谱系还存在数据的颗粒度划分不够细的问题，如学位论文的摘要、题目中所含的大量关键词没有与外部数据集匹配，在一定程度上缩小了学术谱系关联数据集的规模，需要在后续研究中利用 NLP、文献计量等方法做进一步处理与分析。

致谢：感谢我的本科生指导老师盛小平教授对本文的指导！

## 参考文献:

- [1]新华社.中共中央办公厅 国务院办公厅印发《关于进一步弘扬科学家精神加强作风和学风建设的意见》[J].中华人民共和国国务院公报,2019(18):20-24.
- [2]盛怡瑾,赵勇.科学家学术谱系的内涵、构建与测度研究述评[J].图书情报工作,2023,67(14):109-118.
- [3]周积明,雷平.清代浙东学派学术谱系的构建[J].学术月刊,2004,(06):40-46.
- [4]韩天琪,樊小龙,袁江洋.唐敖庆谱系与福井谦一谱系比较研究[J].科学与社会,2013,3(1):110-123.
- [5]程豪杰,袁勤俭.科学家学术谱系繁衍的影响因素研究——以情报学家严怡民为例[J].现代情报,2021,41(5):77-86.
- [6]刘俊婉,杨波,王菲菲,等.基于 LDA 主题模型的学术谱系内知识传承研究——以谈家桢为核心的遗传学学术谱系为例[J].图书情报工作,2018,62(10):76-84.
- [7]黄娟,唱婷婷,魏群义.基于科学学术谱系文献著者智库的阅读推广研究——以重庆大学图书馆为例[J].图书馆杂志,2018,37(08):43-48.
- [8]黄剑.论钱伟长大学体育教育思想[J].运动,2014,(3):80-81.
- [9]张丹华,陈海青.钱伟长爱国主义政治教育思想形成的内外因渊薮[J].社科纵横,2009,24(2):139-141.
- [10]嵇醒.应用数学和力学的结合-II——钱伟长的代表作[J].力学与实践,2014,36(5):670-672.
- [11]刘炜.关联数据:概念、技术及应用展望[J].大学图书馆学报,2011,29(2):5-12.
- [12]赵夷平.基于关联数据的机构知识库资源聚合与知识发现研究[D].长春: 吉林大学,2018.
- [13]牛永晏.学者科研学术记录数据集构建研究[D].南京: 东南大学,2019.
- [14]刘丹.面向关联数据的学位论文元数据语义化研究[J].大学图书馆学报,2015,33(5):74-82.
- [15]袁满,李明轩,张维罡,等.基于本体和关联数据的知识集成模型研究[J].吉林大学学报(信息科学版),2023,41(1):67-75.
- [16]中国科学技术协会.中国科学技术专家传略:工程技术编.力学卷[M].北京:中国科学技术出版社,1993.
- [17]李冰.尹双增.忆恩师钱伟长: 师恩如山终生难忘[EB/OL].(2010-10-21)[2024-6-1].<https://www.tsinghua.org.cn/info/1014/10119.htm>.
- [18]侯西龙,谈国新,庄文杰,唐铭.基于关联数据的非物质文化遗产知识管理研究[J].中国图书馆学报,2019,45(2):88-108.
- [19]Skevakis G,Makris KC,Kalokyri V,et al.Metadata management,interoperability and linked data publishing support for natural history museums[J].International Journal on Digital Libraries,2014,14(3):127-140.
- [20]Xu B,Xu Y,Liang J,et al.CN-DBpedia:A never-ending chinese knowledge extraction system[C]//International Conference on Industrial,Engineering and Other Applications of Applied Intelligent Systems.2017:428-438.
- [21]乌云其其格.日本诺贝尔物理学奖获奖谱系的反思[J].科技导报,2009,27(7):106-106.

## Construction of Scientists' Academic Genealogies Based on Linked Data——The Case of Chien Weizang's Academic Genealogy

Zhang Hao

(School of Cultural Heritage and Information Management, Shanghai University)

**(Abstract)** Constructing the academic genealogies of scientists is particularly important for commemorating the academic honors and achievements of scientists and their academic descendants, as well as preserving the social memory that belongs to the scientific community and the academic collective. This paper proposes a construction path for academic genealogies based on linked data and divides it into four levels; secondly, taking Chien Weizang's academic genealogy as

an example, this paper designs the ontology of the academic genealogy by organizing relevant information of Chien Weizang's academic genealogy. The construction of Chien Weizang's academic genealogy is completed through key steps such as entity RDFization, ontology semantic mapping, entity association, and linked data publication. Finally, the paper prospects the application from three perspectives: knowledge association, knowledge discovery, and dynamic updating. Constructing scientists' academic genealogies based on linked data can effectively avoid the drawbacks of previous constructions, enrich the knowledge content of academic genealogies, and has significant practical significance for understanding scientists' educational thoughts and academic inheritance.

**〔Keywords〕** Academic Genealogy   Linked Data   Ontology   Chien Weizang

**〔作者简介〕** 张昊（2002—），男，本科生，上海大学文化遗产与信息管理学院，oula\_@shu.edu.cn，研究方向：信息组织、档案著录；