

关联数据的可视化技术研究与实现^{*}

■ 陈涛¹ 夏翠娟² 刘炜² 张磊²

¹ 中国科学院上海生命科学信息中心 上海 200031 ² 上海图书馆 上海 200031

摘要: [目的/意义] 目前,关联数据领域的研究正从如何生产数据向如何消费数据的方向转变,可视化已成为该领域研究的新热点。通过关联数据可视化可以在有限的空间(如一张图)传达无限的信息,更直观地展示数据之间的内在联系,利于信息的发现。[方法/过程] 针对关联数据系统中实际产生的实例数据和本体结构这两类 RDF 数据,分别介绍一些适用于这两类数据形式的可视化技术和实现方法,并引用实际开发和研究中的成功案例进行阐述。[结果/结论] 本文所提到的可视化实现方法具有一定的通用性,可以为其他关联数据应用系统的可视化提供指导和借鉴。

关键词: 关联数据 SPARQL RDF 数据可视化 R

分类号: G203

DOI:10.13266/j.issn.0252-3116.2015.17.017

1 引言

“数据的真实价值就像漂浮在海洋中的冰山,第一眼只能看到冰山一角,而绝大部分则隐藏在表面之下”^[1],这句话以最直接、最朴实的方式道出了数据的重要性以及数据再利用的无穷潜力。分散的、孤立的、杂乱无章的数据很容易造成资源的极大浪费,使大量的数据资源得不到充分共享、利用,语义网与关联数据技术的出现很好地解决了这一难题。语义网(Semantic Web)是一种数据的网络(the Web of data),旨在让数据得以共享,而不仅仅是被应用程序束缚。关联数据是语义网的一种轻量级的实现方法,它以机器可读的方式实现不同数据源之间数据的关联与交互。

关联数据从 2007 年提出到现在,已逐渐成为数据发布的通用标准,得到了生命科学、图书情报、新闻媒体、社交网络以及政府数据等多个领域的强烈关注^[2],越来越多的机构及组织融入了开放数据发布的浪潮。据统计,截至 2014 年 8 月,开放的关联数据云(LOD cloud)中已包含 1 000 多个开放的数据集、超过数以百亿计的 RDF 三元组^[3]。面对如此庞大的数据资源,在积极寻求如何融入到开放数据发布浪潮的同时,更应该去思考怎样合理、有效地利用这些已有的开放数据。

目前,国内对关联数据的研究热度已从理论研究转向应用研究,相关文献有丁楠等的“基于关联数据的政府信息聚合研究”^[4]、张永娟等的“基于关联数据的专家系统库的构建”^[5]、仝召娟等的“基于关联数据的非遗数字资源聚合研究”^[6]等。但对关联数据的可视化技术研究相对较少,洪娜等的“关联数据中关系发现的可视化实践”^[7]是为数不多的研究之一。而国外,对关联数据的可视化研究则相对较多,如:J. Klímek 等对关联数据分析和可视化框架^[8]以及对 LOD 云图中的捷克数据集集中的政府数据和地理信息做了可视化分析^[9]。

可视化是数据描述的图形表示,旨在一目了然地揭示数据中的复杂信息。数据可视化技术主要是指借助于图形化手段,清晰有效地传达与传递信息。可视化技术本身并不是一个单独的概念,而是与信息图形、信息可视化、科学可视化以及统计图形密切相关。当前,在研究、教学和开发领域,数据可视化乃是一个极为活跃而又关键的方面。关联数据可以被看成是数据的一种新的表现方式,单一数据源数据的可视化方法对关联数据的可视化而言同样适用;但关联数据也有区别于一般数据的独特之处,如关联数据具有统一的

^{*} 本文系国家自然科学基金项目“W3C 的 RDB2RDF 标准规范在关联数据服务构建中的应用”(项目编号:13CTQ008)和国家社会科学基金项目“关联数据的理论和应用研究”(项目编号:11BTQ041)研究成果之一。

作者简介: 陈涛(ORCID:0000-0002-6609-4914),工程师,E-mail:chentao01@sibs.ac.cn;夏翠娟(ORCID:0000-0002-1859-6979),高级工程师;刘炜(ORCID:0000-0003-2663-7539),研究员;张磊(ORCID:0000-0002-6384-4783),高级工程师。

收稿日期:2015-07-25 **修回日期:**2015-08-14 **本文起止页码:**113-119 **本文责任编辑:**刘远颖

数据结构(RDF)和交换格式(N3、Turtle等),可以实现跨域的数据整合与分析,因此与一般数据的可视化方式也有所差异。此外,关联数据系统一般都需要构建在本体之上,本体主要用来描述概念与概念之间的关系。通过本体的使用,可以赋予关联数据系统一定的语义,实现智能的推理过程。本体数据也常用 RDF 数据结构进行表达,但由于本体是一种元数据,并非实际系统产生的实例数据,因此在其可视化方式上与实例数据的可视化有所不同。

本文首先简单地介绍了在关联数据系统构建中常用的一些存储及发布方法,随后以“上海图书馆家谱知识库”为例,具体阐述关联数据可视化技术在家谱知识库以及本体构建中的实现方法。总体来说,在关联数据应用领域,本文率先对其可视化做了研究与归纳,文中涉及的关联数据的可视化实现方法及工具,具有一定的通用性,可为其他应用提供参考。

2 关联数据的存储与发布

关联数据的主要特点是以机器可读的形式,实现不同数据源之间数据的开放与共享,这就要求不同的数据源之间以某种标准的数据格式进行存储与交互。RDF(Resource Description Framework,资源描述框架)已成为语义网与关联数据系统构建中的通用标准。因此,在进行关联数据的可视化之前需要对数据进行 RDF 结构化(结构化后的数据称为 RDF 数据),并进行存储与发布。数据要想转变为关联数据,RDF 结构化是首要的,也是最基本的步骤。不同的数据类型有不同的 RDF 结构化方法,如:Excel2RDF、XML2RDF、Marcmods2RDF、Bibtex2RDF、Any23 等。

2.1 关联数据存储方式

有了 RDF 数据后,需要对这些数据进行持久化存储,在实际系统构建中,主要有以下几种存储方式:

2.1.1 使用关系型数据库进行存储 通过 JenaSDB 或者 Sesame 可以实现 RDF 在关系型数据库中的存储,如 MySQL、Oracle、PostgreSQL 等,这种方式主要依赖于传统的数据存储方式,因此在使用中,应用较少。

2.1.2 使用 NoSQL 数据库进行存储 NoSQL,泛指非关系型的数据库,主要有键值对、列存储、文档型、图形(Graph)等几种^[10]。使用 NoSQL 数据库可以实现 RDF 数据的关系存储,但是一般的 NoSQL 数据库不支持 SPARQL 语法的查询,如 HBase、MongoDB、Neo4j 等,因此在实际系统构建中,同样应用较少。

2.1.3 使用 TripleStore 数据库进行存储 TripleStore

数据库是关联数据专门的存储方式,可以将其看成是图数据库的一种类型。相比 NoSQL 数据库,TripleStore 数据库支持 SPARQL 语法查询,不同数据库之间具有统一的数据格式,可以方便地实现数据的交互。

对于 TripleStore 数据库的选择,可以简单依据如下原则:①从商用与开源方面进行选择。商用的关联数据库主要有 AllegroGraph、OpenLink Virtuoso、OracleNoSQL 等;开源的则可以选用 Jena TDB、Sesame、BigData、4Store 等。②从数据量的大小进行选择。对于三元组数量达到千万级别以上的 RDF 数据,有可能的情况下建议使用商用数据库;对于数据量不大的应用系统,可以使用轻量级的开源的数据库。

2.2 关联数据发布方式

在关联数据发布之前,还可以对 RDF 数据进行一定的关联,例如可以采用基于模式的关联、基于属性的关联、基于相似度的关联(半自动)等方法^[11]。关联后的 RDF 数据需要进行发布后,才能实现数据的共享与重用,发布时可以依据关联数据的四原则进行^[12]。常用的关联数据发布方式有以下几种:

2.2.1 使用 SPARQL Endpoint(端点)进行发布 使用 SPARQL Endpoint 是关联数据应用中最常见的一种发布方式。对于关系型数据库中的数据,可以采用 D2R 或者 Virtuoso 数据库进行发布,这两种方式都带有 SPARQL 端点;对于没有 SPARQL 端点的情况,可以采用 Pubby 或者 SPARQLEditor 构建 SPARQL 端点。

2.2.2 使用 RDFa 方式进行发布 RDFa 的方式可以看成是在系统前端进行关联数据的发布。RDFa 是 W3C 的推荐标准,它扩充了 XHTML 的几个属性,将 RDF 的三元组嵌入在 XHTML 文档中,从而可以使符合标准的使用端从 RDFa 文件中提取出 RDF 三元组,例如 Apache Any23。

2.2.3 构建静态文件进行发布 直接使用 RDF 文件进行下载的方式也可以看成是数据发布的一种,对于小的 RDF 数据集或者已经归档的 RDF 三元组可以采用这种简单的方式。

3 关联数据实例的可视化

关联数据经过存储和发布后,可以根据不同的应用场景进行数据的可视化,这里以上海图书馆家谱知识库系统为例,介绍关联数据可视化的实现方式。上海图书馆是国内外收藏中国家谱(原件)数量最多的单位,这些家谱是中华文化的重要体现,是研究人文历史和地域文化的重要资源^[13]。这里的可视化主要以

家谱知识库中的实例数据为例。

3.1 实例数据的可视化实现框架

家谱知识库系统采用关联数据开源框架 (Apache Jena) 构建,系统产生的实例数据的可视化框架见图 1。家谱作品原始数据存放于 CSV 中,首先需要 对原始数据进行 RDF 结构化。转换时,可以采用 OpenRefine 工具,该工具主要用来管理杂乱数据,进行数据的整理与扩展,使用该工具可以将 CSV 文件导出为 RDF 格式的语义数据。转换后的 RDF 数据可以直接导入到 Open-linkVirtuoso 数据库中,该数据库自带了 SPARQL 端点,因此可以很容易地进行关联数据的发布。不过采用 SPARQL 的访问方式比较专业,对家谱中的规范数据 (人名、地名等),可以采用关联数据结构框架 (Jena) 进行重新发布,这里不再赘述。

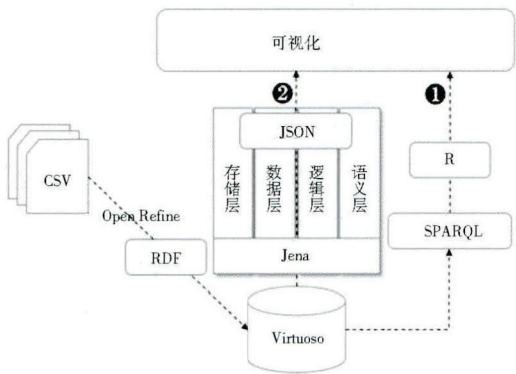


图 1 实例数据的可视化实现框架

从家谱知识库系统框架中可以看出,对于 RDF 数据的可视化流程,主要通过两种方式来实现:①直接通过 SPARQL 端点进行数据的获取及可视化 (结合 R 统计工具);②通过 JSON 数据格式进行可视化。

3.2 通过 SPARQL 端点进行数据的可视化

有很多可视化工具可以通过 SPARQL 方式进行 RDF 数据的可视化,如 RDF-spark、Sgvizler、Visual-Browser 等。使用这类可视化工具,不需要修改系统后台的程序代码,只需要在前端页面的 HTML 中引入相应的 JS 包,并在 HTML 中加入相关代码即可,不过可视化效果有限,而且很多需要依赖于 GoogleAPI。

在统计软件方面,常用的统计软件有 SPSS、SAS、STAT、R、S-PLUS 等。R 是一门用于统计计算和作图的语言,它不单是一门语言,更是一个数据计算于分析的环境,现已成为国内外众多研究人员和统计学者喜爱的数据分析工具^[7]。R 可以在诸多领域进行应用,比如数据挖掘、机器学习、社交网络、生物信息、金融数据分析等。关联数据具有数据的开放性 & 规范性,R 则

具有强大的数据分析能力,如将关联数据集和 R 统计结合起来,将会赋予关联数据系统更加强大的计算及统计分析能力^[14-15]。在家谱知识库的可视化中,使用统计工具 R 进行 RDF 数据的统计分析,这种统计分析可以完全独立于系统单独进行,也可以将 R 工具嵌入到应用系统中^[16]。图 2 为家谱本体简图,这里统计家谱知识库中各省市家谱作品的分布情况,家谱作品和谱籍地的三元组由属性 bf:place 连接,谱籍地中含有省市信息。对应的 R 脚本为:

```
endpoint <- 'http://localhost:8890/sparql'
options <- NULL
sparql_prefix <- "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX bf: <http://bibframe.org/vocab/>
PREFIX shl: <http://gen.library.sh.cn/vocab/>"
q <- paste(sparql_prefix,
"select ?prov (count(distinct ?s) as ?count)
where { graph <http://gen.library.sh.cn/graph/work> {
?s a bf:Work ; bf:place ?place.
select ?place ?prov where { graph <http://gen.library.sh.cn/graph/place> {
?place shl:province ?prov. |||
|| order by desc(?count) limit 10
}}
}"
)
res <- SPARQL(url = endpoint, query = q) $ results
... ..
```

其中 endpoint 是 Virtuoso 发布的 SPARQL 端点地址;sparql_prefix 中定义了查询中使用到的命名空间;q 部分为具体的 SPARQL 查询语句,该语句用来查询家谱在各省份的分布情况,并显示馆藏最多的前 10 个省份及其馆藏量。查询时,整合了家谱作品库 (http://gen.library.sh.cn/graph/work) 和谱籍地分布库 (http://gen.library.sh.cn/graph/place) 两个图 (Graph) 共同查询。SPARQL 查询后的结果可以用直方图、饼图、折线图、泡泡图等形式展示,图 3 为其饼图效果。同样还可以对数据进行深层次的分析,例如可以对各省份进行二次分析,分析每个省份家谱所属姓氏的分布情况。

结合 R 统计工具的可视化方式,常用于对数据进行统计分析,如可用于对机构或者研究者发文量的统计。另外,SPARQL 可以实现不同数据源的联邦查询,因此采用这种方式的可视化,还可以对开放的关联数据进行聚合分析。

3.3 通过 JSON 数据格式进行可视化

JSON 作为一种轻量级的数据交换格式,非常适合于服务器与 JavaScript 的交互。因此,只要将 RDF 数

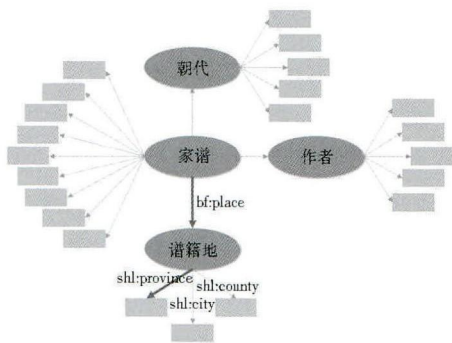


图 2 家谱本体

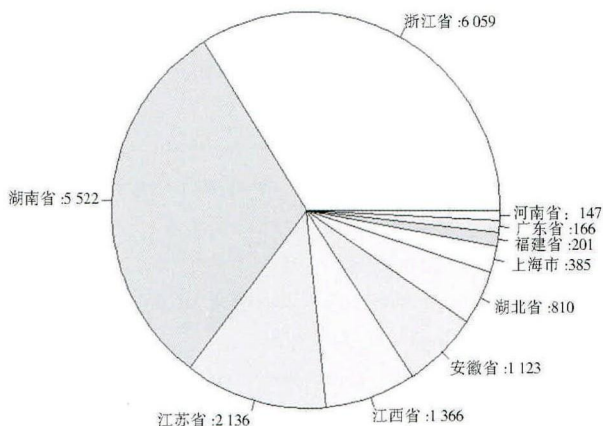


图 3 家谱归属地分布

据进行 JSON 序列化后,就可以采用基于 JavaScript 的可视化工具进行数据的可视化。目前有很多这样的可视化工具,如 D3、D3 Plus、Highcharts、ECharts 等都可以进行丰富的可视化展示。

RDF 数据也可以被看成是一种特殊的数据,以三元组的形式进行数据的组织,只需将 RDF 数据转换成标准的 JSON 数据格式,就可以使用大量的可视化工具进行展示。家谱知识库基于 ApacheJena 框架构建,Jena 模型中已支持对 RDF 数据的 JSON 格式输出,这给数据的可视化提供了更多的选择。这里以家谱中的谱籍地数据为例阐述 RDF 数据的空间位置可视化的实现方法。

OpenLink Virtuoso 数据库已支持多个空间地理数据类型,如点、线、多边形以及他们之间的组合情况。这里以多边形为例说明 Map 可视化在家谱知识库中的应用实例,主要用到多点 (multipoint) 和多边形 (polygon) 这两类的数据类型^[17] (见图 4),多点的情况用于在地图中查询出家谱作品的空间地理位置,并用符号 (如:小旗子) 进行标识;多边形可以用来在地图上进行交互式操作。例如,通过选择一定的区域范围 (以江苏省为例),来查看该区域内的家谱作品分布情况,对

应的 SPARQL 语句为:

```
SELECT DISTINCT ?place ?label ?long ?lat FROM <http://gen.library.sh.cn/graph/place> WHERE {
    ?place owl:sameAs ?as ; bf:label ?label.
    {SELECT ?as ?long ?lat FROM <http://www.cba.ac.cn/graph/geography> WHERE {
        ?as geo:geometry ?geo ; geo:long ?long ; geo:lat ?lat.
        FILTER (bif:st_within (bif:st_point (?long, ?lat), bif:st_geomfromtext (' POLYGON (( 121. 87759232213814 31. 824315567090196, 120. 47134232213816 30. 999221249295417, 119. 15298294713816 31. 300094548720026, 118. 18618607213816 32. 04807720251023, 119. 06509232213816 33. 01137858723608, 117. 04360794713816 34. 327965771467674, 118. 80142044713814 34. 90660823717434, 120. 64712357213814 33. 81835628931579, 121. 87759232213814 31. 824315567090196))'))))}
    {SELECT ?place FROM <http://gen.library.sh.cn/graph/work> WHERE {
        ?work bf:place ?place ; dc:title ?title.
        FILTER CONTAINS(?title, '陈') }}
```

Multipoint	MULTIPOINT(3 7, 4 2, 8 6)
Polygon	POLYGON((1 2, 6 1, 9 3, 8 5, 3 6, 1 2))

图 4 家谱知识库中的空间地理数据类型

这里涉及到 3 个视图 (Graph) 的联合查询:

- (1) 谱籍地 <http://gen.library.sh.cn/graph/place>
- (2) 地图库 <http://www.cba.ac.cn/graph/geography>
- (3) 家谱作品 <http://gen.library.sh.cn/graph/work>

查询时,首先从家谱作品库中查询标题中含有“陈”姓的家谱作品的谱籍地 (?work bf:place?place);其次从地图库中查询地图中所圈范围内的地域资源 (?as geo:geometry ?geo)。这里用到了 Virtuoso 数据库的内置空间函数 bif:st_within (in g1 any, in g2 any, in prec double precision),该函数表示查询的所有空间点 g1 都在 g2 范围内;其中 POLYGON 为页面中所选择的多边形区域,也就是 st_within 函数中的 g2,而 g1 为地图仓库中的地点资源。最后在谱籍地库中选择所选多边形区域中含有家谱作品的地方,这里将谱籍地和地图库中的地点资源做了对应 (?place owl:sameAs ?as)。

查询出来的结果见表 1(这里仅显示 5 条记录),查询结果最终将以 JSON 格式返回到前端 Web 页面,JSON 数据片段为:

```
[{
  "label": "泰兴",
  "long": "102.02",
  "lat": "32.17",
  "place": "http://gen.library.sh.cn/Place/1021"
}, {
  "label": "常州",
  "long": "119.95",
  "lat": "31.78",
  "place": "http://gen.library.sh.cn/Place/244"
}, ... {
  "label": "常熟",
  "long": "120.75",
  "lat": "31.65",
  "place": "http://gen.library.sh.cn/Place/183"
}]
```

表 1 SPARQL 语句查询结果

place(资源)	label(标签)	long(经度)	lat(纬度)
http://gen.library.sh.cn/Place/1021	泰兴	102.02	32.17
http://gen.library.sh.cn/Place/244	常州	119.95	31.78
http://gen.library.sh.cn/Place/157	无锡	120.3	31.57
http://gen.library.sh.cn/Place/182	江阴	120.27	31.9
http://gen.library.sh.cn/Place/183	常熟	120.75	31.65

有了 JSON 数据,就可以调用多种可视化工具(JavaScript 类库包)进行数据的可视化。显示时,将经纬度在地图中相应的地点做上标记即可(见图 5),这里地图可视化采用的 OpenLayers。OpenLayers 是一个专为 WebGIS 客户端开发提供的 JavaScript 类库包,用于实现标准格式发布的地图数据访问。



图 5 江苏省陈氏家谱地域分布图

获取了地域的经纬度后,可以结合时间线,设计出

更加丰富的时空图,如图 6 为 1870-1880 年间家谱作品的空间分布图,拖动时间轴时,地图会动态显示家谱作品随时间变化的分布情况,这里使用的是 timemap.js 类库包。

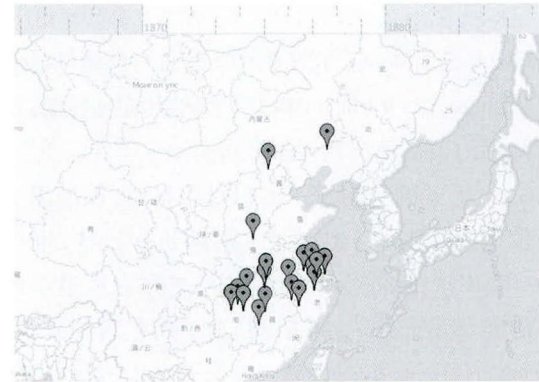


图 6 1870-1880 年间家谱作品分布

4 本体词表结构的可视化

本体主要用来描述实例数据的体系结构,它是一种元数据,以词表的形式定义知识的组织结构,虽然也是 RDF 数据,但是不适用于上述实例数据的可视化。对于本体的存储同样可以使用关联数据数据库,本体的编辑,一般在线下通过第三方的本体构建工具进行本体的设计,设计好的文件可以保存为 OWL、RDF/XML 或其他格式,这里称为本体源文件。对本体的存储及可视化框图如图 7 所示:

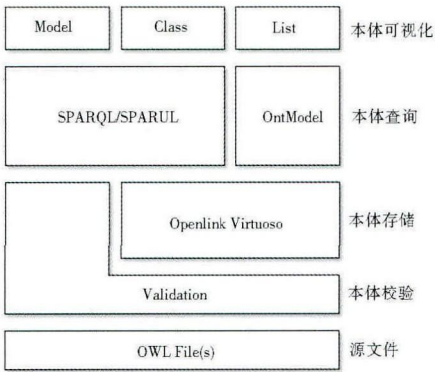


图 7 本体源文件可视化

具体的本体可视化流程为:

- (1) 本体校验。结合 W3C 的 Validation 工具对本体文件进行语法校验。
- (2) 本体存储。校验后的本体文件,经解析后存储到 Virtuoso 数据库。
- (3) 本体查询。为了更有效地进行本体的可视化,对于本体类及属性的查询采用了 SPARQL 查询和 Jena 本体模型(OntModel)操作相结合的方式;除了查

询外,这里还支持 SPARUL 对本体中的 RDF 进行修改。

(4) 本体可视化。对于本体的可视化,主要有 List 列表展示、Class 树形展示和 Model 模型展示 3 种方式。

List 列表形式比较简单,可以借助于 SPARQL 对 OWLGraph 进行查询,分别显示所有的类及属性,如 select ?s where { ?s a rdfs:Class } 将查询出所有的类,查询出的结果用表格的形式显示即可。

Class 形式的结构展示也相对容易,可以采用 Jena 本体模型 (OntModel) 对类及其属性进行操作,也可以借助 SPARQL 查询完成。图 8 中,左侧为家谱本体中所有的类及其子类,右侧为具体的类所包含的属性,如 sh1:Person 类用来描述家谱中的先祖名人,包含的属性主要有 rel:childOf (父)、sh1:birthday (生于)等。

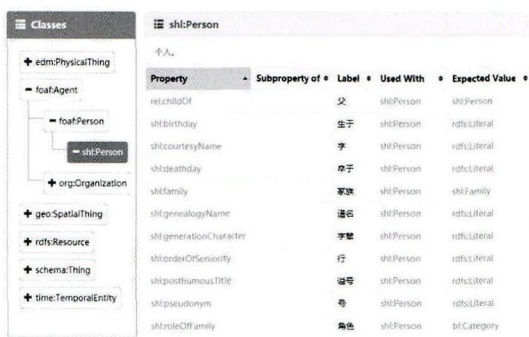


图 8 家谱本体 Class 形式

以上两种方式的显示效果比较简单,对于本体可视化,同样可以借助第三方工具进行,如 RDF Gravity、Welkin;当然如果能将第三方的工具嵌入到系统中,将会使应用系统具有更佳的灵活性和扩展性。HP RDF 视图可视化 (HP RDF Graph Visualization) 就是这样一种轻量级的工具,可以很方便地集成多种关联数据应用系统。家谱知识库系统中对于模型 (Model) 的可视化浏览采用的就是 HP RDF 可视化工具。

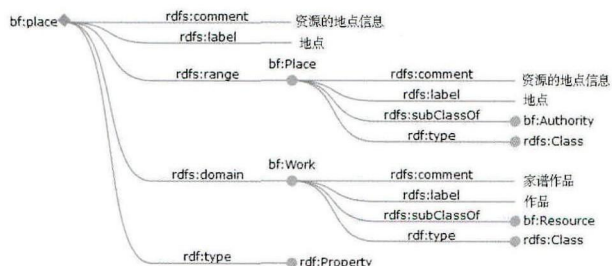


图 9 地点属性 bf:place 的 Model 结构可视化

图 9 为属性 bf:place 的可视化结果,菱形表示属性 (Attribute),圆表示资源 (Resource);可以看出,bf:place 表示资源的地点信息,它的类型是属性 (rdf:

Property),应用领域 (rdfs:domain) 是家谱作品类 bf:Work,取值范围 (rdfs:range) 是地点类 bf:Place。采用这种可视化方式将更加直观地展示本体中概念之间的层级关系,可视化效果最佳,例如,从图 9 中可以看出 bf:Place 与 bf:Work 以及 bf:Authority 之间的关系,即 bf:Place 是 bf:place 的取值范围,bf:Work 是 bf:place 的应用领域,bf:Authority 是 bf:Place 的父类。

5 结语

数据可视化技术主要有以下几个特点:①交互性。用户可以方便地以交互的方式管理和开发数据。②多维性。用户可以灵活地查看对象或事件的数据的多个属性或变量。③可视性。结合图形、时间轴、空间地理等多种形式反映数据的内在联系。针对关联数据的可视化,本文从实际应用出发,介绍了针对实例数据和本体结构这两类 RDF 数据的可视化技术和具体的实现方法。采用 SPARQL 端点的数据可视化方法,充分利用关联数据开放及互联的特点,完全将 RDF 可视化与应用系分离,仅依赖 SPARQL 查询,这就需要对 SPARQL 语法有深入的了解。JSON 数据格式的可视化方法则完全脱离 SPARQL 语法,应用现有的编程技术即可实现 RDF 数据的可视化,并且可视化的工具种类多,可视化效果丰富,缺点是不能进行跨域数据的联合分析。而本体结构的可视化则相对简单,可视化的效果也没有实例数据的可视化丰富和多彩。

当然,除了文中涉及到的一些可视化工具外,还有很多其他优秀的可视化工具可供选择,如 Simile Exhibit、Gephi、Cytoscape、Map4rdf 等。在实际应用中,可依据工具的可视化程度、易用性、可扩展性以及数据的要求等多方面进行选择,有时候往往需要将多个可视化工具相结合,共同进行数据的可视化。

参考文献:

- [1] 舍恩伯格. 大数据时代 [M]. 周涛,译. 杭州:浙江人民出版社, 2012, 74-80.
- [2] 陈涛, 张永娟, 陈恒. Web 数据到 RDF 数据的框架实现 [J]. 现代图书情报技术, 2015(2):1-6.
- [3] State of the LOD Cloud 2014 [EB/OL]. [2015-08-10]. <http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/>.
- [4] 丁楠, 王钰, 潘有能. 基于关联数据的政府信息聚合研究 [J]. 情报理论与实践, 2015,38(7):76-79,85.
- [5] 张永娟, 孙晓丽, 管兴华, 等. 基于关联数据的专家系统库的构建 [J]. 图书馆杂志, 2015,34(5):80-84.
- [6] 全召娟, 许鑫, 钱佳轶. 基于关联数据的非遗数字资源聚合研究 [J]. 图书情报工作, 2014,58(21):21-26.

[7] 洪娜,钱庆,范炜,等. 关联数据中关系发现的可视化实践[J]. 现代图书情报技术,2013(2):11-17.

[8] Křímek J, Helmich J, Necaský M. Payola: Collaborative linked data analysis and visualization framework [C]//The Semantic Web: ESWC 2013 Satellite Events. Montpellier: Springer Verlag, 2013:147-151.

[9] Křímek J, Helmich J, Necaský M. Application of the linked data visualization model on real World data from the Czech LOD Cloud [C/OL]. Linked Data on the Web (LDOW2014). [2015-08-10]. http://events.linkedata.org/ldow2014/slides/ldow2014_slides_13.pdf.

[10] 王林彬,黎建辉,沈志宏. 基于 NoSQL 的 RDF 数据存储与查询技术综述[J]. 计算机应用研究,2015,32(5):1281-1286.

[11] 沈志宏,黎建辉,张晓林. 关联数据互联技术研究综述:应用、方法与框架[J]. 图书情报工作,2013,57(14):125-133.

[12] How to publish linked data on the Web[EB/OL]. [2015-08-10]. <http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/>.

[13] 夏翠娟,刘炜,张磊,等. 基于书目框架(BIBFRAME)的家谱本体设计[J]. 图书馆论坛,2014(11):5-19.

[14] R 语言的优劣势是什么? [EB/OL]. [2015-08-10]. <http://www.itongji.cn/article/042144392015.html>.

[15] SPARQL for R tutorial-linked open piracy [EB/OL]. [2015-08-10]. http://semanticweb.cs.vu.nl/R/sparql_lop/sparql_lop.html.

[16] SPARQL with R in less than 5 minutes [EB/OL]. [2015-08-10]. <http://www.r-bloggers.com/sparql-with-r-in-less-than-5-minutes/>.

[17] Virtuoso Geo spatial enhancements [EB/OL]. [2015-08-10]. <http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VirtGeoSPARQLEnhancementDocs>

作者贡献说明:
陈涛:可视化技术实现、论文撰写及最终版本修订;
夏翠娟:系统本体设计;
刘炜:研究思路框架设计;
张磊:文献调研与方案验证。

Research and Implementation of Visualization Technology for Linked Data

Chen Tao¹ Xia Cuijuan² Liu Wei² Zhang Lei²

¹ Shanghai Information Center for Life Sciences, Chinese Academy of Sciences, Shanghai 200031

² Shanghai Library, Shanghai 200031

Abstract: [Purpose/significance] At present, the research direction has changed from “how to produce it” to “how to consume it” in linked data field. And the data visualization has become a hot topic in this area, through which, people not only can gain infinite information in a limited space, such as using a simple picture, but also understand the inner relationships among different data more intuitively and discover others. [Method/process] In this article, some visualization technologies and tools are introduced which can be utilized in instance data generated in daily useand ontology structure. Thepaper also shows some cases of work done in application development and research process. [Result/conclusion] In conclusion, the method and process of visualization mentioned in this article is a good guidance and reference in linked data visualization, which can be applied to other linked data application systems.

Keywords: linked data SPARQL RDF data visualization R

《图书情报工作》2015 年增刊(2) 征稿启事

为了给图书情报工作者提供更多的学术交流机会,使更多作者的优秀科研成果得以发表,《图书情报工作》杂志社定于 2015 年下半年出版《图书情报工作》增刊(2)。内容涉及基础理论研究、信息资源管理、信息服务、信息技术与人才培养等。

征文要求:

1. 主题明确,数据可靠,文字通顺,且一稿专投(即未在他刊上发表);
2. 请登录本刊网站 www.lis.ac.cn 在线投稿(投稿请注明“2015 年增刊(2)”字样),并留下详细联系方式;
3. 如稿件在 30 天内未收到录用通知,稿件即可自行处理;
4. 投稿前请按照本刊要求自行检查中文标题、作者姓名、单位及职称、中文摘要、关键词、分类号等要求项是否齐全,并请按照本刊体例格式著录参考文献。

截止日期:2015 年 10 月 20 日 联系电话:010-82623933 010-82626611-6638

联系人:赵 芳 E-mail:tsqbgz@vip.163.com