

# 基于 Wikidata 的机构类目范畴树构建与优化\*

贾君枝 叶壮壮

**摘要** 本文以 Wikidata 知识库为数据来源,抽取其中的机构数据,利用各机构的 subclass of 和 instance of 属性构建机构类目范畴树,以定量分析方法揭示其在类目、实例以及深度上的数量特征,并发现机构类目范畴树存在的三个问题,即机构类实体与其他实体类型重叠、机构类目层级混乱以及类目不平衡。在此基础上,提出对机构类目范畴树的优化方法,以设计适用性较强的机构类目体系框架与结构,为解决机构名称层级混乱以及知识共享等问题寻求一条行之有效的途径。图 5。表 5。参考文献 12。

**关键词** Wikidata 机构范畴 信息组织

---

## Construction and Optimization of Organizational Category Tree Based on Wikidata

Jia Junzhi Ye Zhuangzhuang

**Abstract:** Based on the organizational data extracted from Wikidata, we construct an organizational category tree by two property subclass-of and instance-of of organizations, then reveals its quantitative characteristics in the classification, instance and depth by means of quantitative analysis, and finds out such three issues as an overlap of organizations with other entity types, a rather convoluted scheme of subclasses and an imbalance of categories. Furthermore, this paper puts forward the optimizing methods of the existing categories, so as to design a classification framework and structure of organization with strong applicability. Thus an effective way to solve the problem of inconsistencies in the hierarchy and promote knowledge sharing can be found. 5 figs. 5 tabs. 12 refs.

**Keywords:** Wikidata; Organizational Category; Information Organization

---

### 1 引言

机构泛指高等院校、科研院所、公司企业、政府机关等实体对象<sup>[1]</sup>。该类实体对象数量庞大、层级关系复杂,且名称变化形式多异、衍化多样,一方面给以机构名称为联接点的信息检索、统计分析、计量评价等活动造成严重困扰,同时也影响检索效率以及统计分析、计量评价的可信度;另一方面,机构名称著录混乱加重了科研成果归属问题,进而降低了对机构进行绩效管理 with 评价结果的准确度<sup>[2]</sup>。因此,机构名称的规范化是当前亟待研究的课题。对机构的类型、属性、特

征等进行系统的研究不但能为机构的准确辨识提供依据,而且能为机构规范文档的应用奠定基础。若要对机构的规范化管理以及有效控制,势必要按照一定的分类依据构建机构名称范畴树,从一定程度上有效地区分各种不同名称形式。

Wikidata 因其开放协作、多语言以及结构化数据存储等特点,受到科研人员的广泛重视。其数据由用户手动输入或通过机器人从维基百科和其他维基媒体姐妹项目中提取,以实现维基百科的自动化更新<sup>[3]</sup>。Wikidata 的内容以免费许可

---

\* 本文系国家自然科学基金重点项目“基于关联数据的中文名称规范档语义描述及数据聚合研究”(项目编号:15ATQ004)研究成果之一。

证提供,因此可以被消费并链接到数据网络上的其他数据集。它扩大了任何一种单一语言版本的维基百科的覆盖范围,使用户不再受制于语言版本或者资源稀少的困惑。结构化数据使得数据可以在不同的应用程序、企业和社区之间共享和重用<sup>[4]</sup>。截至2017年2月7日,Wikidata 条目已达25,700,304个<sup>[5]</sup>,其中机构名称总计1,080,833个,占总条目数的4.3%,可见,Wikidata 中的机构数目已经具有一定规模。

针对当前机构范畴树的缺乏,本文旨在以 Wikidata 为机构数据的来源,按照一定的分类原则对其进行构建,基于代表机构实体的条目及相应的类型属性构建一个通用的机构范畴树。通过构建机构层级体系的方式,将机构实体进行有序集中,对其类型属性进行系统描述,对不同机构实体之间的关联关系进行清晰地明确揭示,以应对机构名称混乱、机构间关系模糊、机构信息资源无序繁琐等问题,从而提升信息检索查准率和查全率,提高知识组织和服务水平。

## 2 机构类目范畴树的构建与分析

Wikidata 是一个开放、协作编辑的知识库,与维基百科直接相关,旨在作为维基百科所有语言版本的中央数据库。新数据由用户手动输入或通过机器人从维基百科和其他来源提取并输入到 Wikidata 知识库中。

Wikidata 的数据模型主要由两个部分组成,即条目和属性。存储库主要包含条目和语句。条目用于“表示人类知识中的所有东西,包括主题、

概念和对象”,并且给出唯一的标识符、标签和描述。语句用于“记录关于条目的数据”和“由至少一个属性-值对组成”。与其他语义技术相反,条目既表示实体又表示类,每个条目被分配一个与语言无关的唯一标识符,并且由以字母为 Q 的 ID 开头。同时,属性以字母 P 开头的 ID 表示,除了具有与条目相同的多语言概念之外,该属性与 RDF 属性紧密对应,即它们将条目链接到其他条目或值<sup>[6]</sup>。

Wikidata 以条目为中心创建了类的层次结构,类成员由属性 subclass of 以及 instance of 进行连接<sup>[7]</sup>,这样构建好的类目体系将条目类及其属性组成了立体网络,既实现了对条目的识别区分,又对条目之间的关系进行了明确揭示,便于机器实现语义层面理解。

### 2.1 机构类目范畴树的生成

Wikidata 中的机构名以“subclass of”和“instance of”两个属性与顶级类“organization”相关联。SPARQL 查询是获取 Wikidata 数据的主流方式,本研究以维基数据查询服务测试版为工具,对机构类(organization)进行逐级类目的抽取<sup>[8]</sup>。表1所示为抽取 Wikidata 中机构的二级类目的 SPARQL 语句,其余层级类目的抽取以此类推。表2所示为抽取 Wikidata 中机构的直接和间接子类的 SPARQL 语句。表3所示为抽取机构实例数量的 SPARQL 语句。将这些抽取的条目按照各自的“subclass of”属性组织成树形结构,图1所示为机构类目范畴树的部分类目。

表1 Wikidata 中机构二级类目的 SPARQL 查询语句

```
SELECT DISTINCT ? s ? sLabel WHERE {
  ? s wdt:P279 wd:Q43229.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
} ORDER BY ? sLabel
```

表 2 Wikidata 中机构直接和间接子类的 SPARQL 查询语句

```

SELECT ? item ? itemLabel ? Size ? e ? eLabel {
  # number of indirect subclass
  SELECT ? item (count (distinct ? e) as ? size) {
    ? item wdt:P279 * wd:Q43229
  }
  OPTIONAL { ? e wdt:P279 ? item }
  } GROUP BY ? item
  }
  OPTIONAL { ? e wdt:P279 ? item }
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
  }
    
```

表 3 机构类目范畴树各类目的实例数量的 SPARQL 查询语句

```

SELECT ? item ? itemLabel ? size {
  # number of instances
  SELECT ? item (count(distinct ? element) as ? size) {
    ? item wdt:P279 * wd:Q43229
  }
  OPTIONAL { ? element wdt:P31 ? item }
  } GROUP BY ? item
  }
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
  }
    
```

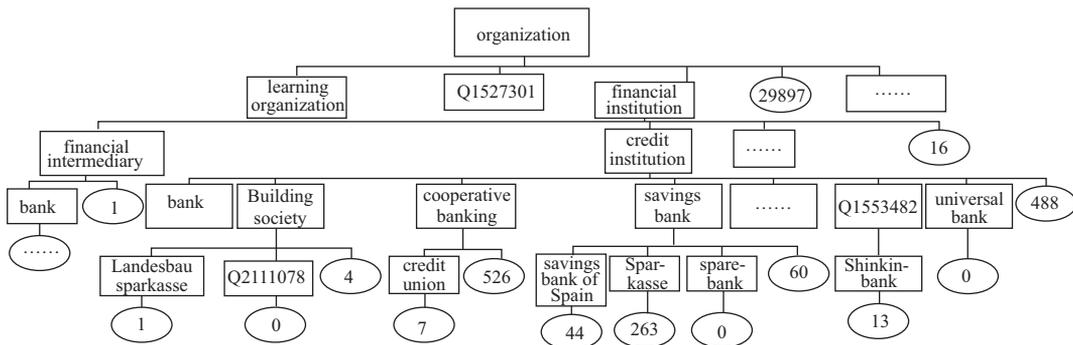


图 1 机构类目范畴树类目(部分)

注: □表示类, ○表示实例, ○中的数字表示实例数量

## 2.2 机构类目范畴树的特征揭示

### 2.2.1 机构实体类目分布

采用 SPARQL 语句对处于机构类目范畴树各个层级的子类 and 实例进行抽取,并统计相应的子类和实例数目。为方便描述,做以下定义,将机构类目范畴树中 organization 的二级类称为“机构的直接子类”,机构类目范畴树中层数大于 2 的所有条目称为“机构的间接子类”。同理,organi-

zation 的 instance of 属性值称为“机构的直接实例”,机构类目范畴树中除 organization 之外其余各层类目的 instance of 属性值称为“机构的间接实例”。截至 2017 年 2 月 7 日,机构的直接子类数为 210 个,间接子类 4720 个,直接实例 29,897 个,间接实例为 1,046,006 个,共计 1,080,833 个条目。机构类目范畴树中平均每个子类有 203 个实例。Wikidata 实例丰富,较为全面。以点到面

的形式全面扩展,几乎包含所有机构,如高等院校、政府、企业等。而 Dbpedia 这一与 Wikidata 一样以维基百科为数据来源的知识库中有 275,000 个机构实体,包括 67,000 个公司、53,000 个教育机构 and 155,000 个其他机构实体<sup>[9]</sup>。相较而言, Wikidata 的机构实体数量已具有一定的规模。在组成机构类目范畴树的 4930 个条目中,有 910 个条目仅有标识符,而没有相应的标签及声明。其中,210 个二级类中有 41 个这样的条目,机构类目范畴树的其余层级类目中,这样的无标签条目有 869 个。对于本研究而言,这部分数据没有实际意义,故剔除。

### 2.2.2 机构类目范畴树的实例分布

分类应遵循现实性原则,即类目所代表的事物客观存在,同时必须有一定数量的符合类目特征的数据集,类的实例集合一定程度上可以检验

类目的应用程度,可以以实例的数量为指标衡量某类目的客观性。根据机构的实例总量和机构的子类总量可以得出,机构类目范畴树中平均每个子类有 203 个实例。实际上,机构类目范畴树中各类目的实例数量分布很不均匀:一,以 instance of 属性与机构类目范畴树的顶级类 organization 相连的实例数量高达 29,897 个;二,包含机构各级子类的 4020 个有效条目中,实例数量为 0 的条目有 1181 个,占到总条目数的 29.4%,其中 1063 个无实例的条目处于机构数的最深一层;三,这些实例分布在类目的各个层级,其中实例分布于最深一层的条目有 2001 个,占到 49.8%;四,实体的实例分布明显不均,高达 1000 个实例数的条目有 134 个,如 business enterprise 的实例数已达到 110,394 个,具体如表 4 所示。

表 4 机构类目范畴树中拥有不同实例数量  $i$  的条目的分布情况

实例数范围	$i=0$	$0<i\leq 10$	$10<i\leq 100$	$100<i\leq 1000$	$1000<i\leq 10000$	$i>10000$
相应条目数	1181	1474	865	366	111	23

### 2.2.3 机构类目范畴树的动态更新

Wikidata 的一个独特的特征就是其数据处于持续更新中<sup>[10]</sup>。随着机构实体本身以及所处环境的不断变化,机构数据不断变动,因此需对机构类目范畴树进行必要的更新与管理维护,这也为后续应用与服务质量提供重要保障<sup>[11]</sup>。而基于开放知识库获取机构实体以及关系,一定程度上保证了数据的准确性与修订及时性,同样,利用 Wikidata 丰富的数据来源,构建机构类目范畴树的可持续建设方案变得切实可行。

## 2.3 机构类目范畴树存在的问题

基于 Wikidata 构建的机构类目范畴树并不是完美无瑕的。通过对机构的二级类进行定量和定性相结合的分析,可以发现机构类目范畴树存在的一些问题。

### 2.3.1 机构各子树的深度参差不齐,类目处于不平衡状态

借用数据结构中“树的深度”概念,定义一棵树的根结点层次为 1,其他节点的层次是其父结

点层次则加 1。一棵树中所有结点的层次的最大值称为“这棵树的深度”<sup>[12]</sup>。为了便于分析,这里将 organization 类的直接子类看作树的根节点,对 169 棵机构子树的深度进行分析。各机构子树的深度分布如图 2 所示。机构的直接子类,即机构类目范畴树的基本大类,是对机构范畴进行得最概括、最本质的划分。而各直接子类的深度这一变量的分布情况可以较好地反映机构类目范畴树基本大类划分的合理性。

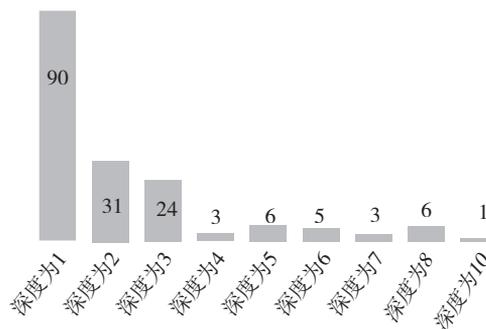


图 2 机构 169 棵子树的深度分布

由图 2 可知,机构子类树的深度范围从 1 到 10 不等,深度为 1 的子树最多有 90 棵,占到总数的 53.25%,86%的子类的深度是 1—3 级。深度达到 10 的子树只有 1 棵,深度为 1 或 2 的子树的数量是深度大于 2 的子树的数量的 2.4 倍,反映了机构类目范畴树的类目处于高度不平衡的状态。

### 2.3.2 机构类目的多从属问题

Wikidata 中的每个机构实体由以 Q 为前缀的 ID 唯一标识,实体之间通过属性相连接。实例通过 instance of 属性与对应的实体连接,实例与实体之间存在多对一的关系。因此,我们考虑机构类目范畴树中存在这样的情况:通过 subclass of 连接的实体的上位类与实体之间也存在多对一的情况,即一个条目的多从属特征。为验证该想

法,我们以机构类目范畴树中作为二级类的 169 个条目为例,用表 5 所示的 SPARQL 语句来获取 169 个条目的上位类数目和相应的上位类标签以及标识符。这里所指的是除 organization 之外的上位类,将条目的除 organization 之外的上位类称为“额外父类”。经过检测,在 169 个条目中,有 38 个条目存在多上位类的情况,占总数的 22.5%。对这 38 个条目的额外父类进行统计分析,发现这些条目的额外父类有两种基本情况:一种情况是条目的额外父类是机构范畴之外的条目,这样的条目有 27 个;另一种情况是条目的额外父类处于机构范畴内,并处于机构类目范畴树较深的层次,这样的条目有 11 个。相对应地,得出机构类目范畴树存在以下两个问题:一是机构实体与其他实体重叠;二是机构子类层级混乱。

表 5 机构实体的上位类情况的 SPARQL 查询语句

```
SELECT ? item ? itemLabel ? parents ? broader ? broaderLabel {
  {# number of additional superclasses
  SELECT ? item (count (distinct ? parent)-1 as ? parents) {
    ? item wdt:P279 * wd:Q43229
    OPTIONAL { ? item wdt:P279 ? parent }
  } GROUP BY ? item }
  OPTIONAL { ? item wdt:P279 ? broader }
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
}
```

#### (1) 机构实体类与其他实体类的重叠问题

上述 38 个拥有多父类的机构二级类目中, circus 的额外父类无标签,不做讨论。有 27 个条目的额外父类不在机构范畴内,即这些条目的额外父类为其他类型的实体。由于 Wikidata 对机构类定义及划类标准不一,导致其与其他实体存在重叠问题。Wikidata 中,机构被定义为具有共同目标的社会实体,依此定义,重叠的实体类有人群实体(如 collective, musical ensemble, team)、地理实体(如 state, cemetery, geopark, water board, Landeskirche)、政治经济实体(如 anarchy、organization related to nonviolence、undenominationalism, resistance movement, alternative trading system, exchange, institute)及物理实体(如 medical facility, recording

studio, newspaper, magazine)。除一级类外,构成机构类目范畴树类目的 4020 个条目中,有 509 个条目存在着多级从属问题,即条目具有多个上位类,且至少有一个上位类在机构体系内。如 state 及其部分下位类的多从属情况如图 3 所示,图中加粗斜体指不在机构范畴内的额外父类。

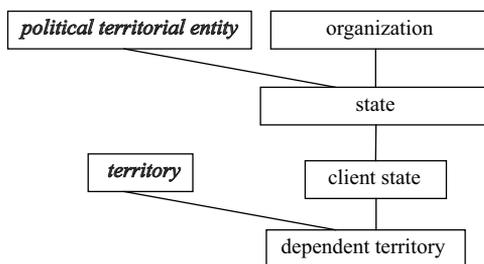


图 3 机构类实体与其他实体类重叠

(2) 机构类目范畴树中类目从属多造成类目层级混乱

38 个拥有多父类的二级类目中, 有 11 个条目的额外父类是处于机构类目体系中深度为  $n(n > 1)$  的条目。将这些条目按照它们各自的 subclass of 属性整理成图 4 所示的层级图, 可以直观地显示机构类目范畴树中子类层级混乱的问题, 主要表现为以下三种: (1) 既与机构顶级类 organization 为同位类, 又是其下位类, 如 committee, council, curatorial platform as organization 三个条目; (2) 既属于 organization 子类, 又是其同位类的子类, 如 fraternity; (3) 既是 organization 的直接子类, 同时又是机构的间接子类, 如 library association 既在 organization 类下, 又在 nonprofit organization 类下, 重复入类尽管可以体现多维揭示的特征, 但易于导致类目体系冗余。经统计, 在构成机构类目范畴树类目体系的 4020 个条目中, 有 431 个条目存在上述问题, 即条目多从属造成机构类目范畴树的层级混乱。

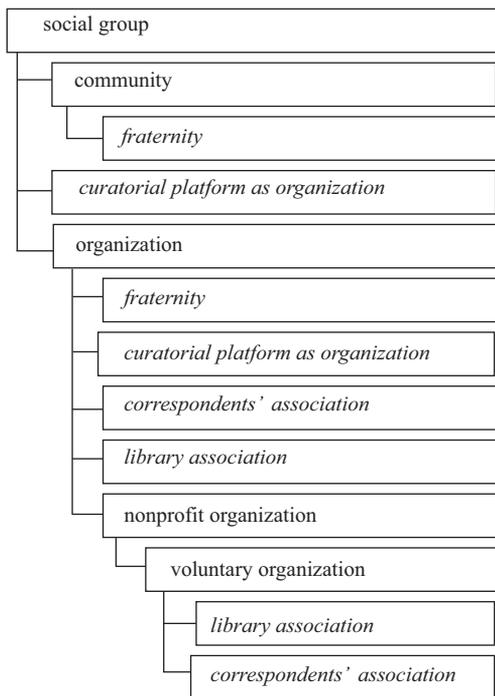


图 4 层级混乱示意

### 3 机构类目范畴树的优化方法

通过对机构类目范畴树进行的定量及定性分析, 发现 Wikidata 中的机构实体不仅数量庞大, 而且各实体具有明确的类型属性, 对于准确梳理世界范围内各种机构的上下位关系, 从而应对机构名称混乱、机构关系模糊、机构信息资源无序繁琐等问题具有重要参考价值。为了实现这样的目标, 构建一个更加科学、通用的机构类目范畴, 需要对机构类目范畴树中诸如机构类目的多从属等不合理的部分进行优化。

分类法的一个基本原则就是各类之间要均衡, 能够体现出对称性, 整体要协调。我们已经知道, 机构的 169 个直接子类处于不平衡状态, 所以改造机构类目范畴树的第一步, 是以机构的 169 棵子树作为整体进行剪枝和合并, 从而形成较为合理的基本大类, 并使得机构类目范畴树中的每一条目都对应唯一的上位类。第二步就是从每个条目的子类和实例数量出发, 对处于各层次的类目进行处理优化。由于每个条目与其上下位类的关系是既定的, 所以在处理时都是将被处理条目与其下位类看作整体进行的, 即处理单位是以被处理条目为根的子树。

#### 3.1 基本大类的优化

##### 3.1.1 无标签条目的子类处理

将没有标签的条目作为类目是不合理的, 因为其实含义不明确, 不能确切表达类目的实际内容范围, 内涵和外延不清楚。对这一类条目的基本处理原则是将其剪掉。具体到以无标签条目为根节点的子树中, 处理原则应该视无标签条目的子树的深度而定。以无标签条目为根的树深度为 1 的, 则意味着它没有下位类, 直接剪枝即可; 以无标签条目为根的树深度大于 1 的, 则删掉该无标签条目后, 将其下位类中的有标签条目的层次提升一级。

##### 3.1.2 层级混乱条目的处理

层级混乱是指同一条目在同一子树内或不同子树间重复出现。对于这类条目的处理, 我们要考虑条目在机构的子树中所处的层次。对于条

目在同一子树内或不同子树间层次值不相同的情况,处理的基本原则就是使得条目在机构类目范畴树中的层次取得最大值,剪掉其余位置的该条目及其子树。对于条目在同一子树内或不同子树间层次值相同的情况,处理的基本原则就是对条目的同位类的数量进行比较,同位类数量相对较少则保留条目及其子树,剪掉处于其余位置的该条目及其子树。这样的处理方法使得整个机构的分类遵循系统性原则,类目的层层划分、层层隶属可以更加严密有序。这里的剪枝处理实为对重复入类的条目的剪枝,因此不会发生条目数量和实例数量的变化。

### 3.1.3 与其他实体类型重叠的条目的剪枝

对于与其他实体类型重叠的条目,分析其各自的声明,将可以划分为机构实体的条目进行保留,将不能作为机构实体的条目进行剪枝。具体如下:(1)与人群实体及政治、经济实体重叠的条目,与机构实体内涵一致,所以予以保留;(2)将与地理实体重叠的条目及其相应的下位类进行剔除,分别是 cemetery、state、geopark、water board 和 Landeskirche 及相应下位类共 236 条;(3)与物理实体重叠的条目,视条目的声明而定。若声明显示其为具体的某个物理实体,应将这些条目及相应下位类剪枝,如 recording studio 及其下位类共 3 个条目,而 newspaper 和 magazine 的声明显示两者为出版物,因此剪掉 newspaper 和 magazine 及其相应的下位类共计 108 个条目;若条目的声明显示其具有机构的特征,则不做处理,如 medical facility 的声明是常规行医的位置,其下位类为医院、诊所等,故保留。

## 3.2 机构各类目的优化

经过第一步处理形成的机构类目范畴树,层次上存在的问题得到优化。然而,处于每一层级的条目仍存在以下几种不合理的条目:第一是无实例的条目;第二是机构类目范畴树中只有一个下位类的条目;第三是既有直接子类又有直接实例的条目。将机构类目范畴树中所有条目以有无实例为标准分为两类,对不合理情况进行优

化。由于每个条目的上下位关系是既定的,所以进行相应处理的前提依然是:将问题条目及其下位类作为整体来考虑。

### 3.2.1 实例数为 0 的条目

实例数为 0 的条目,处理原则应该视该条目所处的层级而定。无实例条目作为二级类且没有间接子类,类目不具有客观现实性和稳定性,因此应将这类无实例条目剪枝,共计 25 个。同理,剪掉处于机构类目范畴树最深层的无实例条目共计 1063 个。处于机构类目范畴树中其他层级的无实例条目共计 93 个,处理原则应该视其子类数而定。无实例条目只有一个子类,表明两者具有相同的内涵和外延,应该将两个条目做合并处理,即保留其中一个条目。合并处理视无实例条目的间接子类和间接实例数而定。无实例条目的间接子类数不为 0,则剪掉无实例条目,其直接子类和所有间接子类提升一级;无实例条目的间接子类数为 0,间接实例数不为 0,则剪掉无实例条目,保留其子类和相应的实例并提升一级;无实例条目的间接子类数和间接实例数均为 0,则将无实例条目保留,剪掉其子类。无实例条目有多个子类,则将其保留不做处理。

### 3.2.2 实例数不为 0 的条目

实例数不为 0 的条目,处理原则视其直接子类的数目而定。直接子类数为 0,即条目处于机构类目范畴树的最深层,这类条目不做处理,共计 2001 个;直接子类数为 1,将条目与其子类合并,并合并两者的实例,这类条目共计 315 个。具体合并原则如下:条目无间接子类,则保留两者中实例数较多的条目并将另一条目的实例合并至被保留的条目;条目有间接子类,则剪掉该条目,并使其相应的直接和间接子类提升一级,将该条目的实例合并到其子类的实例中。直接子类数大于 1 的条目有 523 个,对于这类条目的处理,考虑增列“未识别”类作为其下位类,将条目的直接实例置于“其他”类下。

### 3.2.3 “未识别”类的处理

在 523 个有多个子类的多实例条目中,有 83 个条目的实例数量超过 1000。实例数目过多表

示类目的外延过大,不能明确地指示及类分资源,因此考虑将这些实例进行聚类,然后以 instance of 属性分别连接至其所属的更细化的类目。

以图 1 中 financial institution 为例,其优化后的层次结构如图 5 所示。处理结果如下:剔除无标签条目 Q2111078 和 Q1553482, Q1553482 的子类 Shinkinbank 提升一级;剪掉 credit institution 下的 bank 分枝,保留 financial intermediary 下的 bank 分枝;Landesban sparkasse、sparebank 和 universal bank 实例为 0 且均处于最深层,直接剔除;

financial intermediary 只有 bank 一个子类,表明两者具有同样的内涵和外延,剔除 financial intermediary,将其实例与 bank 的实例合并作为 bank 的实例,bank 及其所有的直接和间接子类均提升一级;cooperative banking 只有一个子类 credit union,剔除 credit union,将其实例合并到 cooperative banking 的实例中。最后,对实例数量不为 0 且不处于机构类目范畴树的最深层的类目进行处理。为这些类目增设“unidentified”类,将他们的实例置于“unidentified”类下。

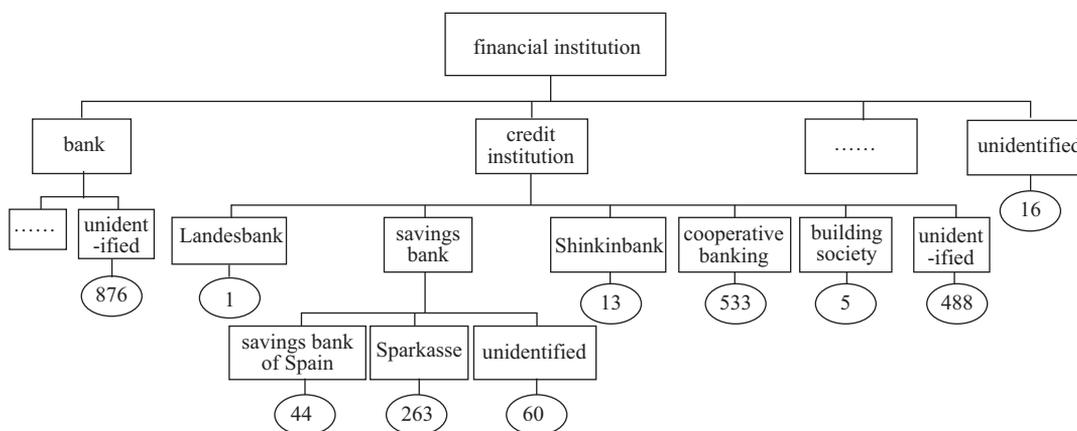


图 5 优化后的 financial institution 范畴树

#### 4 结论

为了理清各种机构的上下位关系和应对机构名称混乱、机构关系模糊、机构信息资源无序繁琐等问题,对机构信息进行合理的组织变得尤为重要。因此,以构建一个更加科学的通用的机构类目范畴树为目标,本文将 Wikidata 中的机构实体作为数据来源,对通过机构实体的 subclass of 和 instance of 两个属性构建的机构类目范畴树进行了定量分析,发现机构类目范畴树存在以下几个问题:机构类目不平衡,机构实体与其他实体有重叠现象,机构实体的多从属导致机构类目范畴树层次混乱。针对机构类目范畴树存在的问题,提出相应的优化方法,对于实例数量大于 10000 的类目,提出以聚类的方式将其划分到更

专属的类下。然而合适的聚类方式还有待探讨,希望在今后的研究中能弥补这一缺憾。

#### 参考文献

- 1 贤信.机构规范文档结构及构建方式研究 [D].北京:中国科学技术信息研究所,2015.
- 2 曾建勋,王立学.面向知识评价的规范文档建设方法 [J].图书情报工作,2012(10):101-106.
- 3 Vrandecic D. The Rise of Wikidata [J]. Intelligent Systems IEEE,2013,28(4):90-95.
- 4 Brasileiro F, et al. Applying a Multi-Level Modeling Theory to Assess Taxonomic Hierarchies in Wikidata [C]//International Conference Companion on World Wide Web.2016:975-980.

- 5 Wikidata: Statistics [ EB/OL ]. [ 2017-02-07 ].  
<https://www.wikidata.org/wiki/Wikidata:Statistics>.
- 6 Wikidata Introduction [ EB/OL ]. [ 2017-02-07 ].  
<https://www.wikidata.org/wiki/Wikidata:Introduction>.
- 7 Voß, J. Classification of Knowledge Organization Systems with Wikidata [ C ] // Proceedings of 15th European Networked Knowledge Organization Systems. 2016: 15-22.
- 8 Wikidata Query [ EB/OL ]. [ 2017-02-07 ].  
<https://query.wikidata.org/>.
- 9 DBpedia 2016-04 Statistics [ EB/OL ]. [ 2017-04-13 ].  
<http://wiki.dbpedia.org/dbpedia-version-2016-04>.
- 10 Spitz A, et al. State of the Union: A Data Consumer's Perspective on Wikidata and Its Properties for the Classification and Resolution of Entities [ C ] // Workshops of the Tenth International AAAI Conference on Web and Social Media. 2016: 88-95.
- 11 李慧佳,等.中文机构名称规范库建设的实践与分析——以“中科院机构名称规范库”建设为例 [ J ].图书与情报, 2016(1): 133-139.
- 12 邓朵朵.数据结构中树结构和图结构的比较 [ J ].科技信息, 2013(26): 164-165, 167.

(贾君枝 教授 山西大学经济与管理学院, 叶壮壮 山西大学经济与管理学院情报学专业 2016 级硕士研究生)

收稿日期: 2017-05-08