

## 关联数据的知识发现研究

李 俊<sup>1,2</sup> 黄春毅<sup>2</sup>

(1. 成都信息工程学院 图书馆, 四川 成都 610225; 2. 四川大学 公共管理学院, 四川 成都 610064)

**摘 要:**首先对关联数据的发展现状进行介绍,详细描述了W3C和LOD社区总结出的9条发布和提供数据服务的最佳实践。论述了对关联数据进行知识发现的必要性,同时提出基于关联数据进行知识发现的模型,并对模型中每一阶段的工作和目标进行详细介绍。

**关键词:**关联数据; 知识发现; 数据挖掘

**中图分类号:**G302      **文献标识码:**A      **文章编号:**1007-7634(2013)03-76-06

## Knowledge Discovery in Linked Data

LI Jun<sup>1,2</sup> HUANG Chun-yi<sup>2</sup>

(1. Library of Chengdu University of Information Technology, Chengdu 610225, China;

2. School of Public Administration, Sichuan University, Chengdu 610064, China)

**Abstract:** This paper introduces the development of linked data, describes the 9 best practices of publishing and providing data recommended by W3C and LOD communities. It also concludes the necessity of knowledge discovery in linked data and the model to implement knowledge discovery in linked data, it discusses the works and objectives each step of the model in detail.

**Key words:** linked data; knowledge discovery; data mining

## 1 关联数据发展现状

自从2006年Tim Berners-Lee提出Linked Data概念至今,经过5年社区和领域专家们的共同努力,至2011年9月19日,关联数据云(LOD)已有295个数据集,310亿条三元组。表1给出了LOD按领域划分数据的统计信息<sup>[1]</sup>。

由表1可见,目前国外政府机构对数据开放的支持度比较高,他们提供的49个数据集只占LOD的16.6%,但信息量占到了42%。大量信息掌握在政府部门手中,他们对数据开放关联的支持度对关联数据和语义Web的发展起到重要的数据支撑作用。

表1 至2011年9月 LOD按领域划分数据的统计信

Domain	Number of datasets	Triples	%	(Out-) Links	%
Media	25	1,841,852,061	5.82 %	50,440,705	10.01 %
Geographic	31	6,145,532,484	19.43 %	35,812,328	7.11 %
Government	49	13,315,009,400	42.09 %	19,343,519	3.84 %
Publications	87	2,950,720,693	9.33 %	139,925,218	27.76 %
Cross-domain	41	4,184,635,715	13.23 %	63,183,065	12.54 %
Life sciences	41	3,036,336,004	9.60 %	191,844,090	38.06 %
User-generated content	20	134,127,413	0.42 %	3,449,143	0.68 %
	295	31,634,213,770		503,998,829	

在关联数据技术发展的初始阶段,数据关联往

收稿日期:2011-12-21

基金项目:成都信息工程学院科研基金资助项目(CRF201213)

作者简介:李 俊(1983-),男,江苏人,硕士研究生,馆员,主要从事语义网与关联数据研究。

往是通过第三方组织将已有数据集转换为 RDF 并在网络上发布,或者通过已有的 Web APIs 将数据包装成关联数据。如今,随着关联数据技术的不断成熟和人们对语义 Web 发展重要性认识的不断提高,众多掌握数据资源的人们开始直接使用关联数据技术发布他们的数据并提供对外访问接口。到 2011 年 8 月,有 38.57%(295 个中的 113 个)的数据集是由数据生产者自己发布的<sup>[2]</sup>,较知名的有 BBC Music,英国政府提供的商业、教育数据, DBpedia, Freebase, Open Library 等,另有 61.43%(180/295)由第三方发布,例如 World Factbook, WordNet 等。

关联数据技术的提出是为了使应用能够从一个全球性、相互互联的数据源中发现和整合数据,从而获取更有价值的信息。经过 5 年技术发展, W3C 和 LOD 社区总结出了 9 条发布和提供数据的最佳实践:①提供可参引的 URIs (dereferencable URIs),这代表创建关联数据 4 原则中的前 3 个原则。②提供指向其他数据源的 RDF 链接,这代表第四个原则。这些 RDF 链接将不同数据源中的数据连接成一个全球性的 RDF 图,使得关联数据浏览器和爬虫能够在数据源之间跳转和导航。表 1 显示出 LOD 中链接超过 5 亿条,其中生命科学领域提供了 38.06% 的 URI 链接。在目前 LOD 的 295 个数据集中, URI 链接数超过百万的数据集有 29 个,一万到一百万之间的有 88 个,一千到一万之间的有 58 个。③使用广泛部署词汇表中的词汇 (terms)。为了使应用能更容易理解关联数据,数据提供者在任何可能的地方应该使用已经被广泛部署词汇表中的词汇来描述数据。LOD 中几乎所有的数据集都使用 W3C 提供的基本词汇表(如 RDF, RDF Schema, OWL)中的词汇,另外有 64.75%(191/295)的数据集使用了一些非专有词汇表,其中 31.19%(92/295)的数据集使用了 Dublin Core, 27.46% 的数据集使用了 foaf, 19.66% 的数据集使用了 skos。④使专有词汇表可参引。广泛部署的词汇表往往不能完全满足数据集在 Web 上发布所有内容所需要的词汇,数据提供者经常需要自己定义一些专有词汇作为补充。目前有 64.41%(190/295)的数据集使用了专有词汇表,剩余的 105 个数据集没有使用专有词汇表。为了使应用能够自动从 Web 上获得词汇表的定义,专有词汇表的 URIs 定义应该是可参引的。如何使专有词汇表 URIs 可参引, W3C 提供了一些最佳实践教程<sup>[3]</sup>。在使用了专有词汇表的 190 个数据集中, 83.68%(159/190)的数据集已使它们的专有

词汇表 URIs 可参引。⑤在专有词汇表和其他词汇表建立映射关系。专有词汇表应该与其他已存在、被广泛使用的词汇表中相似词汇建立映射联系,这样可以使应用能够尽可能多的理解数据并转化成它们自身的图式 (schemata)。W3C 的推荐规范中为表达这种映射关系定义了一些词汇,如 owl:equivalentClass, owl:equivalentProperty, rdfs:subClassOf, rdfs:subPropertyOf 等。在使用专有词汇表的 190 个数据集中, 7.89%(15 个)的数据集建立了它们的词汇表与其他词汇表的映射关系。⑥提供出处元数据。为了使应用能够确切地知道数据的来源,同时能够评估数据质量,数据源在发布原始数据的同时应该同时发布其出处元数据。一个通用方法是将其作为文档层次元数据描绘出来。描述出处信息被广泛使用的词汇表是 Dublin Core,它能够描述建立者、发布者、发布时间等信息。目前 36.63%(89/295)的数据集提供了出处信息。⑦提供许可元数据,一般使用 Dublin Core 中的 dc:rights 为发布的数据添加数据许可证。⑧提供数据集层次上的元数据。为了使实例数据更具自我描述性,鼓励数据发布者提供元数据来描述整个数据集的信息,如数据集的主题、数据集统计信息等。一种方法是使用 void vocabulary,其提供了完整的词汇表来描述这些元数据。另还可以使用语义站点地图来描述数据集元数据。有 36.95%(109/295)的数据集采用了 void 或语义站点地图提供了数据集信息。⑨描述访问方法。在 Web 上发布关联数据最原始的方法是使标识数据的 URIs 在 RDF 描述中可参引,而访问 LOD 数据集的两种基本方法一是通过 SPARQL 端点,二是使用数据集提供者提供的 RDF dumps 完整导出数据。有 68.14%(201/295)的数据集提供了 SPARQL 端点, 39.66%(117/295)的数据集提供了 RDF dumps 服务。void vocabulary 提供了词汇 void:sparqlEndpoint 和 void:dataDump 来帮助描述上述两种访问方法的链接地址。

## 2 对关联数据进行知识发现的必要性

语义 Web 发展十年间产生了许多标准语言,如 OWL, SPARQL, SKOS 等,出现了众多从事语义 Web 相关产业的公司。但语义 Web 的发展大大低于预期和 Web 本身的发展,部分原因在于语义 Web 理念和技术的推动早期依赖于 W3C 和国外一些高校,注

重技术发展和规范的制定,而缺少对数据实际使用的重视。同时由于语义 Web 的实现涉及众多学科领域如数据采集、信息检索、知识表达、机器学习、数据库、自然语言处理等,增加了开发者构建语义 Web 应用的成本,阻碍了数据的利用和开发。

关联数据的发展为实现语义 Web 提供了新方向和技术手段。五年来,伴随着关联数据技术的发展,出现了很多开放领域数据集,并有机地联系成了一个整体的数据云图。从发展现状和趋势来看,关联数据技术作为语义 Web 实现的一种技术手段已经得到众多社区、机构和政府部门的广泛支持。由于政府部门往往是大量信息的创建者、使用者和拥有者,他们对数据的发布和使用在未来一定会有越来越广泛的空间,同时对数据蕴含的潜在信息也将有更深层的挖掘需求。LOD 所包含的统计信息和专业领域信息量也在不断地增长,对于相关机构和组织,除了显式的使用关联数据开发应用之外,对数据内在信息的需求日益强烈。同时,人们意识到,隐藏在这些海量数据之后的更深层次、更重要的信息能够描述数据整体特征,可以预测事物发展趋势,这些信息在决策生成过程中具有重要的参考价值。

LOD 目前拥有 295 个数据集,310 多亿条三元组。目前关联数据集在一些应用领域是有价值的。在 QA、生物医药等专业领域应用、关联数据驱动的 Web 应用<sup>[4]</sup>等方面取得了显著成效,但与整个 LOD 310 亿条三元组的数据量相比,这些开发利用只是冰山一角。大量的信息给人们带来方便的同时,也带来了一系列的问题,比如信息量过大,超过了人们掌握消化的能力;一些信息真伪难辨,给信息的正确运用带来困难;信息组织形式的不一致性,增加了对信息进行有效、统一处理的难度等。因此,关联数据高效利用的需求与海量数据充分利用的困难性成了学者们需要解决的问题。

在 2010 年 4 月的 LDOW2010 大会上,关联数据领域的先驱们将关联数据挖掘作为了一个讨论主题<sup>[5]</sup>,为解决上述问题提供了新思路。

### 3 关联数据知识发现模型

#### 3.1 模型设计

基于数据库的知识发现(Knowledge Discovery in Database)是指从大量数据中提取有效、新颖、潜

在有用、最终可被理解的模式的非平凡过程<sup>[6]</sup>。KDD 是由多个步骤构成的处理过程,包括数据预处理、模式提取、知识评估及过程优化。所谓非平凡是指具有一定的智能性和自动性,而不仅仅是简单的数值统计和计算。KDD 是一个反复迭代的人机交互处理过程,该过程需要经历多个步骤,并且很多决策需要由用户提供。KDD 过程主要由三个部分组成,即数据清理、数据挖掘(Data Mining)和对挖掘出模式的解释及评估。DM 是 KDD 中最重要的一个环节。

KDD 及其相关的 DM 理论和技术已经相当成熟,随着语义 Web 的发展,语义数据挖掘成为知识发现一个新课题,相关挖掘算法的研究文献<sup>[7-9]</sup>逐渐增多。为了能够调和关联数据高效利用的需求与海量数据充分利用困难性之间的矛盾,本文将传统知识发现的模型<sup>[10]</sup>加以修改,以满足在关联数据上进行知识发现的需求。图 1 给出了模型的架构(KDLD, Knowledge Discovery in Linked Data)。下文将对此模型的每一步骤作具体阐述并给出实现模型的技术建议。

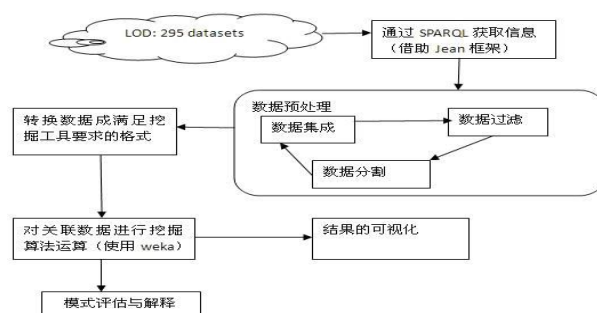


图1 关联数据知识发现模型

#### 3.2 通过 SPARQL 获取信息

LOD 295 个数据集中有 201 个提供了 SPARQL 端点提供数据查询服务。KDLD 这一阶段的任务可与传统 KDD 过程中数据选取类比,主要是通过查询数据集来获取数据,为了准确选取数据,最好有领域专家参与。

在查询信息之前,用户须先明确自己的信息需求,大体估计哪些数据源可能会包含自己所需要的数据。例如,如果想知道一个国家 GDP 增长率对其电影产量的影响<sup>[11]</sup>,数据源可以选择 World Factbook<sup>[12]</sup> 和 Linked Movie Database<sup>[13]</sup>。

SPARQL(Simple Protocol and RDF Query Language)是 RDF 数据模型定义的查询语言和数据获取协议<sup>[14]</sup>。SPARQL 将 Web 2.0 和语义 Web 联系起



来,是网络数据库的查询语言和数据获取标准<sup>[15]</sup>。它定义了RDF查询语言的语法和语义,可用于跨数据源的查询,数据既可以存储在本地,也可以由中间件提供。SPARQL提供可选图模式以及表示合取和析取的能力,其查询结果可以看成RDF图的集合。

SPARQL查询包含三元组模式的集合,称为基本图模式。一个简单的SPARQL查询如图2<sup>[16]</sup>。

```
1 SELECT DISTINCT ?author ?phone WHERE {
2   <http://data.semanticweb.org/conference/eswc/2009/proceedings>
3     swc:hasPart ?pub .
4   ?pub swc:hasTopic ?topic .
5   ?topic rdfs:label ?topicLabel .
6   FILTER regex( str(?topicLabel), "ontology_engineering", "i" ) .
7
8   ?pub swrc:author ?author .
9   { ?author owl:sameAs ?authAlt } UNION { ?authAlt owl:sameAs ?author }
10
11   ?authAlt foaf:phone ?phone
12 }
```

图2 查询作者和电话,该作者在ESWC2009做了 ontology engineering 相关报告

在实际应用中,可以借助一些框架工具帮助实现SPARQL查询,如Jena<sup>[17]</sup>。为了能够从多个数据源同时查询数据,满足信息多方面整合的需求,还可以使用Bastian Quilitz和Ulf Leser提供的工具库DARQ,此工具库可进行联合SPARQL查询。图3展示了其工作原理<sup>[18]</sup>。

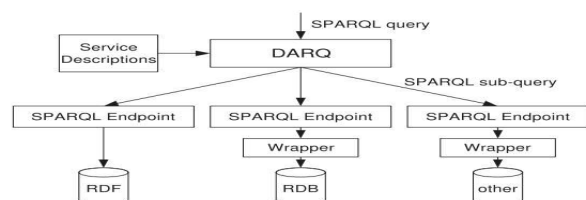


图3 DARQ的工作原理

### 3.3 数据预处理

把数据从数据集检出后,KDLD进入到数据预处理阶段。数据预处理是数据分析、数据挖掘等前期准备工作,在这一阶段需完成数据完整性和一致性的检查、消除噪声、过滤与数据挖掘无关的冗余数据,这一阶段工作的好坏关系到是否能挖掘出高质量的数据模式,工作量也比较大,传统KDD中这一阶段往往会占据整个KDD过程六成时间。数据预处理包含三个具体任务,即数据集成、数据过滤和数据分割。

运用SPARQL能够从一个或多个数据集中检索数据,运用一些开源框架,数据检索过程变得非常灵活而方便。数据集成需要将不同数据源的数据根据一些共同属性和关联进行合并。具体操作中可分为两部,先是隔离,保证为实体的每次出现指

派唯一标识符,然后是调和,即确认相同实体并将该实体的各次出现合并在一起。例如上文提到GDP与电影生产的关系,从World Factbook检索到各国GDP数据,从LinkedMDB检索到电影数据,根据结果中共同的属性国家,即可将来自不同数据源的数据集成起来。

数据过滤阶段完成数据清理、对不需要数据的删除等,从而排除一些不必要数据。例如将两个数据源的数据根据共同属性集成后,会产生一个重复的国家属性数据,就需要对重复数据进行删除处理。另外,如果只想研究GDP不少于某个值的那些国家,在此阶段就可以将那些不满足条件的数据排除掉。数据过滤阶段还需要完成对噪声数据的处理,可以使用传统KDD中分箱、聚类、计算机与人工检查结合、回归等技术进行操作。对于数据集成中存在的的不同数据也须在这阶段进行处理。

数据分割阶段等同与传统数据挖掘过程中的数据概化,使用概念分层用高层次的概念替换低层次原始数据,创建低层概念集到高层概念集的映射,在各种抽象级别上处理数据,从而在多个抽象层上发现知识。通过数据概化还可以压缩数据,在压缩的数据集上进行数据挖掘更为有效。实际操作中,如果需要做统计分析,可以把同一属性的数据分割成一些抽象类。例如,通过SPARQL查询到国家信息,如果包含国家人口数量的属性,使用数值属性的概念分层,把人口数量小于一千万的归于较小人口数量,一千万至一亿的归于平均人口数量,大于一亿的归于较高人口数量。这样便把人口属性概化成不同的类,为后续统计分析做准备。

数据预处理阶段的工作对于数据挖掘的结果非常重要,因此在实际应用中,需把数据集成、数据过滤和数据分割作为一个整体步骤来考虑,需要根据需求不断调整数据模型以满足具体挖掘算法的要求。

### 3.4 转换数据格式

数据挖掘是知识发现过程中最重要的步骤。数据挖掘包括机器学习技术、基于数据库和统计分析三个主要方向。机器学习技术的特点是用户不需要关心算法如何实现、如何计算,只需构造适合算法的数据模型,其他由数据挖掘软件去操心。这非常适合于对相关领域业务非常精通但对数据挖掘算法不精通的用户。

Weka是由新西兰怀卡托大学开发的免费开源软件,全名是怀卡托智能分析系统。Weka是由Java写成,可以运行于几乎所有的操作平台,汇集了当今最前沿的机器学习算法及数据挖掘工具。它为数据挖掘实现的整个过程,包括准备要输入的数据,评估学习方案,以及可视化输入数据及学习结果等提供了广泛支持。

在KDLD的过程中可采用Weka来进行关联数据的数据挖掘工作。故此阶段需要将已检出、经过数据预处理的关联数据转换成满足Weka要求的格式,如Arff文件格式。这一阶段的任务类似与传统知识发现过程中的数据转换。

### 3.5 对关联数据进行挖掘算法运算

经过前期数据检索、数据预处理和数据格式转换,已经将关联数据处理成满足挖掘工具要求的形式。这一阶段将数据输入挖掘系统,便可进行分类、聚集、发现数据关联规则等挖掘算法操作。

关联规则反映了一个事物与其他事物之间的相互依存性和关联性。如果两个或多个事物之间存在一定关联关系,那么其中一个事物就能通过其他事物预测到。LOD包含了295个不同领域的数据集,就数据数量和数据内容来看,必然存在许多未被发现的相关性。例如将数据集World Factbook与LinkedMDB的特定数据进行关联分析,也许会得到一个国家GDP数据与电影产量的关联性,将数据集Open Library与World Factbook的数据进行关联分析,也许会得到一个国家文化产业发展与经济发展之间的关联性。

Weka对聚类、分类与预测等挖掘算法都有很好的支持度,只要用户对自身领域业务比较精通,能够正确选择相关数据集进行数据检索和预处理,得到满足算法要求的数据模型,再根据需求选择适合的挖掘算法,就能方便快捷地挖掘到自己感兴趣的模式,从而达到充分、高效地利用海量数据进行相关领域的研究利用。

### 3.6 结果的可视化和模式评估

数据挖掘结果可视化是指将数据挖掘后得到的知识和结果用可视化形式表示出来。包括关联规则的单维关联规则可视化、多维关联规则可视化,分类的决策树可视化、分类3D散点图,聚类的聚类树等。将挖掘结果可视化有助于用户对数据、数据处理结果的了解和认识。Weka系统能够实现

挖掘结果的可视化,虽然可视化技术比较简单,但Weka是开源系统,在实际应用中可以自主开发一些可视化程序。Weka手册介绍了如何扩展Weka,添加可视化插件实现可视化定制<sup>[19]</sup>,从而将挖掘结果更清晰地展现给用户。

在得到挖掘结果后对挖掘出的模式进行评估是知识发现过程中不可或缺的步骤。挖掘结果是否有价值、是否新颖需要精通专业领域知识和对业务敏感的专家级用户进行评价。本文构建的挖掘模型,在最初的数据检索阶段和模式评估阶段,都需要专家级用户的参与来保证关联数据高效充分的利用。

## 4 结 语

随着关联数据集数量不断增多,关联数据高效利用的需求与海量数据充分利用困难性之间矛盾的逐渐显现。传统对关联数据的消费方式还只停留在浏览、检索和显式利用数据构建应用的层次上。对数据高效和具有深度的使用只局限与某些专业领域数据集(如生物、医药),为了对海量数据高效利用和对隐藏在数据背后的知识有效发现,本文将传统知识发现模型引入到关联数据领域,对其进行修改以适应关联数据数据结构化、数据分散的特点。本文构建的模型适用于不同学科领域,使得各领域用户在开发利用关联数据的过程中具有统一的使用模式,满足不同用户对隐藏在数据背后知识同样的需求。文中构建的知识发现模型仍有改进和发展空间,统计分析和人工智能技术也可应用于知识发现过程,根据不同用户对不同数据集的使用偏好和挖掘习惯,自动推荐最适合的机器学习算法,从而优化知识发现过程,以达到充分利用关联数据获得有价值知识的最终目的。

### 参考文献

- 1 State of the LOD Cloud[EB/OL].<http://www4.wiwiwss.fu-berlin.de/locloud/state/>,2011-12-01.
- 2 Search-the Data Hub[EB/OL].<http://thedatahub.org/dataset/search?q=tags:published-by-producer+AND+%28groups:lodcloud+AND+-tags:lodcloud.unconnected+AND+-tags:lodcloud.needsfixing%29>,2011-12-01.
- 3 Best Practice Recipes for Publishing RDF Vocabularies [EB/OL].<http://www.w3.org/TR/2008/NOTE-swbp-vocab-pub-20080828/>,2011-12-01.

- 4 沈志宏,张晓林. 关联数据及其应用现状综述[J]. 现代图书情报技术, 2010,(11):1-9.
- 5 Linked Data on the Web(LDOW2010) [EB/OL].<http://events.linkedata.org/ldow2010/>,2011-12-04.
- 6 李雄飞. 数据挖掘与知识发现[M]. 北京:高等教育出版社, 2010:1-3.
- 7 王志愿,夏士雄,张 磊. 语义驱动的关联规则挖掘算法研究[J]. 计算机工程与设计,2011,32(3):936-939.
- 8 Haishan Liu. Towards Semantic Data mining[EB/OL]. <http://ceur-ws.org/Vol-660/paper6.pdf> 2011-12-01.
- 9 Ziawasch Abedjan, Felix Naumann. Context and Target Configurations for mining RDF Data[EB/OL].[http://www.hpi.uni-potsdam.de/fileadmin/hpi/FG\\_Naumann/publications/2011/smer-abedjan.pdf](http://www.hpi.uni-potsdam.de/fileadmin/hpi/FG_Naumann/publications/2011/smer-abedjan.pdf) 2011-12-01.
- 10 U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. The KDD process for extracting useful knowledge from volumes of data [J]. Communications of the ACM, 1996,39(11):27-34.
- 11 Venkata Narasimha, Pavan Kappara. LiDDM: A Data Mining System for Linked Data[EB/OL].<http://events.linkedata.org/ldow2011/papers/ldow2011-paper07-pavan.pdf>,2011-11-05.
- 12 CIA-The World Factbook[EB/OL].<https://www.cia.gov/library/publications/the-world-factbook/>,2011-12-01.
- 13 Linked Movie Data Base[Start Page][EB/OL].<http://linkedmdb.org>,2011-12-01.
- 14 SPARQL Query Language for RDF[EB/OL].<http://www.w3.org/TR/rdf-sparql-query/>,2011-12-01.
- 15 高志强. 语义 web 原理及应用[M]. 北京:机械工业出版社, 2009:2-3.
- 16 Olaf Hartig, Christian Bizer, Johann-Christoph Freytag. Executing sparql querying over the web of linked data[EB/OL]. <http://data.semanticweb.org/pdfs/iswc/2009/paper301.pdf>, 2011-12-01.
- 17 ARQ-SPARQL Tutorial[EB/OL].<http://openjena.org/ARQ/Tutorial/index.html>,2011-12-01.
- 18 Bastian Quilitz, Ulf Leser. Querying Distributed RDF Data Sources with SPARQL[EB/OL].<https://www.ki.informatik.hu-berlin.de/wbi/research/publications/2008/DARQ-FINAL.pdf>,2011-12-01.
- 19 weka\_manual[EB/OL].<http://nchc.dl.sourceforge.net/project/weka/documentation/3.6.x/WekaManual-3-6-5.pdf>,2011-12-01.

(责任编辑 孙晓明)

(上接第75页)

促进信息流通的枢纽作用,有利于不同学术群之间的信息和思想的交流。

## 4 结 语

索引事业在中国索引学会的带动下,已经逐步形成自己的作者群,多倾向于作者合作的方式撰写论文,核心作者比较集中,经验证每年段均符合洛特卡定律,这对科学组织科研生产团队,加快学科发展速度都起到积极作用。

索引学20年来新老交替的作者团队,高频作者几乎都是各大名校的学科带头人,他们主要在六个方面为索引事业的发展作出巨大贡献:①XML数据管理与数据挖掘;②数据库系统理论与技术;③多媒体信息检索;④空间索引;⑤应用索引技术的逆向工程;⑥科学计量与情报检索。

## 参考文献

- 1 邱均平,杨思洛,周春雷. 改革开放30年来我国情报学研究的作者分析[J]. 情报学报,2009,(4):626-633.
- 2 顾立平. 学科化知识服务的类型研究——以区域经济的学科竞争力分析为例[J]. 科学学研究,2011,(2):186-191.
- 3 蒋 颖,金碧辉,刘筱敏. 期刊论文的作者合作度与合作作者的自引分析[J]. 图书情报工作,2000,(12):23-28.
- 4 马悦英. 分形论在情报学中的应用[J]. 图书情报工作,1998,(5):12,15-16.
- 5 邱均平. 信息计量学[M]. 武汉:武汉大学出版社,2007:169.
- 6 许雯燕,康平立,龚 勋. 科学生产率的分布及其产生机制[J]. 现代情报,2007,(9):212-214.
- 7 王 陆. 典型的社会网络分析软件工具及分析方法[J]. 中国电化教育,2009,(4):95-100.
- 8 Otte E, Rousseau R. Social network analysis: A powerful strategy, also for the information sciences[J]. Journal of Information Science, 2002,(6):441-453.

(责任编辑 孙晓明)