

# 基于大数据的主动科研管理模式与优化决策机制

罗瑞丽, 王元卓

中国科学院计算技术研究所, 北京 100190

## 摘要

在科研管理过程中,从科研人员、项目、成果产出等多维度可获取大量数据。科研管理大数据背后蕴含着丰富的、复杂关联的知识。科研管理大数据中信息与知识的有效、丰富,依赖于科研管理知识图谱的建立。采用开放知识网络模型方法,以构建“专家图谱”为例,融合分析了科研管理大数据中的关联知识和潜在信息,并探讨基于这些技术手段,如何将传统被动的管理模式转变为主动的科研管理模式,进而建立基于大数据的新型管理模式与决策机制。

## 关键词

大数据;开放知识网络;科研知识图谱;主动科研管理;科研管理决策

中图分类号:TP391

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2019038

## *Active scientific research management model and optimization decision mechanism based on big data*

LUO Ruili, WANG Yuanzhuo

Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

## *Abstract*

During the scientific research management, a large number of data could be accessed from researchers, projects, output and other multi-dimensional. The big data of scientific research management implicitly contains tremendous highly-interconnected knowledge. The effective obtaining of rich knowledge from big data, depends on the building of the large-scale knowledge graph. An expert graph was built up using open knowledge network, and the associated knowledge and potential information were analyzed. Based on these analysis results, how to change the traditional passive management model into active scientific research management mode was discussed, and then a new management model and decision-making mechanism based on big data was established.

## *Key words*

big data, open knowledge network, scientific research knowledge graph, active scientific research management, scientific research management decision

## 1 引言

大数据已成为社会各界关注的新焦点。大数据时代的到来,使得科学研究的组织和方式发生了极大的变革。继实验科学、理论科学、计算科学之后,科研范式(即数据密集型范式)进入了大数据驱动的第四范式,各项科学活动产生的数据正在日益增加<sup>[1]</sup>。与此同时,由这些科学活动、科技成果等构成的科研数据的价值也逐渐受到各领域的关注。对于科研领域,公开科研数据一方面可以帮助科研人员提升自身的科研影响力,另一方面也可以避免由于信息不对等造成的科研活动与科研资金的重复与浪费;对于政府部门,公开科研数据可以为政府响应国家战略号召与满足社会应用需求提供强大的数据支撑。

现今,科研数据的大量产生、科研数据价值的迅速凸显等问题迫切需要科研管理者采取措施对其进行妥善管理。大数据具有不同于传统数据工程的5个明显的特征,即5V特性,具体为:大数据规模巨大;大数据类型繁多;产生速度快,处理能力要求高;数据真伪难辨,可靠性要求更严格;价值大,但密度低,挖掘难度大。目前对于科研大数据的管理主要有两种方式,即技术手段和人文手段。其中,人文手段包括制定科研数据管理政策,为科研数据管理提供明确框架和规范,保障科研数据管理的科学性<sup>[2]</sup>。

根据数据产生的时间,科研管理大数据可以分为历史数据和动态数据。其中,历史数据以成果数据(如论文、专利、著作、奖励等)、项目数据为代表,动态数据以科研人员行为(如参加学术会议、进行学术报告等)为代表。使用分类、聚类、统计分

析、关联分析等方法对科研管理大数据中的历史数据进行深度挖掘和分析,可以发现数据中隐藏的线索、规律、模式。比如对科研人员影响力的分析可以帮助发现有价值的科研方向和科研项目。如果将这些时间敏感的数据看成一个随机序列,那么根据历史序列值和当前序列值进行合理推断,就可以预测未来序列值,并使用数学模型对这种序列进行近似描述。因此,如果想要从历史数据中挖掘重要信息,并预测未来数据,就需要对历史数据进行因素分析、模式识别、参数估计、假设检验。

科研管理大数据的直接表现形式往往是分散的数据和链接,将这些反映相互关系的链接整合起来就是一个网络<sup>[3]</sup>。这个网络中的数据大多为非结构化数据,且数据具有来源多样化、实时性强、突发性强、高噪声等特点。该网络中包含数据信息、关联信息、网络结构等信息。有效利用科研管理大数据可以有效地对数据进行消歧、去重、挖掘以及深入分析。

近年来,工业界和学术界都致力于开放大数据知识库的研究和开发工作。目前,国内外相关知识库和应用系统已达上百种,包括KnowItAll<sup>[4]</sup>、TextRunner<sup>[5]</sup>、NELL<sup>[6]</sup>、Probase<sup>[7]</sup>、Satori、Prospera<sup>[8]</sup>、Sofie<sup>[9]</sup>以及一些基于维基百科等在线百科知识构建的知识库(如DBpedia<sup>[10]</sup>、YAGO<sup>[11]</sup>、Omega<sup>[12]</sup>、WikiTaxonomy<sup>[13]</sup>)。除此之外,在对开放大数据的管理、计算和分析方面,国内外已经发布许多相关的知识搜索计算平台,如Evi公司的TrueKnowledge知识搜索平台、Google公司的知识图谱(Knowledge Graph)、Facebook公司的实体搜索服务(Graph Search)等。其中影响力较大的有Probase和Google公司的知识图谱。Probase拥有千万级概念,是拥有概念数最多的知识库。Google公司的知识图谱拥

有亿级实体和百亿级关系。

本文分析了科研管理中的大数据,根据获取方式可以分为历史数据、生成数据和开放数据,并提出了应用开放知识网络对科研管理大数据中的知识进行建模的方法,进一步以科研论文合作为案例,构建了“专家图谱”,通过数据对比,分析了利用科研管理知识图谱可以进行的一系列有效的主动科研管理和优化科研决策等具体应用。

## 2 科研管理中的大数据

根据获取方式的不同,可以把科研管理中的大数据分为历史数据、生成数据、开放数据3个类型,在科研管理过程中依据不同的应用需求对这些数据进行融合<sup>[14]</sup>。

### 2.1 历史数据

自20世纪末推行高校和科研机构信息化建设以来,很多机构内部建立了科研管理系统,在科研人员或科研管理人员将相关数据输入系统的被动阶段已经积累了大量的原始数据,如科研课题的立项和结题(含纵向课题、横向课题等)、科研合同经费(含配套经费、经费预算、经费到账等)等信息的录入、修改、查询、统计报表。科研活动进行过程中会产生项目信息(如申报书、审批书、中期报告、结项报告等)、学术论文、学术著作、演示文稿、研究报告、试验数据、软件代码、多媒体资料以及用数字形式表达的想法、观点、经验及诀窍的总结等科研成果和信息资源。尽管管理人员只能通过简单的操作(如排序、统计)获取一些表面信息,但这些原始数据的积累为大数据时代的到来奠定了基础。这些数据使得科研成果的查询和管理更加方便、快速,还可以帮助管理人员掌握科研人

员的整体科研情况和重点研究方向等。

### 2.2 生成数据

数据的另一个产生途径是科研人员主动发布数据。然而,科研人员很少主动推送数据,因此,科研管理部门需要承担起这个职责,建立相应平台,促进科研人员间的信息交互。具体地,科研管理部门可以在现有系统的基础上,添加信息采集模块,供科研人员自主填写信息。这些信息可以为以后的科研活动和项目提供宝贵的数据基础。除此之外,科研管理部门也要组建学术交流平台,促进科研人员间的学术交流和合作,如组建跨学科科研小组、推广个人科研成果等,避免在研究中出现“闭门造车”的现象。

### 2.3 开放数据

开放数据近年来的快速增长主要得益于两个方面。一方面是数字化平台和设备的快速发展,首先学术性社交平台(如学术博客、微博)的快速发展促进了科研人员在公开网络平台中展示个人学术成果、交流学术心得;其次新型移动设备(如智能手机)具有易于携带、可随时接入网络等特点,便于科研人员随时在开放网络中发表个人学术见解。这个阶段数据的产生方式是主动的<sup>[15]</sup>。另一方面是科研管理系统与其他数字化平台关联度的增加,不同部门、不同功能的管理系统之间的数据共享和自动匹配可以帮助科研管理系统更方便地进行信息收集和管理。

## 3 科研管理知识图谱构建

科研管理知识图谱的构建基于科研管理数据的获取和融合。在构建科研管理知识图谱时,首先要构建知识图谱的基本构

成要素,包括概念、实体、属性和关系<sup>[16]</sup>。知识图谱的构建可以分为手工构建和自动构建。手工构建比较典型的例子是知网(HowNet)<sup>[17]</sup>、同义词词林<sup>[18]</sup>、概念层次网络(hierarchical network of concepts, HNC)<sup>[19]</sup>和OpenCyc<sup>[20]</sup>等。本节应用开放知识网络(open knowledge network, OpenKN)对科研管理数据进行建模表示,以科研学术数据为例,对专家图谱中的隐性知识挖掘和合作关系推荐等进行研究,为主动科研管理和优化决策提供支持。

### 3.1 科研管理知识的表示

本文采用开放知识网络统一表示科研管理知识。首先构建一个以多种类型的科研实体为节点、实体间关系为边的异构开放知识网络。其中,节点类型包括作者(A)、论文(P)、会议(C)、组织(O)和关键词(K),相应地,节点间边的关系类型包括作者间的学术合作(如论文合作、项目合作等)、论文间的引用关系等。网络中的每一个节点和边都包含特定的时间和空间属性信息。在科研学术网络中,这些时空属性信息包括作者的工作地点、合作的产生时间等。具体地,已知时间集合 $T$ 和空间集合 $S$ ,已知节点的类型集合为 $N$ ,边的类型集合为 $R$ ,在 $T$ 、 $S$ 时空属性下的开放知识网络 $G_{T,S}$ 可以表示为如下的八元组形式:

$$G_{T,S} = (V, E, \phi, \psi, \theta, \tau, \lambda, \eta) \quad (1)$$

其中, $V$ 是网络中节点的集合, $E$ 是有向边的集合,即一系列关系对 $(u,v,r)$ 的集合,其中 $u,v \in V, r \in R$ ,即每对点都被赋予了一个或多个关系。 $\phi, \psi, \theta, \tau, \lambda, \eta$ 为映射函数,在构建科研管理开放知识网络的过程中,使用映射函数表示集合间的关系。

图1展示了一个小规模科研领域的开放知识网络结构。具体地,网络中不同颜色的节点表示不同类型的实体,该网络

的点集合为 $V=\{a,b,c,d,e,f,g,h,i,j,k,l,m,n\}$ ,时间集合为 $T=\{t_1,t_2,t_3,t_4\}$ ,空间集合为 $S=\{s_1,s_2,s_3\}$ ,图1中列举了该科研管理网络中点集合和边集合上的映射函数以及相应的函数值。在科研管理网络中, $\phi:V \rightarrow N$ 为点集合到点类型的映射函数,对于 $V$ 中的任意元素 $v$ ,都可唯一确定其类型 $\phi(v) \in N$ ,图1中等式 $\phi(a)=A$ 表示点 $a$ 的类型是 $A$ 。 $\psi:E \rightarrow R$ 为边集合到关系类型的映射函数,其中边用三元组的形式表示。图1中等式 $\psi\{a,b,coauthor\}=coauthor$ 和 $\psi\{a,b,citation\}=citation$ 表示在节点 $a$ 、 $b$ 之间存在合作和引用关系,因此可知每对点之间的关系类型最多有 $|R|$ 个。 $\theta:V \rightarrow 2^T$ 为点集合到时间集合的映射函数,可通过计算点的时间戳来描述点的生命周期,其中 $2^T$ 是集合 $T$ 的幂集,图1中等式 $\theta(a)=\{t_1,t_2\}$ 表示在点 $a$ 上的时间映射函数。 $\tau:E \rightarrow 2^T$ 为边集合到时间集合的映射函数,可通过计算边的时间戳来描述边的存在时间,图1中等式 $\tau\{a,b,coauthor\}=\{t_1,t_4\}$ 表示在边 $(a,b,coauthor)$ 上的时间映射函数。 $\lambda:V \rightarrow 2^S$ 为点集合到空间集合的映射函数,可通过计算点的空间信息来描述点的活动轨迹,图1中等式 $\lambda(a)=\{s_1,s_3\}$ 表示在点 $a$ 上的空间映射函数。 $\eta:E \rightarrow 2^S$ 表示边集合到空间集合的映射函数,可通过计算边的空间信息来描述边的存在轨迹,图1中等式 $\eta\{a,b,coauthor\}=\{s_2\}$ 表示在边 $(a,b,coauthor)$ 上的空间映射函数。

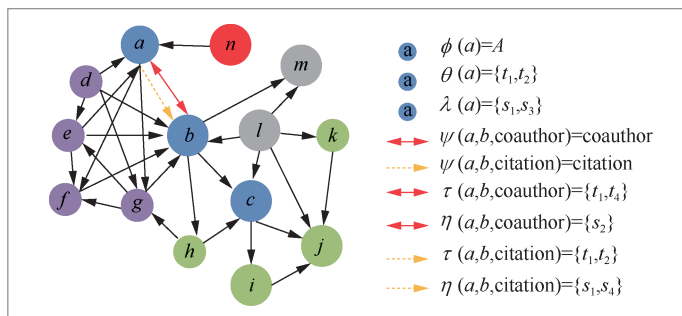


图1 科研领域开放知识网络结构示意图

需要指出的是,由于OpenKN的知识库可以是与领域相关的,所以开放知识网络也可以是与领域相关的,即点集到点类型的映射函数 $\phi$ 与边集到关系类型的映射函数 $\phi$ 都与领域相关。

### 3.2 专家图谱

本节内容阐释学术网络领域的开放知识网络的自增长过程。笔者从互联网上的开放学术页面中爬取数据,寻找开放知识网络的5类节点(作者、文章、组织、会议、关键词)。笔者从DBLP(database systems and logic programming)书目网、微软学术搜索、Google学术搜索等多种资源中开放式地获取数据。笔者从在1928年到2012年信息检索、数据挖掘、人工智能、机器学习和计算机科学这5个领域的顶级会议上发表超过3篇论文的作者中选取一个子集,作为种子使用者;接下来,寻找这些作者的合著者、隶属的组织以及在他们文章中提到的关键词。这些作者、文章、组织、会议以及关键词的总数分别是2 030、857、920、260和7 004。将这些数字相加得到学术开放知识网络的11 071个顶点。针对这个网络的边,笔者在表1中列举了这些概念间的所有有向关系的类型,其中前两列是概念对和它们间的双向关系,第三列表示关系存在的时间,其中, $t$ 表示开始时间, $t'$ 表示结束时间, $\infty$ 表示无穷长。数据集中有如下关系的事例对的总数是29 332。

表1 概念间的有向关系

实体对	关系	时间
A,P	write, written by	$[t, \infty]$
C,P	publish, published by	$[t, \infty]$
P,K	mention, mentioned by	$[t, \infty]$
A,O	employee, employer	$[t, t']$
P,P	cite, cited by	$[t, \infty]$

构建好科研学术开放知识网络后,继续确定这个原始的知识网络。事实上,笔者已经证实这个过程可以通过从开放知识网络中抽取特定的三角形(网络中由节点和边构成的三角结构)来完成。因此,笔者从科研学术开放知识网络中找到所有的三角形,从中随机选取150个三角形,并保证这些随机选取的三角形的所有的顶点包含了全部的节点类型。接下来,对于每个三角形,任意去除其一条边,并且使这些去除的边的关系尽可能地不同,并记录下这些去除边的三角形。在表2中列举了一小部分去除一条边的三角形。第三列展示了三角形的3个顶点和剩下的两条边,这使得在第一列中的两个概念之间的关系可以在第二列中被推断出来。例如,如果两个作者写了同一篇文章,那他们就是合著者关系。特别地,部分关系后面的数字(2)表示这些关系被记录了两次,并且它们是无向的。另外,笔者在最后一列注明了这些被推断的关系的持续时间,其中, $t_1$ 、 $t_2$ 表示关系的开始时间, $\text{sta}(t_1 \cap t_2)$ 表示开始时间为 $t_1 \cap t_2$ , $\infty$ 表示无穷长。在整个数据集含有这些被推断关系的事例对的总数是106 882。

通过这些原始网络上使用基本的规则,可以推断出这些隐性关系,并构成了科研管理所需要的“专家图谱”,如图2所示。

在图2中,给出了学者间合作关系推断的图解。不同颜色的实线表示不同类型的关系,有时间期限的点线表示被推断的关系。从图2可以看出,这个专家图谱中采用知识演化模型推断出了80%的隐性关系,换言之,该模型可以预测出科研活动中专家之间的绝大多数隐性关系和潜在合作。

## 4 主动科研管理模式

从专家图谱的具体案例可知,对科研

管理大数据建立开放知识网络可以深入挖掘出科研人员、学术成果、项目合作等多方面的关联信息以及很多数据中没有明确表达出来的隐性信息,这将极大地促进科研管理模式的转变,使其从原有被动的事后审批、被动管理向主动的、事前的、有针对性的管理模式发展。

### 4.1 提高主动性和数据敏感度

传统的科研管理更多的是完成资料的上传和分发、发送和接收等任务,属于被动管理,这种管理无法适应大数据时代的新需求。科研领域开放知识网络提供了科研数据存储、管理和挖掘的新方式。构建包含动态时空属性的科研领域开放知识网络,可以实现对大增量科研数据的动态实时更新,避免出现由于数据更新不及时造成科研管理出现问题的情况。构建基于开放知识网络的科研管理大数据分析模型,将科研数据的获取、管理、挖掘、创新等过程进行融合和深入探索,对科研数据进行深度挖掘、推断和分析,将传统的“被动管理”方式转化为“主动服务”方式,从而提高数据敏感度,提供及时的、优质的服务。科研管理部门也应化被动为主动,积极获取更多的科研数据信息,以保证科研管理数据库的及时性和有效性;根据数据特征进行整合和管理,以保证数据的延续性;主动关注科研领域的前沿信息、发展动态、研究方向,为组织机构提出相关发展意见,为科研人员提供全面服务,促进科技的进步与发展。

### 4.2 提供基于数据的个性化服务

科研管理旨在提升科研人员的积极性。因此,科研管理人员需要为科研人员提供高效的服务,创造便利的科研条件。

表 2 原始学术开放知识网络

实体对	关系	三角形	时间
A,A	coauthor(2)	$A \triangleleft P \triangleleft A$	$t_1$
P,P	same mention(2)	$P \triangleleft K \triangleleft P$	$[sta(t_1 \cap t_2), \infty]$
A,A	college(2)	$A \triangleleft O \triangleleft A$	$[sta(t_1 \cap t_2), \infty]$
A,A	cite, cited by	$A \triangleleft P \triangleleft A$	$[sta(t_1 \cap t_2), \infty]$
A,C	contribute(2)	$A \triangleleft P \triangleleft C$	$t_1$
O,O	collaboration(2)	$O \triangleleft P \triangleleft O$	$t_1$
A,P	cite, cited by	$A \triangleleft P \triangleleft P$	$t_1$
A,A	coauthor(2)	$A \triangleleft C \triangleleft A$	$t_1$

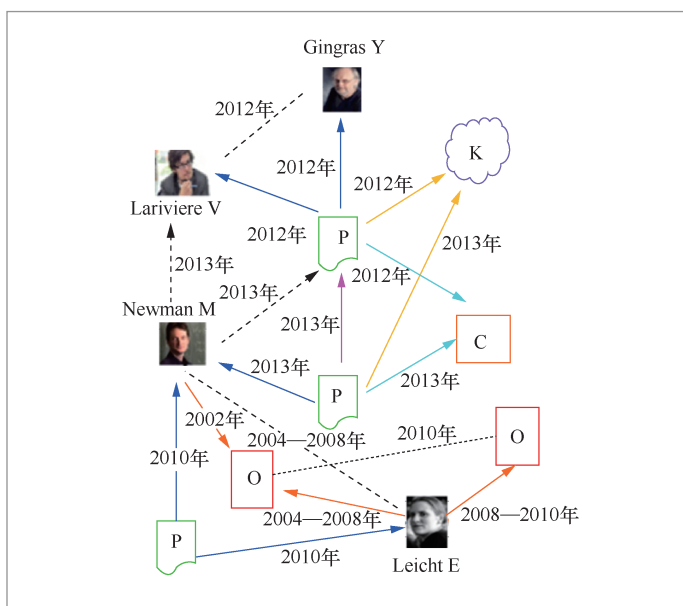


图 2 专家图谱

现今科研数据向综合化、学科多样化方向发展,而科研人员的研究方向则越来越精细化、专业化。这就需要科研管理者根据科研人员的特点和方向,提供适合科研人员的个性化的服务。通过构建科研领域开放知识网络,科研管理人员可以清晰地发现每个科研人员的研究方向和科研实体间的各种关系,科研领域开放知识网络为提供高效的个性化服务提供了良好的数字化数据基础。同时,构建基于开放知识网络的科研数据分析、预测方法,可以帮助科研管理人员智能化地分析已掌控的海量

数据,主动积极地收集科研人员的科研数据,并通过数据融合建立多种分析图谱,分析预测研究热点和发展动态,为科研人员提供学术帮助,推荐与之相匹配的科研项目、合作伙伴等,为科研人员提供个性化服务。

## 5 优化决策机制

### 5.1 由精细化的单项管理走向趋势化的复合管理

大数据时代的科学研究呈现出数据密集和数据驱动的特点。例如在个人的科研行为角度,越来越多的科研人员开始采用数字化的数据获取、挖掘和服务。在科研管理者的决策角度,传统的科研管理对单项管理要求精细化和专业化,而在大数据时代,则希望对多源异构的实时数据和历史数据进行挖掘与分析。开放知识网络通过构建异构的网络结构,实现不同来源、不同学科甚至不同语言之间知识的融合与存储,可以满足大数据时代对数据与信息多样化的需求。同时,开放知识网络中包含了大量的时空信息,通过对这些时空信息进行动态实时的更新,构建基于开放知识网络的动态数据分析与挖掘方法,可以为科研管理者提供更加准确、实时的决策与管理支持。

### 5.2 对外信息公开,加快科研成果转化

在大数据时代,科研管理人员的任务不仅是收集数据,还包括数据的对外输出。把科研成果推向社会,促进社会发展已经是科研管理工作中一项重要的工作。近年来,国家对高校和科研机构的科研成果技术转移和转化越来越重视,但科研成

果转化为现实生产力的速度较为缓慢。通过开放知识网络技术,将不同形式的科研管理大数据融合到同一个网络中,对该网络进行深度的分析、挖掘与预测,可以将科研人员、项目成果等信息对外进行有针对性的推荐,同时也可以将外界的信息融合到网络结构中,把社会需求、企业关注信息、项目合作需求及时推荐给科研人员,并借鉴社交网络的信息传播方式推广个人成果,提高科研成果的关注度和影响力。开放知识网络的构建可以帮助科研管理人员清楚地发现不同科研实体间的关系,帮助不同单位的科研管理人员建立良好的合作关系,利用社交媒体、学术会议等多种形式及时发布科研成果和创新研究。

## 6 结束语

大数据极大地促进了科研创新,同时科技成果大爆发也给高校和科研机构的科研管理提出了新的、更高的要求。如何利用好积累的数据资源,并通过各种信息化技术手段,不断获取科研人员、科研活动中的大数据资源,将科研管理大数据高效运转起来,建立以大数据为驱动,与国家知识创新体系发展相适应的科研管理生态系统,真正实现紧跟科技发展的脚步,发挥大数据时代信息共享、智慧决策等优势,是现阶段科研管理的主要发展方向。同时,要强化基础环境的综合应用和服务管理能力,使科研管理人员切实体会到大数据驱动的科研管理的优势,从而带动科研管理的数字化、信息化发展。未来,可以在信息化的基础上向智能化方向转变,使用开放网络技术对信息进行智能化管理,从而提供高效率、高质量的科研管理服务。

## 参考文献:

- [1] 李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. 中国科学院院刊, 2012, 27(6): 647-657.  
LI G J, CHENG X Q. Research status and scientific thinking of big data[J]. Bulletin of the Chinese Academy of Sciences, 2012, 27(6): 647-657.
- [2] 王元卓, 靳小龙, 程学旗. 网络大数据: 现状与展望[J]. 计算机学报, 2013, 36(6): 1125-1138.  
WANG Y Z, JIN X L, CHENG X Q. Network big data: present and future[J]. Chinese Journal of Computers, 2013, 36(6): 1125-1138.
- [3] 王元卓, 贾岩涛, 刘大伟, 等. 基于开放网络知识的信息检索与数据挖掘[J]. 计算机研究与发展, 2015, 52(2): 456-474.  
WANG Y Z, JIA Y T, LIU D W, et al. Open web knowledge aided information search and data mining[J]. Journal of Computer Research and Development, 2015, 52(2): 456-474.
- [4] ETZIONI O, CAFARELLA M, DOWNEY D, et al. Web-scale information extraction in knowitall: (preliminary results)[C]//The 13th International Conference on World Wide Web, May 17-20, 2004, New York, USA. New York: ACM Press, 2004: 100-110.
- [5] BANKO M, CAFARELLA M J, SODERLAND S, et al. Open information extraction from the web[C]//The 20th International Joint Conference on Artificial Intelligence, January 6-12, 2007, Hyderabad, India. New York: ACM Press, 2007: 2670-2676.
- [6] CARLSON A, BETTERIDGE J, KISIEL B, et al. Toward an architecture for neverending language learning[C]//The 24th AAAI Conference on Artificial Intelligence, July 11-15, 2010, Atlanta, Georgia. Menlo Park: AAAI Press, 2010: 1306-1313.
- [7] WU W, LI H, WANG H, et al. Probase: a probabilistic taxonomy for text understanding[C]//The 2012 ACM SIGMOD International Conference on Management of Data, May 20-24, 2012, Scottsdale, USA. New York: ACM Press, 2012: 481-492.
- [8] NAKASHOLE N, THEOBALD M, WEIKUM G. Scalable knowledge harvesting with high precision and high recall[C]//The 4th ACM International Conference on Web Search and Data Mining, February 9-12, 2011, Hong Kong, China. New York: ACM Press, 2011: 227-236.
- [9] SUCHANEK F M, SOZIO M, WEIKUM G. SOFIE: a self-organizing framework for information extraction[C]// The 18th International Conference on World Wide Web, April 20-24, 2009, Madrid, Spain. New York: ACM Press, 2009: 631-640.
- [10] AUER S, BIZER C, KOBILAROV G, et al. DBpedia: a nucleus for a web of open data[C]//The 6th International The Semantic Web and 2nd Asian Conference on Asian Semantic Web Conference, November 11-15, 2007, Busan, Korea. Piscataway: IEEE Press, 2007: 722-735.
- [11] SUCHANEK F, KASNECI G, WEIKUM G. YAGO: a core of semantic knowledge[C]//The 16th International Conference on World Wide Web, May 8-12, 2007, Banff, Canada. New York: ACM Press, 2007: 697-706.
- [12] PONZETTO S, NAVIGLI R. Large-scale taxonomy mapping for restructuring and integrating Wikipedia[C]//The 21st International Joint Conference on Artificial Intelligence, July 11-17, 2009, Pasadena, USA. San Francisco: Morgan Kaufmann, 2009: 2083-2088.
- [13] PONZETTO S, STRUBE M. Taxonomy induction based on a collaboratively built knowledge repository[J]. Artificial Intelligence, 2011, 175(9-10): 1737-1756.
- [14] 林海伦, 王元卓, 贾岩涛, 等. 面向网络大数据的知识融合方法综述[J]. 计算机学报, 2017, 40(1): 1-27.  
LIN H L, WANG Y Z, JIA Y T, et al.

- Network big data oriented knowledge fusion methods: a survey[J]. Chinese Journal of Computers, 2017, 40(1): 1-27.
- [15] 孟小峰, 慈祥. 大数据管理: 概念、技术与挑战[J]. 计算机研究与发展, 2013(1): 146-169.  
MENG X F, CI X. Big data management: concepts, techniques and challenges[J]. Journal of Computer Research and Development, 2013(1): 146-169.
- [16] 梁家峰, 亓振华. 适应与创新: 大数据时代的高校思想政治教育工作[J]. 思想教育研究, 2013(6): 63-67.  
LIANG J F, QI Z H. Adaptation and innovation: ideological and political education in colleges and universities in the age of big data[J]. Studies in Ideological Education, 2013(6): 63-67.
- [17] 董振东, 董强, 郝长伶. 知网的理论发现[J]. 中文信息学报, 2007, 21(4): 3-9.  
DONG Z D, DONG Q, HAO C L. Theoretical findings of HowNet[J]. Journal of Chinese Information Processing, 2007, 21(4): 3-9.
- [18] 梅立军, 周强, 臧路, 等. 知网与同义词词林的信息融合研究[J]. 中文信息学报, 2005, 19(1): 63-70.  
MEI L J, ZHOU Q, ZANG L, et al. Merge information in HowNet and TongYiCi CiLin[J]. Journal of Chinese Information Processing, 2005, 19(1): 63-70.
- [19] 黄曾阳. HNC理论概要[J]. 中文信息学报, 1997, 11(4): 11-20.  
HUANG Z Y. Summary of HNC theory[J]. Journal of Chinese Information Processing, 1997, 11(4): 11-20.
- [20] LENAT D, GUHA R. Building large knowledge-based systems: representation and inference in the Cyc project[M]. Boston: Addison-Wesley, 1989.

### 作者简介



**罗瑞丽** (1963- ), 女, 中国科学院计算技术研究所副研究员, 科研处处长, 中国科学院网络计算创新研究院筹建办公室负责人, 主要研究方向为计算数学、科研大数据管理, 先后承担国家科技攻关计划项目“KJ8920大型机系统研制”任务、大型商业应用系统等研究项目。



**王元卓** (1978- ), 男, 博士, 中国科学院计算技术研究所研究员、博士生导师、科研处副处长, 大数据研究院院长, 大安全方向总工程师, 大数据分析系统国家工程实验室开放知识网络计算研究方向负责人, 主要研究方向为网络大数据分析、开放知识计算、网络群体行为分析、网络与信息安全、随机博弈模型等。

收稿日期: 2019-07-15

基金项目: 国家自然科学基金资助项目 (No.91646120, No.61572469)

Foundation Items: The National Natural Science Foundation of China(No.91646120, No.61572469)