

基于带权三元闭包的知识图谱的构建方法研究^{*}

孙昊天 杨良斌

(国际关系学院信息科技学院 北京 100091)

摘要 [目的/意义]随着时代的发展,基于关键字匹配的传统搜索方式已经不能满足人们的需求,而知识图谱正是为了满足人们在海量数据中搜索信息而产生的。目前构建社交平台的知识图谱是一个人们极为需求但研究较少的领域。[方法/过程]爬取新浪微博四个官博取其自2017年8月1日至2018年8月1日所发微博的正文及其他相关数据,并以其39 023条数据为样本。在利用共现方式标注共现次数的基础上,实现新的基于带权三元闭包来构建时政类微博的知识图谱,并改变相关参数与人工标注相较得出其最佳参数,从而生成最佳知识图谱。[结果/结论]运用以带权三元闭包为基础构建时政类微博知识图谱的方法,在一定参数约束下,可以生成符合期望的以亲密程度为关系的知识图谱。

关键词 微博 知识图谱 三元闭包

中图分类号 G350

文献标识码 A

文章编号 1002-1965(2019)06-0168-06

引用格式 孙昊天,杨良斌.基于带权三元闭包的知识图谱的构建方法研究[J].情报杂志,2019,38(6):168-173.

DOI 10.3969/j.issn.1002-1965.2019.06.025

Research on the Construction Method of Knowledge Graph Based on Weighted Triadic Closure

Sun Haotian Yang Liangbin

(School of Information Technology, University of International Relations, Beijing 100091)

Abstract [Purpose/Significance] As the world develops, people's needs have changed. The traditional keyword matching search method has not satisfied people's needs. And the use of knowledge graph search has emerged, which enables people to obtain information in massive data. Currently, the knowledge graph that builds a social platform is an area that is extremely demanded but less researched. [Method/Process] The data in this article comes from Sina Weibo, including the microblog content and related data of four of the official Weibo posts from August 1, 2017 to August 1, 2018, and takes its 39 023 data as a sample. On the basis of using the co-occurrence method to mark the co-occurrence times, realizing the knowledge graph of current political Weibo based on the construction of weighted triadic closure, and the relevant parameters are compared with the manual annotation to obtain the optimal parameters. Through such a process, the best knowledge graph is finally obtained. [Results/Conclusion] With the method of constructing political microblog knowledge graph based on weighted triadic closure, and under certain parameter constraints, a knowledge graph that matches the expected degree of intimacy can be generated.

Key words Weibo knowledge graph triadic closure

0 引言

对于搜索引擎来说,网上信息的多样化已成为了巨大挑战。传统的基于关键词匹配的搜索引擎,即利

用向量空间模型(VSM)来获取搜索结果^[1],已经无法满足人们对于模糊搜索的需要。故从自然语言的角度出发,真正理解用户的需求,让搜索引擎能够知道用户搜索的目标,实现人与Web之间的问答、会话、以至让

收稿日期:2018-12-04

修回日期:2019-01-14

基金项目:中央高校基本科研业务经费“社会网络分析视角下多数据融合的知识图谱构建方法研究”(编号:3262018T30);中央在京高校人才培养共建项目:人才培养共建项目-大学生科研训练项目-信息科技学院数据科学与大数据技术科研项目。

作者简介:孙昊天(ORCID:0000-0002-3911-0675),男,1995年生,硕士研究生,研究方向:网络空间安全、数据分析;杨良斌(ORCID:0000-0001-7214-9750),男,1976年生,博士,教授,研究方向:社会网络分析、信息可视化。

通信作者:杨良斌

搜索引擎预测和理解用户的搜索目的,是未来搜索引擎发展的方向,基于此 Google 公司在 2012 年提出知识图谱项目^[2],其本质是一个知识库,基本元素是实体及实体之间的关系,即“实体-关系-实体”三元组。知识图谱在搜索引擎中的应用,不仅能够更有效地向用户展示核心的内容信息,同时也可以为用户展示比较完整的知识体系。

随着社交平台的发展,社交网络领域在人们日常生活中日益重要,社交网络中的搜索业务亦是直线上升,基于关键字匹配的传统搜索方式已经不能满足人们的需求。而在我国,新浪微博不仅是广泛使用的社交媒体,还兼具信息交流的功能,亦是突发事件传播扩散的重要平台。构建其知识图谱最重要的就是其中实体之间的关系的量化,这其中的关系即对于事物而言其联系的紧密程度往往代表了两个事物的相关性,对于事物与事物间的事态发展与变化有着极大的影响。基于此本文提出了一种新的知识图谱的构建思路,即利用带有权重的三元闭包构建新浪微博的知识图谱,为微博社区的搜索提供新的理论和技术支持,并借以此帮助人们实现在社交平台对于信息进行更加完善与深入搜索,推动该领域的发展。

1 国内外研究现状

三元闭包 (Triadic Closure) 作为网络最基本的局部结构和重要的链接生成机制,在网络模型的研究中不断被提到,而其本身是一种非常直观和自然的关系的描述^[3]。同时,三元闭包目前应用广泛,其作为关系强联系与弱联系的一部分,许云峰等人将其利用于社区发现算法的研究^[4],胡昌平等人引用其作为知识网络分析的基本单元^[5],高杨等人将其作为链路预测算法的一部分^[6]。可见三元闭包是一个高效且准确的算法,目前对其的研究应用主要集中于推理分析,但常用的三元闭包原理不涉及权重计算,这对中实体关系的量化是不利的故很少用于关系的量化处理。

而知识图谱目前在各个领域已经得到了广泛的应用,谢刚等人做出了较为完善的知识图谱构建技术综述与细节补充^[7],同时冯新翎等人为区分“科学知识图谱”与“知识图谱”两大概念,对其进行了比较分析^[8],在知识图谱技术快速发展期间,国外各大公司构建了各种知识图谱,这些知识图谱包括 Freebase^[9]、Wikidata^[10]、DBpedia^[11]、YAGO^[12]等,国内赵雪芹等人则利用知识图谱对我国网络传播的影响进行了一定的研究^[13],贾丙静等人利用知识图谱帮助学生更好地完成 C 语言“慕课”学习^[14]。当前国外主要是企业式的批量应用即生成相关知识图谱后推广向大众产生盈利,而国内对于知识图谱的研究偏向于利用传统方法

构建知识图谱后基于知识图谱的基础上来完成其他项目的研究。同时国内外也有一定数量的对知识图谱本身具体构建方法的研究,如王巍巍等人做出了影视方面的知识图谱的构建研究^[15];鄂世嘉等人提出了基于中文百科的端到端自动化构建知识图谱解决方案^[16];国外则在知识图谱数据表示与知识获取方法方面提出了具有价值的思路^[17-18]。

而对于作为社交平台的微博而言,目前杜亚军等人对微博知识图谱构建方法展开了一定的理论研究^[19],但其方法不够成熟且没有实现,尤其关于实体关系抽取部分,在现有的研究中,还没有针对微博帖子中实体之间关系抽取的具体研究,解决该问题实现微博知识图谱构建即是本文所研究的。如何为三元闭包赋予权重,并以时政类新闻作为数据集基于带权三元闭包来构建微博知识图谱,是本文研究的主要问题。

2 研究方法与数据来源

2.1 研究的前期工作 通过爬取新浪微博的时政新闻用于构建知识图谱,具体构建分为多步进行。首先爬取具有连续性且长时段内的微博内容,经数据清洗后进行分词工作,提取有意义的实体,后进行实体关系抽取与确认。通常人们认为,相关事物通常会同时出现,同时出现的事物往往也具有一定相关性^[20]。故共现的关键词之间必然具有一定的联系,往往这种词共现关系的研究具有重要意义^[21],其已被运用于科学知识图谱中^[22]。基于此,本文采用了共现方法来提取实体之间的关系,其有利于带权三元闭包中的权重生成。具体而言,若两个实体在同一个微博帖子中出现,就可认为他们之间是有直接联系的。这里以一个原始博文作为示例,“【收藏! 中国共产党十九届三中全会公报】中国共产党第十九届中央委员会第三次全体会议,于 2018 年 2 月 26 日至 28 日在北京举行。全会由中央政治局主持。中央委员会总书记习近平作了重要讲话…公报全文↓↓转发学习!”此微博经过数据清洗与分词之后,其中有意义的实体包括“中国共产党”“中央政治局”与“习近平”。这其中的所有的实体相互之间必然存在关系,并且互相之间的关系可以量化为“1”,也就是互相共同出现一次。同时因为在其他的微博中也可能存在这些已经出现过的实体再次共现,可以认为共现的越多则可以说明他们的关系越密切,即任意两个有关系的实体可以根据现有数据所出现的共现次数来表示两个实体之间的亲密程度。同时则这种亲密程度也表示实体间联系的紧密与否,事物间联系的紧密程度所代表的意义不再复述。

2.2 带权三元闭包 在得到实体和实体间的关系后,本文利用三元闭包来构建初级知识图谱,以下简述

三元闭包原理。

在现实中,三元闭包所体现的原理很容易从生活中找到与之相关的例子。如生活中常见的三人行组合,即当A与B有一个共同的朋友C,那么A与B就可能成为朋友。其原因在于,一方面,因为共同的朋友C作为A与B认识的桥梁,这就使得他们认识的概率增大,并且C的存在可以为A和B提供了陌生事物之间所缺乏的中介,从而推动互相之间成为朋友。另一方面,C有将A和B撮合成朋友的动机,即A和B不是朋友的事实会成为3人之间建立友谊的潜在推动力。这个原理的思想基础是基于早期的社会网络分析提出的相关理论。

如果两个人有同一个朋友,随着时间的推移,那么这两个人成为朋友的可能性就会增大。如果基于3人友谊构成一张社交网络图,即如图1所示:

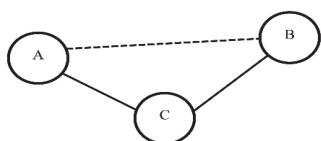


图1 三元闭包图例

在该社交网络中,A和B有一个共同的朋友C,因此A和B成为朋友的概率就会增加从而建立某种联系,即形成如图1中所示的“虚线”的情况。在网络结构中,三角形结构被认为是最稳定的。AB边在该三角形结构中起到了辅助闭合的作用。也正是基于此,该理论被称为“三元闭包”或者“三元闭合”^[23]。那么我们可以得到,在社交平台的事物中,3个事物若存在如上关系,那么他们3个互相之间必然也存在一定联系,同样在微博中也适用该原理。

综上所述,在原始的三元闭包原理中其中的边是无权重边,不过依据共现方法,可以利用共现次数的形式来展示其中的两个实体之间实际存在联系的边的权重,用以标注两个实体之间的亲密程度,以此进行实体之间关系的量化。例如,基于本文的数据集中,现提取了具有代表性的7个实体同时抽取了其中的关系,并且关系也已经基于共现量化为次数,以实体作为节点,将实体之间的关系作为边,并规定这些表示微博中具有的直接联系的边称为实边,基于此构成只含实边的图2,该图的实体共现次数由表1表示,如图2所示。

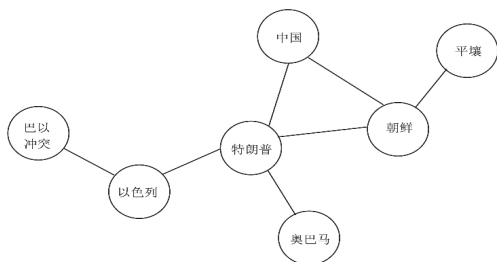


图2 仅含实边图例

如下为仅包含图2实体共现次数的表1:

表1 共现次数表

	特朗普	中国	朝鲜	平壤	奥巴马	以色列	巴以冲突
特朗普	-	113	62	0	8	12	0
中国		-	44	0	0	0	0
朝鲜			-	17	0	0	0
平壤				-	0	0	0
奥巴马					-	0	0
以色列						-	3
巴以冲突							-

根据三元闭包的原理,可以推理出图中部分实体也存在一定联系,由此联系形成新的边可以帮助实体间形成稳定的三角形^[24]。由此产生出并非原始存在的边,即前文的“虚线”,本文称为“推理边”且规定推理边关系值应为“1”,这是因为推理边不应大于实边。同时由于在数量关系的约束下,共现次数与实体之间的亲密程度并非线性关系,即在存在一定共现的次数之后数值的增长已经不能线性表示实体之间的亲密程度,故应利用降低线性增长的对数方法重新计算实体间关系值^[25],并且研究中对其中底数设置为一个参数进行调整来优化并根据结果取最优值。该参数本文称为“X”,这里取“X”设为自然对数来进行计算,即实体间共现次数将以e为底计算对数,四舍五入取整,同时数值加1,并且规定最小值为2,这是因为实边不应小于推理边,以计算后的值作为实边的关系值,这些值称为初级权重,至此形成了带有初级权重三元闭包,如图3所示:

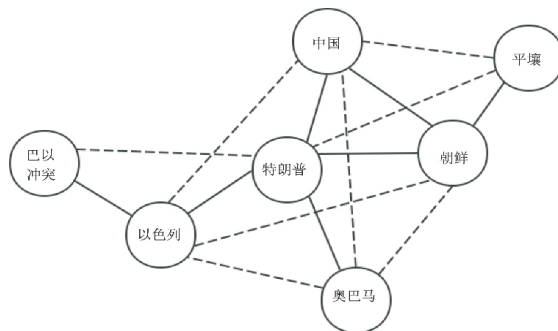


图3 含推理边图例

并得到一个更新后的带有初级权重三元闭包的表2:

表2 初级权重表

	特朗普	中国	朝鲜	平壤	奥巴马	以色列	巴以冲突
特朗普	-	6	5	1	3	3	1
中国		-	5	1	1	1	0
朝鲜			-	3	1	1	0
平壤				-	0	0	0
奥巴马					-	1	0
以色列						-	2
巴以冲突							-

根据前文所述,形成三角形的图是相对稳定的。

故在图中如果存在3条边能够组成一个三角形, 本文将这种情况定义为闭合。

同时, 规定完备闭合至少包括两条实边, 闭合则只需3条边形成三角形, 如此区分是由于推理边应包含所有推理而实边只需包含自身推理导致的。并且由图3中可以看出, “特朗普”“中国”“朝鲜”仅依靠实边已经构成了一个三角形。因此, 当计算实体之间关系权重的时候, 如果完备闭合的3条边都是由实边组成, 即不包含由三元闭包原理所生成的推理边, 对于计算这样边的权重, 其权重值应该相应的增大。基于此可以得出一条边在图中的真实权重。故综上可根据公式(1)重新计算权重:

$$W_{ij} = \frac{w_a + w_b * 2}{W_A + W_B * 2} \quad 0 < W_{ij} < 1 \quad (1)$$

其中 W_{ij} 代表所计算边的权重。分母部分只包含完备闭合, 即 W_A 表示整个图中包含有推理边的完备闭合的边并将其初级权重相加, W_B 表示整个图中只包含了3条实边的完备闭合的边并将其初级权重相加。类似的, 公式上半部分即是包含了所计算边的闭合, 当计算边为实边时应基于完备闭合, 推理边应基于闭合, 即 w_a 表示所计算边包含具有推理边的完备闭合或闭合的数量与其初级权重相乘, 同时 w_b 表示只包含了3条实边的完备闭合数量与其所计算边的初级权重相乘, 在计算过程中产生的权重值称为权重。由所示示例中, 可以看出完备闭合的总数为具有推理边的有8个, 只有3条实边的有一个, 总权重为:

$$8 * 1 + (6 + 5 + 5) * 2 + 3 + 3 + 3 + 2 = 51$$

	A	B	D	E	F
31468	2017/8/10 22:40	【地震后, 他们在路边免费分发食物和水】#四川九寨沟地震#后, 各方救援力量投入到5	1335	1156	8564
31469	2017/8/10 22:30	【太平洋金色晚霞如画】8月10日, 我国新一代远洋综合科考船“科学”号在西太平洋行驶	39	20	176
31470	2017/8/10 22:10	【人民微评: 返回结账更可贵】#四川九寨沟地震#来袭, 逃生比付账更重要; 化险为夷	696	616	7915
31471	2017/8/10 21:16	#九寨沟七级地震# 【九寨沟七级地震转移人员约7.1万人 灾险排查加快推进】四川九寨沟7	6	6	45
31472	2017/8/10 20:57	#精河地震# 【新疆精河县地震灾区由救援转入保障和重建工作】记者从新疆精河县了解	10	5	61
31473	2017/8/10 20:13	【信任中国! 菲律宾外长这样反击南海问题, 恐遭外媒】2日至8日, 菲律宾主办的东盟	108	144	522
31474	2017/8/10 19:06	【信任中国! 菲律宾外长这样反击南海问题, 恐遭外媒】2日至8日, 菲律宾主办的东盟	534	781	3254
31475	2017/8/10 19:03	#九寨沟七级地震# 【首批灾区震后高分辨率影像图用于应急救援】国家测绘地理信息局	9	15	43
31476	2017/8/10 18:52	【地震“逃单”后续: 震后不少顾客返回结账】四川九寨沟地震发生后, 西安震感强烈, 某	90	118	1120
31477	2017/8/10 18:40	#九寨沟七级地震# 【滑坡不断 道路疏通任务艰巨】8月10日, 九寨沟县至九寨沟景区的道	9	7	39
31478	2017/8/10 17:50	#九寨沟七级地震# 【第77集团军出动地面及空中力量搜救被困群众】8月10日, 西部战区	13	15	83
31479	2017/8/10 17:03	#人民直播# 【#四川九寨沟地震#发布会 关注最新情况】截至8月10日12时, 经初步核	333	554	1377
31480	2017/8/10 16:46	【中国科学家第一次实现量子卫星全部科学目标, 一段视频带你看懂“墨子号”的神奇使命	3	3	11
31481	2017/8/10 16:41	【最高法: 严厉打击以互联网金融名义进行的违法犯罪】记者10日从最高人民法院获悉,	13	9	20
31482	2017/8/10 16:37	【地震“逃单”后续: 震后不少顾客返回结账】#四川九寨沟地震#后, 西安震感强烈, 某商	2127	1525	18090
31483	2017/8/10 16:35	【地震数万游客大转移 他们说: 一个不能少!】据央视, 8月8日, 四川九寨沟7.0级地震	87	63	655
31484	2017/8/10 16:13	#九寨沟七级地震# 【九寨沟地震致20人遇难 431人受伤】记者从四川省政府新闻办获悉,	50	23	22
31485	2017/8/10 16:13	【欧洲“毒鸡蛋”风波愈演愈烈】受杀虫剂氟虫腈污染的“毒鸡蛋”风波在欧洲愈演愈烈, 不	4	10	5

图4 部分数据展示图

后对这些数据进行初步数据清洗, 主要采用与时政新闻有关的文本, 并综合考虑文本长度, 以及其发送时间、评论数、点赞数、转发数等, 进行完善的人工数据标注与清洗, 同时去掉了部分类似“#早安, 中国#”的伪时政类文本, 最后得到14 653条可用的包含时政新闻的文本数据。

现在举例计算“中国”与“特朗普”边的权重。由图中所示或由矩阵可以推导出, 该边存在一个包含了只有3条实边及2个具有推理边的完备闭合, 故该边权重为:

$$6 * 2 + 6 + 6 = 24$$

小数部分采取四舍五入后取两位小数, 故该边权重为 $24/51 = 0.47$, 利用该方法可以计算出所有边的权重。

示例中计算所有边权重后根据每条边的权重所形成一个带权三元闭包的表3:

表3 权重表

	特朗普	中国	朝鲜	平壤	奥巴马	以色列	巴以冲突
特朗普	-	0.59	0.49	0.04	0.18	0.24	0.02
中国		-	0.49	0.04	0.06	0.06	0
朝鲜			-	0.12	0.06	0.06	0
平壤				-	0	0	0
奥巴马					-	0.06	0
以色列						-	0.04
巴以冲突							-

至此, 带权三元闭包构建工作完成, 基于此思路就可以构建一个时政新闻的微博知识图谱。

2.3 数据来源 利用“八爪鱼”爬虫软件^[26]爬取微博中四个较为主流的官博, 即“人民日报”“新华观点”“紫光阁”“中国网”的数据, 数据包括自2017年8月1日至2018年8月1日所发微博的所有博文以及其发送时间、评论数、点赞数、转发数等, 共39023条数据, 示例如下:

3 实验与分析

3.1 实验过程 试验主要基于Python语言的环境下实现基于带权三元闭包的微博时政新闻知识图谱构建, 并将其与人工标注比较得出准确率。

清洗文本数据集后, 利用Python开源分词工具

“jieba”分词^[27]对数据集进行分词处理,并基于此提取出相关实体,将完成后的结果与数据集利用基于经典词袋模型^[28]进行共现查找来计算其共现次数。并将实体部分含有如“中国”“中华人民共和国”等表示同一意思但以不同表达出现的实体合并,对此的处理称为共指消解^[29],由于微博热点事件属于新闻语料,其语义和语法资源尚未完善,所以用语义和语法结合的方法效果会受到影响,因此采用大规模语料库与数据统计分析方法相结合的方法进行共指消解^[30]。同时将相同实体的共现次数归为一个实体,如“中国”与“中华人民共和国”和“特朗普”共现出现113次,那么共指消解后“中国”与“特朗普”的共现次数即为113次。

由于所选用数据集均为时政新闻相关数据,故在仅存在实边的基础上便没有出现孤立实体或多个实体集,即所有实体间必能经过数条实边联系,后基于编译软件JetBrains PyCharm Community Edition利用Python语言完成前文所述的带权三元闭包的实现。

同时,由于三元闭包会导致大量边的加入,即使关系存在但过于稀疏会使得知识图谱过于冗杂,本文基于改进PFNET算法^[31]来计算其稀疏程度并去除稀疏边,计算前后其中边权重不变。最后,基于在数据清洗阶段只留下了时政新闻文本数据的原因,最终知识图谱中所有实体都能经过一定的边形成联系即没有出现数个完全无关的独立的图谱,综上所述,形成了完备的“实体-关系-实体”的三元组结构,完成了微博时政新闻知识图谱的构建。

3.2 结果与分析 为分析结果,现根据所抽取的实体,依据微博所带的话题分类模式与百科知识,依照人工标注实体之间的关系,并将其关系的亲密程度即联系的紧密程度标注为“上”“中”“下”3个等级,并以此来作为评价标准。其中权重的评价以以下标准为主:

$$\begin{cases} 0.4 < W_{ij} < 1 & \text{上} \\ 0.2 < W_{ij} \leq 0.4 & \text{中} \\ 0 < W_{ij} \leq 0.2 & \text{下} \end{cases}$$

基于该标准,可以得出带权三元闭包与人工标注相比较的准确率,并改变前文所述的“X”值,得出相对最优的准确率,由图5表示准确率的变化图:

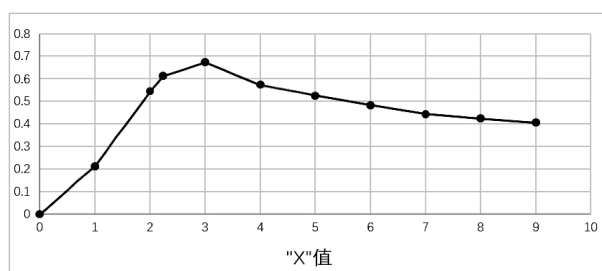


图5 准确率图

由图5可知,“X”为3时为其最优值,同时此处给出与前文3.2小节“带权三元闭包”中表3相同类型的表5,即取最优值“3”时的权重表,从中也容易得出基于此最优值所得出的知识图谱更加合理。

表5 最优值权重表

	特朗普	中国	朝鲜	平壤	奥巴马	以色列	巴以冲突
特朗普	-	0.53	0.53	0.04	0.21	0.09	0.02
中国		-	0.43	0.04	0.06	0.06	0
朝鲜			-	0.13	0.06	0.06	0
平壤				-	0	0	0
奥巴马					-	0.06	0
以色列						-	0.04
巴以冲突							-

可见,基于带权三元闭包原理表示微博中实体亲密程度以表示关系并基于此构建知识图谱是较为可靠的,在合理数值的约束下与人工标注相比较准确率达67.28%。

4 结论与展望

本文由特定领域知识图谱构建的角度出发,提出了基于带权三元闭包构建微博时政新闻知识图谱的方法,并探究了这一方法的合理性与可靠性。结果表明该方法在合理数值的约束下是一个较为可靠的方法。同时该方法扩展性极好,该方法构建的知识图谱有利于利用社会网络分析的方法展开相关研究,限于篇幅,后续研究将对此进行讨论。

试验研究结论接近68%,其准确率符合预期,不过应该会有更好的设计思路,且特定领域的知识图谱作为一个知识库,有助于帮助用户从海量信息源中迅速找到完整的有价值的信息,来帮助用户发现实体之间深层次的关系。用本文方法构建的知识图谱表示了实体间的联系即实体间关系的亲密程度,但对其他内容包含的不够充分,这也是作者今后工作的兴趣点,希望有更多的学者可以关注该领域。

参考文献

- [1] Sidorov G, Gelbukh A, Adorno H G, et al. Poster: Soft similarity and soft cosine measure: Similarity of features in vector space model[C]. Taller de Tecnologías del Lenguaje 2014: 1-14.
- [2] 朱木易洁, 鲍秉坤, 徐常胜. 知识图谱发展与构建的研究进展[J]. 南京信息工程大学学报(自然科学版), 2017, 9(6): 575-582.
- [3] Easley D, Kleinberg J. Networks, crowds, and markets: Reasoning about a highly connected world[M]. Cambridge, UK: Cambridge University Press, 2010: 32-54.
- [4] 许云峰, 赵宁, 郝雪君, 等. 基于三元闭包和会员闭包的社区发现算法研究[J]. 河北科技大学学报, 2014, 35(1): 103-108.
- [5] 胡昌平, 陈果. 层次视角下概念知识网络的三元关系形态研

- 究[J]. 图书情报工作 2014 58(4): 11-16.
- [6] 高 杨, 张燕平, 钱付兰 等. 基于三元闭包的节点相似性链路预测算法[J]. 计算机科学与探索 2017 11(5): 822-832.
- [7] 谢 刚. 知识图谱精化研究综述[J]. 电子技术应用 2018 44(9): 29-33 38.
- [8] 冯新翎, 何 胜, 熊太纯 等. “科学知识图谱”与“Google 知识图谱”比较分析——基于知识管理理论视角[J]. 情报杂志, 2017 36(1): 149-153.
- [9] Bollacker K, Evans C, Paritosh P, et al. Free base: a collaboratively created graph database for structuring human knowledge [C]. New York: ACM 2008: 1247-1250.
- [10] Vrandeć D, Krotzsch M. Wikidata: A free collaborative knowledge base[J]. Communications of the ACM 2014 57(10): 78-85.
- [11] Lehmann J, Isele R, Jakob M, et al. DBpedia—a large-scale multilingual knowledge base extracted from Wikipedia[J]. Semantic Web 2015 6(2): 167-195.
- [12] Rebele T, Suchanek F, Hoffart J, et al. YAGO: A multilingual knowledge base from Wikipedia, Wordnet and GeoNames[M]. Berlin Heidelberg: Springer International Publishing 2016: 177-185.
- [13] 赵雪芹, 王少春, 刘雅宁 等. 我国网络传播研究的知识图谱分析[J]. 情报探索 2018(9): 123-129.
- [14] 贾丙静, 葛 华, 李德胜. “慕课”时代 C 语言知识图谱的构建[J]. 绵阳师范学院学报 2018(11): 100-103.
- [15] 王巍巍, 王志刚, 潘亮铭 等. 双语影视知识图谱的构建研究[J]. 北京大学学报(自然科学版) 2016 52(1): 25-34.
- [16] 鄂世嘉, 林培裕, 向 阳. 自动化构建的中文知识图谱系统[J]. 计算机应用 2016 36(4): 992-996 1001.
- [17] Deroose P, Shen W, Chen F, et al. Building structured web community portals: A top-down, compositional and incremental approach [C]. VLDB 2007: Proceedings of the 33rd International Conference on Very Large Data Bases, New York: ACM 2007: 399-410.
- [18] Niu F, Zhang C, et al. DeepDive: Web-scale knowledge-base construction using statistical learning and inference [EB/OL]. (2014-10-10). http://www.cs.stanford.edu/people/chris-smre/papers/deepdive_vlds.pdf.
- [19] 杜亚军, 吴 越. 微博知识图谱构建方法研究[J]. 西华大学学报(自然科学版) 2015 34(1): 27-35 89.
- [20] 田红梅. 试论图书馆从信息服务走向知识服务[J]. 情报理论与实践 2003(4): 312-314.
- [21] Aalabaf-Sabaghi M. Networks, crowds and markets: Reasoning about a highly connected world[J]. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2012: 175: 1072-1073.
- [22] 储节旺. 国内图书情报领域知识管理研究方法的共词分析[J]. 情报杂志 2012 31(11): 45-50.
- [23] 孟永伟, 王晓英, 沈 茜 等. 基于三角形演化机制的社会网络模型研究[J]. 计算机工程与应用 2016 52(8): 111-114.
- [24] 吴 震. 社会网络建模分析及挖掘算法研究[D]. 杭州: 浙江大学 2017: 15-120.
- [25] Blondel V D, Senellart P P. Automatic extraction of synonyms in a dictionary [EB/OL]. (2013-06-07). <http://pierre.senellart.com/publications/blondel2002automatic.pdf>.
- [26] 崔玉洁, 廖 坤. 借助八爪鱼采集器实现过刊元数据的自动提取[J]. 编辑学报 2016 28(5): 485-488.
- [27] Bharath Ramesh, Cheng Xiang, Tong Heng Lee. Shape classification using invariant features and contextual information in the bag-of-words model [J]. Pattern Recognition 2015 48(3): 1894-906.
- [28] 郎 君, 秦 冰, 刘 挺 等. 篇章共指消解研究综述[J]. 汉语语言与计算学报 2007 7(4): 227-235.
- [29] Hobbs J R. Resolving pronoun references [J]. Journal of Lingua, 1978 44: 311-338.
- [30] 邢 彪, 根绒切机多吉. 基于 jieba 分词搜索与 SSM 框架的电子商城购物系统[J]. 信息与电脑(理论版) 2018(7): 104-105 108.
- [31] Du Z D. An Efficient algorithm for discovering co-occurrence concepts through pathfinder paradigm [J]. Journal of Donghua University 2006: 153-156.

(责编: 贺小利; 校对: 刘影梅)

(上接第 193 页)

以适应多元化、全球化的工作环境和研究任务。第三, 媒体型智库的资源平台和传播媒介为汇聚和发布富有价值的国际化观点、文章打下较好的媒介基础。

因此, 在当前国家积极倡导提升智库国际影响力的目标引领下, 媒体型智库的成果生产也必须积极倡导“走出去”, 以多元化思维、国际化视野去开拓新的智库成果, 与国外智库机构开展国际对话与合作, 向世界传递“中国声音”, 努力建成富有国际影响力和竞争力的中国特色新型智库。

参 考 文 献

- [1] 吴 亮, 夏 宇, 马 岩. 媒体融合发展对接中国特色新型智库建设的路径研究[J]. 中国科学院院刊 2016 31(8): 879-

887.

- [2] 刘海陵. 产品定制型媒体智库建设思考 [EB/OL]. [2018-04-13]. <http://media.people.com.cn/n1/2018/0413/c419052-29925066.html>.
- [3] 吴 亮. 媒体与智库融合路径与机理研究——以瞭望智库为例[J]. 新闻战线 2018(2): 19-23.
- [4] 毅 欧. 新闻出版业首个新型智库联盟在京成立 [EB/OL]. (2017-12-01). <http://yuqing.people.com.cn/n1/2017/1201/c209043-29679686.html>.
- [5] 孙 筠. 基于协同创新理念的产业智库联盟构建及其运行模式研究[D]. 长春: 吉林大学 2018.
- [6] 甘 恬. 瞭望智库: 媒体型智库的成长逻辑[J]. 传媒评论, 2015 265(11): 20-23.
- [7] 王 玥, 夏 宇. 智库姓党 智库有智[J]. 中国报业 2018 450(17): 20-22.

(责编: 王 菊; 校对: 刘影梅)