

社会化标注系统资源聚合的实证分析

毕 强 赵夷平 孙中秋 (吉林大学管理学院 长春 130022)

摘要 文章选取豆瓣网图书情报学科的图书标签作为实证样本数据,在对豆瓣资源聚合现状进行分析的基础上,借助社会网络分析法和主题词表等相关工具和技术,挖掘标签间的潜在共现关联关系,明晰标签语义和层级关系,在广度和深度方向实现标签资源聚合。

关键词 社会化标注系统 分众分类法 资源聚合

Empirical Analysis of Socialization Tagging System Resource Aggregation

Bi Qiang Zhao Yiping Sun Zhongqiu (School of Management, Jilin University, Changchun, 130022)

Abstract The Paper set tags of library and information science books in Durban as the empirical sample data. Based on the analysis of current situation of resources aggregation of Durban, with the help of related tools and techniques of social network analysis and thesaurus, this paper mines the potential co- occurrence relationships of tags, makes semantic and hierarchy relationships clear of tags, and achieves resources aggregation in breadth and depth directions.

Keywords social tagging system folksonomy resources aggregation

网络信息环境的变化与用户信息需求的多样化要求社会化标注系统必须实现资源聚合,而聚合方法的互补融合是资源聚合维度扩展的有效途径。因此,本文尝试融合分众分类和社会网络分析这两种资源聚合方法,拓展社会化标注系统的资源聚合维度,实现社会化标注系统资源聚合。

1 相关研究工作

1.1 社会化标注系统相关研究

国外学者对社会化标注系统的研究起源于2004年,研究时间较长,成果较丰富,而国内的研究始于2008年,时间相对较短,研究成果较少。Del. icio.us、Flickr、豆瓣等社会化标注系统的推广和流行,使社会化标注系统研究受到学者的关注,出现了研究热潮。目前国内外学者在社会化标注系统领域的研究较多集中于标签、用户和资源三个方面。

(1) 标签研究

国内外学者主要从以下三个方面研究社会化标注系统中的标签资源。

①签聚类。将用户感兴趣的资源聚合到一起,优

化标签组织体系,有效解决标签的随意性和非规范性等问题,便于用户标注和检索。Begelman 编制标签聚类算法,实现标签自动聚类,提升标签服务的用户体验^[1]。Cui 等将随机游走方法应用于标签间潜在关联关系挖掘和语义距离测量中,并在此基础上提出新的标签聚类方法^[2]。张帆提出PLSA 标签聚类算法,实现标签自动聚类,挖掘资源主题,用以支撑用户建模^[3]。窦永香等人在提取根标签时运用Porter 算法,提出依据用户的需求实现标签聚类的方法,提高资源标注的准确性^[4]。李慧宗等提出基于标签共现的标签聚类方法,用于缓解标签向量空间模型构建过程中的高维稀疏现状^[5]。

②标签推荐。将用户迫切需要的标签和资源提供给用户是社会化标注系统用户服务的宗旨与核心。Adrian 等通过计算标签间的语义关联度,提出构建基于语义关联的标签推荐系统^[6]。Rendle 等利用因式分解模型PITF 形象地展现了资源、用户和标签之间的交互过程,并依据BPR 标准构建标签推荐系统^[7]。Zhao 通过分析多种传统的标签排序算法,用异构图的形式展现标签之间的关联关系,并突出新的标签推荐算法“GRoM”^[8]。丛维强等构建立体化的张量结构模

本文系国家自然科学基金项目“语义网络环境下数字图书馆多维聚合与可视化研究”(编号:71273111)的研究成果。

型,该模型将社会网络分析法运用到潜在语义的加权统计中^[9]。孙甲申等提出一种用于降低标签混淆和提高标签准确率的主题模型^[10]。

③标签本体化。本体高度形式化和语义化的特征有助于标签规范管理,也有助于高效挖掘标签之间的潜在语义关系。Djuana等通过分析 Wordnet 标签间的层级和语义关系,构建标签本体库,用于提高标签推荐的准确度^[11]。Lee等提出大数据环境下标签本体构建的方法论“TBiSon”,用于绘制标签知识地图,发现标签间的语义关联^[12]。王翠英从理论角度分析了本体与分众分类法的优势和不足,并指出两者具有极大的互补性,将本体用于社会化标注系统资源聚合能推动标签的语义化进程^[13]。张云中则提出“Folksonomy 局部本体”的概念,剖析分众分类法和本体之间进行语义映射的机制,用以重构标签语义体系^[14]。

(2) 用户研究

用户是社会化标注系统的核心要素,为用户提供高质量的资源和服务是社会化标注系统资源聚合的最终目的。用户是研究者永恒关注的研究热点,其中用户行为、用户需求、用户模型研究较受学者关注。

用户行为研究。用户标注、检索、浏览、评论等行为是对用户行为进行分析的重要依据。Farooq等以 CiteULike 为研究样本对象,根据用户标注历史日志,从六个方面衡量标签的权重,完善 CiteULike 的标注功能^[15]。张宁在聚类分析的基础上对用户在社会化标注系统中的检索和标注行为进行详细分析^[16]。潘婵等通过计算关键词和领域标签之间的相似度,分析两者之间存在差异的原因,进而分析用户的行为特征^[17]。

用户需求研究。用户需求研究能帮助社会化标注系统挖掘用户兴趣和偏好,从而为用户提供有针对性的服务。Zhou等提出“User Recommendation”框架,用来构建用户兴趣模型,并进行个性化推荐,促进兴趣小组内成员的信息交流^[18]。余臻等借助叙词表、BP神经网络模型等工具,词频统计和权重算法,构建标签权重矩阵,用于表明用户的需求偏好^[19]。杨丹和曹俊构建名为“social bookmark”的标签推荐系统,并根据标签的相似性来判定用户的偏好^[20]。田莹颖根据 TF-IDF 算法,借助后控词表,按照时间顺序赋予标签不同的权重值,根据协同过滤算法,聚类并匹配具有相同偏好的用户兴趣小组及其关注资源,以为用户提供个性化推荐服务^[21]。

用户模型。模型是研究思路的精华与体现,同时也能指导实证研究。Kim等提出基于标签的用户模型构建研究,并将其用于用户推荐系统,因为标签能反映用户的特征和需求^[22]。Mika将社会维度引入本体,将传统本体的平面维度扩展到三维立体维度,构建社区语义网模型,用于解释用户交流过程中基于社区的语义聚合现象^[23]。Fu提出了一种分布式认知框架下理性模型,

用于解释标签在知识获取和探索性任务中对用户认知的影响,并指出标签聚类结果能清晰地体现用户的行为和认知特征^[24]。武慧娟通过分析用户间的网络关系,从动态和静态两个角度建构信息推送模型^[25]。

(3) 资源研究

国内外研究者一般不将资源作为单独的研究对象,通常情况下,资源往往与标签或用户结合起来作为统一的研究对象。Goh等采用文本分类技术对 del. icious 网站中的标签资源分类,并利用支持向量机判定文本与其标签是否匹配,从而发现关键资源,为用户推荐合适资源^[26]。Bi等根据“用户-标签-资源”三元组聚合社会化标注系统的资源,提出社会化标注系统资源检索框架,该框架极大地提高了用户检索效率^[27]。宣云干等将信息熵引入社会化标注系统资源聚合,利用信息熵重新定义权重计算公式,构建基于熵的资源聚合模型,从而提高资源的语义度^[28]。杨萌尝试将专家分类法与大众分类法结合,提出社会化标注系统资源聚合新方式——taxonomy 与 folksonomy 混合的资源聚合模型,较大程度地改善了资源聚合效果^[29]。Zhang等提出基于“用户-资源-标签”三元组的资源聚合与推荐方法,用于个性化信息推荐系统,使系统服务的准确性、新颖性和个性化水平获得显著提高^[30]。

1.2 资源聚合相关研究

作为网络资源再组织的一种重要方式,“资源聚合”一经提出,就备受国内外学者的关注,国内外学者对于网络资源聚合的研究视角较为相似,研究内容主要围绕以下几个方面展开。

①相关理论研究

资源聚合的含义一直困扰着众多研究者,国内外学者也给出不同的理解。国外学者 Oren 指出资源聚合是处理异构信息间存在的语义关联不清问题,并实现基于语义和利用关联关系的资源聚合过程^[31]。杜晖认为资源聚合是借助相关技术、手段聚集和整合网络资源,实现网络资源的高效组织、合理揭示、有效存取、最佳利用,并将结果以用户最迫切需求的方式呈现给用户的过程^[32]。杨萌等将资源聚合定义为基于资源组织、分析理论,借助聚合方法和技术,聚集和融合异构资源,以获取资源全貌,并满足不同用户需求的知识再组织过程^[33]。

资源聚合机理和模型。Liang等提出一个资源容量模型,并辅以资源定位和应用信息服务功能,用于批量计算网络资源^[34]。Ma等构建网络资源发现模型,该模型引入 P2P 技术,分为“Ordinary peer”、“Super peer Agent”和“Super Peer”三层,有效弥补了已有的资源发现方法的不足^[35]。李劲等提出馆藏资源的语义聚合模型,该模型能有效解决馆藏资源多样化和异构性问题^[36]。

②关键方法和技术研究

方法与技术是实现资源聚合重要支撑,科学有效的资源聚合方法和技术对聚合效果的影响极大,对于

③聚合成果的应用研究

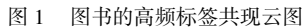
比较国内外研究现状可以发现,国内外学者对于社会化标注系统研究均从社会化标注系统构成要素——用户、标签、资源角度展开研究。国内外学者关于资源聚合研究大多从理论、方法与技术、应用角度入手,说明国内外学者研究的大体方向一致。但与国外研究相比,国内研究还存在研究时间短、体系不健全、偏重理论、缺乏实证研究等问题。而且,国内外鲜有学者从聚合方法层面研究社会化标注系统资源聚合,鉴于此,本文尝试融合分众分类和社会网络分析这两种资源聚合方法,拓展社会化标注系统的资源聚合维度,实现社会化标注系统资源聚合。

2.1 数据来源与筛选

表1 吴丹统计的图书情报领域关键词 中文部分

2.2 豆瓣图书“标签云”

表2 豆瓣网图书情报学图书高频标签 前30



由于有些图书的标签数量较多,因此,本文在标

签,验证了社会化标注系统标签聚合与推荐猜想。表2中,标注频次较高的豆瓣网图书情报学图书标签为搜索引擎(search-engine)、信息检索(IR、Information Retrieval)、交互设计、UCD(以用户为中心的设计)等,他们均是社会化标注系统的标签资源中的关键标签,在资源聚合中要尤其注意这些标签,发挥用户标注的作用。

图1形象地展现了豆瓣网图书情报学的部分高频图书标签,此标签云用R语言绘制,标签字体大小代表标签的重要程度,简单形象,是许多社会化标注网站推荐热门标签的首选方式。通过结合表2和图1,可以简单地归结出图书情报学科的关注热点:搜索引擎、信息检索、用户服务(用户体验、用户界面设计、人机交互等)、社会网络分析、自然语言处理与机器学习、算法研究、信息组织等。虽然社会化标注系统利用分众分类法从标注频次维度实现标签聚类,为用户提供热门标签,但单一维度的聚合并不能满足用户的多样化需求,且分众分类法不能揭示标签之间的关联和语义关系,因此,需要引入社会网络分析法,辅以本体和主题词法,从而兼顾聚合的层级和语义效果。

2.3 豆瓣图书标签的社会网络结构

(1) 标签预处理

由于标签数量较大,数据处理时间有限,本文选取排名前50的高频标签作为分析对象,编制标签邻接矩阵,两标签间是否存在关联的判断依据为两者是否同为豆瓣网热门标签,以及是否为豆瓣网推荐的热门标签的相关标签,据此,编制标签矩阵。

(2) 网络密度

网络密度(Network Density),指社会网络中个体之间相互交流的积极程度。若网络中有A个行动者,理论上存在的最大关联数为 $A(A-1)/2$,而实际上网络中存在B个关联,则网络密度即为实际关联数与理论最大关联数之比。比值大,说明网络密度大,网络成员间关联关系较多,交流频繁,个体受整体影响大;比值小,说明网络密度小,网络成员间关联关系少,交流匮乏。网络密度应用于社会化标注系统标签资源的网络密度计算,证实选取的样本数据确实存在关联关系少的现象。利用Ucinet(中文版)计算网络密度的步骤为:网络→凝聚力→密度,网路密度一般用P来表示,P值一般介于0~1之间,越接近1表明网络密度越大,个体间的关联越多,反之相反。通过计算,豆瓣网图书情报学图书标签网络平均密度 $P=0.0561$,说明豆瓣网图书情报标签资源网络关联稀疏,关联关系有待挖掘。

(3) 网络关联性

网络关联性(Network Relevance)用于衡量网络中个体之间是否存在联系及联系的强度。关联性强的网络信息、权利集中,行动者之间不平等,易形成分派结构,关联性弱的网络信息、权力分散,行动者间关系平等,易形成均匀结构。网络关联性的具体衡量指标为

可达性。

可达性(Reachability),如果网络中两个个体之间存在交流路径,则说两个个体是可达的。可达性在社会化标注系统资源聚合中用来测量一标签关联另一标签的容易度。可达性数值越高,表明标签间的关联关系越强,两标签能较容易实现关联聚合。利用Ucinet计算可达性的具体步骤是:网络→凝聚力→距离。通过计算,豆瓣网图书情报学图书标签间的平均距离为1.022,表明豆瓣网标签间的距离较远,关联性差,需要挖掘和建立标签间的关联。

(4) 网络中心性

网络中心性是对社会网络中个体和组织权利的量化表征。“权利”是指个体和组织影响力和控制力的大小。网络中心性的大小反映个体和组织在整个社会网络中的影响力大小。网络中心性是社会网络分析领域的重要内容,个体和组织在社会网络中居于何种地位,具有多大的权利是学者们对社会网络分析的最初研究内容。网络中心性的具体量化指标包括度数中心度、接近中心度、中介中心度、特征值中心度等。

①度数中心性(Point Centrality),指与该点有直接关联关系的点的数量。若个体与很多其他个体存在直接联系,那么该个体就具有较大的权利,在整个网络中处于中心位置。在社会化标注系统资源聚合中,度数中心性可以表述为若一标签与标签资源网络中的许多其他标签资源都存在直接的联系,那么该标签具有较大的权利,在标签资源网络中居于中心地位。而标签的度数中心度就是与该标签直接关联的标签数量。依据度数中心度实现标签资源聚合能发现标签资源网络中处于中心地位的标签,中心标签在标签网络中影响力较大,是知识传递扩散的重要节点。有利于识别资源聚合中的重要标签,这些标签与其他标签联系紧密,有利于促进知识交流与共享。利用Ucinet计算度数中心度的步骤为:网络→中心度→度。

②中介中心性(Betweenness Centrality),指个体对社会网络中资源的控制力。若个体处于多条两个个体交往的网络路径上,那么个体在整个网络中居于重要地位,其通过影响信息传递从而影响个体间的交往。中介中心度的衡量指标是“中间性比例”,具体为链接两点且经过个体的路径与经过两点路径总数之比。中介中心性用于社会化标注系统资源聚合中,可以表述为,如果一标签位于多条两标签交往的路径上,那么此标签体现媒介特征,在标签网络中扮演者枢纽和桥梁的角色,具有较大的控制力。利用Ucinet计算中介中心性的步骤是:网络→中心度→Freeman中间度→节点中间度。

③接近中心性(Closeness Centrality),衡量个体不与其他个体接近,不依赖和影响其他个体的程度。若个体与其他个体之间的距离都比较短则说明该个体

具有较高的接近中心度。个体与网络其他个体的捷径距离之和就是接近中心度。接近中心性用于社会化标注系统资源聚合中,可表述为,若标签到其他标签捷径距离之和最短,则此标签的接近中心度越高,独立性越强。利用接近中心性指标实现社会化标注系统资源聚合,能发现独立的重要的标签体系。利用 ucinet 分析标签的点度中心度的步骤为:网络→中心度→接近性。表3为豆瓣网图书情报学科图书标签网络中心性计算情况。

表3 豆瓣图书情报标签网络中心性指标计算

由表3可知,“互联网”、“计算机”、“信息检索”、“计算机科学”等标签的点度中心度值较高,说明这些标签在标签网络中居于中心位置,与其他标签之间的关联关系较多,影响力较大。“互联网”、“计算机”、“信息检索”、“管理”等标签的中介中心度值较高,表明这些标签在标签网络中处于多个其他标签交往路径上,是信息传递的重要纽带。“互联网”、“计算机”、“信息检索”、“计算机科学”等标签的接近中心度值较高,表明其受网络中其他标签的影响较小,可能形成单独的网络分支。综合三个指标发现,“互联网”、“计算机”、“信息检索”、“交互设计”、“计算机科学”等标签的点度中心度、中介中心度和接近中心度值都比较高,充分证明这些标签在标签网络中的重要地位,在标签资源聚合过程中要给予充分的重视。而点度中心度、中介中心度和接近中心度值都比较高的标签之间的关联,尤其是中介中心度高的标签间的关联则是标签网络中的关键关联,其影响和控制着整个标签网络的信息交流与传递。

对比高频标签和网络中心度高的标签(表2和表3),不难发现,高频标签虽然是用户使用频次最多的标签,但并不是标签资源体系中最重要标签,单纯依靠标注频次的方式推荐标签资源,已不能满足用户的确切需求。在分众分类法的基础上,利用社会网络分析法,挖掘标签资源网络中的重要标签迫在眉睫。此外,社会网络分析法属于浅语义挖掘,因此在实现标签资源聚合过程中,需要借助主题词表,分析标签

之间的层级和语义关系,最终完成广度延展和深度语义的聚合。

2.4 豆瓣网标签资源聚合效果展示与评价

为构建层级化、语义化的标签资源体系,本文以分众分类法为融合基础,借助本体的同义、继承、相关关系,利用主题词表对抓取的标签资源划分层级,标注语义,最后,利用社会网络分析法展现标签之间的层级和语义关联。具体的思路为:利用《中国图书馆图书分类法》这一工具,对标签间语义关系进行规范,构建层级关系。利用社会网络分析法,编制高频标签矩阵,从可达性、互惠性、点的度数三个方面凝聚标签子群,实现标签资源合理高效的聚合。

(1) 标签层级处理

由于要兼顾聚合效果的深度和广度两个方面,因此,为弥补分众分类法和社会网络分析法在语义挖掘上的劣势,本文借用主题词表这一深度语义聚合方法,参照《中国图书馆图书分类法》的层级规则,规范标签间的语义关系,构建标签的层级关系。对于不存在层级关系的标签,本文的处理依据是通过统计两标签是否为同一本书的标签,若是同本书标签则表明两标签存在关联关系,否则不存在关联关系。利用社会网络分析法,绘制关联关系矩阵,实现豆瓣网图书情报学科标签资源聚合。本节沿用标注频次较高的50个标签为样本数据,通过将高频标签和《中国图书馆图书分类法》《汉语主题词表》对照,提取如下具有语义和层级关系的标签体系:

①社会学(C91)→社会网络、复杂网络(C912)→社会网络分析→SNS

②信息传播学(G2)→图书馆学(G25)→信息组织、信息检索、信息管理、信息构建

(2) 标签矩阵

根据标签是否存在语义层级关系,以及两标签在同一本书标注结果中是否共现,判定两标签是否存在关联关系,从而编制标签邻接矩阵。其中,“1”代表两标签存在关联关系,“0”表示两标签无关。表4是高频标签邻接矩阵表。

(3) 标签子群凝聚

网络子群(Network Subgroup)指具有同质特征的网络个体形成的网络群体,是基于相同属性关联的行动者集合。在社会化标注系统资源聚合中,子群能够用来划分标签资源网络,有助于研究者掌握标签资源的结构特性。子群存在多种的量化方法,因此,具有多

表 4 高频标签矩阵

[illegible]

种类型,从可达性、互惠性和中心性角度划分,主要包括“可达性”子群、“互惠性”子群和“点度数”子群。

①“可达性”子群

可达性是判断网络关联度的指标,“可达性”子群用于考察网络个体间是否存在关联关系,能否可达,成员之间不必邻接。“可达性”子群能快速发现标签之间的关联关系。借助 Ucinet 凝聚“可达性”子群的具体步骤是:网络→子组→n- 丛。

②“互惠性”子群

互惠性是判断社会网络个体是否存在相互选择的指标,若两个体相互选择,则两者是互惠关系,否则不存在互惠联系。“互惠性”子群要求网络成员间必须是邻接,关联关系必须为“双向选择”。“互惠性”子群用于社会化标注系统资源聚合,有助于发现标签网络中的强联系团体。借助Ucinet凝聚“互惠性”子群的具体做法为:网络→子组→派系。

③“点度数”子群

点的度数是衡量个体在网络中重要程度和影响力大小的指标,可用群组内成员与其他成员间关联频次计算。“点度数”子群能发现标签系统中的有影响力的重要标签,这些标签是社会化标注系统资源聚合中的重要标签节点。借助 Ucinet 凝聚“点度数”子群的具体做法为:网络 \rightarrow 子组 \rightarrow K- 丛。

由于本文构建的矩阵为简单邻接矩阵, 因此, 标签资源子群凝聚适合选择“互惠性”子

群。将标签矩阵导入 Ucinet, 运行软件, 得到如图 2 所示的标签子群划分图。

如图 2 所示, 依据互惠性聚合标签资源, 可将标签资源划分为 33 个小子群, 通过分析能够发现, “搜索引擎” (search-

engine)、‘信息检索’(IR、Information+ Retrieval)、‘数据挖掘’、‘互联网’等标签多在一个聚合子群中。‘交互设计’、‘设计’、‘用户体验’、‘互联网’等标签多在同一子群。‘信息检索’、‘计算机科学’、‘图书馆学’、‘互联网’等标签也可划分到一个子群。而如‘数据挖掘’、‘互联网’、‘信息检索’(IR、Information+ Retrieval)”等出现在多个子群中的高频标签,则是标签网络联系的纽带,清晰其关联关系,有助于标签体系的优化。此外,许多标注频次不是很高的标签,如“图书馆学”、“数据分析”、“编程”等标签,也出现在多个子群中,表明这些标签在标签资源网络中的重要地位。豆瓣网向用户推荐标签时,可考虑这些标签。



图2 豆瓣网图书情报图书高频标签“互惠性”子群划分

社会网络分析法实现资源聚合的优势之一就是其具备强大的可视化功能,通过知识图谱的方式能生动形象地展现标签的重要程度以及标签之间的关联关系。本文在绘制知识图谱时,为显现标签间地位不同,按照点度中心度值的高低将标签的图例设置成不同的大小,图例越大表明此标签的点度中心度值越高,说明其在标签体系中的地位越重要。连线代表标签间存在关联关系,箭头代表方向。利用社会网络分析工具 Ucinet 绘制标签聚合图谱的步骤是:可视化→Pajek→file→open 标签矩阵。图 3 为豆瓣图书情报学图书高频标签资源聚合图。

①密度和关联性。网络密度和关联性能反映网络个体及其关系的紧密程度,利用 Ucinet 软件计算聚合后的豆瓣网图书情报学图书标签资源网络密度,得到密度 P 为 0.2407,距离为 1.896,两指标数值较聚合前(分别为 0.0561 和 1.022)均有明显提高,说明聚合后的标签资源网络中个体间的交流较多,关联关系较为密切。

③实用性和可视化。社会化标注系统资源聚合应是基于分众分类法和社会网络分析法两者优势,并借助相关理论方法、工具弥补两者不足的聚合,因此,在聚合过程中以高频标签作为聚合样本数据,体现用户智慧结晶,具有较好的实用性。此外,Ucinet在可视化效果上优于标签云,且具备强大的视图处理功能。

本文以豆瓣网图书情报学图书标签资源为研究对

参考文献

- [1] Begelman G. Automated tag clustering: Improving search and exploration in the tag space[J]. Edinburgh: Proleedings of the Collaborative Web Tagging Workshop at WWW'06, 2006: 1- 15.
- [2] Cui J, Liu H, He J, et al. Tag Clus: A random walkbased method for tag clustering[J]. Knowledge & Information Systems, 2011, 27(2): 193- 225.
- [3] 张帆. 基于链接分析和聚类的标签语义分析方法研究[D]. 大连: 大连理工大学, 2013: 1- 66.
- [4] 窦永香, 苏山佳, 赵捧未. 基于 Porter 算法的英文标签聚类方法研究[J]. 现代图书情报技术, 2009(9): 40- 44.
- [5] 李慧宗, 胡学钢, 何伟, 等. 社会化标注环境下的标签共现谱聚类方法[J]. 图书情报工作, 2014, 58(23): 129- 135.
- [6] Adrian B, Sauermann L, Roth- berghofer T. Contag: A semantic tag recommendation system[C]. Graz: Proceedings of I- Semantics, 2007: 297- 304.
- [7] Rendle S, Schmidt- Thieme L. Pairwise interaction tensor factorization for personalized tag recommendation[C]. Wsdm Proceedings of the Third Acm International Conference on Web Search & Data Mining, 2010: 81- 90.

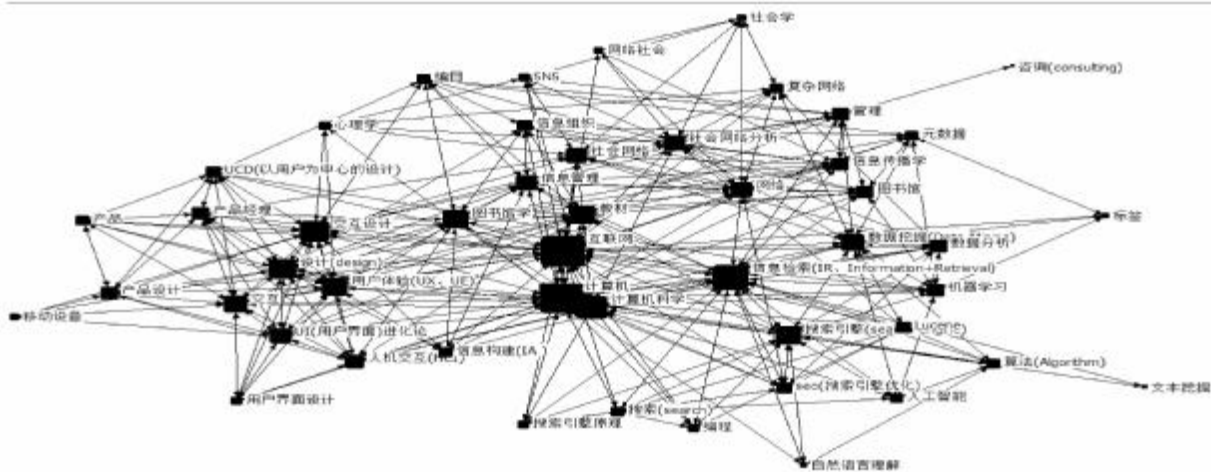


图3 图书高频标签聚合知识图谱

- [8] Zhao W, Guan Z, Liu Z. Ranking on heterogeneous manifolds for tag recommendation in social tagging services[J]. Neurocomputing, 2015, 148 : 521- 534.
- [9] 丛维强,李保珍,王 逊. 基于加权元组潜在语义分析的社会标签推荐[J]. 情报科学, 2015(1) : 57- 62.
- [10] 孙甲申,王小捷. 一种用于社会化标签推荐的主题模型[J]. 北京邮电大学学报, 2014(3) : 38- 41.
- [11] Djuana E, Xu Y, Li Y. Constructing tag ontology folksonomy based on Wordnet[C]. Shanghai: Proceedings of the IADIS International Conference on Internet Technologies & Society, 2011: 1- 9.
- [12] Lee H J, Sohn M. Tag- based integrated semantic ontology construction and evolution[C]. Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. IEEE Computer Society, 2013: 221- 227.
- [13] 王翠英. 本体与 Folksonomy 的比较研究[J]. 图书馆建设, 2008(5) : 85- 88.
- [14] 张云中. Folksonomy 中的语义关系呈现: Folksonomy 局部本体[J]. 图书情报工作, 2012, 56(20) : 127- 131.
- [15] Farooq U, Kannampallil T G, Song Y, et al. Evaluating tagging behavior in social bookmarking systems: Metrics and design heuristics[C]. New York: Proceedings of the International ACM Conference on Supporting Group Work, 2007: 351- 360.
- [16] 张 宁. 分众分类系统的用户行为特征分析[D]. 太原: 山西大学, 2013: 1- 105.
- [17] 潘 婵,冯利飞,丁婉莹,等. 基于标签- 关键词的用户行为分析[J]. 情报杂志, 2010, 29(3) : 139- 142.
- [18] Zhou T C, Ma H, Lyu M R, et al. UserRec: A user recommendation framework in social tagging system[C]. Atlanta: Proceedings of the National Conference on Artificial Intelligence, 2010: 1486- 1491.
- [19] 余 臻,宁宣熙,李保珍,等. 社会化标注系统中用户需求偏好的一种获取方法[J]. 南京航空航天大学学报, 2009, 41(1) : 139- 144.
- [20] 杨 丹,曹 俊. 基于 Web2. 0 的社会性标签推荐系统[J]. 重庆工学院学报(自然科学版), 2008(7) : 52- 53.
- [21] 田莹颖. 基于社会化标签系统的个性化信息推荐探讨[J]. 图书情报工作, 2010, 54(1) : 50- 54.
- [22] Kim H, Alkhaldi A, Saddik A E, et al. Collaborative user modeling with user- generated tags for social recommender systems[J]. Expert Systems with Applications, 2011, 38(7) : 8488- 8496.
- [23] Mika P. Ontologies are us: A unified model of social networks and semantics[J]. Lecture Notes in Computer Science, 2007, 5(1) : 5- 15.
- [24] Fu W T. The microstructures of social tagging: A rational model[C]. Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work. ACM, 2008: 229- 238.
- [25] 武慧娟. 社会化标注系统中个性化信息推荐模型研究[D]. 长春: 吉林大学, 2014: 1- 162.
- [26] Goh D H, Chua A, Lee C S, et al. Resource discovery through social tagging: A classification and content analytic approach[J]. Online Information Review, 2000, 33(3) : 568- 583.
- [27] Bi B, Shang L, Kao B. Collaborative resource discovery in social tagging system[C]. New York: Cikm Proceeding of Acm Conference on Information & Knowledge Management, 2009: 1919- 1922.
- [28] 宣云干,朱庆华. 基于熵的社会化标注系统资源模型研究[J]. 计算机应用与软件, 2011, 28(3) : 89- 90.
- [29] 杨 萌. 基于 Taxonomy- folksonomy 混合模型的社会化标注系统资源聚合研究[D]. 长春: 吉林大学, 2014: 1- 149.
- [30] Zhang Z, Zhou T, Zhang Y. Personalized recommendation via integrated diffusion on user item tag tripartite graphs[J]. Physica a Statistical Mechanics & Its Applications, 2010, 389(1) : 179- 186.
- [31] Oren Etzion. The World Wide Web: Quagmire or gold mine[J]. Oren Etzioni Communications of the ACM, 1996(11) : 65- 68.
- [32] 杜 晖. 基于耦合关系的学术资源深度聚合[D]. 武昌: 武汉大学, 2013(5) : 1- 212.
- [33] 杨 萌,张云中,徐宝祥. 社会化标注系统资源聚合与导航研究综述[J]. 情报理论与实践, 2014(3) : 140- 144.
- [34] Liang F, Liu H, Liu Y, et al. A Capability- based matchmaking mechanism supporting resource aggregation with in large- scale distributed computing infrastructure[J]. Lecture Notes of the Institute for Computer Sciences Social Informatics & Tele Communications Engineering, 2014(133) : 145- 154.
- [35] Ma S, Sun X, Guo Z. A resource discovery mechanism integrating P2P and Grid[C]. Xinxiang: IEEE International Conference on Computer Science and Information Technology, IEEE, 2010: 336- 339.
- [36] 李 劲,程秀峰,宋红文,等. 基于语义的馆藏资源深度聚合模型探讨[J]. 湖北民族学院学报: 自然科学版, 2013, 31(2) : 212- 215.
- [37] Studer R, Benjamins V R, Fensel D. Knowledge engineering: Principles and methods[J]. Data & Knowledge Engineering, 1998, 25(97) : 161- 197.
- [38] 邱均平,杨 强,楼 雯. 资源本体构建理论与实证研究[J]. 情报理论与实践, 2014(5) : 1- 6.
- [39] Daouadi A, Nguyen K, Lemay M, et al. Ontology- based resource description and discovery framework for low carbon grid networks[J]. IEEE International Conference on Smart Grid Communications, 2010, 23(3) : 477- 482.
- [40] 欧石燕,胡 珊,张 帅. 本体与关联数据驱动的图书馆信息资源语义整合方法及其测评[J]. 图书情报工作, 2014, 58(2) : 5- 13.
- [41] Gruber T. Ontology of folksonomy: A mash- up of apples and oranges[J]. International Journal on Semantic Web and Information Systems(IJSWIS), 2007, 3(1) : 1- 11.
- [42] Sternitzke C, Bartkowski A, Schramm R. Visualizing patent statistics by means of social network analysis tools[J]. World Patent Information, 2008, 30(2) : 115- 131.
- [43] 游 毅,成 全. 试论基于关联数据的馆藏资源聚合模式[J]. 情报理论与实践, 2013, 36(1) : 109- 114.
- [作者简介] 毕 强,男,1954年生,吉林大学管理学院教授,博士生导师。
赵夷平,男,1981年生,吉林大学管理学院博士研究生。
孙中秋,女,1991年生,吉林大学管理学院硕士研究生。
收稿日期:2015- 07- 15