

Folksonomy与SNA资源聚合类比研究*

孙中秋, 陈晓美, 周珊珊

(吉林大学管理学院, 长春 130022)

摘要: 随着国内外数字资源多维度聚合研究的不断深入, 资源聚合方法的改进和融合也成为学术研究热点。在对分众分类和社会网络的基本结构进行简要阐述的基础上, 文章对Folksonomy和SNA两种资源聚合方法, 从聚合理念、维度、实现技术和具体应用四个方面进行比较分析, 指出Folksonomy和SNA在资源聚合中的优势和不足, 并分析了资源聚合中Folksonomy和SNA的融合情景, 以期对相关研究和后续的实践应用工作提供理论指导。

关键字: 分众分类法; 社会网络分析; 社会化标注系统; 资源聚合

中图分类号: G253

DOI: 10.3772/j.issn.1673—2286.2014.06.004

引言

随着数字资源的多维聚合成为国内外研究热点, 资源聚合方法的多视角考察与探索引起业界的广泛关注。目前, 较为成熟的资源聚合方法主要包括主题词表、Folksonomy、Ontology (本体)、文献计量法和SNA (Social Network Analysis) 等。

国外学者较早开展资源聚合方法的比较融合研究, 主要包括Folksonomy与Ontology^[1,2]、SNA与文献计量法^[3,4]的比较融合。国内对资源聚合方法的融合研究开展较晚, 但成果颇丰, 主要有Folksonomy与主题词表^[5,6]、Ontology^[7]、概念格^[8], 主题词表与Ontology^[9], SNA与文献计量法^[10]的比较分析, 但至今还鲜见有学者将Folksonomy与SNA进行比较和融合研究。鉴于此, 本文在借鉴性吸收已有研究成果的基础上, 从资源聚合的理念、维度、实现技术和具体应用四个方面对Folksonomy和SNA进行比较分析, 试图寻找Folksonomy与SNA在资源聚合中的融合点, 以期为后续研究提供理论指导。

1 分众分类和社会网络的基本结构

1.1 分众分类的基本结构

Folksonomy, 或称分众分类法, 在分类学中又称为社会化标注系统, 最早由Thomas Vander Wal和Gene Smith提出, 是“Folk”和“Taxonomy”两个单词的组合, 意为“以用户为中心的基于大众一致意见产生的分类体系”^[11], 即用户按照自己的分类标准, 依托标签共建共享平台, 赋予信息资源特征标识 (标签), 并按照用户标注频率进行分类的知识组织方法。马费成^[12]指出Folksonomy是由用户、资源和标签构成的三元组。三者的关系如图1所示。

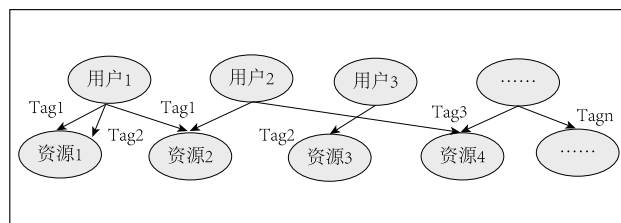


图1 分众分类的基本结构

由图1可以看出Folksonomy的三层体系: 资源、标签、用户。从横向角度看, 资源与资源之间、用户与用户之间都存在潜在的联系, 这必然导致标签之间出现潜在语义关系。从纵深角度理解, 信息资源位于基础层, 提供可发现和利用的信息资源; 用户位于核心层, 是分众分类体系结构的灵魂所在; 标签位于中介层, 是沟通

* 本研究得到国家自然科学基金项目“语义网络环境下数字图书馆资源多维度聚合与可视化展示研究” (编号: 71273111) 和国家社会科学基金重大项目“基于语义的馆藏资源深度聚合与可视化研究” (编号: 11&ZD152) 资助。

用户和资源之间的桥梁,三者之间相互影响,相互完善,形成完整的社会标注系统。

1.2 社会网络的基本结构

社会网络是因个体成员间有效互动形成的较稳定的社会关系和社会团体的总和。笔者认为,SNA是社会网研究过程中形成的方法,从资源聚合的角度,可将社会资源网络看成是由知识单元、知识关联和知识群落构成的三元组。如图2所示。

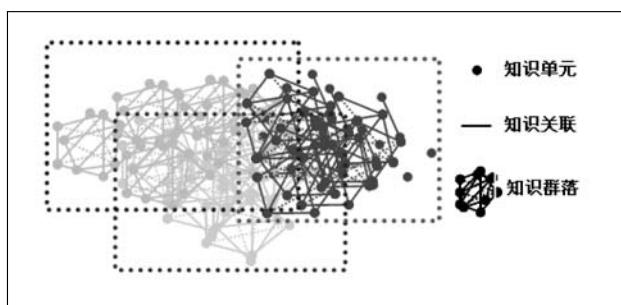


图2 社会网络的基本结构

图2中,知识单元是社会资源网中的资源节点,知识关联是资源节点之间的关联关系,知识群落则是依据关联关系划分的知识单元的集合,而不同知识群落之间的交叉关联则构成多维立体的社会资源网络体系。

2 Folksonomy与SNA资源聚合比较分析

Folksonomy与SNA两种方法虽然都是实现资源聚合的重要手段,但二者在资源聚合的理念、维度、实现技术和具体应用等方面却有着显著差异。

2.1 聚合理念类比

2.1.1 以用户为中心 vs 以资源为中心

Folksonomy作为实现资源聚合重要手段的优势在于将“以人为本”的理念贯穿资源聚合的全过程,允许用户对发布和使用的信息资源进行自主分类和标注,打破了传统的、单纯以资源属性特征为依据的资源聚合方式,强调用户认知和需求对于资源聚合的重要性,是以用户为导向的资源聚合方法。SNA则是将“以资源为中心”的理念融于资源聚合中,这里的资源不仅仅指用户

资源,还包括如标签资源、关系资源、评论资源等,其中以关系资源最为重要,关系资源指用户、标签和资源间的关联关系,从这个意义来说,SNA是以资源为中心的资源聚合方法。

2.1.2 社群认知聚合 vs 语义关联聚合

Folksonomy是依据用户自定义标签进行协作分类的方法,是基于自然语言的社群认知聚类^[13]。Folksonomy实现资源聚合的思路为:高频标签统计、标签分类、标签聚合。如以豆瓣网为例,用户添加标签后,网站依据后台词库或统计算法计算标签频次,并按预先设定的分类标准对标签进行分类和聚合。故用户可以完全凭借兴趣爱好,依据标签检索到自己感兴趣的资源,方便快捷。标注形式和内容的多样化固然有利于研究者的多视角探索,实现资源的多维度聚合,但大多数网站的标签资源聚合仅仅依据用户标注的频次,忽视标签之间、用户之间和资源之间的潜在语义关系,不利于资源的语义聚合。

SNA主要用于资源间关联关系的分析,其将资源间的关系看成“网”或“网络”,并对此“网络”进行结构划分,形成不同的子群,是基于语义关联的资源聚合方法。SNA实现资源聚合的思路为:识别重要节点(高频、高中心度)、判定关键枢纽(中介中心度)、划分网络结构(凝聚子群)。如以学术团体社会网络分析为例,通过统计学术人员互引情况,形成矩阵或知识图谱,识别有影响力作者或团体。关联关系的挖掘有利于实现资源的语义聚合,但却限制用户标注的自由性,不利于资源的认知聚合。

2.1.3 聚合全面性 vs 聚合准确性

Folksonomy是通过用户自由标注和相互协作的方式实现资源聚合^[14],充分体现出用户的群体智慧在资源聚合过程中的能动性,从资源聚合效果上看,资源聚合的泛指度高。但由于用户多为草根阶层,注册主体多样化,且无门槛限制,标注主体的素质和认知存在差异,导致无实义标签、近义标签、同义标签、心理概念模糊标签^[15]大量存在,标签之间“各自为政”,关联度低,聚合的准确性低。

SNA是通过专业人员对资源之间的关联关系深入揭示实现资源聚合,从资源聚合效果上看,资源聚合的

专指度高。由于在资源聚合过程中, 无意义或关联度低的标签必将消失, 影响资源聚合的全面性。

2.2 聚合维度类比

2.2.1 Folksonomy资源聚合维度

Folksonomy实现资源聚合主要是围绕社会化标注系统的构成要素——标签、用户、资源三个维度展开。

(1) 标签。标签是连接用户和资源的桥梁, 在某种程度上既体现了用户需求、认知和情感, 也蕴含了资源的某些特征, 因此, 标签聚合重视标签的聚类以及标签间语义关系的挖掘。实现标签资源聚合的主要依据是元数据、主题词表、本体等。目前, 标签聚合的成果已经应用于许多网站, 如热门标签等, 有些网站还在此基础上推出“关联标签”服务, 进一步挖掘标签之间的语义关系, 但大多为浅语义关系挖掘, 挖掘深度不够。

(2) 用户。用户是社会化标注系统的标注主体, 是资源的生产者和利用者, 基于用户视角的资源聚合着重于用户分类研究, 根据用户的喜好将具有相近情感体验的用户归为一类, 在此基础上向用户提供各种个性化推荐服务。实现以用户为中心的资源聚合的主要依据是标签、本体、同义词环等。目前, 用户资源聚合的成果主要应用于个性化服务, 如豆瓣网“喜欢该资源的用户有……”等。但由于用户的认知和情感可能会随时间推移而改变, 所以用户资源聚合具有不确定因素。

(3) 资源。指狭义的资源, 是社会化标注系统中除去标签和用户之外的资源, 如图书资源、评论资源等。资源聚合是按照标签、用户和资源之间的关联关系聚类资源的过程, 在此基础上, 实现资源的优化重组。实现资源聚合的主要依据是本体、文献计量学等。目前, 资源聚合成果主要应用于学科分类导航、相关资源推荐等, 诚然, 在资源聚合的逻辑性、结构性、层级性等方面还有待提高。

2.2.2 SNA资源聚合维度

社会网络分析法实现资源聚合主要是围绕网络中心性、网络群聚性和网络关联性三个维度展开。

(1) 网络中心性。网络中心性主要用于衡量节点在整个社会网络中的地位和影响力, 从而决定此节点在资源聚合中的取舍。网络中心性的具体衡量指标

包括: 点度中心性 (Degree Centrality)、中介中心性 (Betweenness Centrality)、接近中心性 (Closeness Centrality)。

(2) 网络群聚性。网络群聚性主要用于实现节点分类和定位, 用于网络结构和层级划分, 能判断资源的群组归属。网络群聚性的具体衡量指标主要有凝聚子群 (Cohesive Subgroups)、派系 (n-派系、n-宗派、k-丛) 等。

(3) 网络关联性。网络关联性主要用于判断网络节点间关系的亲疏远近、节点在网络中枢纽作用的发挥程度等, 并以此为依据判定节点的重要程度。网络关联性的具体衡量指标包括关联度 (Connectedness)、等级度 (Digraph Hierarchy)、效率 (Digraph Efficiency) 和最近上限 (Least Upper Bound)。

由于在后续研究中, 笔者将利用社会网络分析法实现资源聚合, 因此, 本文对社会网络分析法的资源集合维度只简要介绍, 不再赘述。

2.3 实现技术类比

Folksonomy与社会网络分析均能实现资源聚合, 但聚合技术的不同使得聚合结果不尽相同, 表1列举了Folksonomy和SNA实现资源聚合的主要技术和工具。

表1 Folksonomy与SNA资源聚合实现技术

比较点 比较对象	实现技术	实现工具
Folksonomy	词频统计标签技术 (过滤、分类、推荐) 等	R语言、标签云
SNA	矩阵、知识图谱等	社会网络分析软件 (Ucinet、Pajek、Netdraw)、C#编程、矩阵生成工具 (matlab) 等

由表1可知, Folksonomy实现资源聚合的主要通过词频统计、R语言分词和聚类以及标签技术 (标签过滤技术、标签分类技术、标签推荐技术) 形成标签云得以实现。SNA则主要通过c#等编程算法获取和处理数据, 形成矩阵, 并在此基础上利用社会网络分析软件如Ucinet、Pajek、Netdraw等实现资源聚合。

(1) 标签云 (Tag Cloud)。网站内标签的集合, 利用“统计上浮”原理将代表用户认知的标签形成云图呈现给用户。标签一般按照热门程度的高低显示出不同

的颜色或大小,但也有按照字典顺序排列标签的组织方式。用户可以依照热门程度或者字典来搜寻需要的信息,而这些标签均与资源之间建立超链接,点击标签后直接链往分类页面,简单方便,但不能显示标签之间的关联关系以及群组分类。

(2) 矩阵(Matrix)。纵横排列的二维表格,资源聚合中用于数据处理。社会网络中涉及的矩阵主要有邻接矩阵、发生阵、隶属关系矩阵、有向关系矩阵和多值关系矩阵^[16]等,通过矩阵运算,可以得出可达性、距离等资源聚合指标,但矩阵生成与计算难度较大,耗时较长。

(3) 知识图谱(Mapping Knowledge Domain)。显示资源体系发展进程与结构关系的一种图示,其数据分析功能强大、聚合效果生动形象。资源聚合中主要用于分析著者与机构合作、引用与被引、学科发展脉络等。知识图谱生成软件如Ucinet、Pajek等操作简单,对资源和属性之间的关联关系挖掘透彻,是图书情报学者实现资源聚合的首选工具。

(4) R语言。免费的、源代码开放的统计分析软件。资源聚合中主要用于统计分析和制图,分词和绘图功能强大。随着研究人员对R研究的深入,R也被用于矩阵计算,其分析速度可以媲美Matlab等矩阵生成工具。R较为简单,可从网上直接下载数据包,根据数据处理的需要改编源代码,实现数据分析和制图功能,但需要研究人员具备计算机专业知识和编程技术。

2.4 聚合应用类比

2.4.1 Folksonomy资源聚合应用

目前,Folksonomy实现资源聚合的成果主要用于完善社会化标注网站的服务功能,如热门标签、关联标签等。

(1) 热门标签推荐。社会化标注网站的热门标签推荐功能,是Folksonomy实现资源聚合的典型应用。如用户对Flickr网站资源进行标注时,会自动显示其他用户的标注情况,并于网站右侧列出最热门的标签(Top Tags),图3为Flickr网站的热门标签推荐服务^[17]。

如图3所示,热门标签推荐功能极大地扩展了用户标注和检索的范围,提高了检全率,节约检索时间,简单方便,与用户的贴合度高,但仅以标注频次作为判定标签重要程度的标准有失科学性和说服力。



图3 Flickr热门标签推荐



图4 豆瓣读书个性化服务

(2) 个性化推荐服务。个性化信息的推荐和订阅服务是Folksonomy资源聚合的另一典型应用,国内社会化标注网站中,豆瓣网的个性化推荐服务开展较好,“豆瓣猜你会喜欢”和“喜欢这本书的人还喜欢……”等服务均是个性化推荐的具体应用。图4是豆瓣读书的个性化服务功能。

如图4所示,不同于热门标签推荐,个性化推荐服务是通过对标签、资源和用户之间关联关系的挖掘,建立关联标签,寻找用户感兴趣的资源,用户友好程度高。但关联关系挖掘深度不够,资源聚合混乱。

2.4.2 SNA资源聚合应用

SNA实现资源聚合主要用于图书情报界的学术研究,如学术博客圈划分、人际关系网分析、学术资源聚合等。

(1) 社交平台人际关系分析。SNA最早用于社交网络如Twitter、Facebook的人际关系资源聚合。通过计算中心性、关联性和群聚性等指标,分析人际网密度,显现意见领袖或有影响力团体。图5为郭金龙^[18]对图书情报界学者博客互引情况的社会网络分析结果。

如图5所示,学者数量显示网络密度,学者间关联数量代表点度中心度、中介中心度等,两学者间隔节点

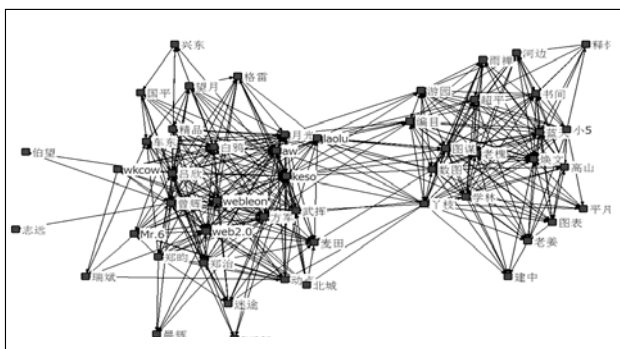


图5 图书情报界学者博客互引情况社会网络分析

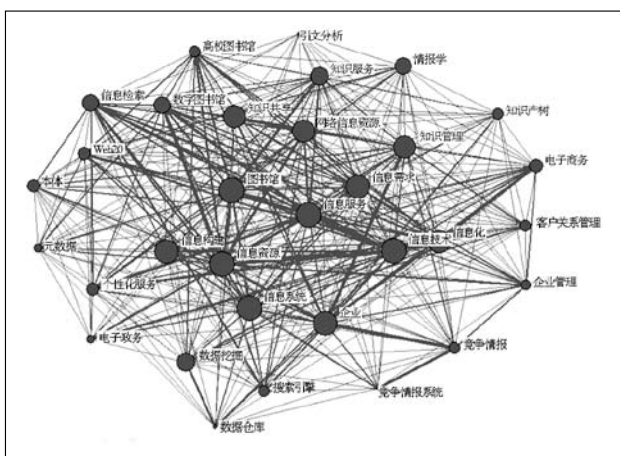


图6 图书情报学硕士学位论文关键词共现知识图谱

数用于衡量可达性和效度, 学者间距离远近及联系多寡用于划分子群。

(2) 学术资源聚合。通过将作者、机构、关键词、参考文献等看成各自资源网络中的节点, 构造关系矩阵, 研究合著、合作、互引、学术分支和团体等学术现象, 剖析学术交流网络的结构和特点。图6为魏群义^[19]利用社会网络分析软件Pajek对国内近10年图书情报学硕士学位论文关键词进行统计的知识图谱。

如图6所示, 关键词节点大小用于显示重要性, 关键词间连线显示相关程度, 距离远近和连线数量用于判定研究热点和学科分支。

3 Folksonomy与SNA资源聚合优劣势和融合分析

3.1 Folksonomy与SNA资源聚合优劣势分析

由上述比较分析可知, Folksonomy与SNA在资源

聚合中既有优势也存在不足。

3.1.1 Folksonomy资源聚合优劣势分析

聚合理念上, Folksonomy强调用户认知和需求在资源聚合中的决定作用, 追求全面性、时效性和实用性, 突出资源聚合的传递扩散功能。但资源聚合的关联性和结构性差, 规范程度低。

聚合维度上, Folksonomy从标签、资源和用户三个维度实现资源聚合, 以词频作为资源聚合的子维度, 聚合过程简单, 有助于识别关键资源。但聚合维度较为单一, 形成的资源聚合网络体现出平面化特征, 缺乏说服力。

聚合应用上, Folksonomy主要用于社会化标注系统资源聚合, 贴近普通用户需求。但由于其存在语义模糊、结构性差等缺点, 应用领域的扩展面临困境。

3.1.2 SNA资源聚合优劣势分析

聚合理念上, SNA强调资源间影响和交流对资源聚合的影响, 追求准确性和规范化, 突出资源聚合的语义关联关系挖掘功能, 但资源聚合的全面性、时效性差, 实用性有待检验。

聚合维度上, SNA从中心性、关联性和群聚性三个维度实现资源聚合, 以点度中心度、中介中心度、子群、可达性等作为资源聚合的子维度, 聚合维度多样化, 说服力高, 有助于形成多维立体的资源聚合网络, 但聚合过程复杂。

聚合应用上, SNA主要用于社交平台和学术资源的聚合, 兼顾普通用户和研究人员的需要, 应用领域的扩展前景较好, 但其规范化和关联性等缺点可能限制用户进行自由标注。

3.2 Folksonomy与SNA资源聚合融合分析

Folksonomy与SNA的融合是基于两者优点和缺点的一种全新整合, 应该扬长避短, 根据需求不同进行不同形式的融合^[20]。本文将Folksonomy与SNA实现资源聚合的优劣势置于直角坐标系中, 分析两者的融合情形, 图7为Folksonomy与SNA资源聚合融合分析图。

由图7可知, Folksonomy与SNA资源聚合可分为四种情形:

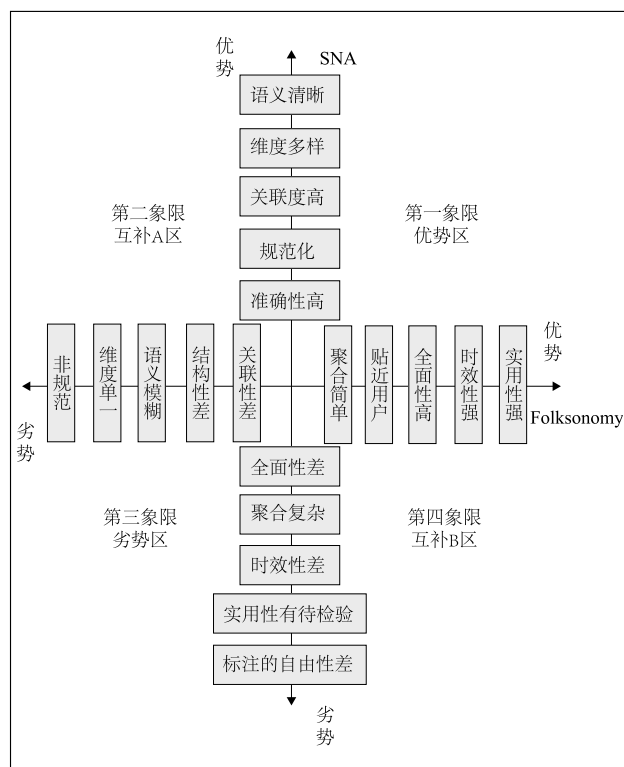


图7 Folksonomy与SNA资源聚合融合分析图

优势区（第1象限）：Folksonomy与SNA的双向融合，即结合两者资源聚合的优势。

此类融合既不以Folksonomy作为融合的落点，也不以SNA作为融合的受体，而是将两者放在同等的层面上，充分结合Folksonomy与SNA的优势实现资源聚合，使得资源聚合兼具实用性与规范性、全面性与准确性、时效性与关联性特征。

互补A区（第2象限）：SNA向Folksonomy的融合，即利用SNA资源聚合优势弥补Folksonomy资源聚合的不足。SNA资源聚合的规范度高、语义性强、关联度和准确性高，聚合维度多样化等特征，有助于解决Folksonomy资源聚合中的规范度低、语义模糊、关联性和准确性低及聚合维度单一等问题。

劣势区（第3象限）：利用其他方法弥补Folksonomy与SNA资源聚合中的不足。通过借助相关的理论、技术或方法，如主题词表、本体、文献计量法等，对Folksonomy和SNA资源融合的不足进行改善，根据实际应用的需要最大限度地消除两者融合的负面效应。

互补B区（第4象限）：Folksonomy向SNA的融合，即借助Folksonomy资源聚合优势弥补SNA资源聚合的不足。Folksonomy资源聚合的实用性、全面性和时效性强、贴近用户需求、聚合难度低等特征，有助于弥补

SNA资源聚合的时效性和全面性差、用户标注自由度低、实用性有待检验等问题。

4 结语

资源聚合并非是一线线索下的知识组织，相反极具多维性，聚合方法的互补融合必能丰富资源聚合网络，优化聚合效果。虽然Folksonomy与SNA两种方法在资源聚合的理念、维度、实现技术和具体应用等方面各有优劣势，但也存在互补融合的可能。Folksonomy与SNA的相互借鉴融合有利于寻求兼顾资源聚合中全面性与准确性、实用性与关联性的平衡点。然而，单纯融合此两种方法并不能使资源聚合达到最优状态，还需借鉴其他较为成熟的方法，如本体、主题词表、文献计量法等。此外，从技术上实现Folksonomy与SNA的融合也是后续研究工作的要点。

参考文献

- [1] 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.
- [2] KIM H L, SCERRI S, BRESLIN J G, et al. The state of the art in tag ontologies: a semantic model for tagging and folksonomies [C]// International Conference on Dublin Core and Metadata Applications, 2008: 128-137.
- [3] 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.
- [4] OTTE E, ROUSSEAU R. Social network analysis: a powerful strategy, also for the information sciences [J]. Journal of information Science, 2002, 28(6): 441-453.
- [5] 李超.一种基于主题和分众分类的信息检索优化方法[J].情报理论与实践,2009(10):108-110.
- [6] 李艳.分众分类与受控词表的集成研究[D].太原:山西大学,2013.
- [7] 熊回香,廖作芳.本体在Folksonomy中的应用研究[J].情报科学,2010(2):274-278.
- [8] 滕广青,毕强,高娅.基于概念格的Folksonomy知识组织研究:关联标签的结构特征分析[J].现代图书情报技术,2012,28(6):22-28.
- [9] 赵东霞,赵新力.基于政务主题词表的本体构建研究[J].现代图书情报技术,2008(3):73-77.

- [10] 邱均平,王菲菲.基于文献计量的国内外社会网络分析研究比较[J].情报资料工作,2011,32(1):33-37.
- [11] PETERS I, STOCK W G. "Power tags" in information retrieval [J]. Library Hi Tech, 2010, 28(1): 81-93.
- [12] 马费成,张斌.图书标注环境下用户的认知特征[J/OL].中国图书馆学报,2014,40(1):4-14 [2014-04-01]. <http://www.cnki.net/kcms/detail/11.2746.G2.20131030.1424.005.html>.
- [13] 滕广青.电子商务网站Folksonomy用户Tag的语义认知分析[J].情报科学,2012,29(12):1871-1875.
- [14] HEYMANN P, GARCIA-MOLINA H. Collaborative creation of communal hierarchical taxonomies in social tagging systems [R]. Stanford: Technical Report, 2006: 1-5.
- [15] SINHA R. A cognitive analysis of tagging [EB/OL]. (2005-09-07) [2014-04-01]. <http://rashmishinha.com/2005/09/27/a-cognitive-analysis-of-tagging/>.
- [16] 刘军.整体网分析讲义-UCINET软件应用(第二届社会网与关系管理研讨会资料)[J].哈尔滨:哈尔滨工程大学社会学系,2007:111.
- [17] 李镜镜.基于标签的网络Folksonomy研究[J].科技情报开发与经济,2009(31):71-74.
- [18] 郭金龙,许鑫.领域博客的社会网络分析:基于图书情报与互联网博客的实证[J].图书情报工作网刊,2012(1):001.
- [19] 魏群义,侯桂楠,霍然.近10年国内情报学硕士学位论文研究热点统计分析[J].图书情报工作,2012,56(2):35-39.
- [20] 张云中.本体与自由分类法的融合机理研究[J].情报理论与实践,2012,35(2):35-40.

作者简介

孙中秋,女,1990年生,吉林大学管理学院硕士研究生,E-mail: 466386531@qq.com。

陈晓美,女,1973年生,吉林大学管理学院副教授。

周珊珊,女,1980年生,吉林大学管理学院博士研究生。

Comparison Study on Folksonomy and SNA in Collections Aggregation

SUN ZhongQiu, CHEN XiaoMei, ZHOU ShanShan
(School of Management, Jilin University, Changchun 130022, China)

Abstract: With the research of multi-dimensional aggregation of digital resources at home and abroad deeply, improving and integrating the polymerization methods of digital resources has become a hot academic research point. Based on elaborating the basic structure of the Folksonomy and social networking briefly, this paper compares Folksonomy with SNA, which are two resource polymerization methods, from four aspects: aggregation concepts, dimensions, implementation technologies and specific applications, pointing out the strengths and weaknesses for Folksonomy and SNA in resource aggregation, analyzing the scenarios for Folksonomy and SNA in resource aggregation, hoping to provide theoretical guidance for related researches and following practical applications.

Keywords: Folksonomy; SNA; Social tagging system; Resource aggregation

(收稿日期: 2014-04-30)