

面向微内容分类组织的复合模式研究*

徐晨

[摘要] 分析 social tagging 与传统分类组织方式各自的比较优势,根据微内容内涵和组织特点,尝试构建微内容三阶段复合式分类组织模式。这三阶段分别是聚合阶段、机制设计阶段和成型阶段。聚合阶段按照信息类型对微内容进行聚合。机制设计阶段含分类机制和组织机制,分类机制从类别、词义、词型、词性角度融合 tags 与分面分类法,组织机制从匹配程度、层次关系和标引深度三个方面比较 tags 与传统主题词。成型阶段采用规范、吸纳、推荐三种方式优化 tags,使其能够更好地管理微内容。

[关键词] 微内容 social tagging tag 分类组织

[分类号] G203

微内容(microcontent)最初是 1998 年由 J. Nielsen 提出的,指那些能快速清晰反映整体信息内容的类似元数据的简短标题或主题词^[1],后来随着 Web 2.0 的发展兴起,微内容扩展到 RSS、Blog、Wiki 等用户聚合生成内容^[2]。本文微内容专指 Web 2.0 环境下用户贡献的网络信息内容,即包含在 microblog、blog、wiki、RSS、SNS、podcasting、social tagging 等 Web 2.0 应用中的文本、视频和图片。Web 2.0 信息管理的主要任务就是要帮助用户管理、维护、存储、分享、转移微内容,但这种主题千变万化、格式灵活多样的“碎片化”内容已很难单纯用传统组织方式管理。

本文针对微内容难以分类组织的难题,综合运用 social tagging 与主题词表、叙词表、分面组配法等传统组织方式,探讨运用最单元化微内容 tags 将两种组织方式取长补短,构建融合应用模式,来分类组织微内容,期望该模式能为 Web 2.0 环境下微内容有效组织管理提供参考。social tagging 及其相关词汇在国内译名较多,但大多尚未有公认的译名,因而此处与之相关的词语皆沿用英文原称。本文采用 social tagging 而非 folksonomy,是考虑到 social tagging 涵盖了 folksonomy,除了静态形成的用户分类方式,还包含了 tagging 这种动态的标引过程。

1 微内容的内涵和组织特点

微内容产生于信息传播领域,因而对于其研究常

见于新闻传播类研究文献中。当前在信息管理领域相关文献不多,主要集中在国内,且多数只将微内容作为一种信息服务辅助工具,探讨其在图书馆管理、本体语义网、信息构建等其他研究领域的应用价值^[3-4],而非视其为研究主体进行分析。国内目前将微内容作为主体观察对象的主要是王国才和蔡淑琴等。王国才将微内容列为研究对象,但用尚限于理论探讨的语义网来分析微内容,目前显然缺乏运用的可行性^[5]。蔡淑琴等关注微内容的序化过程及相关推荐系统构建,从理论上罗列了多种方法手段,不过对具体针对如何运用 social tagging 中的 tags,将其融合已有传统分类组织进行动态序化并未涉及^[6]。

微内容涵盖面较广,从其贡献角度分,有贡献内容和贡献结构两类。贡献结构的微内容指网址、用户点击等不具备独立内容涵义的数据,数量较少且无实际内容涵义;贡献内容的微内容是微内容的主要部分,提供独立内容产品。微内容涵盖了用户生成内容,因为微内容不仅有用户创作产生的网络资源,也有用户根据各自兴趣和目标转发和组织的其他资源。

微内容有三大组织特点,即复杂多类型、片段自组织和群体智能性。复杂多类型由微内容的组成多样性所决定,本文所限定的微内容就包含了文本、音频、视频和图片等多种信息类型。片段自组织是微内容容量决定的,微内容顾名思义,内容信息量较小,甚至有些都是片段化和零星分布的,很难管理,不过因为微内容

* 本文系国家自然科学基金青年项目“多语言环境下 Social Tagging 的内涵机理与应用框架研究——基于比较视角”(项目编号:71103203)研究成果之一。

[作者简介] 徐晨,中南大学商学院讲师,博士,E-mail:xuchenshirley@126.com。

收稿日期:2013-04-01 修回日期:2013-04-21 本文起止页码:113-118 本文责任编辑:王善军

是由用户自发产生的,这种无序碎片化的信息在不断演变聚合过程中,自行组织形成符合用户使用习惯的自组织体系。群体智能性是片段自组织的结果,单个用户的使用方式是无章可循的,但一旦一类或者一群用户以相类似的方式不约而同地使用组织同一主题或相关信息,这种使用模式就慢慢被规定下来并被广泛共享,成为新的被认可的使用方式。最典型的例子就是网络新词的产生以及权威词典对该类词的兼容吸纳,也就是说网络普通大众的群体智能被认可并成为新的使用规范。

早在 Web 2.0 出现前,业内针对网络信息组织就出现过应该采用人工语言还是自然语言的争论,张琪玉等知名学者一针见血地指出,检索用语发展大趋势是检索语言的自然语言化和自然语言的检索语言化,是两者的初级结合到完全融合的过程^[7]。随着网络信息发布主体下放到单个用户,这种融合性的趋势就发展得愈加明显,针对微内容的分类组织也就自然步入 social tagging 与传统组织分类的融合之路。

Social tagging 中 tags 是单元化微内容,同时自身也有内容涵义。不过为了简化研究,本文将 tags 定位为一种功能结构化的微内容,是用来管理微内容的。Tags 兼具分类和标引组织的特征,其简短性、高效链接性和关联特征可以将大量微内容进行分类组织管理。关键是如何建立一个符合微内容自身特点的复合式分类组织管理模式,即融合 social tagging 与传统分类组织方式来分类组织微内容。

2 Social Tagging 与传统分类组织方式融合研究

Social tagging 与传统分类组织的比较融合研究是一个不断递进的过程——从最初预测发现 social tagging 的端倪并分析其概念特征,到批判性地看待其可能的价值作用,再到充分肯定其巨大作用,探讨比较与传统分类组织方式的优劣,并思考两者可能的结合方式。以下按此类研究发展过程进行阐述。

国外这类研究最初是从专家对新生的 social tagging 的不同态度引发的。Social tagging 的出现使一些信息科学领域的专家关注到这个能充分调动用户积极性的信息组织方式,C. Shirky、M. N. Uddin、C. Schwartz 等纷纷撰文分析这种组织方式的价值作用,预测其未来走向^[8-10]。不过 social tagging 的缺陷也是不可忽视的,E. Peterson、M. Guy、A. Noruzi、E. Speller 皆分析过其在语义、表述精确度和组织有效性等方面的

问题^[11-14]。

Social tagging 是一柄双刃剑,蕴含着潜在的利用价值,又存在着不少明显缺陷。唯有客观看待它,充分分析它相对于传统组织方式的优势和劣势,才能更好地运用它。在这个方面,国外部分专家就较理性地进行了探索和尝试。J. T. Tennis 不仅将 social tagging 与 LCSH (Library of Congress Subject Headings) 进行了较为全面的比较分析(尤其是在架构和功能方面),而且还从中提炼出了未来索引所需要的元素^[15]。E. Peterson 更进一步结合实例分析了如何使 social tagging 与传统主题分类法相结合,让它们并行为信息组织和资源整合服务^[11]。K. Yi 则借鉴 social tagging 的优势提出了改进 LCSH 的方法,并用一些统计学指标进行了小范围验证^[16-17]。E. Menard 将 social tagging 与传统分类组织方式的融合对象瞄准图像信息^[18],拓展了过去主要着眼于文本信息的局限。

国内研究人员对 social tagging 与传统分类方法的异同比较最初是伴随着挖掘 social tagging 信息组织机制和应用开发的研究而产生。这类比较研究作为一种独立研究的较少,多数是站在图书馆应用领域进行探讨,分析 social tagging 在图书馆信息组织和服务中的应用可能性^[19-22]。随着 Social Tagging 应用开展,从更广阔视角具体思考 Social Tagging 和传统分类组织方式的融合的文章逐步出现。范炜在开源内容管理系统 Drupal 资源架构中试图将受控词表和 folksonomy 进行结合,在两种分类组织的融合上进行了初步尝试^[23]。陈伟探讨了如何利用分面分类法控制 social tagging,进一步思考了 tags 形成后的控制过程^[24]。滕广青、毕强和高娅则利用关联规则和聚类分析等数据挖掘算法来分析揭示 tags 的导航体系^[25]。

综合国内外研究文献,目前针对 social tagging 和传统分类组织方式融合的研究,在以微内容为研究对象,充分利用两种分类组织方式的特点,对微内容的分类组织全过程进行整体阐述方面比较欠缺。对于微内容分类组织而言,social tagging 中的 tags 同为用户贡献分享,符合大众用户使用模式,是一种平面化的组织方式,弱点主要是较为随意;而传统分类组织由专家制定,具有较好的规范性和权威性,是一种从上至下的组织模式,弱点是不符合普通用户的使用习惯。这两种组织方式的优劣恰好相对,不仅仅对微内容,而且对整个网络信息资源组织都是有借鉴意义的。在 social tagging 的 tags 出现之前,网络信息组织一般常用的是等级结构,这种分类方法等级条目之间无法建立联系,

等级结构过于严格,若与 tags 进行融合,很难进行交融。本文建议采用分面分类法这种较为适合超文本链接且相对灵活的知识组织方式,也贴近人们多角度描述事物的思维习惯。

3 面向微内容的复合式分类组织模式探讨

为了表达的方便简捷,本文将“面向微内容的 social tagging 与传统组织方式结合应用模式”简称为“面向微内容的复合式分类组织模式”。图 1 列出的是笔者提出的面向微内容的复合式分类组织模式。

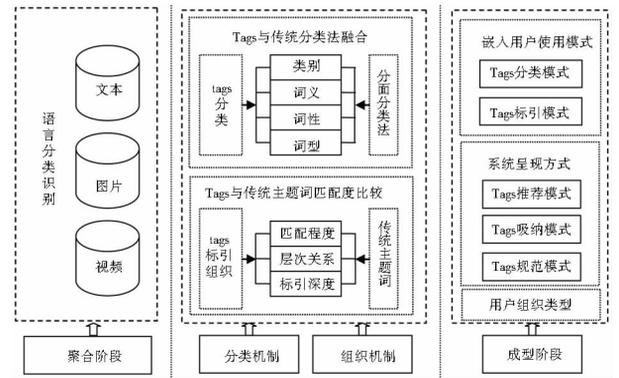


图 1 面向微内容的复合式分类组织模式

复合式分类组织模式分为三个阶段。第一阶段是聚合阶段,按不同的信息类型特点即文本、图片、视频进行聚合。刘军曾借鉴文献分类思想对微内容进行聚合^[26],但考虑到其主要针对图书馆数字参考进行使用,且与本文界定的微内容范畴不同,故未采用。第二阶段是机制设计,从分类和主题标引方式两个角度进行分析,一方面比较聚合后高频 tags 与分面分类法在类别、词义、词性和词型的相融合程度,另一方面比较聚合后高频 tags 与传统主题词在匹配程度、层次关系和标引深度等方面的异同,从而借鉴传统分类组织方法,构建面向微内容的 tags 分类机制和组织机制。第三阶段是成型阶段,基于分类机制和组织机制,构成包含分类模式和组织模式的嵌入面向微内容复合式分类组织成型阶段的用户使用模式,共同指导复合组织模式系统呈现模式,即 tags 的规范、tags 吸纳和 tags 推荐三种应用模式,其中从 tags 规范模式到 tags 吸纳模式再到 tags 推荐模式,依次成递进关系,前一种应用模式是后一种应用模式的基础。

3.1 微内容的聚合

微内容的聚合本质是为了共享交流和再利用开发。不同于纯粹机器程序聚合,微内容聚合更多地是

从用户和社会交流层面进行,在这种交流方式中还隐含着用户聚合和交互的动态过程,将微小的内容从不同侧面聚合,呈现用户的使用方式和认知特征。

微内容的聚合标准和方式是多种多样的。刘军借鉴文献分类的思想,提出按微内容产生顺序进行聚合,不过这种聚合方式主要针对图书馆数字参考提出,偏重服务帮助型和学术学科类^[26]。针对本文所界定的贡献内容的原创性微内容,笔者建议按信息类型进行聚合,区分指向不同信息类型微内容的 Tags 类型。

Tags 依托标注者对于所指向信息的理解感知能力,有目的地提出用于描述该信息的特征和属性,故以 tags 来分类组织微内容,必然需要考虑微内容自身的信息类型。Tags 的类型与指向信息有密切关系。例如 E. Angus 等所概括的指向图片的标注词分类^[27]与 S. A. Golder, T. B. Munk 和 M. E. I. Kipp 等^[28-30]指向文本标注词的分类有明显的区别。指向图片信息的 tags 突出标注的情感表述,对标注内容类型及时间地点和物品类别层次更多更复杂,并增加了版权方面的多种表述等。同时用户的群体性对 tags 分类也会产生较大影响,如同样都指向文本标注信息, M. J. Muller 等^[31]的分类由于分析站点的用户是 IBM 公司内部员工,在分类上就显现出很强的组织自身特点,为此成型阶段引入用户组织类型这一模块。

微内容按信息类型进行内容聚合,保留 tags 与微内容(即指向信息)之间的相互关联链接。不过微内容作为聚合底层数据库保留,作为验证回溯的依据;而聚合后的 tags 供第二阶段进行使用分析,tags 之间的相互关系(如共现、同义替代等)也在聚合中自动存储供使用。

微内容的语言特征是个相对复杂的研究主题,此次主要采用同种语言归类的原则,在聚合之前先将语言进行识别。不过随着全球化进程的加快,多语言混用的情况普遍存在,若出现混搭的多语言 tags 无法识别的情况,笔者建议回溯其产生的网站归属地国家进行检测。

3.2 面向微内容的复合式分类组织模式的机制设计

本阶段的机制设计分为两个部分:其一是基于 tags 与传统分类法融合的分类机制,其二是 tags 与传统主题词匹配度比较。分类机制与组织机制的差异在于,分类机制面向聚合形成一定数量的 tags 类型,是基于长期时间段的,而组织机制则针对用户 social tagging 时选择描述事物的用词,是短期时间点的产物。分类是通过一段时间 tags 的积累,纵向规范分门别类;组织

则是由横向进行关联度的合理界定,穷尽展现描述事物。

3.2.1 Tags 与分面组配法的融合 从分类角度,将 tags 的分类特征及其与传统分类方式中最适合网络信息组织的分面组配法进行比较融合,拟根据分面分类法的理论,分析聚合后 tags 的类别、词义、词性和词型与分面分类法的相容程度。考虑到 tag 的多样性,此次主要以聚合后高频 tags 与分面组配法进行融合,并根据微内容类型及其 Web 2.0 站点应用特征思考 tags 的分类机制设计。

类别是最重要的融合指标。类别指分面组配法对应的类目,主要是理顺 tags 类别与分面组配法类目的从属关系。分面组配法较其他传统分类方法灵活,最大的特点是面的相互独立及互不相关性^[32],相对于 tags 分类又有其概念专属性,依据其合并一些相关程度大的 tags,规范 tags,保证类别独立性。词义、词性和词型则是基于词个体本身进行规范。词义是在 tags 所在语言体系广泛使用涵义,若某个 tags 是多义词或出现隐含义,就需要回溯到该 tags 高频出现的语境下进行分析;若出现同义词,则以词频界定常用词义。词性即词类,是国际通用语言标准和语法规则,从名词、动词等词类特点分析标注词的不同表现和特征,并进一步区别主要词类的分类和组合特点。词型指抛开词义之后词的字面结构,主要是词的变形(单复数等形式)和词语符号搭配情况。

3.2.2 Tags 与传统主题词匹配 从主题标引角度,归纳不同信息类型下的高频主题,拟比较描述同一事物时,tags 与传统主题词表在概念关系、层次关系和标引深度等方面的匹配程度,其中视情况适当采用用户调研方式,思考可融合的主题标引模式。这种模式的最终目的是建立独立于 tags 和传统主题词的主题表,剔除恶意或错误的 tags,选用用户贡献集中的 Tags,计算 tags 共现频率,选择与其有层次关系的主题词,筛选失去使用效率的主题词,界定不同主题的标引深度。

比较指标主要从三个方面进行考虑:

其一是匹配程度。笔者参考 M. Wetterstrom 的文章^[33],又分三个标准:匹配、部分匹配和不匹配。具体选择标准如下:匹配指标注词可以在本语系的受控词表中找到词形和词义都全部匹配的主题词;部分匹配是指通过查询所找到的相应主题词与原标注词,只有词形或词义某一方面能匹配;不匹配是指参照标注词,在词表中进行词形和词义的反复查询对照,仍无法找到相匹配的标注词。

其二是层次关系,即区分是上位词还是下位词。这个主要评价用户 tagging 时与专家 tagging 的概念范畴,通常情况下专家的传统主题词更为笼统,多是 tags 的上位词。层次关系还包括共现关系,即同时出现的 tags。

其三是标引深度。这里借用情报语言学中的概念,是指对文献情报内容进行标引的程度,简单地说,指标引一篇文献所用的标识数量,此处指标注一条网络信息所使用的 tags 数目。这个比较指标主要通过比较 social tagging 与传统主题词描述同一微内容的标引深度,思考能最有效而准确地描述该微内容的 tags 数量。

3.3 面向微内容的复合式分类组织模式的成型阶段

分类机制是 Tags 分类模式的基础,组织机制是 Tags 标引模式的基础,两者共同构成了嵌入面向微内容复合式分类组织成型阶段的用户使用模式,共同指导复合组织模式具体成型阶段的三种呈现模式。

系统呈现方式又分三种具体模式,即 tags 规范模式、tags 吸纳模式和 tags 推荐模式。用户组织类型的设置是考虑到 tags 分类研究中发现 tags 与用户所在组织特性和自身类型有很强的相关性。若聚合站点为专业性站点(如科研交流站点)或企业组织站点等非通用流行站点,则需要根据用户组织类型进行分类组织机制的调整。

3.3.1 Tags 规范模式 Tags 规范模式是基础,是指用受控词与 tags 比较融合的结果规范用户 tagging 时的一些问题。用户 tags 存在随意性和自我特征,不仅词语的词义存在一词多义或歧义的可能,词型拼写错误,倾向采用单一词性抒发自我情绪等分类问题,而且在标引深度和概念层次上也把握不准,不利于重用和共享。Tags 规范模式是存储一些规范标准,如及时纠正用户 tags 的拼写错误;在用户使用情绪词或歧义词时提示其他用户的高频共享词或者与该词共词出现的相关词等。这些规范标准将服务于推荐模式,在推荐模式中具体呈现。

3.3.2 Tags 吸纳模式 Tags 吸纳模式是继承一些与 tags 匹配又能更好地反映针对事物的受控词及聚合时高频出现的新造 tags 或网络词。例如“给力”这个词过去只是北方方言,但随着其近几年在网络中被普遍使用,如今该词已被收入《现代汉语词典》第六版,被人民网、新华网等权威官方站点频繁使用。Tags 吸纳过程需要界定其归属,即将其纳入相同或相近的类和主题,若出现一系列新词又无法归类,应该对分类进行

扩展和延伸,将同现 tags 进行汇总,建立新的分类或将指向新主题标示出来,这个过程中要尽量控制边缘主题过多出现。

3.3.3 Tags 推荐模式 Tags 推荐模式是基于 tags 规范和吸纳后的一种智能阶段,是指在用户进行标注时,站点或应用会提供一种使用建议,这不仅仅是推荐某个 tags 给用户参考,类似搜索引擎那种下拉高频词提醒,而且是一种 tagging 建议,基于规范后的 tags,在吸纳一定 tags 后,根据已有分类组织策略给用户提供一种针对要 tagging 对象的全方面建议,如该词常用的分类、标引深度、词性等。例如年轻人最爱使用 Apple 手机,这个词本义是指一种水果,但随着 Apple 手机广泛流行,如今一提 Apple 人们就会联想到智能手机及其背后强大的苹果公司,所以当出现 Apple 词时,就基于已有分类组织策略和规范标准进入推荐模式环节。首先让用户进行类型选择,即“Apple”究竟是从属于水果类还是科技类,这样自然将一词多义的问题先规避;若选择为科技类,则然后提供高频共词如 iOS、iPhone、iPad 等供用户参考,同时提示针对该词常用适合的标引深度,即 tags 标引数目。

4 结 语

针对主题千变万化、格式灵活多样的“碎片化”微内容难以分类组织的难题,笔者通过界定微内容研究范畴,探讨一种面向微内容分类组织的融合 social tagging 与传统分类组织法(即分面分类法和主题词法)的复合式分类组织模式。这种模式分阶段分层次,融合 social tagging 和传统分类组织方法的各自优点,力求针对微内容的自身特征,构建一种分类组织模式,从而有助于 tags 设计有效性、微内容共享效率及微内容检索查询的准确性等微内容科学合理管理组织的诸多问题。不过这种模式是一种构建方案和设想思路,尚处于尝试成型阶段,笔者将在未来研究中继续将之嵌入 Web 2.0 环境的微内容平台中进行实证应用和评估检测。

参考文献:

- [1] Nielsen J. Microcontent: How to write headlines, page titles, and subject lines[EB/OL]. [2013-03-10]. <http://www.useit.com/alertbox/980906.html>.
- [2] Cmswiki. MicroContent[EB/OL]. [2013-03-10]. <http://www.cmswiki.com/tiki-index.php?page=MicroContent>.
- [3] 范艳君. 微内容及其在图书馆服务中的新应用[J]. 图书馆学研究,2009(11):67-69.
- [4] 邓顺国,宗乾进,杨黎星. 微内容:网络信息服务模式的变革[J]. 图书馆理论与实践,2009(10):36-38.
- [5] 王国才. 基于本体的微内容聚合与查询技术[D]. 合肥:合肥工业大学,2009:28-34.
- [6] 张乾,蔡淑琴,石双元. 微内容序化方法与应用实例[J]. 图书情报工作,2012,56(4):38-41.
- [7] 张琪玉. 网络信息检索用语言的发展趋势[J]. 图书馆杂志,2001(3):5-7,22.
- [8] Shirky C. Ontology is overrated: Categories, links, and tags[EB/OL]. [2013-03-10]. http://www.shirky.com/writings/ontology_overrated.html.
- [9] Uddin M N, Mezbah-ul-Islam M, Haque K M G. Information description and discovery method using classification structure in Web[J]. Malaysian Journal of Library & Information Science, 2006,11(2):1-20.
- [10] Schwartz C. Thesauri and Facets and tags, oh my! A look at three decades in subject analysis[J]. Library Trends,2008,56(4):830-842.
- [11] Peterson E. Beneath the metadata: Some philosophical problems with folksonomy[J/OL]. D-Lib Magazine, 2006,12(11). [2013-03-10]. <http://www.dlib.org/dlib/november06/peterson/11peterson.html>.
- [12] Guy M, Tokin E. Folksonomies tidying up tags[J/OL]. D-Lib Magazine,2006,12(1). [2013-03-10]. <http://www.dlib.org/dlib/january06/guy/01guy.html>.
- [13] Noruzi A. Folksonomies: (Un) Controlled vocabulary? [EB/OL]. [2013-03-10]. http://eprints.rclis.org/archive/00011286/01/Folksonomy,_UnControlled_Vocabulary.pdf.
- [14] Speller E. Collaborative tagging, folksonomies, distributed classification or ethnoclassification: A literature review[J/OL]. Library Student Journal,2007,11(2). [2013-03-10]. <http://www.librarystudentjournal.org/index.php/lsj/article/view/45/58>.
- [15] Tennis J T. Social tagging and the next steps for indexing[EB/OL]. [2013-03-10]. <http://journals.lib.washington.edu/index.php/acro/article/view/12493>.
- [16] Yi K. A semantic similarity approach to predicting Library of Congress Subject Headings for social tags[J]. Journal of the American Society for Information Science and Technology, 2010,61(8):1658-1672.
- [17] Yi K, Chan L M. Linking folksonomy to Library of Congress subject headings: An exploratory study[J]. Journal of Documentation, 2009,65(6):872-900.
- [18] Menard E. Ordinary image retrieval in a multilingual context: A comparison of two indexing vocabularies[J]. Aslib Proceedings, 2010,62(4/5):428-437.
- [19] 乐庆玲. 基于协同机制的 Tag 资源自动分类研究[J]. 现代图书情报技术,2007(9):57-61.
- [20] 羌丽,张学莲,侯汉清. 图书大众标注评介[J]. 图书馆杂志,2009(2):21-26.

- [21] 王翠英. Folksonomy 与主题标引[J]. 情报理论与实践, 2007 (5): 693 - 697.
- [22] 王翠英. 本体与 Folksonomy 的比较研究[J]. 图书馆建设, 2008 (5): 85 - 88.
- [23] 范炜. Drupal 分类组织机制研究: 一种复合信息组织模式[J]. 图书馆杂志, 2010(1): 23 - 26.
- [24] 陈伟. 用户参与的社会性标签信息组织探讨[J]. 图书馆学研究, 2012(6): 79 - 82.
- [25] 滕广青, 毕强, 高娅. 基于概念格的 Folksonomy 知识组织研究[J]. 现代图书情报技术, 2012(6): 22 - 27.
- [26] 刘军. 微内容分层聚合及在数字参考咨询中的应用[J]. 图书馆情报工作, 2011, 55(12): 106 - 109.
- [27] Angus E, Thelwall M, Stuart D. General patterns of tag usage among university groups in Flickr[J]. Online Information Review, 2008, 32(1): 89 - 101.
- [28] Golder S A, Huberman B A. Usage patterns of collaborative tagging systems[J]. Journal of Information Science, 2006, 32(2): 198 - 208.
- [29] Munk T B, Mork K. Folksonomies, tagging communities, and tagging strategies-An empirical study[J]. Knowledge Organization, 2007, 34(3): 115 - 127.
- [30] Kipp M E I. @ toread and Cool: Tagging for time, task and emotion[EB/OL]. [2013 - 03 - 10]. http://www.vraweb.org/conferences/sandiego2008/sessions/session1/Margaret_Kipp.pdf.
- [31] Muller M J, Ehrlich K, Farrell S. Social tagging and self-tagging for impression management[EB/OL]. [2013 - 03 - 10]. [http://domino.watson.ibm.com/cambridge/research.nsf/c9ef590d6d00291a85257141004a5c19/27658d7dcf7e8cce852572330070244d/\\$FILE/TR2006-2.pdf](http://domino.watson.ibm.com/cambridge/research.nsf/c9ef590d6d00291a85257141004a5c19/27658d7dcf7e8cce852572330070244d/$FILE/TR2006-2.pdf).
- [32] 施国良. 国内外分面分类法基本理论研究述评[J]. 图书馆学研究, 2008(12): 2 - 5.
- [33] Wetterstrom M. The complementarity of tags and LCSH-A tagging experiment and investigation into added value in a New Zealand library context [J]. New Zealand Library & Information Management Journal, 2008, 50(4): 296 - 310.

Study on the Integrated Mode for Classifying and Organizing Micro-content

Xu Chen

School of Business, Central South University, Changsha 410083

[Abstract] This paper analyzes the comparative advantages between social tagging and traditional classification organization. Then it tries to develop a three-phase integrated mode to classify and organize micro-content based on its connotation and organizational characteristics, which are aggregation, mechanism design and formation phase. Micro-content is aggregated in the aggregation phase according to information types. Mechanism design includes classified mechanism which integrates tags and faced classification from the perspective of category, meaning, appearance and part of the speech, and organization mechanism which compares tags and traditional subject terms from the perspective of matching degree, hierarchical relation and indexing depth. The formation phase optimizes tags with standardization, enrollment and recommendation to manage micro-content.

[Keywords] micro-content social tagging tag classification and organization

(上接第 102 页)

Study on a Hot Topics Analysis System based on Time Sliced Topic Model

Liao Junhua Sun Keying Zhong Lixia

Institute of Scientific & Technical Information, Shandong University of Technology, Zibo 255049

[Abstract] A Hot Topics Analysis System (HTAS) based on time sliced network data was proposed. HTAS realized the network hot topic data source automatically collected, acquisition and storage. HTAS integrated the google revenue segmentation system IKAnalyzer to batch processing of Chinese documents. HTAS used LDA model to extract and time label to find the evolution of the hot topics on the network. Experiments of Diaoyudao as the hot event show that, the system can effectively acquire, store and analyze this hot topic evolution trend.

[Keywords] topic model topic evolution hot topics LDA