

Tax-folk 混合导航: 社会化标注系统资源聚合的新模型^{*}

张云中 杨 萌

摘 要 Web2.0 环境下,专家分类法与大众分类法呈现出互补和融合的姿态,tax-folk 混合导航已成为优化社会化标注系统资源聚合的一种新思路。本文通过剖析 tax-folk 混合导航模型生成机理,采用形式概念分析方法,构建出包含数据准备、概念格构建、概念格分析、tax-folk 映射、tax-folk 混合导航树和输出与评价六个模块的 tax-folk 混合导航模型。研究发现,tax-folk 混合导航模型兼顾了专家分类法和大众分类法的优势,实现了树状“干强枝繁叶茂”的资源聚合,有效提高了社会化标注系统中资源的可查找性。图 10。表 1。参考文献 11。

关键词 社会化标注系统 专家分类法 大众分类法 资源聚合 混合导航 形式概念分析

分类号 G254.1

Tax-folk Hybrid Navigation: A New Model for Resource Aggregation in Social Tagging System

Zhang Yunzhong & Yang Meng

ABSTRACT With the emergence of web2.0, taxonomy and folksonomy have shown a trend of complementarity and integration, and a hybrid tax-folk navigation structure has become a new way to facilitate the resource aggregation in social tagging system. In this paper, the generation mechanism of tax-folk hybrid navigation model is analyzed, and the tax-folk hybrid navigation model, which consist of six modules: data preparation, concept lattice construction, concept lattice analysis, tax-folk mapping, tax-folk hybrid navigation tree and outputs & evaluation, is constructed with the theory of formal concept analysis. The study finds that tax-folk hybrid navigation model takes into account the advantages of taxonomy and folksonomy, achieves a tree-like resource aggregation, and effectively improves the resources findability in social tagging system. 10 figs. 1 tab. 11 refs.

KEY WORDS Social tagging system. Taxonomy. Folksonomy. Resource aggregation. Hybrid navigation. Formal concept analysis.

1 引言

传统的资源组织和聚合中,专家分类法(taxonomy)扮演着至关重要的角色。专家分类法由权威专家选词制定而成,呈现出层级化的资源组织结构,语义关系规范,符合人们对客观事物分类的认知模式,但其制定费时费力,词表更新缓慢。Web2.0 出现后,社会化标注活动迅速普及,用户的社会化标注活动变得愈发普遍,用户快速、自由地

威专家选词制定而成,呈现出层级化的资源组织结构,语义关系规范,符合人们对客观事物分类的认知模式,但其制定费时费力,词表更新缓慢。Web2.0 出现后,社会化标注活动迅速普及,用户的社会化标注活动变得愈发普遍,用户快速、自由地

^{*} 本文系教育部人文社会科学青年基金项目“社会化标注系统资源多维度聚合与导航研究”(项目编号:13YJC870032)和上海市教育委员会科研创新项目“Web2.0 下本体与大众分类法的互补与融合”(项目编号:14YS007)的研究成果之一。

通讯作者:张云中,Email:zhang-yun-zhong@126.com

对资源进行分类的欲望愈发强烈,随之出现了应用于社会化标注系统的新资源组织形式——大众分类法(folksonomy)。与专家分类法不同,大众分类法是灵活的扁平化平面结构模式,标签之间没有隶属关系,语义模糊稀疏。

Web2.0 环境下,如何选择或合理应用上述两种语义表示和资源组织模式成为当前学界普遍关注和思考的问题。近年来,国内外学者都认为,两种模式各具优劣,存在互补与融合的必要性和可行性。在资源组织过程中,若将二者结合起来,扬长避短,以专家分类法为树干,将各种标签映射到专家分类法中作为树叶,一棵用于资源聚合和导航的知识之树便会破土而出。本文的研究正是基于此思路,旨在探索一种新的 tax-folk 混合导航模型,以更好地促进社会化标注系统资源聚合与导航。

2 国内外研究现状

国内外针对专家分类法和大众分类法混合导航的相关研究先后经历了两个阶段:探索阶段和发展阶段。

(1)探索阶段。Web2.0 网络环境下,大众分类法作为新兴社会化组织工具出现后,与传统专家分类法产生了激烈碰撞^[1],新兴的大众分类法是否会替代专家分类法抑或二者走向融合成为国内外学者关注的热点。于是,有学者试图探讨专家分类法和大众分类法在内涵、特性等方面的差异,辨析二者的相互关系及应用情境^[2-3],也有学者从实践角度探索流行网站使用专家分类法和大众分类法时的优劣与取舍^[4]。这些关于二者异同分析的研究都指向同一结论:专家分类法采用专家选词构建的层级结构组织资源,语义精准规范,但构建费时费力且更新缓慢;大众分类法采用扁平结构,语义稀疏模糊,但自由灵活成本低廉;两者存在显著的优劣互补性,应该扬长避短并有机融合而非相互替代。

(2)发展阶段。发展阶段以探索阶段为基础,其关注的首要问题是大众分类法与专家分类法整合该从何角度切入。学界普遍认为,专家分类法融入大众分类法,大众分类法融入专家分类法,构建

标签层次结构和 tax-folk 混合分类(兼具两种分类法特征的混合分类结构,下同)是四种主流的整合方式^[5]。对应于四种方式,代表性的方法有:专家分类法引导的自由分类方法^[6],tagsonomy 分类法^[7],基于标签获取分类词层次结构^[8]以及 TaxoFolk 混合分类法^[9]。随着理论研究不断深入,tax-folk 混合分类在实践领域(如疾病博客资源组织^[10])的应用中也崭露头角。专家分类法与大众分类法融合呈现出理论与实践并重、方法途径多元的趋势。

当然,国内外相关研究在带给我们启示的同时,也有尚未解决的问题:①研究 tax-folk 混合分类时,关于 tax-folk 混合导航模型生成机理的研究是不可或缺的,这是国内外相关研究没有重视的地方。②受文献[9]启发,形式概念分析(简称 FCA)在实现 tax-folk 混合分类的过程中较其他数据挖掘方法更具简易、高效和可视化的优势,但文献[9]所采用的复杂映射规则及算法却有待改进。形式概念分析在 tax-folk 混合导航中如何更合理、科学、简易地使用,应该成为着重思考、讨论和拓展的方向。

本文正是针对国内外实现 tax-folk 混合导航中存在的两点不足而展开研究,通过剖析 tax-folk 混合导航模型生成机理,构建基于形式概念分析的 tax-folk 混合导航模型,优化基于形式概念分析的 tax-folk 映射规则,从而实现更高效的社会化标注系统资源聚合。

3 形式概念分析视角下的 tax-folk 混合导航模型生成机理

要解决 tax-folk 混合导航问题,首先有必要对 tax-folk 混合导航模型的生成机理进行剖析。本文通过分析专家分类法和大众分类法在社会化标注系统资源聚合与导航问题上的角色和作用,总结出形式概念分析视角下的 tax-folk 混合导航模型的生成机理(见图1)。

图1中,同为资源组织和知识语义表示的专家分类法和大众分类法,往往同时被用来描述社会化标注系统中的资源,在资源组织和语义表示的效果

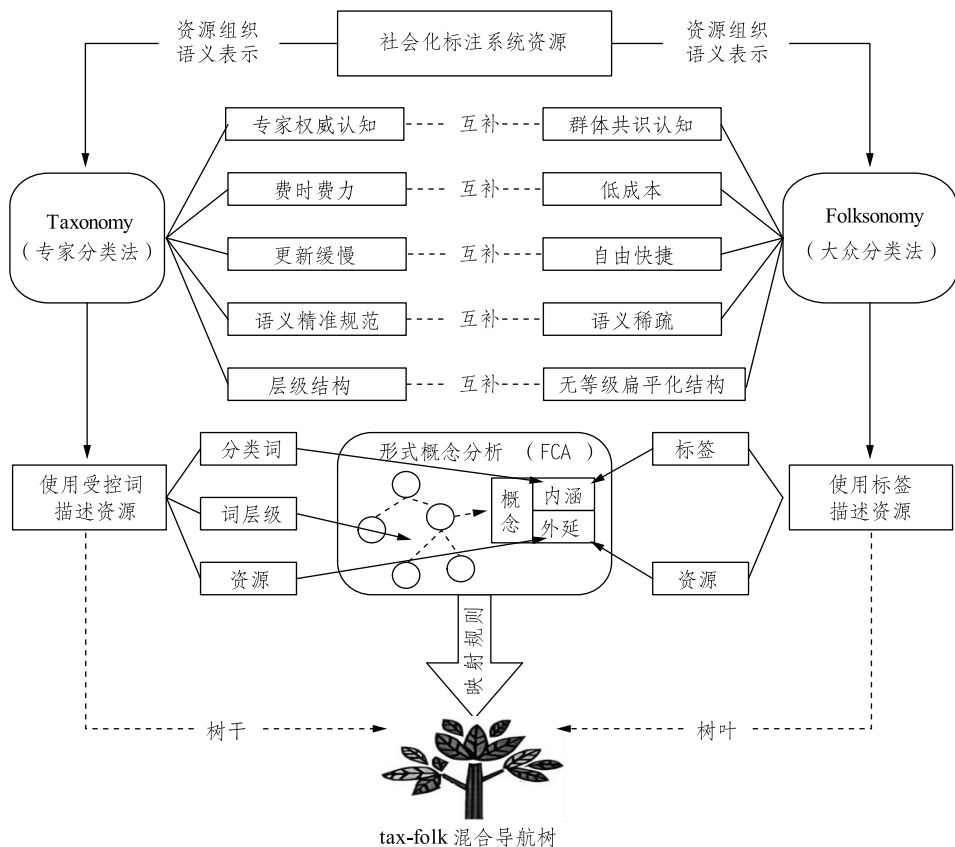


图1 形式概念分析视角下的 tax-folk 混合导航模型生成机理

上,两者各具优劣。专家分类法代表了专家的权威认知,按照人类认识客观事物的分类模式用层级结构组织资源,其语义精准规范,但同时也存在构建费时费力,更新缓慢的缺陷;大众分类法既体现了群体的共识性认知,也保留了用户的个性化认识,其成本低廉,自由灵活快捷,但其语义较为稀疏模糊,无等级扁平化的资源组织结构既不符合人们分类的思维惯性,也弱化了资源间的语义关联。

剖析专家分类法和大众分类法的优劣特点,不难发现二者具有互补性。专家分类法使用受控词和词间关系描述资源,分类词之间有着明确、规范、清晰的层级关系,自顶向下,层层相扣,而且分类词由专家权威制定,词量少但概括性强,好比树从根部理出枝干一样;大众分类法使用标签描述资源,标签之间没有隶属层级关系,自底向上,标签量

多而繁杂,好比树叶、弱枝一样。在社会化标注系统资源聚合与导航的“知识大树”上,专家分类法好似“干强枝粗叶少”,而大众分类法则是“干少枝弱叶茂”。Tax-folk 的混合之道,就是要完成大众分类法向专家分类法的嫁接,本文认为,形式概念分析就是这把嫁接的剪刀。

形式概念分析(FCA)中,概念由内涵和外延共同构成,并以形式背景为数据结构组织概念的两个维度。形式背景转换为概念格后,形成概念节点和概念之间的隶属关系,最终以可视化概念格的形式呈现出概念节点和概念间的关系。分别以“分类词—资源”和“标签—资源”构建形式背景并转换概念格,可以发现,资源可以映射为概念节点的外延,分类词可映射为概念节点的内涵,标签也可映射为概念节点的内涵。在两组概念格中,可基于如

下的基本原理寻找相似节点;对于相同的资源,专家分类法或大众分类法在揭示资源的语义和分类上用词(分类词或标签)存在着相似性和相关性,也就是说,外延相同的概念节点具有相似性。通过相似概念节点的比对,就可以将相似概念节点的标签组嫁接到分类词上,从而实现大众分类法向专家分类法的嫁接。嫁接的过程就是按照映射规则完成映射的过程,详见下文第 4.4 部分。

完成 tax-folk 嫁接后,社会化标注系统资源聚合与导航的知识大树既保留了专家分类法“强干粗

枝”的一面,同时又吸纳了大众分类法“叶茂”的一面,使得社会化标注系统资源聚合与导航成为一棵枝繁叶茂的丰满大树——tax-folk 混合导航树。

4 基于形式概念分析的 tax-folk 混合导航模型构建

以形式概念分析视角下的 tax-folk 混合导航模型生成机理为基础,本文提出了基于形式概念分析的 tax-folk 混合导航模型,如图 2 所示:

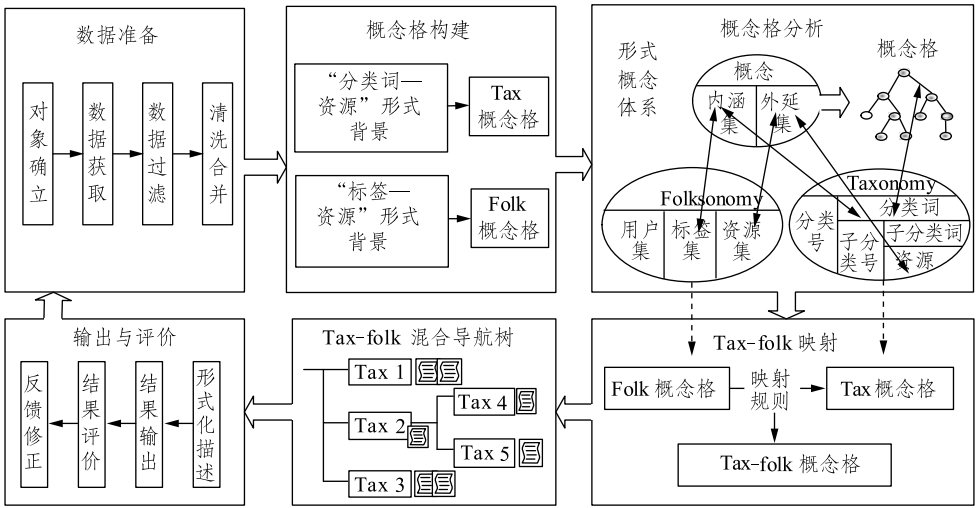


图 2 基于 FCA 的 tax-folk 混合导航模型

基于形式概念分析的 tax-folk 混合导航模型共包括六大部分:数据准备、概念格构建、概念格分析、tax-folk 映射、tax-folk 混合导航树构建、输出与评价。

4.1 数据准备

数据准备阶段是构建 tax-folk 混合导航模型的出发点和基础阶段,至关重要。数据的准备工作充分与否直接影响 tax-folk 输出结果的优劣。该阶段的主要任务是选定待处理的社会化标注系统作为对象,并界定清晰的时间范围和空间范围,进而获取描述社会化标注系统资源语义的两类数据集:专家分类法数据集和大众分类法数据集。专家分类法数据集主要包括两类元素:分类词集和

资源集,专家分类法数据集由专家制定,数据比较规范,因而不需要过多处理;大众分类法数据集细分为用户集、资源集和标签集三类,标签集是该阶段处理的重点。标签集的处理包括清晰、合并和过滤等步骤,具体的操作步骤文献[11]有详细说明,此不赘述。

为清晰阐明 tax-folk 混合导航模型的相关原理和流程,先假定某社会化标注系统资源的两类数据集经过准备过程处理后,专家分类法数据集为 $\{\{t1, t2, t3, t4, t5, t6\}, \{r1, r2, r3, r4, r5, r6\}\}$, 大众分类法数据集为 $\{\{f1, f2, f3, f4, f5, f6\}, \{r1, r2, r3, r4, r5, r6\}\}$, 专家分类法与大众分类法资源组织体系如图 3 所示。

Taxonomy 资源组织体系			Folksonomy 资源组织体系		
t1 (分类词)	资源 r 1		资源 r 1	f1 f6	
t2 (分类词)	资源 r 4		资源 r 2	f1 f2 f4	
t5 (分类词)	资源 r 5		资源 r 3	f3 f4	
t6 (分类词)	资源 r 6		资源 r 4	f1	
t3 (分类词)	资源 r 2		资源 r 5	f2 f5	
t4 (分类词)	资源 r 3		资源 r 6	f1 f2 f4 f6	

图 3 数据准备后的初始数据集

4.2 概念格构建

该阶段的主要任务是依据初始数据分别构建 tax 概念格(由专家分类法数据集生成的概念格)和 folk 概念格(由大众分类法数据集生成的概念格)。

根据数据准备阶段得出的专家分类法数据集,以分类词 t_i 为形式概念的属性和内涵,以资源

r_m 为概念的实例和外延,构建“分类词—资源”形式背景(即专家分类法的形式背景)并将其转换为相应的 tax 概念格。同时,根据数据准备阶段得出的大众分类法数据集,以标签 f_j 为形式概念的属性和内涵,以资源 r_m 为概念的实例和外延,构建“标签—资源”形式背景(即大众分类法的形式背景)并将其转换为相应的 folk 概念格,如图 4 所示:

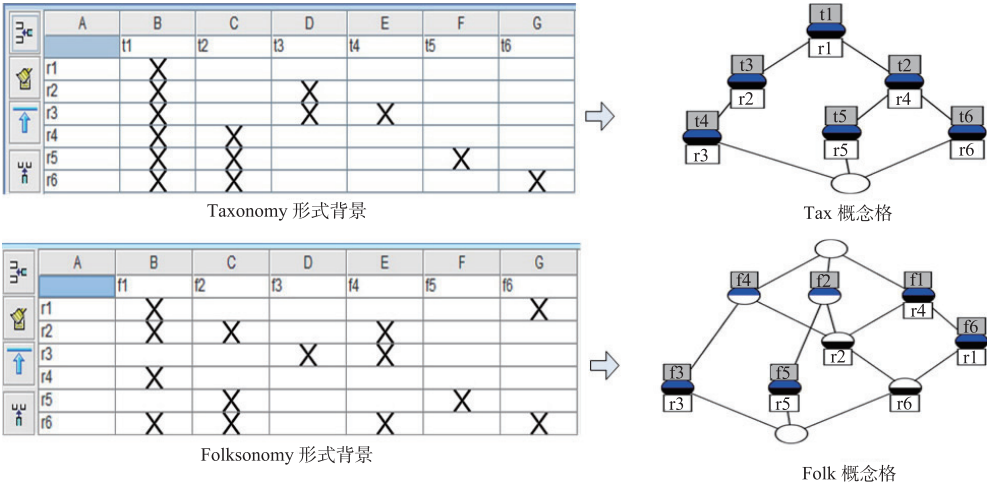


图 4 概念格构建

4.3 概念格分析

根据形式概念分析理论,我们可以得知,概念节点表示一个形式概念,它由两部分组成:概念的内涵集和概念的外延集,形式概念的表达式可采用 $\{ \{ \text{外延集} \}, \{ \text{内涵集} \} \}$ 的方式。

利用概念格理论,我们还可以得知,若概念格中的两个概念 (A, B) 和 (C, D) 存在 $(A, B) \leq (C, D)$

的子概念与父概念的关系,那么:
 $(A, B) \leq (C, D) \iff A \subseteq C (\iff D \subseteq B)$
也就是说,父概念与子概念存在这样的关系:父概念内涵集 D 是子概念的内涵集 B 的子集,同时子概念外延集 A 是父概念的外延集 C 的子集。换言之,子概念在拥有其自有内涵的同时,继承了父概念所有的内涵,父概念在拥有其自身外延的同

时,子概念的所有外延都是父概念的外延。

根据这一条性质,本文做出进一步定义,子概念内涵集中的自有内涵称为主内涵,其继承父概念的内涵称为副内涵,父概念外延集中的自有外延称为主外延,其兼容子概念的外延称为副外延。

4.3.1 Tax 概念格分析

Tax 概念格中共形成了四个要素:概念节点、概念外延、概念内涵、概念节点间的关系,如图 5 所示。概念节点代表的概念涵义为“使用内涵集里的分类词描述外延集中的资源”,概念外延代表节点概念包含的资源,概念内涵代表其外延中的资源所隶属的分类词。概念节点间关系体现了概念格中父概念与子概念的相互关系。

Tax 概念格分析的最核心内容是分类词与分类词之间的关系,也就是子概念属性和父概念属性之间的关系。在形式逻辑中,子概念(即种概念)

与父概念(即属概念)之间存在“种概念=种差+属概念”的客观关系,也就是说,种差是区别子概念和父概念的关键属性。在 tax 概念格中除顶层概念节点外,每一个概念的内涵来自两部分,主内涵(自有内涵)和副内涵(继承内涵),而且副内涵都反映“属概念”的特征,而主内涵扮演的正是种差的角色。种差的作用是对属概念做出进一步限定,因而 tax 概念格中每一个概念节点的主内涵都是对其上层节点内涵的再限定,内涵之间形成了逐级限定的关系。在 tax 概念格中,概念节点的内涵就是分类词,所以经 tax 概念格分析,分类词之间呈现出逐级限定的关系。

一般情况下,由于分类词逐级限定关系清晰, tax 概念格都会展现出清晰的树状架构,这种树状架构与社会化标注系统资源的专家分类法分类体系架构是拟合的。

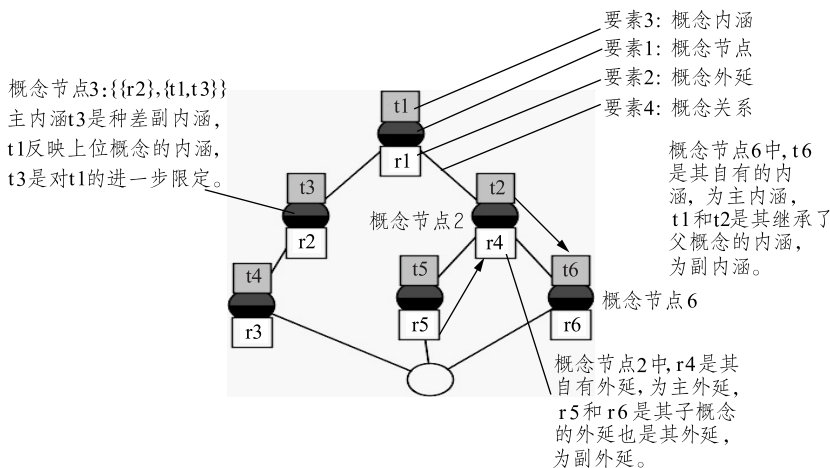


图 5 Tax 概念格分析

4.3.2 Folk 概念格分析

Folk 概念格同样包含四个要素:概念节点、概念外延、概念内涵、概念节点间的关系,如图 6 所示。Folk 概念格中,概念节点代表的概念涵义为“使用内涵集里的标签标记外延集中的资源”。概念外延代表概念节点所包含的资源,概念内涵表示其外延中资源被张贴的标签。概念节点之间也呈现出继承关系,这与 tax 概念格并无差异。

Folk 概念格分析的关键点在于寻找标签与标

签之间的关联。在社会化标注系统中,标签之间呈现出扁平化的组织结构,标签与标签本身并无直接关联,标签关系更是无从谈起。但将标签集—资源集装载入 folk 形式背后,就可利用概念格构造算法构建 folk 概念格。概念生成过程本质上就是概念聚类过程,概念格构造算法的本质就是聚类算法——它将具有相同属性的对象聚集起来,共同构成形式概念。经过概念格构建的过程,标签间关系也经历了从无到有的过程。如图 6 所示,概念节点

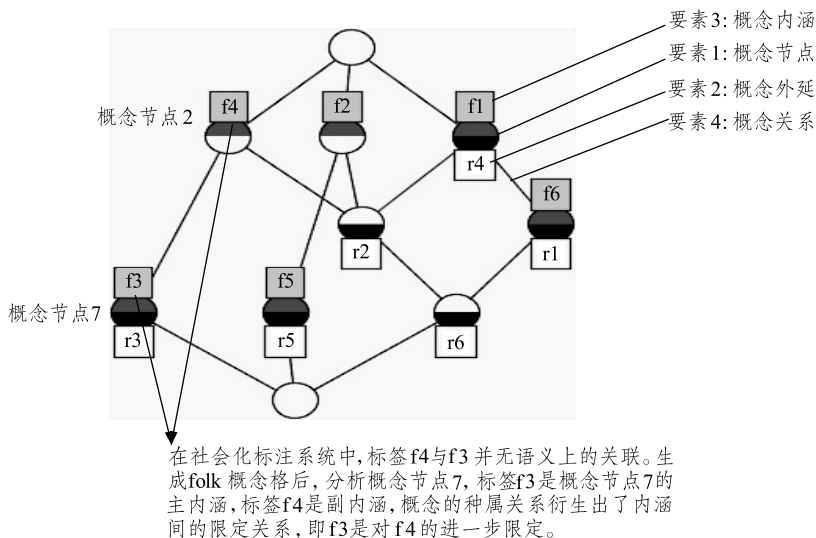


图6 Folk概念格分析

2|{r2,r3,r6}|,{f4}|和概念节点7|{r3}|,{f4,f3}|形成了父概念—子概念的关系,这种父子关系体现出了继承性,即子概念继承了父概念的内涵f4,同时又有其主内涵f3。进一步,从形式逻辑角度看,主内涵f3扮演着种差的角色,而副内涵f4反映属概念的特征,那么,主内涵f3是对副内涵f4的进一步限定。例如上位概念是|{图书1,图书2}|,|文学|,下位概念是|{图书1}|,|文学,小说|,上下位概念间是继承关系,从而衍生出文学→小说的标签限定关系。以此类推,内涵之间逐级形成了限定关系,这就意味着原本并无关联的标签之间也产生了逐级限定关系,而且这种标签的逐级限定关系与tax概念格中分类词的逐级限定关系是相通的。

由于标签关系繁杂,与tax概念格相对比,folk概念格一般呈现为复杂的网状结构。

对tax概念格和folk概念格分析后不难发现,主外延相同的tax概念节点和folk概念节点存在相似性,原因是对于相同的资源,专家分类法或大众分类法在揭示资源的语义和分类上用词(分类词或标签)存在着相似性和相关性。本文认为,tax-folk混合导航的建立,应从主外延相同的概念节点入手,完成大众分类法向专家分类法的嫁接。

4.4 Tax-folk映射

该阶段的主要任务是建立tax-folk映射规则,实现大众分类法向专家分类法的映射。Tax-folk建立在概念格分析的基础上,主导思想是以tax概念格为干和枝,以folk概念格为枝叶向干枝嫁接,具体的映射规则是:

(1)选取tax概念格作为主架构,删除其末端节点,tax概念格呈树状。

(2)将tax概念格的概念节点自顶向下,从左至右以 $C_1, C_2 \dots C_n$ 分别编码。从tax概念格的顶端节点出发,依编码顺序选择具有主外延的概念节点 C_n 作为映射对象,找出其主外延 r_m 。

(3)从folk概念格中找出主外延亦为 r_m 的概念节点,整理出该节点的所有内涵,节点的每条内涵均对应为一个标签 f_j 。

(4)定义数据结构 $f_j(n)$, n 表示folk概念格中标签 f_j 包含的所有资源数;定义数据结构 $f_{jk}(n) \rightarrow f_{j1}(n)$ 表示经聚类后folk概念格中标签 f_{jk} 间的限定关系。

(5)以主外延 r_m 为媒介,将folk概念格中形成的枝叶 $f_{jk}(n) \rightarrow f_{j1}(n)$ 和 $f_j(n)$ 嫁接到tax概念格节点 C_n 的内涵上。

(6)若 C_n 存在多个 r_m ,则对每一个 r_m 同时进

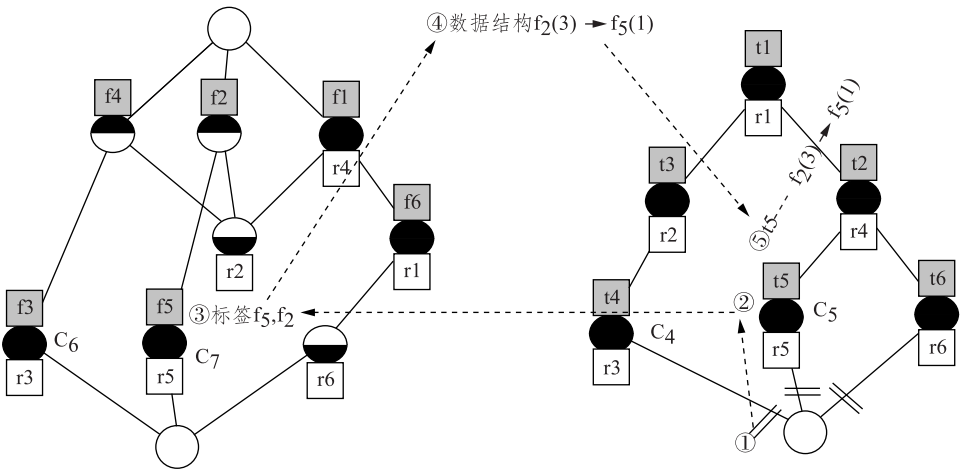


图 7 一个概念节点映射示例

行 3~5 的操作。

(7) 依次遍历 tax 概念格中所有 C_n , 即可获得映射后的 tax-folk 原型。

结合图 5、图 6, 以资源 r_5 为例, tax 概念格其概念节点是 C_5 , 其映射过程如图 7 所示。

4.5 Tax-folk 混合导航树构建

根据 tax-folk 映射规则, 完成所有待映射概念节点的映射后, 就可以构建 tax-folk 混合导航树。Tax-folk 混合导航树由三部分组成, 如图 8 所示。第①部分是导航树的架构部分, 即树干和树枝, 其

原型是专家分类法的分类词及词间关系。第②部分是嫁接到专家分类法上的标签, 即树叶和弱枝 (体现逐级限定关系的标签组合), 其原型是大众分类法中的标签和标签组。第③部分是社会化标注系统资源, 即树的果实, 资源通过两种链接渠道链接到 tax-folk 混合导航树上——tax 渠道和 folk 渠道。Tax-folk 混合导航树吸纳了专家分类法和大众分类法的优势, 既建立了标签与分类词之间的关系, 又聚合出了标签与标签之间的关系, 更重要的是它同时使用两种渠道聚合资源, 并以树的可视化形式展现结果。

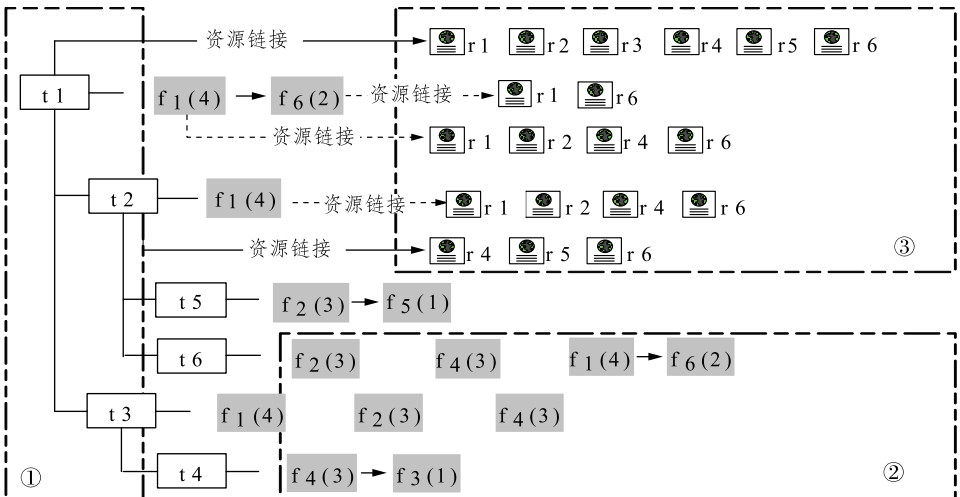


图 8 Tax-folk 混合导航树

4.6 输出与评价

输出与评价是基于形式概念分析的 tax-folk 混合导航模型的收尾阶段,该阶段的主要任务包括:①形式化描述。根据需求,选择合适的数据描述语言将 tax-folk 混合导航树进行形式化的描述。②结果输出。将构建好的 tax-folk 混合导航树应用到社会化标注系统资源聚合平台中,投入使用。③结果评价。成立由专家和用户构成的评价小组,建立合适的评价指标体系对构建过程、输出结果及其应用状况进行科学客观的评价。④反馈修正。在使用和评价的基础上不断反馈、修正构建方法及输出结果。

5 Tax-folk 混合导航的应用:以豆瓣读书为例

豆瓣网是国内社会化标注系统的典型代表,

表 1 数据准备阶段后的实验对象数据(部分)

编号	资源名称	中图分类号	资源的专家分类法	资源标签
r1	情人	I565.45	文学 > 各国文学	杜拉斯(17152)、法国(7491)、外国文学(5339)、文学(4725)
r2	平凡的世界(全三部)	I247.57	文学 > 中国文学 > 小说	路遥(18466)、小说(9088)、中国文学(7678)、文学(3321)
r3	傲慢与偏见	I561.44	文学 > 各国文学	傲慢与偏见(9930)、外国文学(6529)、英国(3467)、文学(2417)
r4	悲伤逆流成河	I247.57	文学 > 中国文学 > 小说	郭敬明(12315)、青春(4335)、成长(1461)、80后(1324)
r5	向左走·向右走	J228.2	艺术 > 绘画 > 中国绘画作品	几米(13858)、绘本(11017)、向左走向右走(7127)、台湾(2437)
...	...	...	...	...
r30	他的国	I247.57	文学 > 中国文学 > 小说	韩寒(11854)、小说(4354)、讽刺(2369)、文学(1435)、中国文学(945)
r31	茶花女	I565.44	文学 > 各国文学	茶花女(4846)、小仲马(4745)、外国文学(4285)、法国(2211)、文学(2174)
r32	黄金时代	I247.53	文学 > 中国文学 > 小说	王小波(13996)、小说(4164)、中国文学(2719)、当代(1529)、文学(1139)

① 豆瓣网.豆瓣读书[EB/OL].[2013-08-07].http://book.douban.com/.
② 中国国家图书馆联机公共目录查询系统[EB/OL].[2013-08-07].http://opac.nlc.gov.cn.

按照概念格构建、概念格分析、tax-folk 映射和 tax-folk 导航树构建的操作流程对上述数据进行操作,即可得出实验结果。Tax-folk 映射的过程片段

如图 9 所示,最终的 tax-folk 导航树的结果片段如图 10 所示。

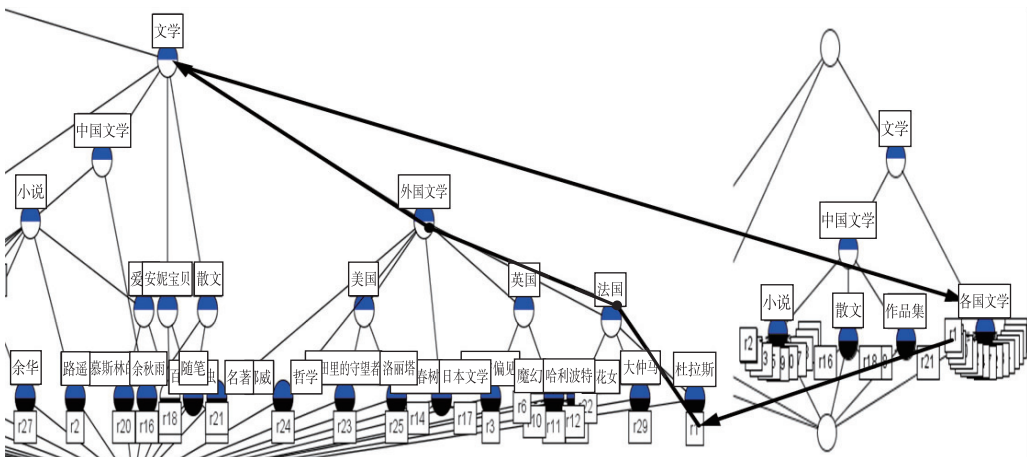


图 9 豆瓣读书的 tax-folk 映射(以 r1 为例)

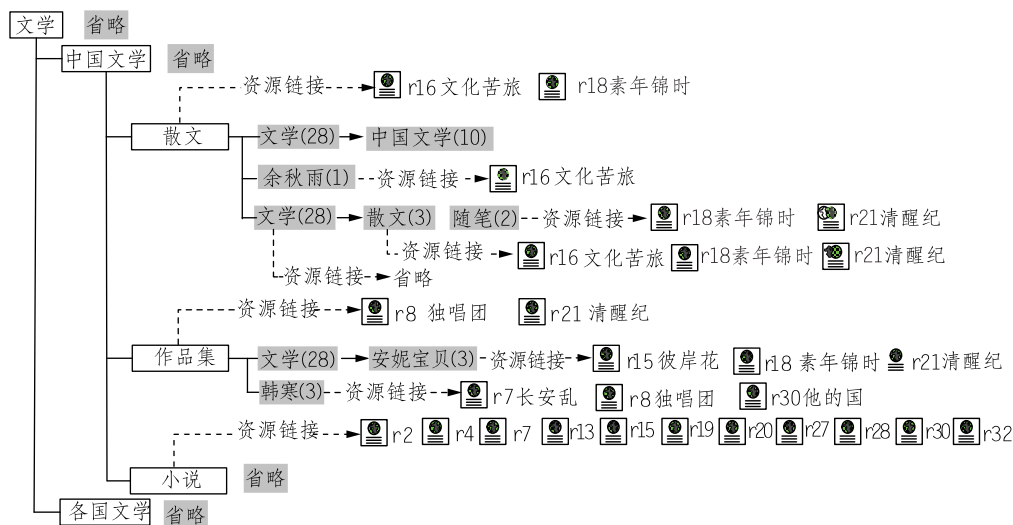


图 10 Tax-folk 导航树的结果片段

针对实验结果准确度的测定,本文认为,社会化标注系统中 tax-folk 模型的根本目的是构建“干强枝繁叶茂”的 tax-folk 混合导航树,实验结果成败的关键在于两点:

(1) 若资源嫁接链(即聚类后产生的“枝”和

“叶”)中的任意标签与分类词架构中目标节点的分类词相同,则说明待嫁接方与受体方是存在共性且契合的,这就是嫁接合理性的体现。可用公式 1 来测定合理性指标。

$$A = \frac{n}{N} \quad (\text{公式 1})$$

公式1中, A 为合理性指标, n 为嫁接链中任意标签与受体方中分类词相同的资源数, N 为参与嫁接的资源总数。 A 越趋近于1,则合理性越高。

(2)若每一条嫁接的“标签组限定关系”与分类词架构中受体方的“分类词限定关系”拟合,则说明这种嫁接不仅是合理的,还是精准的,这就是嫁接精准性的体现。可用公式2来测定精准性指标。

$$R = \frac{FIT(I)}{FIT} = \frac{| \text{标签组限定关系} \cap \text{分类词组限定关系} |}{| \text{分类词组限定关系} |} \quad (\text{公式2})$$

公式2中, R 为精准性指标, $FIT(I)$ 指标签组逐级限定关系与分类词组逐级限定关系交集的数目, FIT 指嫁接受体方分类词组逐级限定关系的总数目。 R 越趋近于1,则精准性越高。

经核算,在本实验中,合理性指标 $A = 29/32 = 0.90625$ (注:主外延为 r_4, r_5, r_7 的节点为不合理嫁接),精准性指标 $R = 3/5 = 0.6$ (注:分子中交集为“文学—中国文学”、“中国文学—小说”、“文学—外国文学”,分母中分类词组限定关系有“文学—中国文学”、“文学—外国文学”“中国文学—小说”、“中国文学—散文”、“中国文学—作品集”),从合理性指标和精准性指标看,实验结果可信。

通过实验可以看出,混合导航既在一定程度上避免了专家分类法仅用分类词描述和聚合资源过于宏观而不能反映用户个性化认知的局限性,又使得标签之间具有了层次关系,弥补了大众分类法扁平化造成弱语义的缺陷。其次,混合导航的两类资源链既有内容上的相关性,同时又各具特色,分类词资源链反映专家对资源聚合的认知结果,标签

资源链反映用户对资源聚合的认知结果。第三,tax-folk 混合导航的“树干—树枝—树叶—果实”的可视化展示模式丰富了用户的视域,为用户构建了极具空间感的资源体系。

6 结论和展望

Web2.0下,社会化标注系统资源聚合与导航问题的求解是一个多值问题,而专家分类法与大众分类法的混合导航就是其中一解。Tax-folk 混合导航能有效提高社会化标注系统中资源的可查找性,同时促进专家分类法管理过程的优化,更重要的是它以较低的成本完成了资源的分类和再组织,用一种全新的资源聚合和可视化形式兼顾了专家分类法和大众分类法的优势,将大众分类法嫁接到专家分类法上,共同构成干强枝繁叶茂的“知识之树”,而完成这种嫁接的工具就是形式概念分析。本文的主要贡献和创新是提出了形式概念分析视角下的 tax-folk 混合导航模型生成机理,据此构建了基于形式概念分析的 tax-folk 混合导航模型,用简易的映射规则实现了大众分类法向专家分类法的嫁接,并将结果用 tax-folk 混合导航形式展示出来。

当然,本文研究也存在不足之处。例如,未能将映射规则转换为机读的算法,从而实现自动化的映射过程。再如,由于概念格的处理容量限制,不易处理大型社会化标注系统的大规模数据,这也亟需改进。另外,对输出的 tax-folk 混合导航树的评价问题也尚未做出详细探讨。这些都是本课题组后续要研究和解决的问题。

参考文献

- [1] 王军,卜书庆.网络环境下知识组织规范研究与设计[J].中国图书馆学报,2012(4):39-45.(Wang Jun, Bu Shuqing. A study and design on the standard for the networked knowledge organization system[J]. Journal of Library Science in China, 2012(4):39-45.)
- [2] 岳爱华,孙艳妹.Taxonomy, Folksonomy 和 Ontology 的分类理论及相互关系[J].图书馆杂志,2008(11):21-24.(Yue Aihua, Sun Yanmei. Research on classification theory and the relationship of taxonomy, folksonomy and ontology [J]. Library Journal, 2008(11):21-24.)

- [3] Gruninger M, Bodenreider O, Olken F. Ontology Summit 2007 - ontology, taxonomy, folksonomy: understanding the distinctions[J]. Applied Ontology, 2008 (3): 191-200.
- [4] Glassey O. When taxonomy meets folksonomy: toward hybrid classification of knowledge? [EB/OL].[2014-05-05]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.7635&rep=rep1&type=pdf>.
- [5] Lemieux S. Hybrid approaches to taxonomy and folksonomy[EB/OL]. [2013-05-06]. <http://www.earley.com/presentations/hybrid-approaches-to-taxonomy-and-folksonomy>.
- [6] Hayman S, Lothian N. Taxonomy directed folksonomies: integrating user tagging and controlled vocabularies for Australian education networks [EB/OL].[2014-05-05]. <http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.150.7623>.
- [7] Sommaruga L, Rota P, Catenazzi N. Tagsonomy: easy access to web sites through a combination of taxonomy and folksonomy[J]. Advances in Intelligent and Soft Computing, 2011,86:61-71.
- [8] Tsui E, Wang W M, Cheung C F, et al. A concept-relationship acquisition and inference approach for hierarchical taxonomy construction from tags[J]. Information Processing & Management, 2010,46(1):44-57.
- [9] Kiu C C, Tsui E. TaxoFolk: a hybrid taxonomy-folksonomy structure for knowledge classification and navigation[J]. Expert Systems with Applications, 2011,38(5):6049-6058.
- [10] Batch Y, Yusof M M, Noah S A. ICDTag: a prototype for a web-based system for organizing physician-written blog posts using a hybrid taxonomy-folksonomy approach[J]. Journal of Medical Internet Research, 2012,15(2):1-23.
- [11] 张云中.一种基于 FCA 和 Folksonomy 的 本体构建方法[J].现代图书情报技术, 2011,27 (12):15-23. (Zhang Yunzhong. A new ontology construction method based on FCA and folksonomy[J]. New Technology of Library and Information Service, 2011,27(12): 15-23.)

张云中 上海大学图书情报档案系讲师,硕士生导师。

通讯地址:上海市宝山区上大路 99 号上海大学图书情报档案系。邮编:200444。

杨 萌 吉林大学管理学院情报学博士研究生。

通讯地址:吉林省长春市人民大街 5988 号吉林大学管理学院。邮编:130000。

(收稿日期:2013-08-08;修回日期:2013-10-23)