

基于 Web 2.0 的用户群体交互分析及其 服务拓展研究*

胡昌平 胡吉明 邓胜利

摘要 为了将用户的群体交互特性运用到网络信息服务中,构建基于用户群体交互的新型服务模式,本研究以“豆瓣电影”为例,采用标签共现的分析方法,对网络用户的群体性及其在交流过程中的关联及强度进行实证检验,分析结论与用户在“豆瓣网”中实际交互程度和对电影的喜好程度大致相符。对基于 Web 2.0 的网络信息服务环境下用户交互的群体性进行分析,从构建用户群体交互环境、设计基于 Web 2.0 的社会化网络服务以及建立用户兴趣小组三个方面拓展网络信息服务模式,旨在推动信息服务机构由单纯的网络信息服务模式向面向用户群体的社会化网络信息服务模式转化。图 5。表 3。参考文献 20。

关键词 Web 2.0 网络信息服务 用户群体性 群体交互

分类号 G250.7

ABSTRACT For applying group interaction characteristic to network information service and structuring new service style on group interaction, the paper takes Douban Movies for example and carries on empirical analysis to network users' group nature using the method of tags co-occurrence. The conclusion is matched up with users' interaction and preference to movies in Douban. The paper expands network information service from construction of interactive environment, design of social network service and establishment of user interest group. The paper aims to promote information services to change their service modes from simple network information service mode to socialized Internet information service mode which is oriented to users group. 5 figs. 3 tabs. 20 refs.

KEY WORDS Web 2.0. Network information service. Users' group nature. Group interaction.

CLASS NUMBER G250.7

Web 2.0 作为一种新生的网络模式,更加注重用户的网络交互,促进互联网向互动性和群体化的方向发展。用户在网络上发布和分享自己感兴趣的知识,同时更期望与具有相同兴趣的其他用户进行关联、互动和协作^[1],表现出强烈的群体性,并影响周围其他用户的网络意识和行为。用户之间的群体交互过程也是其知识学习和创新的过程,可以把其运用到网络信息服务中,使得知识在用户之间更快更好地共享和传播。本文从用户的群体性理论出发,收集社

群网络中用户的交互数据,对使用了相同或相似标签的用户进行聚合,分析其群体关系和强度以及由此产生的群体效应,并尝试将其运用到网络信息服务的拓展中,从构建用户群体交互环境、设计基于 Web 2.0 的社会化网络服务以及建立用户兴趣小组三个方面拓展了网络信息服务模式,旨在推动信息服务机构由单纯的网络信息服务模式向面向用户群体的社会化网络信息服务模式转化。

* 本文系教育部哲学社会科学研究重大课题攻关项目“创新型国家的知识信息服务体系研究”(项目编号:06JZD0032)研究成果之一。

1 Web 2.0 环境下的用户群体性分析

群体心理学研究表明,人们的行为都会表现出一定的群体性^[2]。Web 2.0 作为一种方便社会联系、增强社会协作能力的工具和信息交互平台,使得在 Web 2.0 环境下的用户因为交互而渐渐地聚合在一起,形成具有一定特征的用户群体。

1.1 Web 2.0 环境下用户的群体性

Web 2.0 下的网络服务环境是一个新兴的社交互环境,用户总是在某个特定的服务网络中,以自己的心理取向和志趣进行着特定的网络活动,一旦选择定向,就带来了频繁互动;这种互动一旦形成习惯,具有了某种群体交流结构,就会产生以大多数人的倾向为基础的群体思维,用户就会聚合形成或发现自己所属的群体^[3]。尤其是 Web 2.0 下的服务网络变得更加社交化,给用户提供了广泛交流的平台,用户更容易与朋友联系和交流,使得网络服务环境下的用户群体性更加明显,更容易形成社会化的群体聚合效应。

1.2 基于标签共现的用户群体发现

共现分析是将各种信息载体中的共现信息定量化的分析方法,其方法论基础是心理学的邻近联

系法则和知识结构及映射原则^[4]。心理学的邻近联系法则是指曾经在一起感受过的对象往往在想象中也联系在一起,在想起他们其中某一个的时候,其他的对象也会以曾经同时出现的顺序想起。知识结构及映射原则是指热衷于或从事某一科学热点研究的科学家,无论其社会和知识背景如何,对同一研究课题和概念在很大程度上使用的词汇基本一样;它们高频率的共现说明这些词指示的不同概念之间可能存在关联^[5]。

1.2.1 基于标签共现的用户群体发现机理

标签(Tag)是用来标注知识信息的非线性组织的关键词或术语,能够反映用户的兴趣和知识倾向,使用相同或相似标签的用户则具有相同的兴趣。标签体现了群体的力量,可以让用户体验到一个更加多样化的世界,一个关联度更大的网络服务空间^[6]。

从共现分析的方法论出发,若两个(或多个)对象有一个(或多个)相同的标签,则这两个对象之间必然存在一种潜在的关联,以他们共同出现的标签数量作为测度,称之为共现强度。共现强度越大,表明它们之间的关联越密切。这种分析主要用于通过标签了解对象之间的内在关联及强度,帮助用户发现与自身相似的用户,形成具有特定兴趣的用户群,把用户编织成一个系统而有组织的群体网络^[7],如图1所示。

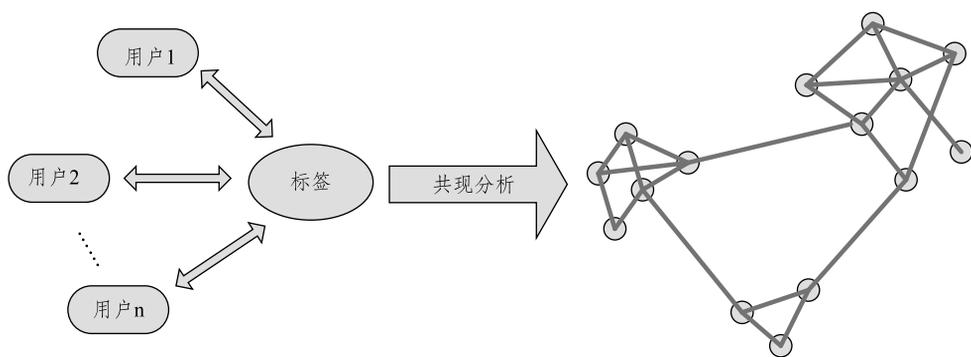


图1 基于标签共现的用户群体关联

1.2.2 标签共现分析方法

标签共现分析是指利用现代多元统计分析

如聚类分析和多维尺度分析等,根据共现强度将一个领域内的用户加以聚类,并形成兴趣共

同体^[8]。

聚类分析是根据研究对象(样本或者变量)的特征,对他们进行定量分类的一种多元统计方法。其基本思想是:同一类中的个体具有较大的相似性,不同类中的个体差异很大。于是可以根据多个观测目标,采用各种聚类算法,将所有的样本或变量分别聚合到不同的类中。聚成的类表示一个关系密切的群体,在兴趣或内容上有着相似性或一致性;类的大小直接反映群体的集中程度和对群体对象的关注程度。

多维尺度分析的原理是通过某种非线性变换,将高维空间转化为低维空间(通常是二维空间),变化后的空间图形仍能近似地保持原图形的集合关系。多维尺度分析图中,各点位置显示了对对象之间的相似性,有高度相似性的点聚在一起,形成共同体。越在中间的对象表明与它有联系的对象越多,在该领域的位置也就越核心;反之,则越独立。Stress 和 RSQ 是多维尺度分析效果的两个测度值;Stress 值越小说明拟合度越好,一般在 0.20 以内是可接受的;RSQ 值越大越理想,一般在 0.60 以上是可以接受的^[9]。

聚类分析和多维尺度分析可以很好地观察到对象的类别及其之间的关系,但是无法直观地表现对象之间关联的强弱以及群体的特征。借助社会网络分析工具绘制用户交互网路图,就可

以清楚地知道用户之间交互的结构和关系^[10]。

2 网络用户群体性实证分析

社群网络为人们在虚拟社群中进行信息交换和知识共享提供了支持,已经成为一种重要的知识交流共享平台。豆瓣网就是一家典型的社群网络,用户可以根据其自身喜好获取喜欢的书籍、电影、音乐等,并可以通过基于这些内容的交流互动来结识志趣相投的朋友。本研究以“豆瓣电影”为例,运用共现分析的方法验证网络用户的群体性并发现其在交流过程中的关联及强度。

2.1 构造标签共现矩阵

选取“好友”列表中出现的 20 位用户,统计其所看过的电影标签^[11]。为了突出用户对所看电影的喜好程度,本文以两两用户相同标签使用次数的交集并取和来表示两者的共现强度,构造用户标签共现矩阵。该矩阵为对称矩阵,非主对角线中的值为用户标签共现强度;关于对角线取值,为了突出亲密关系,用“最大值 + 1”来表示^[12]。20 位用户看过的电影标签部分数据和共现矩阵片段如表 1、表 2 所示。

表 1 电影标签部分数据

ken	舍利姬	西瓜庚	我要西游	l.e.t.o	阿皮
美国 (46)	喜剧 (119)	大陆电影 (116)	爱情 (79)	剧情 (60)	美国 (138)
爱情 (46)	动画 (110)	香港电影 (102)	美国 (36)	恐怖 (57)	剧情 (89)
喜剧 (37)	爱情 (97)	喜剧 (87)	电影 (33)	悬疑 (54)	喜剧 (23)
经典 (35)	日本 (85)	动作 (70)	喜剧 (29)	惊悚 (53)	香港 (21)
日本 (35)	香港 (78)	科幻 (52)	经典 (28)	动作 (50)	中国 (21)

表 2 标签共现矩阵片段

	ken	舍利姬	西瓜庚	我要西游	l.e.t.o	阿皮	宁远
ken	406	335	326	268	171	181	95
舍利姬	335	1194	578	319	434	171	121
西瓜庚	326	578	579	210	314	150	101
我要西游	268	319	210	344	202	177	97
l.e.t.o	171	434	314	202	473	196	141

2.2 数据分析结果

通过多元统计分析得出的结论与他们在“豆瓣网”中的实际交互程度和对电影的喜好程度大致相符,具体分析如下:

(1) 聚类分析。如聚类图 2 所示,结合“豆瓣网”中他们所看电影的实际情况,我们可以将用户分为 5 类。用户群体类别如表 3 所示。

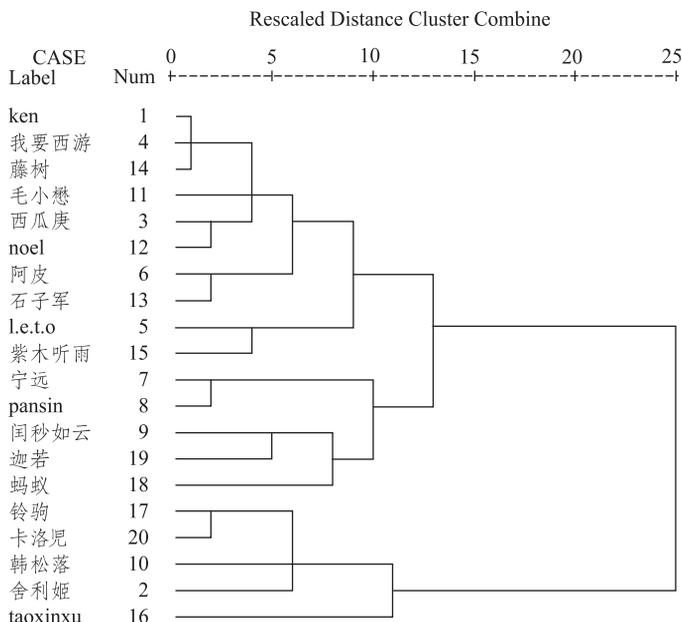


图 2 聚类分析结果

表 3 用户聚合类别

群体	用户	所喜爱的电影
G1	ken、我要西游、藤树、毛小懋、西瓜庚、noel	美国、日本、香港、动画、魔幻
G2	阿皮、石子军	剧情、喜剧、香港
G3	l.e.t.o、紫木听雨	惊悚、动作、科幻
G4	宁远、pansin、闰秒如云、迦若、蚂蚁	爱情、警匪、战争、喜剧
G5	铃驹、卡洛儿、韩松落、舍利姬、taoxinxu	喜剧、动画、爱情、日本、大陆

群体 G2 和 G3 中用户最少,说明这两个群体对电影的关注程度相对于其他群体而言最小,仅仅比较喜欢“剧情”、“惊悚”等类的电影,群体 G1、G4、G5 中用户较多,说明这些群体中用户对电影的关注程度较大且比较集中,尤其是比较喜欢“喜剧”、“美国”、“爱情”等类的电影,这也反映了用户对电影喜好的一个主流趋势,即多集中在“喜剧”、“香港”、“爱情”、“美国”等类别;与“豆瓣电影”标签汇总中显示的电影标签使用情况一致^[13]。

(2) 多维尺度分析。本次分析的 Stress 值等于 0.14972,RSQ 等于 0.91032,说明此次分析的效果很好。从图 3 看出,关联度最大的群体是“藤树”、“ken”、“毛小懋”和“taoxinxu”、“卡洛儿”、“铃驹”、“我要西游”,这也是比较符合聚类分析结果和实际情况的。特别是“taoxinxu”群体非常靠近中心位置,说明他们的互动交流最强,与这一群体有较高关联的用户多,此群体在整个群体中影响力较大;特别是用户“taoxinxu”处于核心地位,说明他与其他用户的互动最多和最

强,对其他用户的影响也最大。实际情况中,“taoxinxu”所看过的电影最多,发表的影评也最

多;群体中的其他用户经常浏览其影评并与其交流,受“taoxinxu”的影响也会看同样的电影。

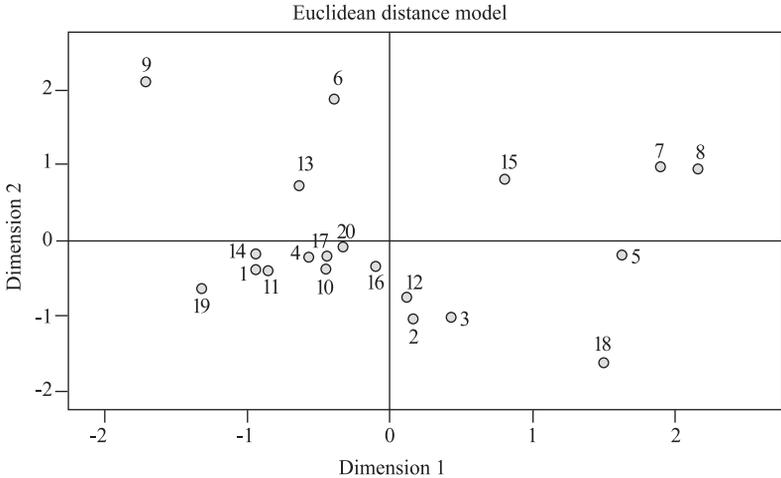


图3 多维尺度分析结果

(3)群体交互网络分析。为了更好地反映用户交互的群体特征,本文借助社会网络分析工具将共现矩阵转化为交互网络;通过网络密度分析网络中用户的紧密程度,通过中心性得出核心用户,通过中介性分析用户交互的流畅程度和单个用户所占网络资源的权力大小,并绘出交互网路图^[14]。

在社群网络中,密度是研究群体中用户之间紧密关系的测度变量,密度越大说明群体交互合作行为越多,对小群体的发展越有利。本文计算得出其交互网络密度为 0.5235,在社会网络中属于密度较高的群体^[15],说明用户之间交互程度较高,具有较高的聚集性。

在交互网络中,如果一个用户的中心度越高,就表明该用户在群体中越具有核心地位。如图 4 所示, Degree 表示绝对中心度, NrmDegree 表示相对中心度;中心度较高的为聚类分析中群体 G5,说明群体 G5 在整个用户群体中属于核心群体;其中铃驹、卡洛儿、taoxinxu 三者的中心度最高,表明其在整个群体中处于核心地位,所看电影最多和发表影评最多以及对其他用户产生的影响也最大。其群体中心度为 18.26%,就群体整体而言,用户的集中趋势不是很大,比

较集中的群体为 G1、G4 和 G5,处于分散或孤立状态的用户也较多。社群网络的中介性越高表明有越多的人需要通过他来与其他用户交互,进而就越能掌握网络中的信息流,对网络的操控性就越高。本文的中介性分析中,得出各个用户的中介性值几乎无差别,说明此群体中用户与用户之间的交互性较高,有小团体存在,但没有单个用户控制群体交互的现象。

其群体交互关联如图 5 所示,可以清楚地看出用户在群体中的交互情况,与上述分析的结果相符。

	1	2	3
	Degree	NrmDegree	Share
17 铃驹	12.960	70.637	0.065
20 卡洛儿	12.790	69.709	0.064
16 toaxinxu	12.493	68.090	0.063
2 舍利姬	11.744	64.010	0.059
18 韩松落	11.662	63.560	0.059
4 我要西游	11.432	62.306	0.057
11 毛小懋	11.370	61.968	0.057
1 ken	11.304	61.610	0.057
12 noel	11.187	60.971	0.056
3 西瓜庚	11.184	60.958	0.056

14	藤树	11.177	60.918	0.056
13	石子军	11.117	60.592	0.056
15	紫木听雨	10.160	55.373	0.051
19	迦若	9.331	50.855	0.047
5	l.e.t.o	9.073	49.449	0.046
6	阿皮	7.229	39.397	0.036
18	蚂蚁	6.755	36.815	0.034
7	宁远	6.113	33.319	0.031
8	pansin	5.537	30.178	0.028
9	闰秒如云	4.297	23.422	0.022

图4 用户交互网络中心度

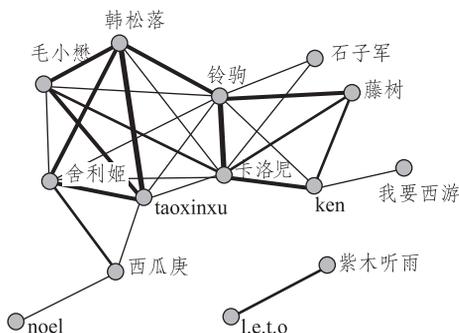


图5 用户群体交互

3 基于用户群体交互的网络信息服务拓展分析

网络信息服务的价值在于提升用户与用户之间交互关系和满足所有群体成员服务需求。而用户群体中的互动是最频繁的,其产生的效果也是最佳的;要想增强网络用户之间的交互,有效地启发和维持网络用户参与的积极性,使用户在交互过程中获得服务需求的满足和知识以及认知和情感的认同,网络信息服务必须面向用户群体开展。而信息服务机构应结合自身的服务优势把用户群体联结起来,将交互的主动权交给用户,并将用户的群体交互特性运用到网络信息服务中,构建基于用户群体交互的新型服务模式。

3.1 用户群体交互环境的构建

构建有利于用户群体交互的分享、互助和

开放的运作平台,建立用户交互网络,能够促进用户之间的交互学习和知识创新,使信息服务机构有效地开展知识信息服务。而利用现代 Web 2.0 的交互性理念和技术,能够建立面向用户群体的信息服务网络,为用户提供更多的参与机会,将“人机交互”的传统网络转变为“群体交互”的社会化网络^[16]。

Web 2.0 环境下的群体交互环境应遵循三层模式的设计思想:首先要充分展示信息服务机构的信息资源,实现信息资源的整合和共享;其次要根据用户的信息服务需求和网络信息行为实现用户潜在需求的发现和服务推荐;最后在资源、服务集成的交互平台上,引导用户之间的互动交流和协作,实现资源和服务的无缝获取和共享,形成具有自服务功能的用户群体。豆瓣网的交互环境就是基于三层模式构建的,这三层模式为:用户信息记录的分享、用户兴趣的发现和推荐、用户之间的会友交流和兴趣小组的建立,充分体现了面向用户需求的个性化推荐和基于用户兴趣与交互体验的群体交流。

3.2 基于 Web 2.0 的社会化网络服务设计

Web 2.0 环境下的网络服务提供普遍采用社会性标签标注用户感兴趣的信息,运用群体智慧形成对各种对象的大众分类。由于标签的共享性,用户通过信息服务平台就可以很容易地发现志趣相投者,通过互动和交流,用户可自主发展社会关系,也可以公开分享社会关系,形成特定的交流圈子。这样,服务资源将会得到更大范围的传播共享,形成社会化的网络服务(SNS)。图书馆可借鉴社会化网络服务设计的模式为读者提供更加个性化的服务。

(1) 标签。标签在数字图书馆中的应用效果较为显著,例如 Tabs。Tabs 是 Lewis&Clark 图书馆专门为读者开发的书签程序,通过“贴标签”、“收藏”、“共享”三个步骤,读者在管理自己喜欢的网页的同时,Tabs 会自动生成收藏同一网页或使用相同标签的读者之间的链接;进而读者通过互动成为好友,建立好友圈子^[17]。

(2) 开展评论服务。开展评论服务来表达用户自己对服务的认识和理解,以及与其他用

户进行思想的碰撞,能够增强用户之间的互动交流。如书评,它是一种知识的表达和传递方式,能够体现用户对图书的阅读体会,满足自己记录学习体会的需求;通过留言和互动,则可以将知识迅速地传播出去,乃至促进隐性知识的显化,达到知识创新的目的^[18]。这样,用户就会在其活动圈子中产生一定的影响力,使得用户具有一定的成就感。

(3)借鉴 Wiki 的协同编辑功能。图书馆可允许用户对信息资源的介绍按 Wiki 的方式编辑(即所有用户都可以根据自己的知识对相关信息进行修改,但只有能得到人们广泛认可的内容才能稳定地继承下去),并允许用户自主添加感兴趣的信息,把数据库的添加和维护权限向用户开放,体现相信用户、相信群体智慧的原则,这样既促进了已有信息资源的使用,又能为信息服务机构增加新的资源;同时支持 RSS 的输出,支持面向 Blog 等的 API 开放接口,充分实现信息推送和定制服务,并为实现系统之间的服务共享提供尽可能的方便等^[19]。

3.3 用户兴趣小组的建立

用户交互的过程就是用户群体形成的过程。建立专门话题或特定兴趣的用户兴趣小组,将会促进用户更加频繁的交流^[14],服务也将得到进一步拓展。在交互环境下,信息服务机构通过收集用户的信息行为和分析用户的交流网络结构,对用户兴趣进行挖掘和分析,划分用户群体,给不同用户推荐其感兴趣的小组;同时每个用户可以自主建立特定的兴趣小组和选择加入自己感兴趣的小组。信息服务机构根据用户兴趣小组的参与情况,把整个系统内的大量动态信息进行初步分类,避免用户的信息冗余,对每个用户只呈现与他们有关及他们感兴趣的信息。

兴趣小组将有效地运用群体智慧,让用户参与到服务中解决其他用户的服务问题,达到自服务的效果。这样,信息服务机构就会从个人独立服务的繁重工作中解脱出来,专注于整体服务水平的提高和优化,进而提高服务机构的整体效益,用户也会因为互动的加强而获得更多的知识。

用户群体交互中网络信息服务的开展过程也是一种社会过程,必然产生虚拟的人际交互网。而人际交互网的整体结构是网络信息服务开展的重要因素,会影响到知识信息服务的创新和发展^[20]。其交互关联的过弱或过强对服务的开展都不利,信息服务机构尤其要注意群体的密度、联结性和阶层性,运用社会网络的原理和分析方法来控制群体网络的结构,使之优化,拓展服务^[14]。

4 结语

信息服务研究的发展方向必然从对信息的处理整合转向对信息服务背后用户特性的研究。在群体交互环境下,通过用户的互动交流,将各个用户的知识资源通过网络连接起来,开发多种用户交互协作方式,使用户从单纯的信息消费者变成信息生产者、知识交流和创新者;通过驾驭集体智慧整合出增值内容,以此促进知识在用户群体内和群体间的共享传递和转换,乃至达到知识创新服务的效果;只有这样才能尽可能突破信息爆炸式增长对信息服务深化拓展的瓶颈,提高信息服务的效率和质量以及带来知识信息服务的创新发展。网络信息服务必将从单纯的信息服务模式转向围绕知识进行的用户群体参与的社会化服务模式,因此研究网络环境下的用户群体特性对网络服务的开展和利用具有很大的作用。希望本文能在基于群体交互的知识信息服务拓展研究方面起到一定的借鉴作用。

参考文献:

- [1] 邓胜利. 基于用户体验的交互式信息服务[M]. 武汉: 武汉大学出版社, 2008:11-13.
- [2] [英]布朗. 群体过程[M]. 胡鑫, 庆小飞, 译. 北京: 中国轻工业出版社, 2007.
- [3] 李其平. 网络传播的社会心理学分析[J]. 理论建设, 2007(6):65-68.
- [4] Kostoff R N K. Database tomography: Multidisciplinary research thrusts from co-word analysis [C]. Proceedings: Portland International Conference on Management of Engineering and Technology, 1991.

国家科学图书馆与 Springer 数字资源 长期保存协议签约仪式举行

2009年9月3日,中国科学院国家科学图书馆与 Springer 科学与商业媒体集团数字资源长期保存协议签约仪式在国家科学图书馆举行,中国科学院副院长、国家科学图书馆馆长李静海院士和 Springer 科学与商业媒体集团首席执行官 Derk Haank 先生分别致辞并代表双方在协议上签字。该协议是国家科学图书馆与国外出版机构签署的第一个数字资源长期保存协议,也是 Springer 签署的第一个由非 Springer 科学与商业媒体集团所在国(德国、荷兰)的国家级图书文献机构对 Springer 现刊数据库进行长期保存的正式协议。

根据协议,国家科学图书馆将成为 Springer 长期保存数据在中国境内(香港、澳门及台湾除外)的保存机构,当出现特殊情况以致 Springer 现刊数据库不能延续服务(双方定义为服务失效)时,在不改变数据完整性、保证数据库署名权、不提供商业盈利服务的前提下,国家科学图书馆可以利用长期保存数据向原订购范围的用户提供检索与获取服务。这标志着国家科学图书馆在建立合法、规范和可靠的数字资源长期保存系统方面迈出了坚实的一步,也是中国图书馆界在数字文献资源长期保存上的突破性进展。(中国科学院国家科学图书馆)

- [5] Rapp R. The Computation of Word Associations: Comparing syntagmatic and paradigmatic approaches [J/OL]. [2008-10-24]. <http://www.aclweb.org/anthology-new/C/C02/C02-1007.pdf>.
- [6] Tag [OL]. [2008-10-13]. <http://en.wikipedia.org/wiki/Tags>.
- [7] 王曰芬,宋爽,苗露. 共现分析在知识服务中的应用研究 [J]. 现代图书情报技术, 2006 (4): 29 - 34.
- [8] 刘林青. 范式可视化与共被引分析: 以战略管理研究领域为例 [J]. 情报学报, 2005 (2): 20 - 25.
- [9] 何晓群. 多元统计分析 [M]. 北京: 中国人民大学出版社, 2004.
- [10] 王运锋, 夏德宏, 颜尧妹. 社会网络分析和可视化工具 NetDraw 的应用案例分析 [J]. 现代教育技术, 2008 (4): 85 - 89.
- [11] 豆瓣电影 [OL]. [2008-10-28]. <http://www.douban.com/movie>.
- [12] 邱均平, 马瑞敏, 李晔君. 关于共被引分析方法的再认识和再思考 [J]. 情报学报, 2008 (1): 69 - 74.
- [13] 豆瓣电影标签 [EB/OL]. [2008-12-03]. <http://www.douban.com/movie/tag>.
- [14] 罗家德. 社会网分析讲义 [M]. 北京: 社会科学文献出版社, 2005.
- [15] [英] 约翰·斯科特, 刘军. 社会网络分析法 [M]. 重庆: 重庆大学出版社, 2007.
- [16] 邓胜利, 张敏. 交互式信息服务环境及其影响因素分析 [J]. 情报科学, 2008 (3): 443 - 447.
- [17] Tabs [EB/OL]. [2009-01-07]. <http://www.lcls.org/tabs/>.
- [18] Shifted Librarian. The perfect library blog example [EB/OL]. [2008-12-28]. http://theshiftedlibrarian.com/archives/2005/07/14/the_perfect_library_blog_example.html.
- [19] 侯丽. Web 2.0 的特性及对信息服务的创新性思考 [J]. 图书馆建设, 2008 (1): 66 - 69.
- [20] Nahapiet A, Ghoshal S. Social Capital, Intellectual Capital, and the rganizational Advantage. The Academy of Management Review, 1998, 23 (2): 242 - 266.

胡昌平 武汉大学信息资源研究中心教授, 博士生导师。通讯地址: 武汉。邮编 430072。

胡吉明 武汉大学信息管理学院信息资源管理专业研究生。通讯地址同上。

邓胜利 武汉大学信息管理学院讲师。通讯地址同上。

(收稿日期: 2009-01-10; 修回日期: 2009-03-13)