

PLSA 在图情领域专家专长识别中的应用*

张晓娟 陆伟 程齐凯

(武汉大学信息资源研究中心 武汉 430072)

【摘要】基于图情领域权威期刊论文数据集,利用概率潜在语义分析(PLSA)算法对表征专家专长的文档进行处理,以此来定位图情领域专家的研究领域。实验结果表明,该方法具有可行性并取得较好的实验结果。

【关键词】专家 专家专长 PLSA 文档主题

【分类号】G353.4

Application of PLSA on Expertise Identifying in the Field of Library and Information Science

Zhang Xiaojuan Lu Wei Cheng Qikai

(Center for the Studies of Information Resources, Wuhan University, Wuhan 430072, China)

【Abstract】Based on the dataset of authoritative journal in the field of library and information science, this paper tries to apply Probabilistic Latent Semantic Analysis (PLSA) algorithm to process documents representing expertise, in order to locate the research areas of experts in this field. The experiment results show that this method is feasible and achieves good results.

【Keywords】Experts Expertise PLSA Topic of document

1 引言

进入21世纪,人类社会正由信息管理社会迈向知识管理社会。有效的知识管理不只是对文档中的知识进行管理,更需要对人们头脑中的知识进行管理^[1]。在企业或组织中,了解“哪些人拥有哪些特定知识”是该企业或组织进行有效知识管理的重点。于是,一些企业或组织正在或者已经建立了专家检索系统。专家专长识别作为专家检索的一方面^[2],是利用和创造知识的重要环节,引起了业界和学界的广泛关注。其中,专家专长识别的任务为:利用企业或者组织内外能够表征专长的各种文档或资源(如电子邮件、报告、数据库文件和网页等),根据专家名识别出其专长(主要包括个人的隐性知识如方法、知识、能力等^[3],如研究领域、技能等)。由此可见,专家专长识别是基于文档的处理方法,是一种特定类型的信息检索方法。

概率潜在语义分析(Probabilistic Latent Semantic Analysis, PLSA)作为LSA的演化形式,利用概率模型来表示“文档-潜在语义-关键词”三者之间的关系,从而能够从语义层面而非单纯字面上去表达和理解文档,可有效解决词语之间的同义与多义等问题。该方法已被应用于信息检索、信息过滤、自然语言处理和机器学习等领域。

鉴于此,本文采用概率潜在语义分析方法来识别图书情报领域专家的专长,其中,所探讨的专家专长为专家的研究领域,即通过专家的研究成果(学术论文)来定位该专家所属的研究领域。

收稿日期:2011-12-20

收修改稿日期:2012-02-02

* 本文系教育部人文社会科学基地重大项目“面向细粒度的网络信息检索模型及框架构建研究”(项目编号:10JJD630014)和国家自然科学基金面上项目“基于语言模型的通用实体检索建模及框架实现研究”(项目编号:71173164)的研究成果之一。

2 相关研究

专家专长识别作为专家检索的其中一项任务,相对于另一任务(即根据专家专长返回专家名)来说,学界对此相关探讨较少。早期的专家专长识别方法主要通过专家本人描述自己的专长领域,并以此构建数据库,再利用传统数据库查询语言来识别专家专长,该方法的主要缺陷是专家参与的主观性以及数据库的操作缺乏灵活性。于是,学者尝试通过分析专家产生的文档(如报告、会议记录、出版物等)以及这些专家所从事的一些事务(如网上行为)来识别专家专长,其常用的技术有社会网络标签分析、博客语义分析、本体技术分析等。如 Budura 等^[4]认为用户所使用的一些社会软件在一定程度上能反应其专长,于是,通过从社会书签中选择该用户所使用的标签来识别专家专长; Dom 等^[5]通过专家之间的发送邮件关系来构建图,再利用 Hits 与 PageRanks 算法来识别专家专长; Tsai 等^[6]认为博客空间作为群体智慧和社会关系的重要来源,是网络用户展示专长知识和表达观点的场所,于是,以用户的博客内容为研究对象进行语义分析和概率建模,实现专长识别和意见挖掘; John 等^[7]通过对用户添加标签的行为进行社会网络分析,从而识别专家专长; Khan 等^[8]针对很难获取网络上有用信息的问题,提出一个基于 P2P 的在线专家专长知识匹配系统,结合本体技术,利用 RDF 语言存储专家专长,最终实现用户之间的知识共享;赵红斌等^[9]借鉴文本自动分类思想,基于文档权重归并法,采用 N 元语言模型,构建了专家研究领域识别实验系统。

综上所述,目前并无统一的专家专长识别方法,而大多研究采用了其他领域的研究方法,但目前仍未有学者将 PLSA 相关算法运用到专家专长识别这一相关研究中。

3 PLSA 模型概述

PLSA 是 Hoffman^[10]针对潜在语义分析存在的缺陷而提出的,即采用概率模型来表示“文档-潜在语义-关键词”三者之间的关系。首先做如下假定:一个文档集合 $D = \{d_1, d_2, \dots, d_M\}$, 文档集 D 中所有词组成的词集合 $W = \{w_1, w_2, \dots, w_N\}$, 文档集 D 中可能包含的潜在语义(即文章主题)集合 $Z = \{z_1, z_2, \dots, z_K\}$ 以及

文档和词的共现频率矩阵 C 。其中, M 表示数据集 D 中的文档个数, N 表示文档集 D 中不同词的个数, 集合 Z 中的主题个数 K 值需人工设定, 矩阵 C 的元素 $c_{i,j}(d_i, w_j)$ 表示词 w_j 在文档 d_i 中出现的频率。概率潜在语义分析假设文档和词之间的概率是条件独立的, 并且潜在语义在文档或词上分布也是条件独立的。在以上假设的前提下, 可使用式(1)来表示“文档-词”的条件概率:

$$p(d, w) = p(d)p(w|d) \quad (1)$$

在“文档-词”条件概率中引入文档主题可得:

$$p(w|d) = \sum_{z \in Z} p(w|z)p(z|d) \quad (2)$$

根据贝叶斯公式将式(2)转换为式(3):

$$p(d, w) = \sum_{z \in Z} p(z)p(d|z)p(w|z) \quad (3)$$

其中, $p(w|z)$ 表示潜在语义在词上的分布概率, 通过对 $p(w|z)$ 排序可以得到潜在语义的一个直观的词表示; $p(d|z)$ 表示文档与潜在语义(即文档主题)分布概率。

概率潜在语义分析使用最大期望(Expectation Maximization, EM)算法对潜在语义模型进行拟合。在使用随机数初始化之后, 交替实施 E 步骤和 M 步骤进行迭代计算。在 E 步骤中计算每个 (d, w) 对产生潜在语义 z 的先验概率, 如下:

$$p(z|d, w) = \frac{p(z)p(d|z)p(w|z)}{\sum_{z'} p(z')p(d|z')p(w|z')} \quad (4)$$

M 步骤(使似然函数最大)使用式(5)至(8)对模型进行重新估计:

$$p(w|z) = \frac{\sum_{d,w} n(d, w)p(z|d, w)}{\sum_{d,w} n(d, w)p(z|d, w)} \quad (5)$$

$$p(d|z) = \frac{\sum_w n(d, w)p(z|d, w)}{\sum_{d',w} n(d', w)p(z|d', w)} \quad (6)$$

$$p(z) = \frac{1}{R} \sum_{d,w} n(d, w)p(z|d, w) \quad (7)$$

$$R = \sum_{d,w} n(d, w) \quad (8)$$

当 L 期望值(见式(9))的增加量小于一个阈值时停止迭代, 此时得到一个最优解, 从而获得 $p(w|z)$ 与 $p(d|z)$ 的分布情况。

$$E(L) = \sum_{i=1}^M \sum_{j=1}^N n(d_i, w_j) \log p(d_i, w_j) \quad (9)$$

4 实验及其结果评测

4.1 实验数据集

根据 Song 等^[11]对专家专长(即专家的出版物质量和数量)的定义, 将专家所发论文作为其专长证据的

主要来源。考虑到因数据集过大而造成进行 PLSA 运算时迭代时间与存储空间的开销太大,则需采用一些具有代表性的数据进行实验。另基于如下事实:期刊的级别越高,所收录的论文更能代表专家的研究方向。于是,本实验采用图书情报领域具有代表性的两权威期刊的论文数据作为实验数据集。

本实验以网络爬虫的数据采集方式从中国知网上下载《情报学报》和《中国图书馆学报》近 30 年的论文数据,并将采集的数据进行处理后存入本地 MySQL 数据库中,其字段包括:Title(标题)、Author(作者)、Address(地址)、Journal(期刊)、Year(年份)、Keywords(关键词)、Abstract(摘要)。存储数据共有 6 564 条记录,包括来源于《情报学报》的 2 382 条记录以及《中国图书馆学报》的 4 182 条记录,通过分析论文的摘要、关键词、题名来获得专家的专长信息。

4.2 实验过程

本实验的主要流程如图 1 所示:

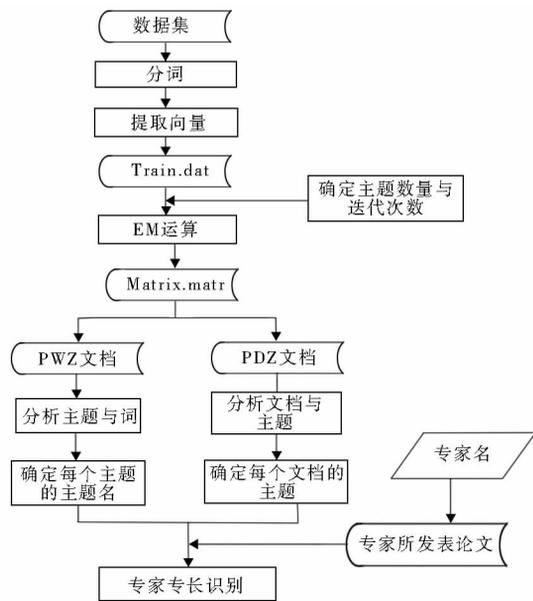


图 1 实验总体流程

其主要思路为:对数据集进行分词处理,建立相关的词向量 Train.dat,再利用 PLSA 的 EM 运算步骤对该词向量进行迭代,从而获得存有主题与词以及文档与主题之间关系的矩阵,通过对此矩阵数据进行解析,获得每个主题下词的分布情况,以及每个文档所属主题的情况,最后统计某专家所发表论文所属主题的频度,将频度高的对应主题定位为该专家的专长。其实验过

程主要有三步:词向量构造、文档-主题-词分析以及专家专长识别。

(1) 词向量构造

本实验的词向量生成过程如图 2 所示:

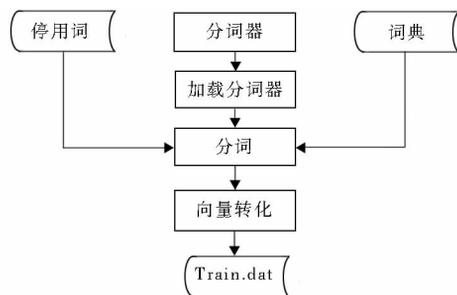


图 2 词向量生成过程

首先对每篇文档(即数据库中的一条记录)中待处理字段进行分词以及去除停用词处理,再统计该文档中每个词在此文档中所出现频次,最后将其存储在 Train.dat 中,其数据格式为: $d_j-w_1:P_1 w_2:P_2 \dots w_i:P_i$ 。其中, d_j 表示文档编号, w_i 表示该词在整个数据集中的编号, P_i 表示该词在该文档中的出现频次。图 3 为 Train.dat 数据的具体存储格式,每一行表示一篇文档的词信息。

3	117:2	130:1	131:1	132:1	133:1	4:11	134:2	135:2	85:9	136:2	83:4	137:1	9:22	138:1	139:2	140:1	139:1	4:1	141:4	142:1	9:1		
4	58:1	94:2	182:3	215:3	216:1	79:5	217:1	83:4	218:4	219:4	220:6	4:7	221:2	222:2	220:1	4:1	182:1	215:1	223:2	224:1	9:5	22	
5	58:1	196:1	249:1	94:2	250:1	245:1	225:2	251:2	252:2	253:2	254:1	98:2	4:4	255:1	256:1	9:12	101:1	102:1	88:1	83:2	257:5		
6	58:1	116:1	83:3	272:2	9:9	9:1	254:1	273:2	274:2	275:5	4:7	79:1	246:3	6:4	276:1	277:1	9:1	278:1	279:1	83:1	278:1	4:1	28
7	58:1	103:1	84:6	297:2	4:7	267:1	6:3	84:1	298:1	299:1	9:5	300:1	83:3	59:5	84:1	301:3	4:1	302:1	6:1	303:1	9:1	62:1	83:1
8	314:1	315:4	316:4	4:6	317:2	103:2	318:1	17:1	207:1	179:1	173:2	319:3	320:6	9:10	321:1	320:1	6:1	115:1	320:1	9:1	322:2		
9	58:1	349:1	51:2	84:6	114:5	4:6	350:2	47:1	59:4	277:1	9:8	72:1	83:1	351:1	6:3	59:1	84:1	114:1	4:1	311:2	352:1	9:1	353:1

图 3 Train.dat 的数据格式

(2) 文档-主题-词分析

当 Train.dat 生成后,调用 PLSA 中 EM 步骤迭代来获取存储主题与词、文档与主题之间关系的 Matrix.matr 矩阵。考虑到迭代结果尽可能包含所有的主题以及迭代时所需要时间和内存空间开销,本实验将主题数设置为 250,迭代次数设置为 500。

① 主题-关键词分析

从 Matrix.matr 中读取主题与词之间的对应关系写入 pwz.ser 文件,得到每个主题下的词分布情况,且这些词按其在该主题下的权重降序排列,数据的存储格式为: $z_1-w_1:P_1 w_2:P_2 \dots w_N:P_N$ 。其中, z_1 表示主题编号, w_1 表示词编号, P_1 表示词 w_1 在该主题下的权重。图 4 为主题号为 z_{145} 下的词分布情况。

考虑到某些主题下所分布词的权重过低,则这些主题并

z_{145} ——情报 竞争 企业 对手 战略 能力 采集 市场 作战 室 论述 策略 实践 竞争力 游戏 战争 动态 市场经济 探讨 激烈 指出 剖析 作用 要素 推动 阐述 重要 财务 态势 工具 竞争性 融合 危机 地位 制定 依据 变革 借鉴 案例 来源 必要性 柔性 有效 内容 明确 企业界 企业经营者 激发 博弈 论 环节 受到 准备 构建 手段 商学院 搜集……

图 4 主题 z_{145} 下的词分布情况

无太大的存在意义,本文通过设定词权重阈值为 $3.0E-4$ 以此来筛选出此类主题,即当某主题下最高的词权重小于该阈值时,将其视为无意义主题。基于如下事实:即某一主题下权重越高的词越能表达该主题。选取每个主题下权重排名前 5 的词,再经过人工剔除与主题名无关的词如“作用”、“准备”等,以及词之间的顺序调换,以此来确定每个主题的主题名。对于无意义主题,并未对其进行主题名提取,最后得到的主题编号与其对应的主题名如表 1 所示:

表 1 主题编号及其对应的主题名

主题编号	主题名	主题编号	主题名
z_0	知识表示与知识共享	z_6	图书馆自动化
z_8	机器翻译	z_{16}	个性化服务
z_{19}	信息资源建设	z_{21}	Web 信息抽取
z_{22}	文献分类法	z_{26}	智能检索
z_{32}	专题索引编制	z_{30}	图书馆服务
z_{31}	信息咨询	z_{35}	叙词表编制与构建
z_{36}	本体与语义网	z_{40}	企业战略与决策
z_{41}	科学评价	z_{42}	信息安全与保障
z_{49}	政务信息资源	z_{45}	XML 索引与检索
z_{51}	信息系统工程	z_{59}	文献资源共享
z_{54}	目录学	z_{65}	信息用户行为
z_{58}	信息资源配置与信息政策	z_{68}	数据挖掘
z_{60}	信息服务与知识服务	z_{74}	高校图书馆与服务
z_{67}	信息资源构建	z_{81}	情报学理论
z_{69}	图书馆体制	z_{86}	信息资源管理
z_{79}	联合目录	z_{94}	信息经济学
z_{82}	图书馆基础理论	z_{102}	数字图书馆
z_{88}	用户信息需求	z_{106}	信息相关性
z_{100}	主题词自动标引	z_{119}	搜索引擎技术
z_{104}	专利分析	z_{124}	电子政务
z_{117}	信息组织与知识组织	z_{133}	信息检索
z_{123}	信息市场营销	z_{137}	非相关文献知识发现
z_{128}	自动文摘	z_{145}	竞争情报
z_{134}	期刊管理	z_{150}	虚拟图书馆
z_{136}	情感分析	z_{165}	信息伦理学
z_{139}	开放存取	z_{171}	公共图书馆
z_{148}	文献编目与著录标准	z_{174}	信息分析与预测
z_{161}	信息可视化	z_{177}	网络计量学
z_{172}	知识产权法	z_{179}	图书馆读者服务
z_{176}	信息产业	z_{181}	信息素养
z_{178}	信息资源集成	z_{188}	社会网络分析
z_{180}	图书馆馆藏与流通	z_{197}	文本挖掘
z_{182}	数字资源整合	z_{200}	信息资源共建与共享
z_{190}	查询扩展与相关反馈	z_{206}	信息资源管理与图书馆管理
z_{201}	图书馆管理思想	z_{215}	元数据描述与标准
z_{224}	电子出版物	z_{226}	数据仓库
z_{228}	信息分布规律	z_{212}	文献计量学
z_{238}	图书馆信息技术	z_{240}	知识管理

因本文采用的数据集有限,则表 1 中的主题不可能包含图情领域所有的研究主题。本实验主要通过某主题下词权重来确定主题名,则提取出的主题名存在着概念粒度不一样以及概念之间交叉重叠现象,而本文主要考虑到主题名如何有效表达主题,则对此类现象并未进行深入探讨。

②文档-主题分析

Matrix. matr 中也存贮了文档与主题的对应关系,通过对其解析,将分析文档与主题的对应关系并存贮在 pdz. ser 中。其格式为: $d_i-z_1:p_1z_2:p_2……z_N:p_N$ 。其中 d_i 表示文档 ID, z_i 表示主题编号, p_i 表示该文档包含主题的概率。图 5 表示编号为 190 文档所包含主题的情况。

D₁₉₀ 28:4.20595453340282E-4 73:9.918820639277547E-4
7:9.848965597895668E-4 ……

图 5 文档与主题之间关系

实验把每个文档下排在前三个主题作为该文档所属的主题,则文档 190 包含的主题为编号分别为 28、73、7 的主题。

(3)专家专长识别

表 2 专家“邱均平”所发论文对应的主题号

文档编号	主题编号	主题编号	主题编号	文档编号	主题编号	主题编号	主题编号
42	216	99	84	72	160	127	181
123	60	150	94	145	224	186	170
178	212	3	170	1290	170	186	86
1905	58	18	240	2412	181	196	62
2100	40	143	206	2109	122	50	228
2390	177	134	94	2410	80	89	110
2431	177	64	185	2569	20	78	40
2893	177	240	218	3004	177	226	246
3432	227	20	174	3562	177	226	116
3421	146	109	165	3632	177	178	179
3681	20	229	240	3879	177	41	216
4209	216	51	41	4276	239	236	235
4527	119	94	210	4635	96	18	0
4666	177	100	54	4890	228	110	236
4909	60	184	168	5003	60	184	168
5678	181	114	26	2230	212	26	240
6034	104	183	218	3321	128	213	183
4232	93	41	228	4456	34	189	176
4476	228	50	124	2703	145	81	184
5210	59	93	136	5990	136	120	216
59	66	146	94	98	212	167	93
158	212	86	206	818	10	170	127
954	186	10	170	1354	18	123	114
1156	172	44	185	1224	172	212	190
1790	177	174	183	2609	212	3	181
2700	229	20	237	2809	2	157	44
2999	212	242	206	3409	136	20	53
3580	210	177	5	3976	210	41	185
4789	93	120	232	4001	20	38	2
5007	41	236	173	5687	120	70	240
3903	120	240	165	6078	177	50	174
6134	240	226	246	6166	110	166	206
6321	66	166	110	6277	53	210	228
6453	166	82	127	6550	166	100	136

专家专长识别的主要思想为:统计专家所发表论文的所属主题频度,将频度高的主题所对应主题名作为该专家的专长。本文以专家“邱均平”为例,其所发论文对隶属的主题如表2所示。经过统计表2中所出现主题的出现频次,排名前10的主题如表3所示,其中,将无意义主题统一标识为“*”。最后识别专家“邱均平”的专长如表4所示。

表3 出现频次前10的主题号及主题名

排名	主题号	主题名
1	177	网络信息计量学
2	240	知识管理
3	212	文献计量学
4	170	*
5	228	信息分布规律
6	41	信息评价
7	166	*
8	20	*
9	94	信息经济学
10	181	信息素养

表4 “邱均平”的专家专长

专家名	专家专长
邱均平	网络信息计量学;知识管理;科学评价;文献计量学;信息经济学;信息素养;信息分布规律

通过调查分析,网络信息计量学、知识管理、信息评价、文献计量学、信息经济学、信息分布规律都是该专家曾经或者现在的研究方向,而对信息素养未曾做过相关研究。

4.3 实验结果评测

本文选取在《情报学报》和《中国图书馆学报》总发文超过5篇的30名专家作为评测对象,其中图书馆学和情报学各15名专家。因本实验的最终结果为:根据专家名返回与之相关的专家专长,其返回结果根据其相关性进行排序,故采用信息检索中的评测指标对专家专长识别结果进行评测,其选用的指标为P@N。另考虑到如下情况:一些专家的研究比较集中在两三个领域中,而另外一些专家的研究领域会更多,则本文分别统计了N=3、5、10这三种情况的实验结果。通过实验将每个专家名返回的专长识别结果反馈给相应专家,使其结合自己的研究领域对实验结果进行评价,最后统计每个专家专长识别结果的P@3、P@5、P@10值,再对这三个值分别求平均值,得到的评测结果如表5所示:

表5 实验评测结果

评测指标	p@3	p@5	p@10
评测结果	0.897	0.825	0.660

5 结语

本文尝试将PLSA引入到专家专长识别中,并取得了较好的实验效果,证明了该方法的可行性。因本实验处于初探阶段,存在以下不足之处:

(1)数据集不全面。专家在学术领域的论文除了该领域的权威论文,其他核心期刊也收录了表征其专长的论文,这将会影响到专家专长识别的全面性,后期研究将会尽可能地从各种级别的期刊中全面获取表征专家专长的论文数据集。

(2)未考虑作者顺序。实际上,第一作者更能代表专家在某一领域的研究专长,而第二、三或者第四作者可能只表示某专家只是涉足某个领域,而非非一定从事该领域的研究,而本文未将不同顺序的作者进行区别对待。

(3)主题名的交叉重复。在为每个主题提取主题名时,只考虑到该主题名尽可能表达主题,而忽略了主题名之间的概念粒度以及概念之间是否存在交叉重复现象。

(4)对照实验的缺乏。本文只是停留在探讨如何将PLSA运用到专家专长识别领域,而该方法未与已有专长识别方法进行比较。

以上不足之处正是笔者后续工作将要深入研究的地方。

参考文献:

- [1] Hendriks P H J, Vriens D J. Knowledge - based Systems and Knowledge Management: Friends or Foes[J]. *Information & Management*, 1999, 35 (2): 113 - 125.
- [2] Seid D Y, Kobsa A. Expert Finding Systems for Organizations: Problem and Domain Analysis and the DEMOIR Approach [J]. *Journal of Organizational Computing and Electronic Commerce*, 2003, 13(1): 1 - 24.
- [3] Expert - Wikipedia[EB/OL]. [2011 - 09 - 30]. <http://www.en.wikipedia.org/wiki/Expertise>.
- [4] Budura A, Bourges - Waldegg D, Riordan J. Deriving Expertise Profiles from Tags[C]. In: *Proceedings of the 2009 International Conference on Computational Science and Engineering*. 2009: 34 - 41.
- [5] Dom B, Eiron I, Cozzi A, et al. Graph - based Ranking Algorithms for Email Expertise Analysis[C]. In: *Proceedings of the 8th*

ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. San Diego, CA, USA: ACM Press, 2003: 42-48.

- [6] Tsai T M, Shih C C, Peng T C. Explore the Possibility of Utilizing Blog Semantic Analysis for Domain Expert and Opinion Mining [C]. In: *Proceedings of International Conference on Intelligent Networking and Collaborative Systems (INCOS ' 09)*, Barcelona, Spain. IEEE, 2009: 241 - 244.
- [7] John A, Seligmann D. Collaborative Tagging and Expertise in the Enterprise [C]. In: *Proceedings of WWW 2006*. 2006.
- [8] Khan S, Nabeel S M. OPEMS: Online Peer - to - Peer Expertise Matching System [C]. In: *Proceedings of the 1st International Conference on Information and Communication Technologies (ICICT*

2005). 2005: 148 - 151.

- [9] 赵红斌, 陆伟. 专家研究领域自动识别研究 [J]. *现代图书情报技术*, 2010(2): 63 - 67. (Zhao Hongbin, Lu Wei. The Study of Expert Research Field Automatic Recognition [J]. *New Technology of Library and Information Service*, 2010(2): 63 - 67.)
- [10] Hofmann T. Unsupervised Learning by Probabilistic Latent Semantic Analysis [J]. *Machine Learning*, 2001, 42(1): 177 - 196.
- [11] Song X, Tseng B L, Lin C, et al. ExpertiseNet: Relational and Evolutionary Expert Modeling [C]. In: *Proceedings of the 10th International Conference on User Modeling, LNCS 3538*. Berlin: Springer - Verlag, 2005: 99 - 108.

(作者 E - mail: zhangxiaojuan624@gmail.com)

Serials Solutions Summon 服务扩展其 STM 的覆盖范围

作为内容管理战略的一部分, Serials Solutions 的 Summon 服务持续扩展其统一索引中 STM(科学、技术和医疗)内容的覆盖范围。扩展包括:

(1) CRCNetBase 有 40 多个学科的 8 000 多本在线图书, 主要是科学和技术类。很快就能在 Summon 服务中对 CRCNetBase 进行检索。除了 CRC 出版社出版的电子书, CRCNetBase 还包括 Auerbach 出版社和 Chapman & Hall 出版公司的电子书。CRCNetBase 平台已被世界各地的学术和企业机构所采用。CRC 出版社是 Taylor & Francis 出版集团内主要的科学和技术图书出版分部。

(2) BCS 信息技术研究所(前身为英国电脑学会)全部的学术会议库 eWIC 将会进入 Summon 服务。eWIC 丛书(Electronic Workshops in Computing)是国际级的, 自 1995 年开始提供。每一卷都是基于一届专家研讨会的结果而出版的, 其目的是提供有关当前最新知识、争论或者研究的信息。

(3) 国际汽车工程师学会的 SAE 技术论文将会在 Summon 服务中进行全文索引。SAE 技术论文是汽车、航天和商用汽车行业的专家所撰写和经过同行评审的, 包含最受欢迎的 SAE 数字图书馆的大部分内容。国际汽车工程师学会出版 9 万 2 千多种有关技术研究和应用工程技术最新进展的技术论文。

Summon 服务索引了 STM 内容中 8 亿多的条目, 通过一个单一的、统一的索引提供比其他任何发现服务更加全面的检索服务。部分已经索引并在 Summon 服务中提供检索的 STM 内容的出版商和来源商包括 IGI Global 信息管理与工程数据库、Sage 出版社、Springer 出版集团、Taylor & Francis 出版集团, 以及 Wiley - Blackwell 出版公司。Summon 对这些内容的大部分都提供了全文检索。

(编译自: <http://www.serialssolutions.com/words/serials-solutions-summon-service-expands-stm-coverage/>)

(本刊讯)