

基于用户的检索服务研究进展

李芳^{1,2} 杨林^{1,3}

(1.中国科学院 国家科学图书馆,北京 100190; 2.昌吉学院,新疆 昌吉 831100;
3.中国科学院 研究生院,北京 100049)

摘要:对近5年来信息检索的相关研究按照基于语义的检索服务、基于检索结果自身特征的检索服务、基于日志挖掘的检索服务和基于用户交互的检索服务这几个方面进行梳理和分析,总结了检索服务研究的特点和趋势。

关键词:语义;自身特征;交互;日志挖掘

中图分类号: G350 **文献标识码:** A **文章编号:** 1007-7634(2012)09-1424-07

The Review of the New Development of the User-based Retrieval Service

LI Fang^{1,2}, YANG Lin^{1,3}

(1. National Science Library, Chinese Academy of Sciences, Beijing 100190, China;
2. Changji College, Changji 831100, China;
3. Graduate School, Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The main target of this paper is to analysis and summarizes the development of the retrieval service, and the retrieval service are divided into four categories according to the semantic, the own characteristics of the retrieval results, the log mining and user interaction. Finally, the characteristics and trends of the retrieval service are summarized.

Keywords: semantic; own characteristics; interactive; log mining

1 基于语义的检索服务

基于语义的检索服务是在语义资源的基础上探索如何帮助用户改善检索效果以提升信息检索服务。基于语义的检索服务可分为基于语义统计特征的检索服务和基于本体的信息检索服务,后者是当前研究的热点。

1.1 基于语义统计特征的检索服务

基于语义统计特征的检索服务是针对检索术语,将检索结果集中所隐含的语义信息进行统计,根据语义统计特征来改善检索服务。基于上下文的词

频评价技术 CTFA4TR 和查询扩展的导航系统 S-node 是较为典型的基于语义统计特征的检索服务应用。CTFA4TR 技术通过考虑被排序文献中查询术语出现的上下文环境和窗体环境,来识别术语的位置环境,然后用术语的位置环境来改善各种文本工具的排序效果。这种技术是在文本的术语位置的上下文环境中,把术语频率作为一个对象进行编码,然后根据检索词在一篇文档中出现的频率以及上下文关系对文档排序,该技术可以在多个文本排序工具中使用^[1]。在新闻或博客的撰写过程中,作者会经常使用一些专业术语和表达,所以在检索时为用户提供合适的查询术语较为困难。为此, Satoshi Shimada 等提出了查询扩展的导航系统

收稿日期: 2011-10-08

基金项目: 教育部人文社会科学研究项目(09XJA870003)

作者简介: 李芳(1982-),女,四川人,博士研究生,讲师,主要从事情报学研究。

S-node。该系统基于重复指数,从文档中抽取两种词来显示穷举性和特殊性,并且构建文档间的超链接,尽可能经过较短的路径到达各种文档。S-node系统将抽取出来的关键词分为特定的词、综合词、边缘词和其他的词四种,通过共现方法,在每个关键词之间、每个文档和关键词之间,以及文档之间产生链接,然后系统根据检索的上下文环境向用户展示相关的文档和关键词,实验表明该系统能够提高检索效率^[2]。

1.2 基于本体的检索服务

基于本体的检索服务主要是利用本体提供的语义信息(语义关系或语义含义),通过检索词扩展、检索词与检索结果集相关性计算等来提高检索精准度。基于本体的检索服务按照本体类型、本体所处领域以及本体数量,又可分为基于已有本体与自建本体的检索服务、基于专业本体和综合本体的检索服务,以及基于单一本体和多本体的检索服务。

1.2.1 基于已有本体与自建本体的检索服务

这种检索服务按照本体类型可分为基于已有本体的、基于自建本体的和既包含自建本体又包括已有本体的检索服务。①Textpresso是基于本体的生物学领域的信息检索和提取系统^[3],它基于已有的Gene Ontology(GO)构建了生物本体^[4],由33个类组成,包括生物学实体概念、生物学实体间关系概念、辅助性词汇三种本体元语。系统将整个文档划分为句子的集合,用户可以根据需要选择其查询术语是与文档全文匹配还是与句子匹配来实现全文和句子的检索^[5]。其中本体对系统进一步扩展,本体中的类或子类可以与关键词结合完成查询术语扩展,有助于用户确定检索词。②IAToNews是基于自建本体的典型应用^[6],是以强大的本体和智能为基础的用户读取和搜索新闻(中国新闻)的信息系统。它由内部的本体数据库提供本体数据,该库包含文章本体和主题本体两个本体,其中文章本体描述新闻的语义内容,主题本体则构建主题领域的层级关系。该系统通过抽取将文章的内容转换为本体格式存储到系统的内容索引数据库,以支持新闻内容语义分析、本体查询处理及结果检索。③HealthFinland^[7]是既包含自建本体又包括已有本体的语义健康信息发布系统,它包括关于健康本体的中心内容结构和工具服务;基于多个健康组织的分布式语义内容渠道;智能语义门户汇总和直观的终端用户展示界面。HealthFinland使用本体和其他控制词汇描述文档的

主题、内容体裁和目标用户,选择通用芬兰上层本体(The General Finnish Upper Ontology, YSO)、医学主题词表(The international Medical Subject Headings, MeSH)、欧洲健康多语种词库(The European Multilingual Thesaurus on Health Promotion, HPMULTI)这三个核心主题领域本体来描述主题。使用了Finland统计局公布的职业分类和UKK Institute数据库出版的较小目标用户分类作为本体的出发点,自建目标用户本体。最后使用The Finnish Ontology Library Service ONKI来发布所有上面讨论的领域本体,整个系统结构如图1。该系统用基于语义的门户网站和集成web服务访问全球信息库,以合作的方式在国家的层面上创建健康内容,减少了重复冗余的工作和成本,除此之外还提供给终端用户基于个人的信息需要找到正确信息或关于健康的概念观点的智能服务。

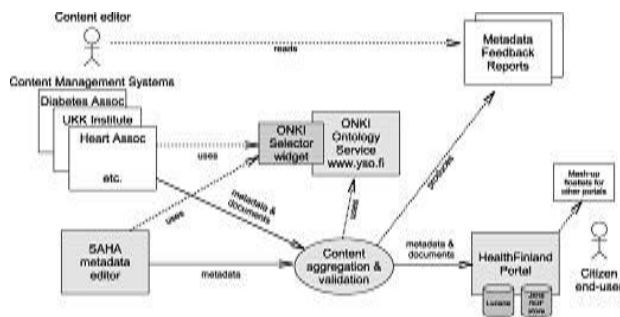


图1 HealthFinland系统结构图^[7]

1.2.2 基于专业本体和综合本体的检索服务

基于本体的检索服务按照本体资源的领域可分为专业本体和综合本体,专业本体如Gene Ontology、MESH等,综合本体如WordNet、HowNet、维基百科等。GoPubMed运用基因本体词表GO和医学主题词表MESH进行文献检索和浏览,对检索结果集进行术语提取和实体识别^[8],从中提取GO术语和MESH主题词^[9-10],通过数据挖掘使检索结果转换成分类类目及对应文献之间的可视化数据,进而抽取特定的模型,即自动生成临时基因本体词表和医学主题词表来对检索进行分类,检索者可以根据这些分类快速找到自己需要的文献【11-13】。Ming-Hung Hsu等提出了典型的粗粒度分类问题变换加权方法。为经过执行扩展激活选出来的候选术语赋予合适的权值。系统的整个处理过程是,在预处理阶段,如果查询词在WordNet或ConceptNet中出现,那么初始查询把标记位置、查询中的n-gram作为一个词语来确定,词和短语以及它们的标记位置在接下来的过程中作为一个概念进行处理。经过

预处理,查询中的概念作为激活源,在 WordNet 和 ConceptNet 中执行扩展激活,词是来自于一组扩展候选术语中的激活分数最高的前 N 个词,对于每一个候选术语,自动标记它相对应的原始查询的权重,并根据全局分析、局部分析、知识基础分析提取它的性质。之后,使用 SVM 方法对候选术语和它们的标记以及性质进行分类训练与测试。候选术语和它们的权重相结合,与原来的查询形成扩展查询,进一步检索。该方法成功地合并了 WordNet 和 ConceptNet,提高了检索性能。类似地, Jianqiang Li 等通过利用 WordNet 的语义特征和文档聚类提出了 FACT(Fully Automatic Categorization approach)方法^[14]。

1.2.3 基于单一本体和多本体的检索服务

基于本体的检索服务从检索服务所涉及的本体数量上,可分为单一本体和多本体融合。Jaap Kamps 等利用维基百科的链接结构来改进信息检索,通过实验验证了维基百科链接和网页内容之间的关系,发现它们之间存在弱相关,利用这种上下文敏感的关系能有效地改进检索效果^[15]。DigLib-CI 是由南非大学数学与信息学院计算机系开发的,以发展方法论和相应软件工具建立的学术数字图书馆本体项目(如图 2)^[16],其主题本体包括基于 2008 年 ACM 和 IEEE/CS 计算机科学课程的计算机科学本体,以及用在 ACM、AIS 和 AITP 信息系统中的信息系统本体。目前它的搜索引擎支持全文、作者、本体、关键词 4 种类型的搜索和文档检索,提供针对学生、教师和研究人員、图书馆员、普通公民四种类型的用户界面,为用户获取 DigLib-CI 中的资源提供较多的选择方式。在系统的本体搜索中,被搜索出的来自于研究或学习材料中的相关的概念扩充了用户查询,用户指定的每一主题本体中更具体的概念被添加到初始的结果查询中,这样,通过由本体搜索与用户查询定义的概念之间的语义关联得到的本体概念,能较快查到文档描述。类似的多本体应用还有 Mohammed Maree 等提出的多元搜索元搜索引擎,用多重本体知识库、结合语义、统计的技术得到用户查询和查询结果集两者的语义关系^[17]。

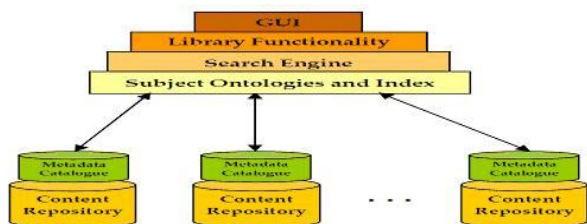


图2 DigLib-CI功能模型^[16]

2 基于检索结果自身特征的检索服务

基于检索结果自身特征的检索服务主要是利用检索结果集固有的特征,挖掘深层信息以提高检索准确度,使检索结果更符合用户需要。以往研究较多集中在挖掘结果集的外部特征,如文献的标题、关键词、被引次数、时间、作者等。近几年则将结果集内部特征、语义特征也纳入其中,主要通过统计、机器学习等方法发现文档的内容特征,并据此判断该文档与检索词之间的相关性。

Scirus 是典型的基于检索结果自身特征来提供检索服务的应用,是 Elsevier Science 和 FAST 合作开发的、较为全面的科学专业搜索引擎^[18],主要包括种子表的产生与维护、集中爬行、数据库载人、分类、索引、提问和检索排序等工作流程,系统的倒金字塔结构(如图 3)能帮助用户准确定位文档。Scirus 对检索结果集主要按照主题和信息类型对文献分类^[19],除此之外,Scirus 还能针对不同的文件格式进行分类。在结果集的排序上,Scirus 既能按照时间排序,又能采用静态排序和动态排序相结合的方式排序,其中静态排序主要根据检索词出现的位置、频率来排序,动态排序主要根据网页的链接数量来排序^[20]。检索结果的内容特征也是主要的关注点。如鉴于每种相似性措施都有各自的优缺点,Hazra Imran 等提出将各种相似性措施组合起来,形成组合相似性措施,基于检索词和文档内容的相关性来排序以实现相关文献排序的优化^[21]。Li 等研究了文章段落语义相关性模型,提出一种自适应的段落检索方法,即根据与给定查询匹配最优的段落对文档排序。具体执行过程就是针对给定查询,按照相关性模型将文档划分为全文、150 字、300 字和 500 字的大小不同段落,并针对这 4 种情况给出相关度分数,并据此分数来确定最优段落。此方法的优点就是在于能够简单地找到文档的最佳段落来改善检索服务^[22]。



图3 Scirus的倒金字塔结构

基于检索结果自身特征的检索技术还出现了一

些新趋势,表现在研究更多地关注检索结果的多样性和新颖性,并将一些其它学科领域的理论应用于检索服务领域中。如针对用户提供的检索词存在模糊性这一问题,Rodrygo L.T.等提出了用于检索结果多样化的新型框架^[23]——多面显式查询多样化(eXplicit Query Aspect Diversification, xQuAD)框架,该框架通过计算初始检索返回的文档集之间的关系及在这一检索要求下可能存在的其他相关方面,加强检索文档集的多样性。该框架主要包括四部分:多面重要性、文档覆盖面、文档新颖性、文档相关性。此框架通过提供跟检索词相关的多元化检索结果,进一步完善了原有的查询,能够帮助用户较好地定位文档。X.Yin等也采用了检索结果多样化^[24],认为相关的涵盖了新的内容的文章比那些相关却含有冗余信息的文章更有价值,将检索过程分为多面探测和多面过滤两个阶段,该方法有效促进了医学领域检索结果的多样性。Rijsbergen 2004年首次提出将量子理论(Quantum Theory)应用于检索领域的观点,此后 Guido Zuccon 做了进一步的研究。2009年,Guido Zuccon 提出将其应用在检索结果排序方面^[25],并于2010年做了量子概率排序规则(Quantum Probability Ranking Principle, QPRP)应用于检索结果排序的试验^[26]。其意义在于认为排序中的文档彼此之间存在干扰,第n个位置的文档的相关性受到它前面的n-1个文档的干扰。研究表明量子理论运用于检索服务能帮助用户较好地定位文档。

3 基于日志挖掘的检索服务

日志包含海量的历史访问信息,其中也蕴藏着潜在的用户反馈,这些信息对于提高检索质量具有很高价值^[27]。基于日志挖掘的检索技术是通过记录用户在检索中的相关数据(检索词、点击行为等),运用挖掘算法来发现用户检索偏好,从而以结果重排、查询扩展和个性化推荐等方式更好地满足用户需求。

针对用户的搜索意图和检索结果之间的差距较大、隐藏在本地网站上用户感兴趣的网页与检索结果并不相关等问题,Su提出了整合关联挖掘和PageRank算法的系统框架PRASE(Personalized Refinement Agent for Search Engine)。该框架的主要贡献在于利用从用户日志中发现的浏览模式进行结果重排和本地推荐,以个性化网页排序和新的推荐方法满足用户需求^[28]。该框架包括两个部分,一是位于

后台的离线挖掘操作,主要在用户浏览行为中挖掘全局和本地的关系;一是位于前端的在线操作,包括结果重排和本地偏好网页的推荐。

查询扩展词的一个重要方法是用文本挖掘或聚类技术从用户日志中找出与当前查询相关的词语。Ma等提出NWS(normalized weight SimRank)方法进行查询扩展。该方法通过实际的搜索引擎收集查询日志,分析查询和URL集之间的关系,创建查询-点击图,再通过多次的转换,建立术语之间的关系图^[29]。Nazli Goharian等通过在查询日志中应用关联规则来丰富文档关联术语的描述,每一个节点存放查询的日志,查询作为节点路线被记录到日志中,节点用这些日志获取术语的关联,并应用该关联进行查询扩展^[30]。类似地,Yin等也将搜索引擎的查询日志和检索结果之间的关联进行扩展查询^[31]。

根据用户日志生成相应的推荐信息,也是检索服务的重要应用。如Yang等提出了分布式农业科学和技术数据的智能检索模型,并将农业本体和用户模型的个性化检索技术应用于检索,其个性化服务模块可以记录和挖掘用户的历史数据,并据此预测用户的兴趣和需求^[32]。Mohamed Kouthear KHRI-BI等应用用户当前的导航历史,利用用户偏好以及学习资源内容的相似性和不同点,进行在线自动推荐^[33]。

4 基于用户交互的检索服务

以往信息检索较多关注如何通过各种算法加强检索的准确度,但随着用户对检索服务要求的提高,研究者逐渐意识到服务终究是以人为本的,并逐渐将用户交互加入到信息检索服务体系中。信息检索中的用户交互旨在通过用户反馈,明确用户检索需求,从文档集中筛选出需求的结果集,并以用户期望的方式加以呈现。当前基于用户交互方面对检索服务的支持,主要通过两种方式,一种是隐性的交互,通过跟踪用户反馈改善检索结果集的显示方式;一种是直接操纵型,用户根据界面在屏上显示某些主题对某一个主题进行交互,目前较为常见的是使用图形来直接操纵。

LAILAPS是检索生命科学数据库的搜索引擎(如图4),是隐性交互的典型应用。该系统结合了新的相关性排序模型,用机器学习的方法对用户相关情况建模,通过跟踪用户反馈改善排序。该方法使用直观精简的用户界面跟踪用户交互,并用一组

属性集抽取每一个数据库中的用户偏好来估计相关度。其排序逻辑是首先计算每个数据库词条的所有匹配位置,然后用9个基础特征类抽取每一个数据库词条的N-维特征向量,最后应用神经网络预测相关度,获得与特征向量相匹配的具体用户相关度分值,并据此来重排检索结果集^[34]。Benjamin Piwowarski等提出了一个基于量子理论的,理论上能够获取完整交互过程的交互式信息检索通用框架^[35]。该研究把检索看作包含两个驱动因素的过程,一个是用户的输入,一个是用户认知改变所带来的用户输入的变化。当前者能转化为标准的概率模型时,后者能采用量子物理的概率理论更好地建模。此框架通过对用户交互进行几何映射,处理单击、相关反馈和查询重构,采用定义的关联子空间提供查询推荐和导航等形式的交互。Mohamed Farah则针对特定用户与用户概况提出了隐含模型,该模型主要特点是根据用户先前检索产生的一组降序检索结果,推理用户未表达出来的偏好。具体的实现过程是在改变对话或交互阶段,用户通过对一组降序的检索文档排序与系统交互;在计算阶段潜在的相关项根据用户反馈和基于一系列相关标准的文档行为进行重组;在最后阶段,调整模型参数,提供最适合的用户反馈,该模型的应用使得信息检索更具个性化^[36]。类似地,Ulises Cervino Beresi等对用户的检索会话编码并加以记录,并将会话作为一种相关标准序列进行可视化,其中每个序列可以看作是对交互会话的一个总结,这种对用户交互行为的可视化应用可为用户推荐更能满足其需要的文档^[37]。

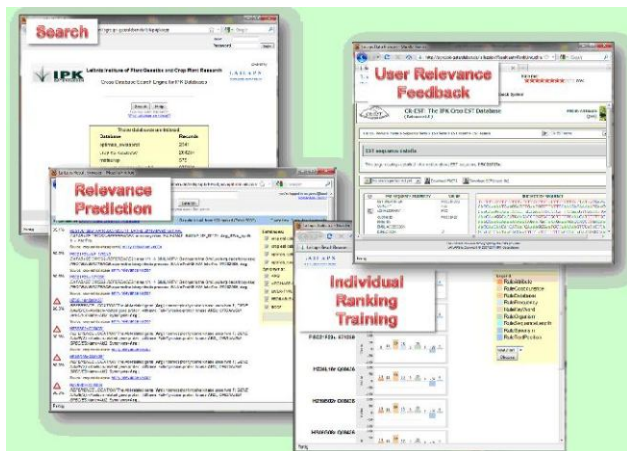


图4 LAILAPS查询会话的例子 四个浏览器窗口 LAILAPS常见的查询工作流程^[38]

直接操纵是检索服务中较为常用的用户交互方式。Gooda Sahib N.等人提出以检索倾向总结的形式向检索用户提供额外的信息,即允许检索者实时

查看他们查询扩展所产生效果的概况,包括更多建议和术语来支持用户做出更好的检索扩展决策。在系统界面上,用户可通过点击术语从系统给出的建议术语中添加10个术语增加到初始查询中,也可以通过单击去除一个已选择的术语,被选中的术语增加到初始查询的后面,在此过程中用户也能手工更改查询。实验结果显示如果用户知道他们的操作所产生的效果,他们就会更频繁地使用系统产生的术语。研究还表明,当用户对信息检索任务不熟悉或存在困难时,该交互式查询扩展方法是高效的^[39]。Jennifer Pearso介绍了一种系统,它包含用户查找提示和通过传统索引构建器建立的索引。其中索引构建器允许用户自己建立索引,根据输入的关键词给出相应文档最相关部分的可视化的概述。在此过程中用户可以自己输入关键词,选择相应的颜色、大小来创建数据的图形显示,系统最后将结果归类显示在同一页上,并通过颜色和大小显示共现。该系统能够帮助用户更好地理解隐含信息,更快地在文档中检索出自己所需的内容^[40]。此外,Neil Newbold认为应该把用户的阅读能力(和动机)纳入到检索中来,用户通过人机交互对年龄和阅读主题进行限制,来获得更贴近他们要求的文档^[41]。

5 结 语

综上所述,可以发现信息检索服务正朝着语义化、个性化、可视化的方向发展。语义化体现在通过挖掘检索词、信息资源的语义关联来提升检索服务质量,采取的主要方式是借助本体等语义资源映射语义关系和利用信息资源自身隐含语义信息提供服务。个性化体现在通过用户检索行为分析和人机交互过程来获取用户使用偏好,有机地将用户的信息需求和信息资源结合起来,提供满足用户个性化需求的检索服务。可视化则是采用可视化技术,丰富信息检索服务呈现形式,获取用户使用反馈意见,从而加强用户在信息检索服务中的参与性。此外,检索服务从利用检索结果外部特征向利用检索结果内部特征演进,主要体现在更加重视文档内容之间的联系,更为关注内容的多样性、新颖性。今后信息检索服务将沿着这几个方面,往更纵深的方向发展。

参考文献

- 1 Rey-Long Liu and Zong-Xing Lin. Improving Text Rankers by Term Locality Contexts[J]. Asia Information Retrieval Sym-

- posium, 2009,(5839):277-288.
- 2 Satoshi Shimada, Tomohiro Fukuhara, and Tetsuji Satoh. S-node: A Small-World Navigation System for Exploratory Search[J]. Asia Information Retrieval Symposium, 2009,(5839): 420-431.
 - 3 吴丹.本体在信息检索中的作用及实例研究[J].情报杂志, 2006,(6):72-75.
 - 4 Hans-Michael Müller, Arun Rangarajan, Tracy K. eal and Paul W. Sternberg. Textpresso for Neuroscience: Searching the Full Text of Thousands of Neuroscience Research Papers[J]. Neuroinformatics, 2008,6(3):195-204.
 - 5 William M. Urbanski, Brian G. Condie. Textpresso site-specific recombines: A text-mining server for the recombinase literature including Cre mice and conditional alleles[J]. genesis, 2009,47(12):195-204.
 - 6 Edward H. Y. Lim, James N. K. Liu and Raymond S. T. Lee. IATOPIA News Channel (IAToNews) An Intelligent Ontological Agent-Based Web News Retrieval and Search System[J]. Intelligent Systems Reference Library, 2011,(8):175-180.
 - 7 Osma Suominen, Eero Hyvönen, Kim Viljanen, and Eija Hukka. HealthFinland A national semantic publishing network and portal for health information[J]. Web Semantics: Science, Services and Agents on the World Wide Web, 2009,7(4): 287-297.
 - 8 Andreas Doms and Michael Schroeder. GoPubMed: exploring PubMed with the Gene Ontology [EB/OL]. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1160231/pdf/gki470.pdf>, 2011-06-24.
 - 9 Heiko Dietze, Dimitra Alexopoulou, Michael R. Alvers. GoPubMed: Exploring PubMed with Ontological Background Knowledge[J]. Bioinformatics for Systems Biology, 2009 ,(5): 385-399.
 - 10 张士靖, 杜建. GoPubMed :基于 Go 和 MeSH 的信息检索与分析研究[J]. 医学信息学杂志, 2009,(7):6-11.
 - 11 李健康. 专业化的语义智能搜索引擎 GoPubMed[J]. 图书馆论坛, 2009 (12):152-155.
 - 12 谢岩岩, 孙继林. 基于数据挖掘技术的知识服务体系以生命科学领域内 GoPubMed 为例[J]. 图书馆杂志 2010 (5):59-65.
 - 13 Ming-Hung Hsu, Ming-Feng Tsai and Hsin-Hsi Chen. Combining WordNet and ConceptNet for Automatic Query Expansion: A Learning Approach[J]. Information Retrieval Technology, 2008 (4993) 213-224.
 - 14 Jianqiang Li, Yu Zhao, and Bo Liu. Fully Automatic Text Categorization by Exploiting WordNet[J]. Information Retrieval Technology, 2009 (5839) :112.
 - 15 Jaap Kamps and Marijn Koolen. The Importance of Link Evidence in Wikipedia[J]. Advances in Information Retrieval, 2008 (4956) 270-282.
 - 16 Maria Nisheva-Pavlova, Pavel Pavlov. Search engine in a class of academic digital libraries[EB/OL]. http://www.math.bas.bg/~nkirov/2010/folk/4pavlova_pavlov.pdf 2011-06-24.
 - 17 Mohammed Maree, Saadat M. Alhashmi, Mohammed Belkhatir. Multi-Search: A Meta-search Engine Based on Multiple Ontologies[J]. Information Retrieval Technology, 2010 , (6458) :110-120.
 - 18 SCIRUS White Paper: How Scirus Works[EB/OL]. http://www.scirus.com/press/pdf/WhitePaper_Scirus.pdf 2011-06-24.
 - 19 于宁, 庞海燕. 科学搜索引擎与学术搜索工具 Scirus 与 Google Scholar 比较研究[J]. 现代情报, 2009,(6): 159-161.
 - 20 程妮. 科学搜索引擎 Scirus 的工作流程初探[J]. 情报科学, 2006,(2):287-293.
 - 21 Hazra Imran and Aditi Sharan. Genetic Algorithm Based Model for Effective Document Retrieval[J]. Intelligent Control and Computer Engineering, 2010,(70):191-201.
 - 22 Xiaoyan Li and Zhigang Zhu. Enhancing Relevance Models with Adaptive Passage Retrieval[J]. Advances in Information Retrieval, 2008,(4956):463-471.
 - 23 Santos, R.L.T., Peng, J., Macdonald, C., Ounis, I. Explicit search result diversification through sub-queries[J]. Advances in Information Retrieval, 2010,(5993): 87-99.
 - 24 Yin, X., Huang, X., Li, Z. Promoting ranking diversity for biomedical information retrieval using wikipedia[J]. Advances in Information Retrieval, 2010,(5993):495-507.
 - 25 Guido Zuccon, Leif A. Azzopardi, Keith van Rijsbergen. The Quantum Probability Ranking Principle for Information Retrieval[J]. Advances in Information Retrieval Theory, 2009, (5766):232-240.
 - 26 Guido Zuccon and Leif Azzopardi. Using the quantum probability ranking principle to rank interdependent document[J]. Advances in Information Retrieval, 2010,(5993):357-369.
 - 27 陈浩然. 基于日志分析的信息检索技术研究 with 实现[D]. 西安: 电子科技大学, 2009.
 - 28 Ja-Hwung Su, Bo-Wen Wang and Vincent S. Tseng. Effective Ranking and Recommendation on Web Page Retrieval by Integrating Association Mining and PageRank[C]. Tainan: Intelligent Agent Technology, 2008.
 - 29 Yunlong Ma, Hongfei Lin, and Song Jin. A Revised SimRank Approach for Query Expansion[J]. Information Retrieval Technology, 2010,(6458):564-575.
 - 30 Nazli Goharian, Ophir Frieder, Wai Gen Yee, and Jay Munderwala. Enriching Peer-to-Peer File Descriptors Using Association Rules on Query Logs[J]. Advances in Information Retrieval, 2010,(5993):532-543.
 - 31 Zhijun Yin, Milad Shokouhi, and Nick Craswell. Query Expansion Using External Evidence[J]. Advances in Information Retrieval, 2009,(5478):416.

- 32 Xiaorong Yang, Wensheng Wang, Qingtian Zeng, and Nengfu Xie. An Intelligent Retrieval Platform for Distributional Agriculture Science and Technology Data[J]. IFIP International Federation for Information Processing, 2011,(344):203-209.
- 33 Mohamed Kouthear KHRIBI, Mohamed JEMNI, Olfa NASRAOUI. Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval[C]. Tunis :Conference Publication, 2008.
- 34 M Lange, K Spies, C Colmsee, S Flemming, M Klapperstück, U Scholz. The LAILAPS Search Engine—A Feature Model for Relevance Ranking in Life Science Databases[EB/OL]. <http://www.biomedcentral.com/content/pdf/1471-2105-11-423.pdf>, 2011-06-24.
- 35 Benjamin Piwowarski and Mounia Lalmas. A Quantum-Based Model for Interactive Information Retrieval [J]. Advances in Information Retrieval Theory, 2009, (5766):224-231.
- 36 Mohamed Farah. Ordinal Regression Based Model for Personalized Information Retrieval [J]. Advances in Information Retrieval Theory, 200, (5766):232-240.
- 37 Ulises Cervio Beresi, Yunhyong Kim, Mark Baillie, Ian Ruthven and Dawei Song. Colouring the Dimensions of Relevance [J]. Advances in Information Retrieval, 2010, (5993): 569-572.
- 38 IPK-Gatersleben. COMPONENT DESCRIPTION [EB/OL]. <http://pgrc-35.ipk-gatersleben.de/apps/lailaps/documentation.html>, 2011-06-24.
- 39 Gooda Sahib, N., Tombros, A., Ruthven, I. Enabling interactive query expansion through eliciting the potential effect of expansion terms[J]. Advances in Information Retrieval, 2010 (5993) 532-543.
- 40 Jennifer Pearson, George Buchanan, and Harold Thimbleby. Creating Visualisations for Digital Document Indexing[J]. Research and Advanced Technology for Digital Libraries, 2009, (5714) 87-93.
- 41 Neil Newbold, Harry McLaughlin and Lee Gillam. Rank by Readability: Document Weighting for Information Retrieval [J]. Advances in Multidisciplinary Retrieval, 2010, (6107) : 20-30.

(责任编辑 刘凤琴)

(上接第 1360 页)

流群体内, 知识水平较高者和知识水平较低者间的比例有待进一步研究。

6 结 语

本文以知识水平的增长量为标准, 通过成员互评的方法来相对客观地测定成员知识水平, 以及利用凝聚分层聚类算法对成员进行群体划分, 量化研究了实践社区成员交流效果问题。借鉴相关实践社区定量研究成果, 比较分析了实践社区中群体中和群体间的交流效果。这种对实践社区交流效果的量化研究有利于提高实践社区成员的交流水平, 对实践社区的维持和发展有很好的参考意义。Wenger 提出实践社区是有共同热情的人们所组成的群体, 他们知道如何做及和谁交流能做得更好^[7]。而通过交流效果的量化研究可以得出最佳的交流群体, 在这个群体内成员清楚地知道和谁交流能够解决问题以及交流的效果最好, 这也是本文研究的重要方面。

参考文献

- 1 Wenger E C., Snyder W M. Communities of Practice :The Organization Frontier[J]. Harvard Business Review, 2000, (3):139-145.
- 2 Ash A., Joanne R. Knowing in action: Beyond communities of practice[J]. Research Policy, 2008, (37):353-369.
- 3 Nen-Ting Huang, Chiu-Chi Wei, Wei-Kou Chang. Knowledge management: modeling the knowledge diffusion in community of practice[J]. Knowledge Management, 2007, (36):607-621.
- 4 Culter R S. Survey of high-technology transfer practices in Japan and in the United States[J]. Interface, 1989, 19(6):67-77.
- 5 Wenger E C. Cultivating communities of practice a quick start-up guide[EB/OL]. <http://www.ewenger.com/theory/index.Htm>, 2011-06-10.
- 6 阳 博, 龙鹏飞. 凝聚分层聚类算法在入侵检测中的应用[J]. 长沙电力学院学报, 2005, 20(3) 57-60.
- 7 Wenger E C. Knowledge Management as a Droughnut: Shaping Your Knowledge Strategy through Communities of Practice[J]. Ivey Management Services, 2004, (2):1-8.

(责任编辑 孙晓明)