

# 基于分众分类的核电专家知识推荐算法研究

◆ 吴 珺

**摘要:** 核电企业搜索引擎系统由于相关领域的专业性, 可以将专家用户浏览过的专业领域的文档进行记录, 作为专家知识推荐给普通用户, 提高搜索引擎的使用效率。本文提出了基于分众分类的核电专家知识推荐算法, 帮助普通用户方便、有效地定位到有价值的知识。

**关键词:** 核电企业搜索引擎系统; 分众分类; 专家知识推荐

企业中常常包含其相关领域的知识, 企业中的员工常常需要对这些知识进行查询, 以更深入了解相关的领域。很多的企业都建立了自己的知识管理系统, 以提高对企业中的知识管理的水平, 通过这种方式来促进企业中知识的共享和创新。核电企业搜索引擎系统不仅能够从企业内外部的不同数据源中抓取各种结构化、非结构化和半结构化的数据, 提供对这些数据的管理, 用户还可以通过搜索引擎方便、快捷地查询到企业内外部相关领域的各种资料、文档等信息。通过搜索引擎的方式, 很容易实现知识在企业中的共享。

核电领域的知识比较专业, 而且涉及的知识方向比较多, 也比较深入。正因为如此, 核电企业中往往会有一些对核电相关各个领域的专家, 而其他的员工则对相关领域的了解不是那么的深入和透彻。对于对某个领域比较了解的专家用户来说, 可以很快速地通过搜索引擎查询、定位到自己专业领域的知识; 而对于普通用户来说, 往往不能很准确地描述自己所要搜索的内容, 或者不能从搜索结果中迅速找到比较有价值的, 自己所需的内容。因此, 普通用户使用搜索引擎时对于关键词的盲目性比较大, 搜索到自己所需内容准确率比较低, 效率也比较低。

为了解决普通用户使用搜索引擎查询时效率低下的问题, 提高知识在企业中的共享速度, 可以在普通用户输入关键字进行搜索时, 提供与该关键字相关领域的专家用户的浏览记录推荐给普通用户进行使用。由于专家用户在某些领域的专业性和权威性, 他们比普通用户更能够快速地从搜索结果中找到有价值的信息进行浏览, 这些信息对普通用户来说有很重要的参考价值。

## 一、分众分类研究

分众分类又叫做“通俗分类法”、“大众分类

法”、“民众分类法”、“自由分类法”<sup>[1,2]</sup>, 英文为 Folksonomy, 是由Folks和Taxonomy组合而创造出的新词汇, 由Tomas Vander Wal于2004年提出<sup>[2]</sup>。Folks表示一群人、一伙人; Taxonomy是指分类法, 是信息架构中的重要部分; 由这两个单词和在一起组成的Folksonomy则是指一种社群参与人运用自定义标签(tag)方式进行协作分类的活动<sup>[3]</sup>。分众分类中用户可以提交自己定义的标签, 通过用户群体定义标签的频率来决定信息的组织方式, 而不是使用传统的严格分类标准, 因此整个分类过程都是完全自发的, 更为方便、灵活, 在当下非常流行的社会性网络服务中得到广泛应用。

分众分类法的概念于2004年8月被Thomas Vadener Wal提出后, Adam Mathes和David N Sturtz等人发表了重要论著, 主要关注在对分众分类概念的普及和推广上; 到了2005年, 对分众分类的研究方向转移到了应用领域, Tony Hammond、Terrell Russell等人发表综述和论文, 围绕分众分类的发展背景、定义、与传统分类法之间的关系、工作原理, 以及对认知权威性的发现和定义上都分别进行了论述; 2006年关于分众分类的研究则侧重于对分众分类法系统构建的研究, Elaine Peterson、Adam Seldow、Chi-Lung、Trant J等人分别在文章中通过回顾传统信息组织方法在实现流程及其不足, 并在教育、图书馆等特定领域引入分众分类, 并充分将分众分类与Web 2.0进行结合; 2007年, 对分众分类的研究重点转移到了分众分类的本质和索引方向, Thomas Gruber<sup>[4]</sup>、Jakob Voss<sup>[5]</sup>等分别在分众分类的社会性, 及其对人工索引方向进行了探讨; 2008年分众分类的研究重点转移到了分众分类的使用范围上, James Sinclair<sup>[6]</sup>发表论文讨论了分众分类的使用范围, 上海交通大学的一部分同学<sup>[7]</sup>则探讨了将分众分类法用于个性化搜索中; 2009年的研究主要在于对分众分类的改进上, Jeff

Z. Pan<sup>[8]</sup>等人发表论文提出在基于分众分类搜索扩展的标签系统中减少模糊性的方案, Trant<sup>[9]</sup>则发表了对steve.museum的研究结果; 2010年对分众分类的研究主要关注在个性化搜索方面, Fabian Abel<sup>[10]</sup>、David Vallet<sup>[11]</sup>等人均对分众分类在搜索中的作用进行了研究; 2011年对分众分类的研究围绕在基于分众分类的推荐上, 并开始将研究方向与云计算进行结合, Jens Illig<sup>[12]</sup>、Iván Cantador<sup>[13]</sup>等人发表的文章都将重点关注到基于分众分类的推荐上, 而Wen-Chung Shih<sup>[14]</sup>则将e-Learning内容检索的基于分众分类的模糊索引创建与云计算环境进行结合。

目前, 分众分类法分为两种, 一种是面向所有使用互联网大众的“广义型”模式, 另一种则是面向某一特定类型用户的“狭义型”模式<sup>[2]</sup>。

广义型分众分类(Broad Folksonomy)面向的是互联网上的所有用户, 用户在知识结构和兴趣方面都有着巨大差异<sup>[15]</sup>。每个用户根据自己的认知, 使用自己的语言对社区中现有的或尚未添加内容提供独特的标签, 每个内容或每种分类都能够使用相当多的标签来进行描述, 会产生大量的各种各样、多种描述内容的标签, 建立起内容广泛的平台, 为用户提供一个较大的信息或知识范围进行信息的浏览和查询。

狭义型分众分类(Narrow Folksonomy)与广义型分众分类最大的不同之处在于用户的特质上<sup>[2]</sup>。狭义型分众分类针对某一特定领域有着类似的知识结构和兴趣的用户提供某一特定领域或专业的信息和知识共享平台。社区中的用户都是基于对同一领域或专业的内容感兴趣加入, 内容的提供者本身可能就是社区中的用户, 为自己上传的内容标注一个标签, 并希望社区中与自己存在着相同兴趣的用户共享相关信息或资源。从一定程度上讲, 狭义型分众分类构建的已不再是面向大众的分类体系平台, 而是在较为狭窄的范围内为具有相同或相似兴趣的用户提供共享专业化信息或知识的平台, 因此也无需数量庞大的用户, 这样更有利于在社区中进行知识共享的实现。

## 二、基于分众分类的知识推荐算法

对于一般的知识管理系统来说, 其知识推荐功能是构建在专家数据库的专家地图上的, 因此系统中专家知识推荐的准确性是和专家数据库中的专家数据有着密切的关系<sup>[16]</sup>。专家数据往往是由企业中专门的员工来维护, 而不是企业中各个领域的专家, 在管理上具有很大的主观性和盲目性, 并且不具备很好的灵活性和时效性, 这样会导致专家数据不够准确, 直接影响到其他非

专家用户的使用。

考虑到核电企业搜索引擎系统的特点, 本文可以将狭义型分众分类与专家知识推荐进行结合, 以提高专家知识推荐的准确性和用户的使用体验。基于分众分类的专家知识推荐逻辑图如图1所示:

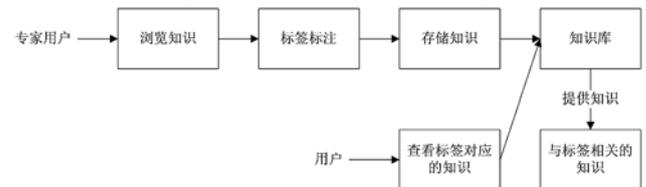


图1 基于分众分类的专家知识推荐逻辑图

与典型的狭义型分众分类不同, 核电企业搜索引擎系统的专家知识推荐中没有使用传统意义上的标签。根据使用搜索引擎的特点, 用户使用关键字来对自己感兴趣的文档资料进行搜索, 因此当用户点击搜索引擎的某个返回文档进行浏览时, 就表明该用户对这个文档感兴趣, 并且该关键字能够表明该文档的某个特征, 因此可以将关键字看作是文档的标签。在内容来源上, 核电企业搜索引擎系统会自动从各个数据源抓取各种数据、文档等信息作为相关搜索结果返回给用户, 所以用户无需自己上传内容。

作为专家知识推荐, 专家用户需要在某个领域具有权威性, 但专家用户的专业性、权威性仅仅只是集中在其所熟悉的领域上, 范围很小。出于研究或者应用的考虑, 专家用户偶尔需要搜索非自己专业领域的资料, 而专家用户并不是这些方面的专家, 其身份和其他普通用户一样, 若把这种情况下专家用户浏览的文档也当作专家用户的推荐信息, 那么专家知识推荐的准确性就会受到影响。

为了解决这个问题, 可以让专家用户创建个性化档案, 输入自己所擅长领域的关键词, 可以作为其专业领域的标签。当然, 每个专业领域的某个定义常常会有多种表述方式, 如分众分类就有大众分类、社会性分类、Folksonomy等多种叫法, 因此专家用户输入的专业领域的关键词的同义词也应该作为专家用户的兴趣标签。为了确保所记录的专家用户的行为轨迹其所熟悉的那个领域的内容, 需要限制只有当专家用户使用自己的个性化档案中的关键词及其同义词进行搜索, 才能对其浏览内容进行轨迹记录。

用户并不是传统意义上的上传自己感兴趣的内容, 并创建对该内进行描述的标签, 并通过标签与其他用户进行分享的方式; 而是专家用户通过输入关键字, 使用系统中的搜索引擎的返回结果发掘与自己专业领域相关且感兴趣的文档, 并将关键字作为该文档的标签, 并在

其他用户通过搜索该关键字或其同义词来得到专家知识推荐的方式来分享信息。

图2是面向核电企业搜索引擎系统的基于分众分类的知识推荐算法的逻辑图：

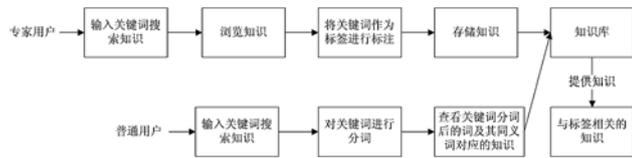


图2 面向核电企业搜索引擎系统的基于分众分类的专家知识推荐逻辑图

根据核电企业搜索引擎系统中知识管理的特点，普通用户和专家用户都常常需要使用搜索引擎查询信息，专家数据库中的专家数据完全可以由专家用户自己提供和管理。系统中用户的创建可以由系统管理员完成，在创建时定义该用户是否为专家用户。系统管理员可以根据该用户在企业中的级别，从事该行业相关领域的年限等具体信息来判断用户是否具有专家用户的资格。因此，基于上述思想提出的面向核电企业搜索引擎系统的基于分众分类的专家知识推荐算法除了具备上述的基于分众分类的知识推荐的优点以外，还具有在核电企业中这个具体的专业领域的实用性，这是其他知识推荐算法所不具备的。

### 三、基于分众分类的核电企业专家知识推荐算法的实现

要想实现面向核电企业搜索引擎系统的基于分众分类的专家知识推荐，首先需要记录专家用户的行为轨迹。其流程图如图3所示：

用户首先需要创建自己的个性化档案，对于专家用户来说，专业领域的关键词能够表明专家所熟悉和感兴趣的领域，相当于专家专业领域的标签。当用户使用搜索引擎搜索文档资料，浏览搜索结果中的某些文档时，说明用户认为该文档与其输入的关键词相关，也即关键字可以看作是对该文档的标签。考虑到普通用户和专家用户对专业领域的了解程度的差异，仅认为专家用户对文档的浏览记录是有价值的。当有用户浏览某个文档时，需要判断该用户是否是专家用户，如果不是专家用户则不对本次浏览轨迹进行记录；如果是专家用户，但专家浏览的不是自己所擅长领域的文档，也不进行记录；只有当专家用户浏览自己专业领域的文档时，才将该专家用户的ID、本次搜索的关键词和所浏览文档的编号存储进数据库中。

在专家用户的轨迹记录完成后，即可根据专家的浏览记录为普通用户进行专家知识推荐。图4显示了知识

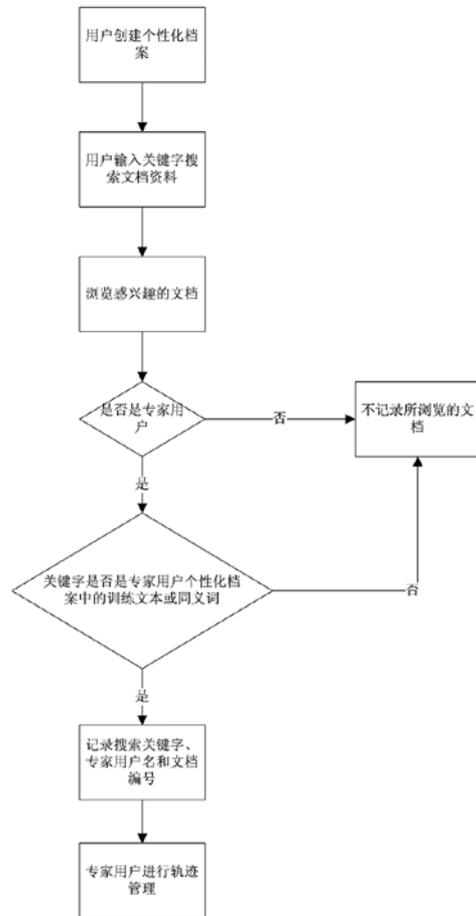


图3 专家轨迹记录流程图

推荐的流程：



图4 知识推荐流程图

当普通用户在搜索框中输入自己感兴趣的关键词进行搜索时，系统就可以为该用户提供专家知识推荐。由于普通用户对所搜索的关键词相关领域的知识不是很熟悉，可能会对所要搜索的内容描述不是很准确，加上很多专业词汇存在同义词现象，因此需要先对关键字进行分词，并获得分词后的同义词信息，以帮助普通用户获得相关领域的尽可能多的相关信息。使用关键词及其同义词在数据库中进行模糊搜索，得到相关专家用户的轨迹信息，即文档编号，对这些编号按照访问次数进行降序排列。一般可以认为，被访问次数越多的文档，其价值也越高。最后根据这些文档编号，在系统中获取文档的相关介绍信息，并最后以搜索结果展示给用户。

### 四、结语

传统的知识推荐算法一般都是以用户对知识评分



或聚类技术等来为用户进行知识推荐；基于分众分类的专家知识推荐，则由专家用户使用标签来对知识进行标注，其他用户在获取知识时，则可以根据这些标签来得到相关的知识推荐。由于分众分类方法中标签是由用户自己创建，因此与传统的方法相比，其灵活性更好，更符合用户的使用习惯和思维习惯。除此以外，基于分众分类的知识推荐算法由于实现起来比较简单，系统结构也不复杂，往往都有比传统知识推荐算法更好的效率。

本文根据分众分类的思想，结合核电企业搜索引擎系统的特点，提出了一种面向核电企业搜索引擎系统的基于分众分类的专家知识推荐方法，不采用传统的分众分类中用户上传自己感兴趣的内容并自定义标签对其进行标注的方法，而是另辟蹊径，认为搜索引擎系统的用户可以将关键词作为标签，以搜索引擎系统搜索出的结果作为自己感兴趣的内容。

虽然可以将用户使用搜索引擎时输入的关键字以及个性化档案中的训练文本作为用户兴趣标签，但这些都难以包含用户在核电领域所有的兴趣，可以通过挖掘用户浏览的文档中的高频关键词，来更进一步拓展更贴近用户真正的兴趣，更好、更准确地描述用户兴趣。



### 参考文献

- [1]毛军.元数据、自由分类法(Folksonomy)和大众的因特网[J].现代图书情报技术,2006(2):1-4,9.
- [2]胡海涛.Folksonomy在网络学习资源管理中的应用研究[D].山东师范大学硕士学位论文,2008,4.
- [3]余金香.Folksonomy及其国外研究进展[J].图书情报工作,2007,51(7):38-40,74.
- [4]Thomas Gruber. Ontology of Folksonomy A Mash-Up of Apples and Oranges[J].Int'l Journal on Semantic Web & Information Systems,2007,3(2).

(上接144页)

5.改革与课程体系相适应的考试考核方法。实施阶段性考核和项目最终性考核与“双证融通”相结合的考核办法，重视培养学生运用所学知识分析解决问题的综合能力，考核的手段和方式更注重适用性、应用性、创新性。软件技术专业的课程体系改革以“就业导向”、“能力本位”、“企业所需”、“强化实践”为研究思路，坚持实事求是的原则，从企业需求、够用即可的要求出发，以JAVA和.NET为两条主线进行课程的制订，以学生兴趣为主，充分调动学生对计算机语言类课程学习的积极性和主动性，从而达到培养学生基本技能和创

[5]Jakob Voss.Tagging, Folksonomy & Co-Renaissance of Manual Indexing[EB/OL]. <http://arxiv.org/pdf/cs.ir/0701072.pdf>.

[6]James Sinclair, Michael Cardew-Hall. The folksonomy tag cloud: when is it useful?[EB/OL]. <http://www.jrsinclair.com/academic/JIS-0449%20-v3%20-%20The%20Folksonomy%20Tag%20Cloud%20-%20When%20is%20it%20useful.pdf>.

[7]Shengliang Xu, Shenghua Bao, Ben Fei. Exploring Folksonomy for Personalized Search[EB/OL]. [http://dmlab.uos.ac.kr/html/lecture/DataMining\(2008-2\)/Exploring%20Folksonomy%20for%20Personalized%20Search.pdf](http://dmlab.uos.ac.kr/html/lecture/DataMining(2008-2)/Exploring%20Folksonomy%20for%20Personalized%20Search.pdf).

[8]Jeff Z.Pan, Stuart Taylor, Edward Thomas. Reducing Ambiguity in Tagging Systems with Folksonomy Search Expansion[EB/OL]. <http://www.abdn.ac.uk/~csc280/PTT09.pdf>.

[9]Jennifer Trant. Tagging, Folksonomy and Art Museums: Results of steve.museum's research[C]. DLIST, 2009.

[10]Fabian Abel, Ricardo Kawase, Daniel Krause. Leveraging multi-faceted tagging to improve search in folksonomy systems[C]. Proceedings of the 21st ACM conference on Hypertext and hypermedia, June 2012: 299-300.

[11]David Vallet, Ivan Cantador, Joemon M. Jose. Personalizing Web Search with Folksonomy-Based User and Document Profiles[C]. ECIR'2010 Proceedings of the 32nd European conference on Advances in Information Retrieval, 2010: 420-431.

[12]Jens Illig, Andreas Hotho, Robert Jaschke, Gerd Stumme. A Comparison of Content-Based Tag Recommendations in Folksonomy Systems[C]. KONT'07/KPP'07 Proceedings of the First international conference on Knowledge processing and data analysis, 2011: 136-149.

[13]I. Cantador, I. Konostas, J. M. Jose. Categorising social tags to improve folksonomy-based recommendations[J]. Web Semantics: Science, Services and Agents on the World Wide Web, 2010, 9(1): 1-15.

[14]Chao-Tung Yang, Shian-Shyong Tseng. Fuzzy folksonomy-based index creation for e-Learning content retrieval on cloud computing environments[C]. Fuzzy Systems (FUZZ), 2011 IEEE International Conference, June 2011: 965-970.

[15]Ciro Cattuto, Vittorio Loreto, Luciano Pietronero. Collaborative Tagging and Semiotic Dynamics[J]. PNAS, 2007.

[16]梁凯春,蔡淑琴,林森.基于分众分类的专家推荐算法研究[J].武汉理工大学学报,2007,29(4):87-90.

(作者单位:厦门大学软件学院)

新能力的目的。



### 参考文献

- [1]王晴.印度软件业崛起原因及其对我国产业结构升级的启示[J].生产力研究,2007(3):104-106.
- [2]王善桃.借鉴北大青鸟ACCP课程体系探索软件技术专业发展之路[J].和田师范专科学校学报,2009(62):173-174.
- [3]赖红清.基于职业教育的校企合作问题的探讨[J].中国教育技术装备,2009(06):160-161.
- [4]刘震.从“校企合作”模式下的课程置换看软件专业师资队伍建设[J].福建点,2008(10):202-203.

(作者单位:黑龙江农业经济职业学院)