

基于用户兴趣特征的知识推送^①

余绅达, 盛步云, 付高财, 吴昊, 鞠帅帅, 谢俊

(武汉理工大学 机电工程学院, 武汉 430070)

摘要: 为弥补产品设计过程中知识检索方式的不足, 同时降低因需求不明确导致知识推送的盲目性, 提出了一种根据用户兴趣特征进行筛选实现设计知识推送的方法。首先阐述了基于用户兴趣特征的产品设计知识推送原理, 给出了设计知识组织管理方法。然后分析了用户兴趣特征集合的构建方法, 采用文本相似度算法, 实现了知识特征与兴趣特征的匹配。最后以某电气产品制造企业为背景, 通过应用实例验证了该方法的可行性和可靠性。

关键词: 产品设计; 知识组织; 知识推送; 兴趣特征; 相似度;

Knowledge Push Based on User Interest Feature

YU Shen-Da, SHENG Bu-Yun, FU Gao-Cai, WU Hao, JU Shuai-Shuai, XIE Jun

(Mechanical and Electrical Engineering, Wuhan University of Technology, Wuhan 430000, China)

Abstract: To offset the shortage of knowledge retrieval methods for product design process, while to reduce the blindness of knowledge push due to uncertain demand, a method according to users' interest feature to achieve knowledge push is proposed. Firstly the theory of design knowledge push based on users' interest feature is expounded, and the method of knowledge organization and management is given. Then the method for establishing user interest feature set is analyzed, and the match between knowledge feature and interest feature is realized by utilizing similarity matching algorithm. Finally case is given to illustrate feasibility and reliability of knowledge push method.

Key words: product design; knowledge organization; knowledge push; interest feature; similarity

产品设计是一个复杂的创新设计过程, 在设计各个阶段需要不同的设计知识^[1]。随着设计知识的不断增加, 产品设计过程中设计人员需要花费大量时间和精力寻找适合自己的设计知识, 这给设计人员造成许多不必要的麻烦, 导致设计效率降低。因此在设计人员接收到设计子任务的同时, 自动挖掘出特定的知识并及时推送给设计人员, 对于新产品设计开发具有重要意义。

知识推送是指将知识库中的知识通过一定的技术标准或协议, 根据接收人员的兴趣和承担任务的需求, 将匹配成功的知识及时传递给用户^[2]。很多学者从不同角度对知识推送进行了研究。文献[3]从知识属性、流程、领域三个维度对企业知识进行了描述, 提出了流程驱动的知识主动推送方式; 文献[4]研究了一种基于工作流引擎和知识引擎双驱动的知识主动推

实现机制, 提出以业务过程为导向、以知识需求为驱动的集人、过程、知识为一体的知识推送服务系统构建方法; 文献[5]研究了设计人员知识需求的生成方法与更新原则以及基于概念语义扩展的设计知识检索方法, 提出面向产品设计人员的知识推送服务。文献[6]通过建立包含目标功能、约束条件和用户兴趣领域的设计意图模型, 提出了一种基于机械设计意图建模的知识主动推送方法。

在实际应用中, 由于知识资源库的不断扩大导致检索不准确, 同时设计人员通常只对与自身工作相关的设计知识感兴趣。为解决知识推送过程中知识检索方式的不足, 提出了基于用户兴趣特征的知识推送方式, 利用用户兴趣特征来弥补仅仅通过设计子任务给定需求信息来检索的不足, 检索的结果更加符合设计人员的需求。

^① 收稿时间:2016-01-18;收到修改稿时间:2016-03-10 [doi: 10.15888/j.cnki.csa.005344]

1 设计知识推送原理

产品设计知识推送是根据产品设计初始阶段定义的约束条件,从产品设计知识库中推送给产品设计人员相应的知识^[7]。现代企业中,由于知识资源的不断增加,另外,不同角色的设计人员对于同一产品、零部件设计知识的需求有很大不同,仅仅根据产品设计初始阶段给定的知识需求信息很难抓住设计人员全部需求,不仅会造成结果数量庞大的问题,而且推送的结果缺乏个性化。因为在说明设计任务知识需求时没有将设计人员的兴趣偏好考虑进来,导致匹配的结果不准确。

基于用户兴趣特征的设计知识推送方法是在设计人员进行相关产品设计时,系统根据设计子任务提供的知识需求信息,结合用户兴趣特征找出设计人员可能需要的知识并推送出去。

用户兴趣特征表示当前用户熟悉和感兴趣的专业领域,对于相同的知识需求,不同的用户由于兴趣特征的不同,最终获得不同的设计知识。另外用户兴趣特征的引入在合理的范围内扩大了设计知识的检索范围,以获得更多符合设计人员需求的知识信息,减少设计人员重复检索的时间。因此,通过引入用户兴趣特征可以更好地实现设计知识推送。

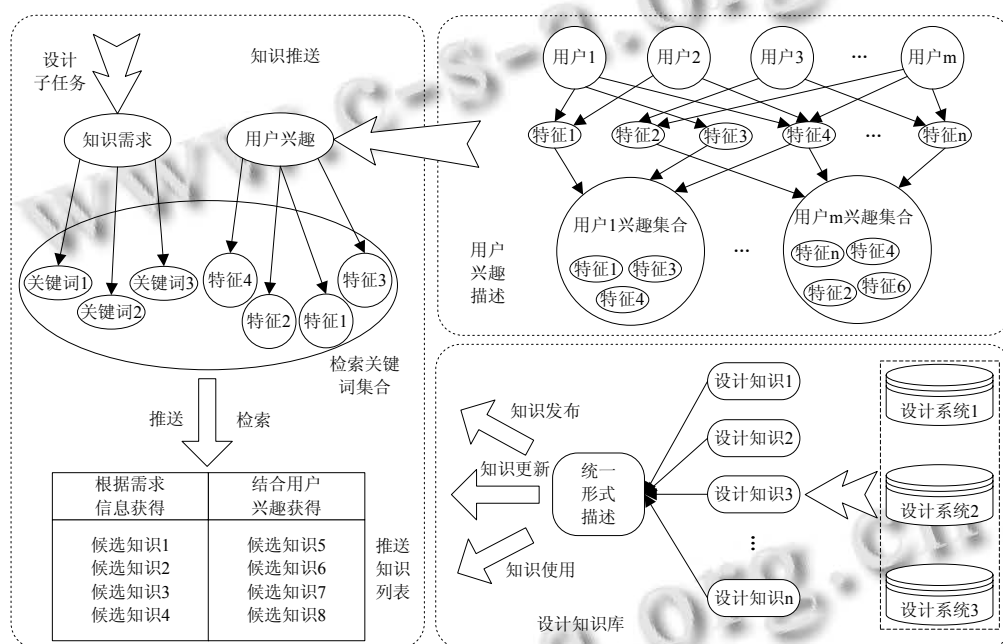


图1 设计知识推送原理

基于用户兴趣特征的设计知识推送原理如图1所示,知识推送过程的实现先要解决两个方面的问题。首先是设计知识库的构建,它是知识推送系统的基础工作。在现有的设计知识管理系统上对设计知识进行重新处理,统一设计知识的描述形式,采用文本的形式将知识关键信息记录下来,便于知识匹配。其次是用户兴趣描述,准确表达用户兴趣能够提高知识推送能力。引入用户兴趣特征使得系统在进行设计知识推送时考虑用户的兴趣,检索的内容在一定范围内扩大,最终得到的候选设计知识不仅包含知识需求提供的知识而且结合用户兴趣得到更多可能需要的知识,达到设计知识推送的目的。

2 设计知识的组织管理

在产品设计中,设计知识作为创新企业的主要智力资产始终贯穿于整个设计流程中^[8]。设计知识主要包括各类设计手册、设计案例、设计图纸、三维模型以及设计标准和规范等。企业现有的知识管理系统缺乏统一的表示,不同类型的设计知识存在于不同设计系统中。随着知识资源库的不断扩大,设计人员常常很难查找到需要的知识。为了更好地实现设计知识推送,需要在现有知识管理系统的基础上,统一设计知识的描述形式。

定义1. 设计知识关键词集合。采用若干个关键词对设计知识的类型、内容、涉及产品等信息进行描述,

这些关键词形成的集合称为设计知识关键词集合,记为 K .集合中的关键词从企业常用关键词汇表 M 中选取,用 k 表示,即 $K = \{k | k \in M\}$.

其中 M 是企业常用关键词汇表.由于设计知识和用户兴趣都需要用关键词来描述,不同设计人员在描述知识和兴趣时会出现用词上的差异.为了避免出现大量语义相同的关键词汇,根据企业实际情况制定关键词汇表,将常用关键词纳入其中.在描述设计知识和用户兴趣时,通过对照关键词汇表保证描述信息的一致性.

根据设计知识的特点以及设计人员的检索习惯,采用文本描述的方式对系统设计知识进行统一处理,包括文档类和非文档类设计知识.设计知识描述文本中包含了关键词集合和映射关系两部分内容.在处理系统原有知识以及新增知识时,设计人员通过提取设计知识关键词完成关键词集合的创建,同时将设计知识实际存放的位置信息添加到文本描述中.图2是推送系统设计知识组织与管理过程.

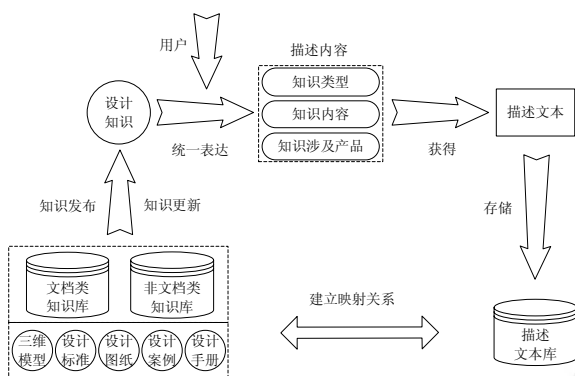


图2 设计知识组织与管理过程

3 设计知识推送方法

第1节中阐述了基于用户兴趣特征的设计知识推送原理,引入用户兴趣特征可以向用户推送感兴趣的设计知识.可以看出推送知识是否符合用户取决于用户兴趣的准确表达和设计知识的准确匹配.下面主要从用户兴趣特征集合的构建与更新、相似度计算、匹配度阈值约束三个方面对设计知识推送方法展开说明.

3.1 用户兴趣特征集合

定义2. 用户兴趣特征集合.采用若干个关键词对用户感兴趣的知识领域或者知识类型进行特征描述,这些关键词形成的集合称为用户兴趣特征集合,记为 J .集合中的关键词称为用户兴趣特征,用 j 表示,

即 $J = \{j | j \in M\}$.

用户兴趣特征为用户专业背景和兴趣领域的直接表达,准确表达用户兴趣是实现设计知识推送的关键.用户兴趣特征集合的建立包括初始兴趣特征集合建立和兴趣特征集合更新两个阶段.使用知识推送系统时,每位用户都拥有自己的账户,用户兴趣特征集合通过个人账户与用户进行绑定.初始用户兴趣特征集合是用户根据自身岗位背景,对感兴趣的知识领域进行提取描述获得的关键词集合.在系统使用初期,主要根据用户初始设定的兴趣特征集合帮助用户完成设计知识推送.

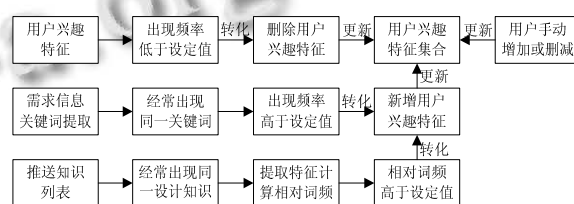


图3 用户兴趣特征集合更新过程

随着用户不断使用,系统将通过增加和删减两种方式对用户兴趣特征集合进行更新.图3表示用户兴趣特征集合的更新过程.用户兴趣特征的增加是通过分析用户对推送知识的访问情况来实现的.如果用户知识推送列表中经常出现同一设计知识,那么系统会认为该设计知识是用户感兴趣的知识类型,将该设计知识的关键词提取出来,计算出关键词的相对词频,将相对词频大于某一设定值的关键词增加到用户兴趣特征集合中.另一种新增用户兴趣特征的方式是当设计子任务提供的知识需求信息中经常出现同一关键词且频率高于设定值,那么系统会认为该关键词可以表达用户兴趣,将其增加到用户兴趣特征集合中.

如果用户使用的推送知识和知识需求信息中长期不包含某一用户兴趣特征,那么系统会认为该特征表达不准确或用户失去对该特征的兴趣,当该特征出现的频率低于设定值时,将其从用户兴趣特征集合中剔除.用户还可以通过手动方式定期对自身的兴趣特征进行修改,将不符合兴趣描述的关键词删除或者新增用户兴趣特征.

3.2 相似度计算

定义3. 检索关键词集合.将知识需求信息中的关键词与当前用户兴趣特征进行合并,剔除重复的关键词,得到一个关键词集合,用于与设计知识进行匹配,

该集合称为检索关键词集合, 记为 L .

图 4 是设计知识推送过程, 首先获取设计子任务所提供的知识需求信息, 同时通过个人账户得到当前设计人员的用户兴趣特征集合, 提取需求信息关键词和用户兴趣特征, 合并得到检索关键词集合. 然后分别计算检索关键词集合和设计知识关键词集合中各个关键词的相对词频, 将检索关键词集合与设计知识关键词集合表示成空间向量, 向量的各个分量由关键词的相对词频来表示, 通过文本相似度算法比较两者间的相似性. 最后将检索得到的设计知识推送给用户.

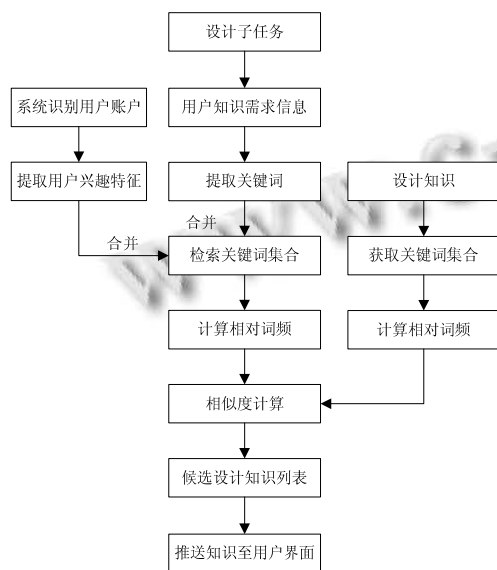


图 4 设计知识推送过程

检索关键词集合是需求信息关键词与用户兴趣特征综合得到的. 考虑到知识推送时需求信息更接近用户当时的要求, 因此其重要程度高于用户兴趣, 故在计算检索关键词的相对词频时, 利用权重值来表示两者的重要程度.

定义 4 权重系数. 用 α 和 β 分别表示检索关键词集合中需求信息关键词和用户兴趣特征的权重值, 以此来表征两者的重要程度, α 和 β 称为权重系数.

因此本文中各个关键词的相对词频可以用改进 $tf-idf$ 算法得到. 设 k_i 是设计知识关键词集合中的一个关键词, 系统设计知识数量为 N , 关键词集合中包含关键词 k_i 的设计知识数量为 n_i , 那么关键词 k_i 在设计知识关键词集合 d_j 中的逆频 idf_i 为:

$$idf_i = \log \frac{N}{n_i} \quad (1)$$

关键词 k_i 在设计知识关键词集合 d_j 中的相对词频 w_{ij} 为:

$$w_{ij} = tf_{ij} \times idf_i \quad (2)$$

其中 tf_{ij} 是关键词 k_i 在设计知识关键词集合 d_j 中的词频.

相同的方式可以得到检索关键词集合中各个关键词的相对词频. 因此检索关键词集合 d_c 中的关键词 k_i 的相对词频为:

$$w_{ic} = \alpha \times tf_{ic} \times idf_i \text{ 或 } w_{ic} = \beta \times tf_{ic} \times idf_i \quad (3)$$

完成全部关键词相对词频的计算之后, 采用 k 维向量 $\vec{d}_j = (w_{1j}, w_{2j}, \dots, w_{kj})$ 和 $\vec{d}_c = (w_{1c}, w_{2c}, \dots, w_{kc})$ 分别表示系统设计知识关键词集合 d_j 和检索关键词集合 d_c , k 是不同关键词个数. 最后系统设计知识关键词集合 d_j 和检索关键词集合 d_c 的相似度计算如下:

$$\text{sim}(d_j, d_c) = \frac{1}{\sqrt{1 + \sum_{i=1}^k (w_{ij} - w_{ic})^2}} \quad (4)$$

3.3 匹配度阈值约束

在完成检索关键词集合与设计知识关键词集合的相似度计算之后, 可以获取所有设计知识与检索关键词集合的匹配度. 为了避免因知识资源库不断扩大而导致检索得到的设计知识数量过多, 在对候选设计知识进行排序之前, 先剔除匹配度过低的设计知识. 匹配度阈值就是限制检索关键词集合与设计知识关键词集合之间的最低匹配度. 当两者匹配度小于匹配度阈值时, 该设计知识将从候选知识列表中剔除. 系统对所有匹配度大于阈值的设计知识按照匹配度大小排序最后推送至用户界面.

匹配度阈值的设置十分重要, 阈值过大或者过小都将影响系统知识推送能力. 如果阈值过大可能导致最终推送给用户的设计知识太少且内容过于相似, 不利于设计人员的创新活动. 如果阈值过小可能导致推送的设计知识与用户兴趣偏离太多, 增加了用户选取设计知识的工作量. 因此, 在设置匹配度阈值时, 需要根据企业知识资源库的真实情况, 结合用户实际的兴趣特点, 以获得适合企业需求的匹配度阈值.

4 实例验证

根据上述提出的基于用户兴趣特征的知识推送方

法,开发了适用于企业电气柜产品的设计知识推送系统.首先采用提出的知识描述方式对设计知识进行统一处理并分类存储,形成新的设计知识管理系统.然后通过个人账户系统建立用户初始兴趣特征集合.

下面以设计某整流模块的垫具为例来验证该方法的可行性.现有知识系统中关于垫具的设计知识数量为10个,系统设计知识数量为200个.首先根据设计知识关键词集合和检索关键词集合确定出现的全部关键词.统计各个关键词被包含的设计知识数量 n_i ,根据公式(1)计算关键词在系统设计知识关键词集合中出现的逆频 idf_i ,表1是垫具关键词逆频 idf_i .

表1 垫具关键词逆频 idf_i

关键词	垫具	零部件	模具	工程图	三维模型	整流模块	逆变模块	控制模块
n_i	10	50	20	60	80	30	25	25
idf_i	1.301	0.602	1	0.5228	0.3979	0.8239	0.9031	1.1249

根据得到的关键词逆频结合公式(2)计算关键词相对词频,表2是垫具设计知识关键词相对词频.

表2 垫具设计知识关键词相对词频

关键词 知识序号	垫具	零部件	模具	工程图	三维模型	整流模块	逆变模块	控制模块
1	1.301	0	1	0	0.3979	0.8239	0	0
2	1.301	0.602	0	0.5228	0	0	0	1.1249
3	1.301	0.602	0	0	0.3979	0.8239	0	0
4	1.301	0	1	0	0.3979	0	0	1.1249
5	1.301	0	1	0.5228	0	0	0.9031	0
6	1.301	0	1	0.5228	0	0.8239	0	0
7	1.301	0	1	0	0.3979	0	0.9031	0
8	1.301	0.602	0	0	0.3979	0	0	1.1249
9	1.301	0.602	0	0.5228	0	0.8239	0	0
10	1.301	0.602	0	0	0.3979	0	0.9031	0

分别以7位兴趣不同的用户来检索垫具.综合需求信息关键词和用户兴趣特征得到检索关键词集合,根据得到的关键词逆频结合公式(3)计算各个用户检索关键词集合中的关键词相对词频.这里 α 和 β 分别取值为0.7和0.3,表3是不同用户检索关键词的相对词频.

表3 不同用户检索关键词相对词频

关键词 用户序号	垫具	零部件	模具	工程图	三维模型	整流模块	逆变模块	控制模块
1	0.7806	0.3612	0.4	0	0	0.4943	0	0
2	0.7806	0.2408	0.6	0	0	0.4943	0	0
3	0.7806	0	0	0.2091	0.2387	0.4943	0	0
4	0.7806	0	0	0.3137	0.1592	0.4943	0	0
5	0.7806	0.3612	0	0	0	0.4943	0.3612	0.4499
6	0.7806	0.3612	0	0	0	0.3296	0.5419	0.4499
7	0.7806	0.3612	0	0	0	0.3296	0.3612	0.6749

根据表2和表3得到的数据结合公式(4)计算检索关键词集合与系统设计知识关键词集合的匹配度.本文采用数值法的方式,通过多组数据多次试验比对发

现,当阈值取0.56~0.58时,获得的知识推送结果较符合企业设计需要,具体试验过程不做详述,这里阈值取0.57,表4是推送给各个用户的知识列表,最后按照匹配度的大小完成知识排序后推送给用户,图5为用户3得到的知识推送结果.

表4 不同用户推送知识列表

用户序号 知识序号	1	2	3	4	5	6	7
1	0.7022	0.7547	0.6391	0.6281	0.5772	—	—
2	—	—	—	—	0.6412	0.6376	0.6995
3	0.7547	0.7022	0.7431	0.726	0.7201	0.67	0.6569
4	—	—	—	—	—	—	0.5772
5	—	0.5953	—	—	—	—	—
6	0.6831	0.7311	0.6281	0.6391	—	—	—
7	0.5793	0.6078	—	—	—	0.5772	—
8	—	—	—	—	0.6569	0.653	0.7201
9	0.7311	0.6831	0.726	0.7431	0.6995	0.6534	0.6412
10	0.6078	0.5793	0.6017	0.5925	0.67	0.7201	0.653

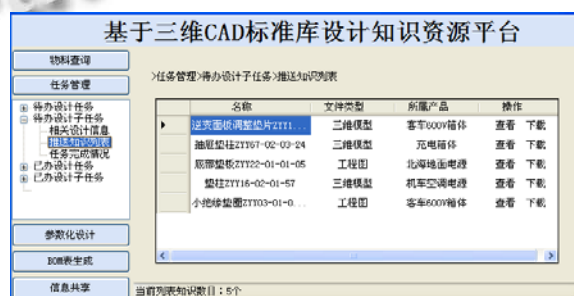


图5 用户3知识推送结果

5 展望

为了避免设计人员的大量检索工作,本文提出了基于用户兴趣特征的知识推送方法.但是本文提出的知识推送方法还不够完善,在检索关键词集合与设计知识关键词集合的匹配算法上需要更深入的研究,从而进一步提高知识推送能力.例如,在进行向量相似性计算过程中,没有考虑到关键词之间的语义关系,导致相似性计算的结果存在一定误差,可以考虑构建同近反义词库来提高相似性计算的精确性,从而提高知识推送性能.

6 结语

设计知识是产品开发过程中的重要资源,知识推送是提高知识获取效率的有效手段.本文在研究了国内外现有知识推送系统的基础上,针对企业电气柜产品设计过程中知识获取方式效率低、匹配得到的设计知识不准确等问题,提出了一种基于用户兴趣特征的设计知识推送方法.通过引入用户兴趣特征,结合用户需求信息可以将更多符合用户需求和兴趣的知识推

送给用户,最后通过实例分析说明了该设计知识推送方法的可行性和可靠性.结果表明,知识推送提高了知识获取效率和准确性,避免了因知识资源库扩大导致的知识利用率不高的问题.

参考文献

- 1 王生发,顾新建,郭剑锋,马军,战洪飞.面向产品设计的知识主动推送研究.计算机集成制造系统,2007,2:234-239.
- 2 柳巧玲.面向企业业务的智能知识推送研究.情报理论与实践,2010,11:76-79.
- 3 乐承毅,代风,吉祥,潘凯,顾新建,彭本红.基于流程驱动的领导知识主动推送研究.计算机集成制造系统,2010,12:2720-2727.
- 4 石美红,王婷,陈永当,毋涛.基于业务过程和知识需求的知识推送系统.计算机集成制造系统,2011,4:882-887.
- 5 蒋翠清,高家飞,李斌生.面向产品设计人员的知识推送服务研究.合肥工业大学学报(自然科学版),2012,3:392-397.
- 6 王占松,田凌,段文睿.基于设计意图建模的知识推送技术.计算机集成制造系统,2015,3:606-617.
- 7 杨洁,杨育,王伟立,赵川.基于粗糙集的产品协同设计知识推送方法研究.中国机械工程,2009,20:2452-2456.
- 8 吉祥,顾新建,代风,乐承毅,徐福缘.基于本体和粗糙集的产品设计知识推送技术.计算机集成制造系统,2013,1:7-20.