

课程知识地图集的研究与应用^①

袁小艳

(四川文理学院 计算机学院, 达州 635000)

摘要: 随着高校教育的发展, 课程和教师增加得越来越多. 为了促进课程和教师知识的共享, 研究了知识地图集的使用. 以知识地图集的理论为基础, 计算了课程间的相似度、教师的隶属度和相似度, 并构建了相应的课程知识地图和教师知识地图. 通过知识地图实现了课程和教师间的相互导航, 并能为用户推荐相应的教师. 最后还把知识地图集应用于查询, 提高了查询的效率. 本文的研究对知识的共享有一定的积极意义.

关键词: 知识地图集; 教师隶属度; 教师相似度; 课程知识; 教师知识

Research and Application of Curriculum Knowledge Atlas

YUAN Xiao-Yan

(College of Computer, Sichuan University of Arts and Science, Dazhou 635000, China)

Abstract: With the development of college education, courses and teachers are becoming increasingly more. To promote the curriculum and the teachers' knowledge sharing, this article studies the use of knowledge atlas. Based on the theory of knowledge map, this paper calculates the similarity of the course, teachers' membership degree and similarity, and to build the curriculum knowledge and teachers knowledge maps. It can realized the mutual navigation between curriculum and teachers through the knowledge map, and it also can recommend the corresponding teachers for users. Finally, we applied knowledge atlas to query, improve the efficiency of the query. The study of this paper has certain positive meaning to the knowledge sharing.

Key words: knowledge atlas; teacher membership degree; teacher similarity; curriculum knowledge; teachers' knowledge

当今是知识经济社会, 组织对知识的管理决定了组织的核心竞争力的优劣. 知识管理不仅仅包含有形资产的管理, 还包含无形资产的管理, 如知识组织、知识设施、知识资产、知识活动、知识人员的全方位和全过程的管理^[1], 并从中发现隐含的知识, 从而扩充知识. 高校是一个知识密集型组织, 它的核心是知识, 更应该进行知识管理. 高校知识管理的核心在于教学, 而教学的核心又在于课程和教师, 所以本文对课程知识和教师知识的管理进行了一定的研究.

知识地图是知识管理的一种有效的、可视化的、对知识进行导航的工具, 它不是知识的集合, 而是知识的一种组织形式, 通过它可以快速、方便地找到需要的知识. 但是随着网络技术的发展, 知识成级数增

长, 单一的知识地图已经很难包含所有的内容了, 这时就可以使用知识地图集. 知识地图集是各种类型知识地图的集合, 通过各种知识之间的联系把知识地图整合起来, 进而实现各种知识的交互导航. 本文通过知识地图集将课程知识和教师知识联系起来, 课程知识和教师知识不再是孤立的, 通过课程可以查询到教师, 通过教师也可以找到相应课程. 关键词库的结合大大提高了信息抽取算法的准确性和通用性, 基于 Web 信息抽取的混合交通出行方案生成与表示系统的成功实验也证明了本文提出的 Web 信息抽取算法的实用性.

1 知识地图集

知识地图集是互相关联的知识地图的集合^[2], 它

^① 基金项目: 国家自然科学基金(61152003); 四川省教育厅一般项目(13ZB0103)

收稿时间: 2014-06-08; 收到修改稿时间: 2014-07-07

不是简单地把知识地图堆在一起,而是建立每个知识地图间的联系,通过这些联系可以交互使用知识地图,进而提高知识地图的价值,相较于知识地图,它更注重于知识间的整体联系.知识地图集不是具体的知识,而是具有联系的知识来源,这些知识来源可以是各种文献和知识库,也可以是具有大量隐性知识的专家,在知识地图集里,不仅能找到显性知识,还可以通过联系得到隐性知识,进而产生新的知识.知识地图集可以包含各种类型的知识地图,如概念图、流程图、专家图、解决问题的方法图等,不拘于一种形式,它是知识地图的一种扩展,通过知识地图之间的联系建立一个统一的整体.从知识地图集的任一入口出发^[2],通过知识的交替使用,都可以找到自己所需要的知识.通过知识地图集,我们可以知道不同活动的流程、发生的原因,可以知道不同领域知识间的关系,还可以知道完成某样工作需要哪些知识.

知识地图集的构建并不是一簇而成的,而是通过知识获取、知识聚类、知识索引、知识地图的形成、知识地图的集成等过程才形成的.

2 课程知识地图集的构建

2.1 课程知识地图构建

课程知识地图的目标是组织课程,建立课程的层次结构,使课程知识更加系统化.课程的关键在于大纲,所以本文研究课程时以课程大纲为研究目标.课程大纲对于我们来说很容易理解,但对于计算机来说却很难理解,所以构建课程知识地图首先应该把课程大纲表示成计算机可以理解的形式,然后再进行课程聚类和课程索引.

2.1.1 课程表示

本文中课程大纲表示先把课程大纲转换为文本文件,然后在其中提取特征词,并为每个特征词赋予权重,计算特征词权重的公式如下:

$$tfreq(w, c) = cf(w, c) * \log\left(\frac{|w|}{wf(c)} + 0.01\right) \quad (1)$$

其中 $cf(w, c)$ 是词语 c 在课程大纲 w 中出现的频率, $wf(c)$ 是指词语 c 在其他大纲出现的频率,从公式中可以看出,一个词语在一个大纲里经常出现,而在其他大纲里很少出现,则认为这个词语有很好的区分大纲的能力.

每个课程大纲的特征词提取出来并赋予权重后,

本文将所有大纲表示成向量空间模型.一个大纲就是一个向量空间,表示为 $V(w) = \{(c_1, d_1(w)), \dots, (c_i, d_i(w)), \dots, (c_n, d_n(w))\}$, 其中 c_i 是特征词, $d_i(w)$ 是它的权值.那么课程大纲间的相似度就可以用余弦距离来表示,公式如下:

$$similar(w_i, w_j) = \cos(w_i, w_j) = \frac{w_i \cdot w_j}{\|w_i\| \cdot \|w_j\|} \quad (2)$$

其中 $w_i \cdot w_j$ 是两个向量的点积, $\|w_i\|$ 是向量的长度.经过这样的计算后,每个课程大纲都用计算机理解的形式表示出来了.

2.1.2 课程聚类

课程大纲能够表示后,就需要对其进行聚类.本文的课程聚类采用 NewMan 的快速聚类算法^[2],将相似课程连接一条边,并赋以相似度为权值,然后将课程划分为不同的组,即集团结构,使得相同集团结构的课程联系紧密,不同集团结构的课程联系疏松.该算法是自底向上的合并类算法,通过计算两个集团合并前后对模块性的影响来选择要合并的集团^[1].模块性定义如下:

$$M = \sum_{i=1}^q (s_{ii} - (\sum_j s_{ij})^2) \quad (3)$$

其中, s_{ii} 是集团 i 内部所有边的权重和在所有集团中所占比例, s_{ij} 是集团 i 和集团 j 边的权重和在所有集团中所占比例, q 是总集团数.两个集团合并到一起时,模块性的变化量如下:

$$\Delta M = s_{ij} - \sum_j s_{ij} \sum_i s_{ij} \quad (4)$$

每次合并时都找模块性增量最多的集团进行合并,直到剩下最后一个集团.具体算法如下:

步骤 1: 每门课程都是一个集团,计算初始 M ;

步骤 2: 合并选择具有最大模块性变化量 ΔM 的两个集团,如果有相同的 ΔM ,就随机选一个;

步骤 3: 如果剩下还有两个及以上的集团,更新因合并集团受到影响的 s_{ij} ,转到步骤 2,否则转到步骤 4;

步骤 4: 输出聚类结果.

2.1.3 课程索引

课程聚类后每个类别都对应对应的课程节点,建立各个课程节点的特征词及其权值,从而选取权值最大的特征词来描述课程节点.课程节点的特征词的权值计算是把该特征词在每一个课程大纲中的权值的和.

通过以上三个步骤就可以把课程进行聚类,并找

到课程间的联系, 形成课程知识地图, 下面以计算机软件技术为例, 如图 1 所示.

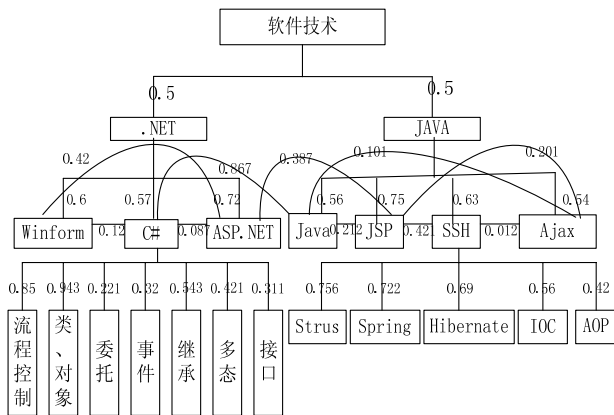


图 1 课程知识地图

2.2 教师知识地图构建

教师在长期的教学实践中, 积累了很多方法和经验, 这些属于隐性知识, 难以表达, 所以本文采用找到了拥有隐性知识的教师就找到了隐性知识, 这就需要建立教师导航的知识地图. 教师导航的前提是要知道教师们被分为多少个不同的教学方向, 每个教师在其所属方向的代表性如何, 以及教师间相互关联程度. 本文中对教师的分析是以教师所授课程的分析为基础的, 通过课程分类可以解决教师分组的问题, 通过课程相似度来计算教师的隶属度和相似度, 从而实现教师的排序和关联, 进而实现教师知识地图.

2.2.1 教师分组

教师分组就是把教师划分到相应的教学方向, 教师能上哪些方向的课程, 那么该教师的教学方向就相应的有哪些. 如图 2 所示, 小圆圈表示课程, 大圆圈表示教学方向, 在图中, 很容易看出每位教师所属的教学方向.

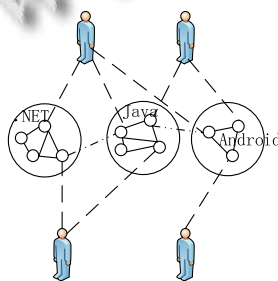


图 2 教师划分

2.2.2 教师排序

在教师层次结构中, 每位教师应该按隶属度的高低进行排序. 隶属度是衡量教师属于某个教学方向的程度, 隶属度越大, 教师在该方向的代表性越强, 顺序越在前面. 用户可以根据教师知识地图的推荐教师顺序, 与相应教师取得联系以获得帮助. 本文根据教师在某个教学方向所讲授课程的情况来确定教师的隶属度, 从两个方面进行分析:

① 在同一个教学方向中, 如果某位教师讲授的课程和其他课程的相似度高, 则该教师对该方向的隶属度越高;

② 当某位教师属于多个教学方向时, 教师在哪个方向上所上课程的比例高, 则该教师对该方向的隶属度比其他方向高.

因此教师 T_i 共上了 K_i 门课程, 他属于教学方向 D_j 的隶属度为:

$$G_{mem}(T_i, D_j) = \frac{\sum_{b=1}^{C_{i,j}} \sum_{d=C_{i,j}+1}^{N_j} Similar(h_b, h_d)}{C_{i,j} \times (N_j - C_{i,j})} \times \frac{C_{i,j}}{K_i} \quad (5)$$

$$= \frac{\sum_{b=1}^{C_{i,j}} \sum_{d=C_{i,j}+1}^{N_j} Similar(h_b, h_d)}{K_i \times (N_j - C_{i,j})}$$

其中 $C_{i,j}$ 是教师 T_i 在方向 D_j 中所上的课程数量, 方向 D_j 共有 N_j 门课程 (h), $Similar(h_b, h_d)$ 是两门课程的相似度, 课程的相似度采用课程大纲的相似度来计算.

2.2.3 教师关联

本文采用教师相似度来衡量教师间的关联程度, 相似度越高, 教师的教学方向越相似, 教师的关联度就越高. 教师的相似度采用所上课程的相似度来计算, 算法如下:

步骤 1. 整理每位教师所上课程;

步骤 2. 计算每个教学方向中教师间的相似度. 由于教师的相似度采用课程大纲的相似度衡量, 所以只需要把某教师每门课程的教学大纲和另一个教师每门课程的教学大纲的相似度取平均值就可以了. 教师 T_i 与教师 T_j 在课程节点 D_k 的相似度为:

$$Similar_k(T_i, T_j) = \frac{\sum_{p=1}^{C_i} \sum_{q=1}^{C_j} Similar(h_{i,p}, h_{j,q})}{C_i \times C_j} \quad (6)$$

其中 C_i 是教师 T_i 所上课程的数量, C_j 是教师 T_j 所上课程的数量, 是教师 T_i 和 T_j 的课程相似度.

步骤 3. 采用 D 个矩阵 T 来表示每个教学方向所有教师的相似度. 假设教师的总数为 M, 在矩阵中, 若某位教师不在某个知识节点下, 则他和其他教师的相似度为 0, 教师和自身的相似度是 1. 矩阵如下所示:

D _k	P ₁	P ₂	P ₃	...	P _M
P ₁	1	p ₁₂	p ₁₃	...	p _{1M}
P ₂		1	p ₂₃	...	p _{2M}
P ₃			1	...	p _{3M}
...				...	p _{4M}
P _M					1

步骤 4. 计算教师在所有教学方向的总体相似度. 将所有矩阵中各元素相加, 从而得到一个 M*M 型的矩阵 F_{M*M}, 矩阵中每个元素就是教师间的总体相似度, 公式如下:

$$F_{M * M} = \sum_{n=1}^U T_n \tag{7}$$

步骤 5. 把教师每个方向的相似度和总体相似度标准化. 标准化函数为:

$$F_x = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{8}$$

标准化方向的相似度时, x 是第 k 个矩阵的元素值, 标准化总体相似度时, x 是矩阵 F 中的元素值. 通过标准化后的矩阵就可以找到每个教师最相关的教师了. 通过前面计算出的教师的隶属度和相似度就可以形成教师知识地图了. 下面同样以计算机软件技术为例, 如图 3 所示.

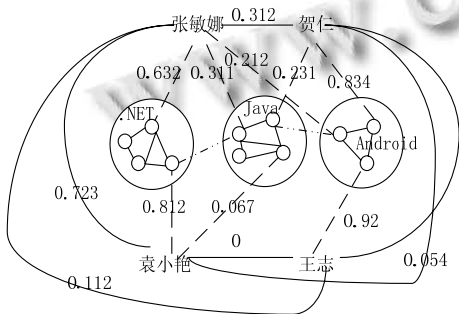


图 3 教师知识地图

2.3 知识地图的集成

到目前为止, 课程知识地图和教师知识地图的构建已经完成, 接下来的工作就是给两个知识地图建立

联系, 只有相互有联系的知识地图的集合才可以被称为知识地图集. 课程知识地图与教师知识地图间的联系很容易建立, 课程属于教师, 教师讲授课程, 它们之间的这种联系在教师讲授课程时就已经产生, 所以只要找到两者中的一个, 就可以通过所属联系找到另外一个, 进而实现课程和教师的相互导航.

3 基于知识地图集的查询

一般的关键字查询算法只能查找确定性的知识, 像“C#课程 教师”或者“袁小艳教师 课程”这种关联式的查询却无法做. 知识地图集将各种知识联系在了一起, 恰好能够解决这种关联式的查询, 因此本文提出了这种基于知识地图集的查询算法.

3.1 基于知识地图集的查询扩展

在上一节中, 已经为课程知识和教师知识定义了一整套权值方案, 根据这些权值就可以把权值高的知识加入到查询关键字中, 比如用户搜索“C#课程 教师”, 通过查询扩展可以找到“袁小艳”, 那么通过“袁小艳”扩展得到的查询结果可以更符合用户的要求, 而原始查询中的索引词有可能没有在结果文档中. 查询扩展的算法如下:

```

For all qei ≠ 0 in QE
If FREj ≠ Φ
Remove qei from QE
QE = QE + δqeiFREj
EndIf
EndFor

```

其中, QE 是查询向量, QE=(qe₁,qe₂,...,qe_n), qe_i 是查询的索引词 r_i 对应的权值, 初始值为 1, r_i 是查询关键词 T=(t₁,t₂,...,t_n) 的第 i 个索引词, T 对于的相应概念是 N=(n₁,n₂,...,n_m), 每个概念 n_j 都关联了一个特征向量 FRE_j=(fre_{j1},fre_{j2},...,fre_{jm}), fre_{j_i} 是索引词 r_i 对应的权值, δ 是调节因子.

假设查询条目是“C#设计”, 转化为查询向量 QE=(1.0C#, 1.0 设计), 通过知识地图查询, 发现“C#”是其中定义的知识, 需要扩展, 而“设计”是一般的词, 不需要扩展. 通过前面所述的查询扩展算法和课程知识地图-图 1, 可以把查询向量扩展为 QE=(0.6C#, 0.51 流程控制, 0.58 类和对象, 0.13 委托, 0.19 事件, 0.32 继承, 0.25 多态, 0.19 接口, 0.52Java, 0.07Winform, 0.05ASP.NET, 1.0 设计).

从QE可知,词语“设计”有较高的权值,“C#”的权值较低,这样包含“设计”的词语并且和“C#”相关的知识会有较高的排序分数。

3.2 基于知识地图集的查询算法

基于知识地图集的查询算法步骤如下:

步骤1. 采用基本查询算法找到初始知识集,并存入优先队列;

步骤2. 搜索节点*i*的相邻节点*j*,并计算节点*j*的输入值,公式如下:

$$IN(j) = \sum_i OU(i)wn_{ij} \quad (9)$$

其中OU(*i*)是节点*i*的输出,OU(*i*)=f(IN(*i*)),一般采用线性函数,wn_{*ij*}是知识*i*和知识*j*的关联权值;

步骤3. 如果知识*j*不在优先队列里,则将知识*j*加入到优先队列,否则更新知识*j*的输入值,同时把*j*加入到结果队列,并更新相应的权值;

步骤4. 不满足算法的终止条件,返回步骤2. 算法的终止条件有三个:当前已经没有相关知识,设定数值的知识低于出度数,知识点的距离高于设定的阈值;

步骤5. 满足算法的终止条件,得到最终的结果知识集,并按照对应的输入值进行排序。

4 实验比较

为了衡量基于知识地图集的查询算法的性能,本文需要计算该算法的查准率和查全率,这就需要采用常见的信息检索系统评价方法来计算,并且还需要和传统的基于关键词的查询进行比较. 实验中采用某高校的课程教师数据库系统进行查询,其比较如图4所示。

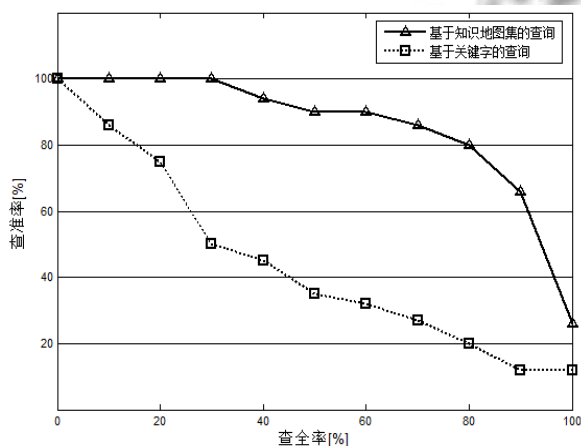


图4 两种算法比较

从图4中可知,本文所提出的查询算法比传统的基于关键词的查询方法的查全率和查准率都要好得多。

5 结语

随着高校的发展,专业越来越多,课程也越来越多,教师也增加得快,所以课程管理越来越受到高校的关注和重视. 课程管理不能仅仅局限于数据和信息的管理,应该上升到知识管理的程度,挖掘各种课程和教师的知识,进而实现各种课程和教师知识的共享. 本文采用知识地图集,把课程知识和教师知识有效地组织起来,实现了课程和教师知识的导航,并在其基础上实现了查询,有一定的借鉴意义。

参考文献

- 唐平.基于知识管理的高校教学管理系统[学位论文].重庆:重庆大学,2004.
- 刘芳莉.科研成果管理中知识地图集的研究[学位论文].大连:大连理工大学,2007.
- 杨晶晶.电子学习中的知识地图构建研究[学位论文].临汾:山西师范大学,2013.
- 张凌,乔晓东.基于组织认知地图的隐性知识共享的实现.情报理论与实践,2013,36(7):81-84.
- 张凌.基于认知地图的隐性知识表达与共享—广义认知地图与狭义认知地图.情报理论与实践,2012,(2):72-76.
- 聂建国.图书馆专家知识地图构建与应用.曲靖师范学院学报,2013,32(3):113-116.
- 蒋霞,蒋文波,马军华.知识管理在教育技术学领域中的应用.软件导刊·教育技术,2011,(9):71-73.
- 赵京,徐少同.知识地图的关键技术与典型应用.情报理论与实践,2012,12(35):101-105.
- 蔡荣啸,高强,陈刚.云计算视野下构建个性化的知识地图.中国教育技术装备,2013,(15):7-9.
- 马青云.医药文献的知识地图构建研究与应用[学位论文].广州:广东工业大学,2013.
- 李伟,陈毓芬.基于用户的空间知识地图可视化探析.测绘科学技术学报,2013,30(6):638-642.
- 熊奇.基于知识地图的知识检索与推进方法研究[学位论文].上海:上海交通大学,2009.