

基于领域本体的语义查询扩展^①

胡川洌¹, 符云清^{1,2}, 钟明洋¹

¹(重庆大学 软件工程学院, 重庆 400044)

²(重庆大学 计算机学院, 重庆 400044)

摘 要: 在信息检索研究领域, 资源与查询词的匹配决定信息检索质量。现有检索方法的检索结果存在过多不相关信息, 不能很好满足用户需求。针对传统信息检索存在的问题与当前语义查询扩展方法的特点, 本文在分析各种语义查询扩展方法及其相关研究的基础上, 提出一种改进的基于领域本体的语义查询扩展方法。该方法论通过本体模型和概念相似度的计算对检索信息进行检索意图树的构建并扩展; 然后在资源本体中以最短路径的方式搜索资源。实验结果表明, 本文方法相较其他查询扩展方法能得到更好的检索结果。

关键词: 领域本体; 应用本体; 查询扩展; 概念相似度

Semantic Query Expansion Based on Domain Ontology

HU Chuan-Lie¹, FU Yun-Qing^{1,2}, ZHONG Ming-Yang¹

¹(College of Software Engineering, Chongqing University, Chongqing 400044, China)

²(College of Computer Science, Chongqing University, Chongqing 400044, China)

Abstract: In the field of information retrieval, the match between resources and query words determines retrieval quality. The search results using current query methods exist too much irrelevant information and cannot satisfy customer needs. Considering the defects of traditional information retrieval and current feature of semantic query expansion, an improved ontology-based semantic query expansion method based on the analysis of various query expansion algorithms and related research has been proposed. The method constructs and expands the user search tree according to the query message basing on the ontology model and concept similarity computation. The experiment results show that the method can get better query results comparing to other query expansion method.

Key words: domain ontology, application ontology, query expansion, concept similarity

1 引言

自 Tim Berners-Lee^[1]等人提出语义网的概念之后, 语义已成为 Web2.0 走向 Web3.0 过程中的研究热点, 对语义和本体的研究与应用也逐步延伸到各个领域。检索是语义应用中, 用户与数据交互的重要环节。目前的 Web2.0 在用户的参与下, 网络的数据呈现高数量级的增长, 且数据多而无序, 用户检索得到的数据很大部分是无用甚至错误的, 这并不满足检索的查准性要求。出现这种情况的主要原因是网络上的资源缺乏语义信息, 不能被计算机所理解。传统的检索是以关键字匹配为基础的检索技术, 无法从语义的层面理

解用户输入的检索词所表达的含义。在此契机下, 学者将查询扩展^[2]引入到信息检索中, 即通过有效方法对关键词进行扩展得到比初始关键词更丰富的词汇, 避免同义词、近义词和缺词的情况, 再送入搜索引擎中进行检索, 从而得到满足用户检索要求的结果。

针对传统检索方式存在的不足并结合当前查询扩展方法的相关研究成果, 提出了一种改进的基于领域本体的语义查询扩展方法。该方法是在领域本体的指导下对用户检索请求进行扩展, 并运用到 E-Learning 环境下的资源检索, 构建以课程知识作为领域本体的教育资源数据检索模型。最后通过实验, 结果表明改

① 基金项目: 重庆市自然科学基金(CSTC, 2010BB2248); 中央高校基本科研业务费科研专项(CDJZR10090002)

收稿时间: 2011-10-18; 收到修改稿时间: 2011-11-16

进的检索模型能克服传统检索方式的不足，提高检索的查全率和查准率。

2 查询扩展研究综述

查询扩展是指利用计算机语言学、信息学等多种技术，把与原查询相关的语词添加到原查询，得到比原查询更长的新查询，然后检索文档，以改善信息检索的查全率和查准率，解决长期困扰信息检索领域的词不匹配问题，弥补用户查询信息不足的缺陷。

目前，常用的查询扩展方法包括：全局分析法^[3]、局部分析法(LCA)^[4]和基于语义知识的方法。全局分析法是对全部文档中的词或词组进行相关分析，计算每对词或词组间的关联程度，根据预先计算的词间相关关系将与查询用词关联程度最高的词或词组加入原查询以生成新的查询。例如，对全部文档进行分析，建立检索词同文档的关联矩阵T，再构造词的关联矩阵 $S=T \cdot T^T$ (T的转置)，文档集出现的检索词与查询信息相似性用权重值的和进行衡量，最后按权值降序排列，考前的检索词作为扩展词。局部分析法(LCA)是为克服全局分析法中构建全局词汇表的巨大工作量问题而提出的，该方法利用初次检索得到的最相关的n篇文档作为查询扩展来源，但扩展效果高度依赖于初次检索的结果及文档内容的词量大小。语义查询扩展是将本体中所反映出的语义关系应用到检索信息的扩展中^[5]，利用概念间语义关系消除语义偏差，通过概念的相似度计算确立可扩展的语义边界。

在本体的指导下，为信息融入了语义，因此概念相似度计算是进行语义扩展的重要步骤，其精度也是提高检索质量的关键。但目前分词及普通词汇的相似度计算仍依赖于《知网》^[6]。基于距离的语义相似度计算模型、基于内容的语义相似度计算模型和基于属性的语义相似度计算模型是三种主要的用于计算领域本体的概念相似度的模型，针对这些方面，国内外学者也有了大量的研究工作^[7,8]。

查全率(Recall)、查准率(Precision)及度量值(F-measure)是评价检索结果优劣的三个度量指标。上述扩展方法和相似度计算模型及算法都致力于在语义数据检索中得到较好结果。但在利用领域本体时，却没有将领域与领域之下构建的应用分开理解，在计算相似度时大多是针对上下位关系，而忽略了其他关系及应用程序层面的本体建立。

3 基于领域本体的语义查询扩展

3.1 检索框架组成

基于领域本体模型的资源检索框架如图1所示，根据用户的检索意向在知识本体中去掉毫无相关的概念并保留与检索条件相关的概念以达到较高的查准率和查全率。由于匹配到的概念可能出现于知识本体中各个节点(如一个词在多个概念中出现)，因此需要在匹配到的概念中寻找一棵最能表达用户检索意图的子树以避免在被过多的知识点索引的资源中进行应用本体下的检索，这既可减小检索过程的时间复杂度又能带来一定程度上的性能提高。根据子树及原始检索信息精确化检索信息，最后在应用本体中采用基于关键字的语义数据检索算法获取满足检索要求的资源。

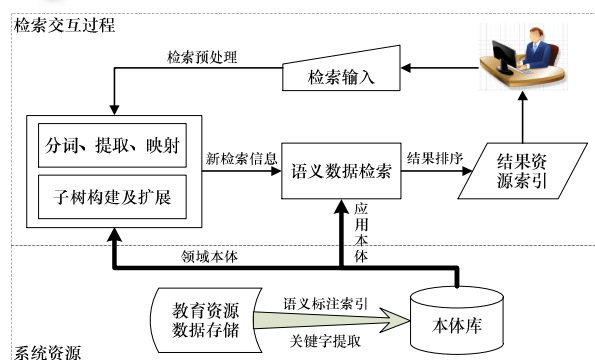


图1 检索模型框架

3.2 准备工作

该过程是在执行检索算法之前所需要做的准备工作，包括对教育资源的标注、索引及专业词汇词典的构建。

在教育资源应用本体的指导下可对具体教育资源扩充更多、更全的描述，尤其是对非文本资源如视频、PPT等。为提高查准率，要对资源的内容进行标注而非仅关键字提取，这需要某种语义标注算法^[9]，对非文本对象可以利用教育资源本体中对该资源的描述进行标注。对标注后的资源需将其标注结果同所属的知识点关联索引并存储起来。

在标注及检索中都会用到分词工具，但面向特定领域的专业词汇一般会比较复杂且难以构建义原、义项，若没有一个关于领域词汇的词典而只使用“知网”分词会对标注或检索结果造成影响，如对“马尔科夫链约瑟夫环 AOV-网”分词结果会是“马尔科 | 夫 | 链 | 约瑟夫 | 环 | AOV- | 网”。这样的分词结果很难映射

到领域本体中，在应用本体中进行检索时会使原始词汇的语义发生改变。

3.3 改进的查询扩展方法

本文的查询扩展方法主要是针对在领域本体中对检索信息精确化并扩展，在应用本体中的语义数据检索则采用基于关键字的语义数据检索算法^[10]。算法流程如图 2 所示。

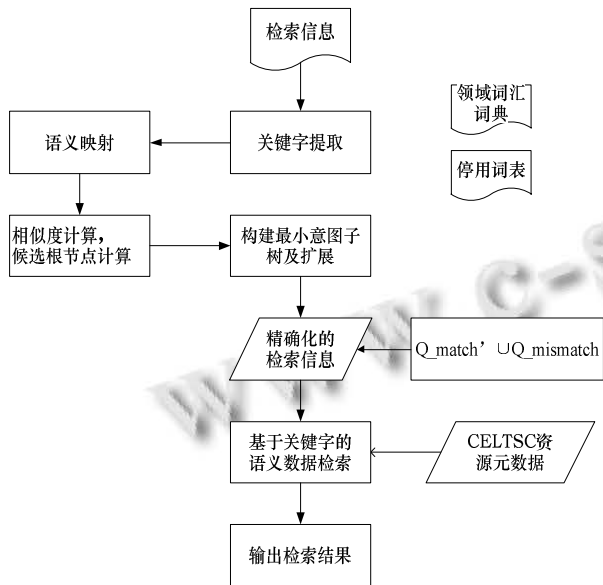


图 2 检索算法流程

该检索算法主要包含六个步骤：

- Step1. 关键字提取；
- Step2. 概念映射；
- Step3. 构建表达用户检索意图子树；
- Step4. 扩展子树；
- Step5. 关键词合并；
- Step6. 基于关键字的语义数据检索。

本文在步骤 2 至步骤 5 进行了改进，融入了概念相似度的计算方法，并将概念在检索信息中的比重作为检索意图的标量加入相似度计算模型中，更能体现用户的检索目的。

Step1 关键字提取

该步骤对用户输入的查询信息进行分词，使用停用词表去掉多余词汇及符号，形成有检索意义的关键词集合，并对集合中的每一个词汇进行同义扩展，得到扩展的查询关键字集合， $Q=\{q_1, q_2 \dots q_i\}$ 。同理，对领域本体中的概念或具有语义的属性值，也需要同义扩展以获取最高概率的匹配，扩展后得到概念词集合

$K=\{k_{1c1}, k_{2c1} \dots k_{ic1}, k_{1c2}, k_{2c2} \dots k_{jc2}, \dots, k_{1cx} \dots k_{mex}\}=\{K_{c1}, K_{c2} \dots K_{cx}\}$ 。

定义 1. 基概念： $\exists c_i \in C$ 且 $K_{c_i} \cap Q \neq \emptyset$ ，则 c_i 为基概念。

如果某个概念及其同义描述在检索关键字集合中存在，或者通过“知网”相似度计算且计算结果大于预定义阈值，则将此概念定义为基概念。这样原始的查询关键字会拆分为两个子集： Q_{match} 和 $Q_{mismatch}$ 。

Step2 概念映射

定义 2. 概念占有率 $O(c)$ ：设概念 c_i 及同义扩展后的词汇集合的元素个数为 α ，在集合 K_{c_i} 中词汇与集合 Q_{match} 中词汇匹配到的词汇个数为 β ，则：

$$O(c_i) = \frac{\beta}{\alpha} = \frac{|K_{c_i} \cap Q_{match}|}{|K_{c_i}|} \quad (1)$$

该公式对于从检索关键字到本体概念的映射过程中可以很直观的表现用户检索的意图。基于用户认知的角度，当 O_c 的值越大 ($O_c \in [0,1]$)，即概念 c 在检索关键词中所占的比重越大，说明用户检索该概念或其子概念下的数据的可能性就越大，因此 O_c 能从一定程度上反应用户的检索意图。

概念映射的过程通过定义 1 与定义 2 可以使查询关键字集合映射到本体的概念中，命中的概念分散在本体树中各个节点。例如，如图 3 所示的领域本体，假设检索条件分词后结果为 $Q=\{b2, C4, D6, C1, d2, c2\}$ ，为方便说明，此处暂不考虑不能映射的词汇集合。经过步骤 2 后将用户检索信息映射到图中灰色标记的概念中。

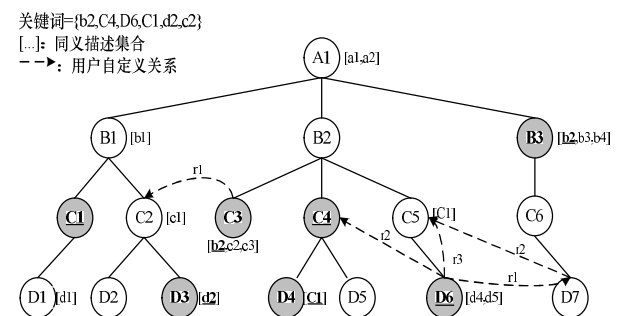


图 3 概念映射

Step3 构建用户检索意图子树

由于有些词可能在不同概念中出现，表达不同语义或语境，依据用户检索习惯，检索目的一般偏向于

某一方面（例如用户不会去查询两个知识跨度比较大但是在知识描述上会出现相同或者相似词汇的知识点，例如图 3 中概念 C1、D4，因此可以构造一棵最能代表用户查询要求的子树以用于表示用户检索意图的最小概念集合。

定义 3. 概念节点的深度和宽度：基于树的概念，记概念节点 c_i 的深度记为 $Dep(c_i)=Dep(Parent(c_i))+1$ ，根节点深度为 1；概念节点 c_i 的宽度表示为 $Wid(c_i)$ ，即概念 c_i 的分支数。

定义 4. 概念权重 $\omega(c_i)$ ：概念节点的深度和宽度在本体树中都与其在树中的权重有影响，权重 $\omega(c_i)$ 可由下式计算：

$$\omega(c_i) = \frac{\omega(Parent(c_i))}{Dep(c_i) + Wid(c_i)} \quad (2)$$

特别的，规定 $\omega(\text{root}) = 1$ 。利用概念权重可以计算出两个概念间的语义距离 $D(c_i, c_j)$ ，即在本体树中连通两个概念的最短路径上所有概念的权重之和，其计算公式为：

$$D(c_i, c_j) = \sum_{m=0}^n \omega_m \quad (3)$$

定义 5. 概念节点的度 $Deg(c_i)$ ：表示该概念的子节点个数。若 c_i 为叶子节点，则 $Deg(c_i)=0$ 。

在计算概念语义相似度时仅考虑层次关系并不全面，对象属性对于概念之间的相似度也有着同样重要的影响，在此引入关系相似度的计算公式，设本体中用户自定义对象属性集合为 $R=(r_1, r_2 \dots r_n) \subseteq \mathbb{P}$ ，定义概念 c_i 和 c_j 在关系 r 上的关系值 $RV_r(c_i, c_j)$ ：

$$RV_r(c_i, c_j) = \begin{cases} 1, & \text{从 } c_i \text{ 到 } c_j \text{ 存在自定义关系 } r \\ 0, & \text{从 } c_i \text{ 到 } c_j \text{ 不存在自定义关系 } r \end{cases}$$

由于考虑到关系的方向性（由 OWL 描述属性控制），因此 $RV_r(c_i, c_j)$ 与 $RV_r(c_j, c_i)$ 不一定相等，根据关系的定义可用如下公式计算概念 c_i 、 c_j 的关系相似度 $Sim_r(c_i, c_j)$ ：

$$Sim_r(c_i, c_j) = \frac{\sum_{k=1}^n RV_{r_k}(c_i, c_j)}{m} \quad (4)$$

在公(4)中， n 为集合 R 元素个数， m 为概念 c_i 与其它概念存在自定义关系的且不重复的关系的数量。不同领域不同本体中的不同关系对于概念相似的影响程度及权重不同，为了使公式独立于应用，避免对每

种关系设立权重值，上述对概念间的关系相似度计算公式做了简单处理。

综合公式(3)、(4)，概念 c_i 、 c_j 的相似度 $Sim(c_i, c_j)$ 的计算公式为：

$$Sim(c_i, c_j) = \lambda \left(\frac{\alpha}{\alpha + D(c_i, c_j)} \right) + (1-\lambda) Sim_r(c_i, c_j) \quad (5)$$

公式(5)中，调节因子 α 与 λ 是针对具体应用设置的参数，因为概念相似度是基于个人认知的具有强主观性概念。

构造最能代表用户查询要求的子树其目的就是要使这棵树能最大化的包含基概念，而构造的关键点就是要找到树的根节点，作为根节点的概念不能是定义 1 中的基概念，否则可能出现很多节点都可以作为根节点的情况，在本体层次结构中，节点 c 与其直接父节点也是语义相关的，所以子节点也会反过来影响父节点，要确定根节点，需要利用子树中的基概念对根节点的影响作为考察标准并将其量化，以选出所有候选根节点集合中的最优节点。

利用概念相似度的计算方法及概念占有率对检索意图的意义，可以给基概念 c_i 设置一个基概念值 ρ 作为量化对根节点影响的标量。对基概念集合 M 中的概念 c_i ，有：

$$\begin{aligned} \rho(c_i) &= O(c_i) + \sum_{j=1 \cup j \neq i} Sim(c_i, c_j) \times O(c_i) \times O(c_j) \\ &= O(c_i) \times [1 + \sum_{j=1 \cup j \neq i} Sim(c_i, c_j) \times O(c_j)] \end{aligned} \quad (6)$$

$c_i, c_j \in$ 基概念且 c_i 到 c_j 存在关系 $r, r \in R$ 。

以图 4 为例，公式(5)中调节因子 $\lambda=0.5, \alpha=1$ ，概念 D6 关联到其他概念的自定义关系的数量 $|R|=3$ ，权重 $\omega(D6)=1/80, O(D6)=1/3$ ，概念 D6 只与基概念集合中的概念 C4 具有关联关系 r_2 ，利用公式(5)可得到 $\rho(D6)=O(D6) \times (1 + Sim(D6, C4) \times O(C4)) \approx 0.51685$ 。

在公式(6)的基础上，可以构造子节点 c_i 对父节点 c_p 的影响值 $I(c_p)$ 的计算公式：

$$I(c_p) = \left\{ \sum \rho(c_i) \times \frac{Deg(c_i) + 1}{Deg(c_p) + 1} \mid \begin{matrix} c_i \in \text{基概念} \\ c_p \notin \text{基概念且 } Parent(c_i)=c_p \end{matrix} \right\} \quad (7)$$

公式(7)分子分母各加 1 是为了避免叶子节点的度为 0 而出现计算错误。如图 5 所示，节点 A1, B1, B2, C2, C5 的概念占有率 $O(c)$ 均为 0，其直接后继都包含有基概念，因此这些节点都是备选根节点，因此要分别计算各个备选节点的影响值 $I(c_p)$ ，例如，其余

备选根节点均可按公式(7)进行计算 $I(c_p)$ 值的计算。

$$I(B2)=IF_{B2}(C3)+IF_{B2}(C4)+IF_{B2}(D4)+IF_{B2}(D6)=\rho(C3)*\frac{0+1}{16+1}+\rho(C4)*\frac{2+1}{16+1}+\rho(D4)*\frac{0+1}{16+1}+\rho(D6)*\frac{0+1}{16+1}\approx 0.64527.$$

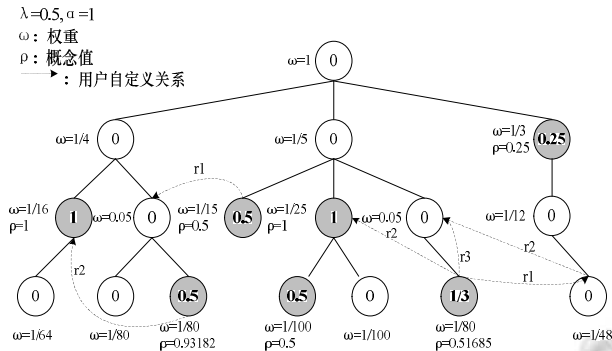


图 4 权重及基概念值的计算

IF: 影响因子
I: 影响值
: 用户检索树

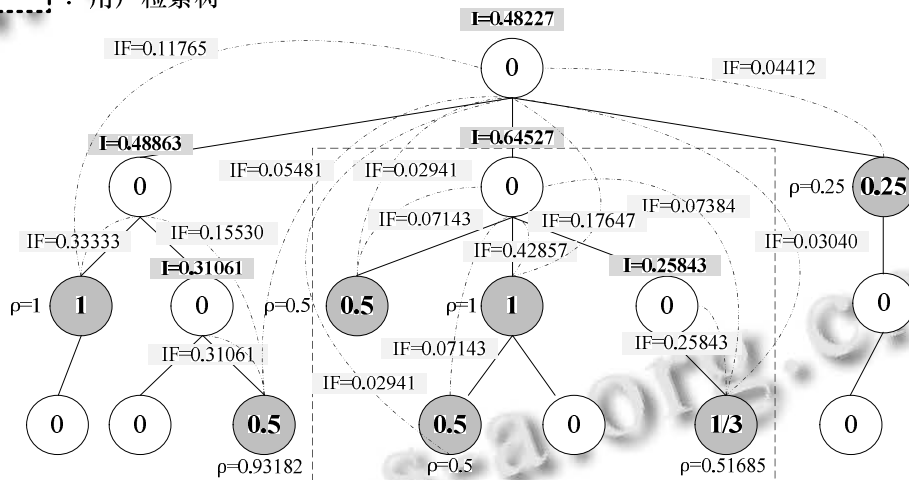


图 5 用户检索树的构建

Step4 扩展子树

由于用户检索可能存在盲目性，输入的检索信息会和他的真正意图相关但不明确的信息，而构建的检索树仅根据查询输入的产生的，这样就有可能得到错误的查询结果，即使输入的查询目的明确，给出与检索目的相关的信息也可能是用户关心的。因此，用户检索意图形式化后，可以扩展检索树以获取更多相关信息。在本体树深较大情况下，构建出的检索树可能也是一棵深度较大的树，因此可以根据具体应用环境作一定层次的扩展，例如，本文实验约定扩展层次为 2(Expand Level = 2，即在领域本体层次树中距离为

从基于认知的角度看，公式(7)所表达的含义是如果由某个节点及其子节点构成的子树中所包含的与用户检索意图相关的概念越多，那么这些概念对根节点的影响就越大，并且越贴近检索目的。当某两个知识点中存在相同或相似词汇但该词汇表达了不同的语义，即有歧义存在时，利用构建子树的方法就能在一定程度上消除歧义。如概念 B3 与 C3 都有相同的同义描述 b2。通过相似度计算和构建用户检索树的方法就可排除 b2 的歧义，得到符合用户检索要求的语义。

得到所有备选父节点的影响值后，需要遍历备选根节点，选出拥有最大影响值的节点作为最终根节点，以该节点为根的子树也就成为最能表达用户检索意图的检索树。

2)，首先是由于本文使用的领域本体深度较小，其次是考虑到资源的类型比较多，诸如视频、语音类的资源可能在一个资源中就包含的很多的知识点而不得不将其索引在层次比较低的节点中。可以利用公式(5)计算相似度并设置阈值以寻找相关知识点。

Step5 关键词合并

通过步骤 4 后，检索关键词集合 Q_{match} 被极大的简化和精确化，不在子树中的词汇和有歧义的词汇被去掉，在子树中的其他的相关概念词汇被加入进来，从而得到新的关键字集合 Q_{match}' 但集合 $Q_{mismatch}$ 中无法进行概念映射的关键字却不能做任何处理，也许该

集中的某个元素只是作者的名字，或是某篇文档的日期等等，这些信息虽与领域无关但具备语义信息且对检索结果也有很重要的影响，所以需要应用程序本体的帮助，在其上进行检索。将精确化后的关键字集合 Q_{match} 与在关键词提取阶段不能匹配的关键字集合 $Q_{mismatch}$ 合并，得到最终关键词集合 Q' 。

Step6 基于关键字的语义数据检索

在检索范围被缩小，得到最终关键词集合 Q' 后，根据教育资源应用本体与领域本体的关系，可在教育资源应用本体中检索相关资源，而此部分采用检索算法是文献[10]提出的语义数据检索算法，该算法利用最短路径的思想，在语义层面上进行检索，使检索结果具有明确的语义，表现了查询关键字间的语义关联性以及关联程度，且该算法的时间复杂度为多项式时间。

4 实验与分析

对本文所提出的检索方法应用于教育资源的检索，其检索质量通过查全率(recall)、查准率(precision)及度量值(F-measure)来衡量，并同传统基于关键字的检索相比较。

$$\text{查全率} = \frac{\text{检索到的相关资源的数量}}{\text{所有相关资源的数量}} * 100\%$$

$$\text{查准率} = \frac{\text{检索到的相关资源的数量}}{\text{检索到的所有资源的数量}} * 100\%$$

$$\text{度量值} = \frac{(1+\phi) * \text{查准率} * \text{查全率}}{\phi * \text{查准率} + \text{查全率}} * 100\%$$

度量值是查全率和查准率的加权调和平均数，用于综合评价系统的检索质量，调节因子 ϕ 可根据系统对查全率、查准率的不同侧重而调节。本文侧重于查全率，因此在实验中取调节因子 $\phi=2$ 。

实验使用前文提到的 ACM CC2005 教程体系中的课程“数据结构”作为领域本体。数据集由文档、视频、PPT 素材构成，其中文档集合选自多个版本的数据结构教程和网络相关话题网页共计 788 篇，视频和 PPT 资源共计 63 个，每个知识点下的资源分配比较平均，以便检索提问可较全面的覆盖领域本体中知识。由于对资源的标注没有相应的自动标注算法，因此实验前进行了手动标记。实验中拟采用的查询类别包括：

- 1) 数据结构的基本概念；
- 2) 线性表的实现；
- 3) 树；

- 4) 图的搜索；
- 5) 内部排序。

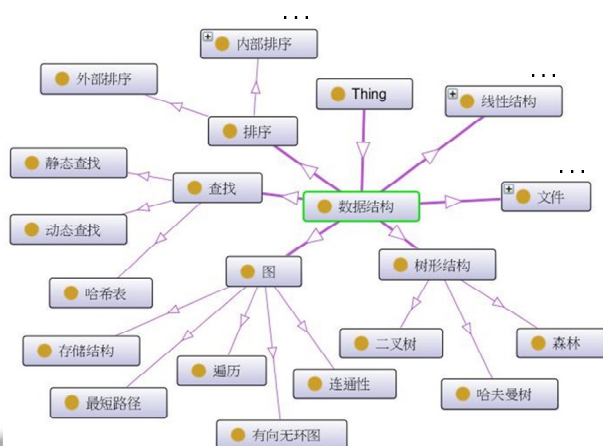


图6 “数据结构”课程领域本体部分概念层次结构图

表 1 是各个检索类别所使用的提问信息 S。表 2 是构建用户检索树后形成的查询关键字集合。实验将用基于关键字的全文检索方法(Method1)及现有基于本体的 LCA 查询扩展方法(Method2)同本文方法(Method3)进行对比，其结果如图 7、8、9 所示。

表 1 各检索类别的原始提问信息

类别	提问信息
(1)	{数据结构的基本概念和存储空间}
(2)	{链表的实现方式}
(3)	{二叉树的遍历}
(4)	{图的深度优先搜索实现}
(5)	{快速排序的时间复杂度}

表 2 构建用户检索树后形成的查询关键字集合

类别	提问信息
(1)	{数据结构,基本,概念,存储空间}
(2)	{线性表,链表,实现,方式,线性链表,循环链表,双向链表}
(3)	{树,二叉树,遍历,线索二叉树}
(4)	{图,遍历,深度优先,广度优先,搜索,实现}
(5)	{排序,内部排序,快速排序,时间复杂度}

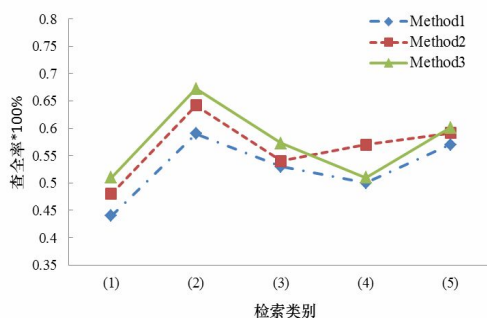


图7 查全率对比图

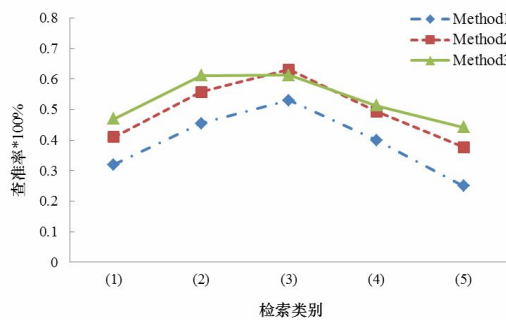


图 8 查准率对比图

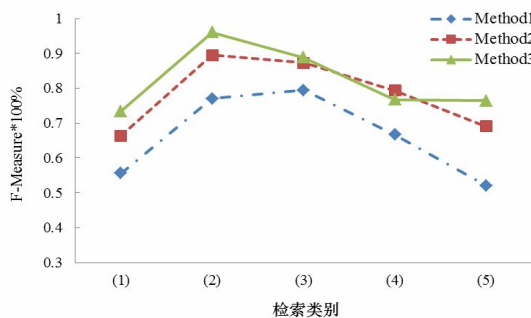


图 9 度量值(F-Measure)对比图

如图 7、8、9 中可以看出, 基于本体的 LCA 查询扩展方法和本文的查询扩展方法不论查全率还是查准率均优于基于关键字的全文检索方法, 因为关键字检索会得到太多不相关信息。而本文方法由于对用户检索意图进行了计算并扩展, 因此查准率比方面总体上比基于本体的 LCA 查询扩展方法高。也由于对用户检索意图进行推测, 导致可能丢弃少部分基概念, 所以第四个查询的查全率并不比基于本体的 LCA 查询扩展方法高, 但仍优于全文关键字检索。然而从应用本体本身来讲, 进行对比的两种方法都只能局限于纯文本对象的检索, 而本文的目标是在多类型、多格式的教育资源中完成检索, 这是另外两种方法是无法完成, 因此本文方法的应用面更为广泛。

5 结语

本文针对传统信息检索模型缺乏语义以及现有语义检索模型的不足, 提出了改进的基于领域本体的资源检索方法。从用户的检索意图的角度进行检索信息的精炼与扩展, 提高检索提问的质量, 进而对领域本体的概念相似度的计算方法和现有应用本体层的语义数据检索方法开展了研究。最后通过实验对本文的检

索方法进行了验证, 并同传统检索模型和现有基于本体的 LCA 查询扩展方法在查全率、查准率和度量值进行了比较, 上述三个指标均有所提高, 表明检索结果的质量有了改善且检索效果有了优化。

本文提出的方法仅是一个开始, 还需要做进一步的研究工作。根据领域可以自下而上整体的理论, 在领域本体向上整合的过程中, 本体间知识的映射会给用户检索树的构建带来困难, 需要对现有计算模型进行进一步优化。

参考文献

- 1 Lee TB, Hendler J, Lassila O. The Semantic Web. <http://www.sciam.com/article.cfm?id=the-semantic-web>.
- 2 Rijsbergen V. A new theoretical framework for information retrieval. Proc. of AMC Conference on Research and Development in Information Retrieval 1998. 194-200.
- 3 Dai JH. Fuzzy cluster-based query expansion[Master Thesis]. SunYat-sen University, Taiwan. 2004. 2-5.
- 4 Xu JX, Croft WB. Query Expansion Using Local and Global Document Analysis. Proc. of CAN-SIGIR Conference Retrieval Zurich, Switzerland. 1996.4-11.
- 5 Qiu YG, Frei HP. Concept based query expansion. Korfhage R, Rasmussen E, Willett P, eds. Proc. of the 16th Annual Int'l ACM SIGIR Conference on Research and Development in Information Retrieval. Pittsburgh: ACM Press. 1993. 160-169.
- 6 江敏, 肖诗斌, 王弘蔚, 施水才. 一种改进的基于《知网》的词语语义相似度计算. 中文信息学报, 2008, 22(5): 84-89.
- 7 Batet M, Sanchez D, et al. An ontology-based measure to compute semantic similarity in biomedicine. Journal of Biomedical Informatics, 2011, 44(1): 118-125.
- 8 Segura A, Salvador-Sanchez, et al. An empirical analysis of ontology-based query expansion for learning resource searches using MERLOT and the Gene ontology. Knowledge-Based Systems. 2011, 24(1): 119-133.
- 9 威欣, 肖敏, 孙建鹏. 基于本体知识库的自动语义标注. 计算机应用研究, 2011, 28(5): 1742-1747.
- 10 Jin H, Ning XM, Jia WJ, Yuan PP. Practical and effective IR-style keyword search over semantic web. Information Processing and Management, 2009, 45(2): 263-271.