韩其琛,李冬梅.基于叙词表的林业信息语义检索模型[J].计算机科学与探索,2016,10(1):122-129.

ISSN 1673-9418 CODEN JKYTA8
Journal of Frontiers of Computer Science and Technology
1673-9418/2016/10(01)-0122-08
doi: 10.3778/j.issn.1673-9418.1502017

E-mail: fcst@vip.163.com
http://www.ceaj.org
Tel: +86-10-89056056

基于叙词表的林业信息语义检索模型*

韩其琛[1,2],李冬梅[1+]
1. 北京林业大学 信息学院,北京 100083
2. 中国科学院大学 工程科学学院,北京 100049

Semantic Model with Thesaurus for Forestry Information Retrieval*

HAN Qichen[1,2], LI Dongmei[1+]
1. School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China
2. School of Engineering Science, University of Chinese Academy of Sciences, Beijing 100049, China
+ Corresponding author: E-mail: lidongmei@bjfu.edu.cn

HAN Qichen, LI Dongmei. Semantic model with thesaurus for forestry information retrieval. Journal of Frontiers of Computer Science and Technology, 2016, 10(1): 122-129.

Abstract: With the speedy development of the Internet, keyword-based retrieval method has failed to meet the needs of people. The semantic relationship within the thesaurus can improve recall ratio and precision ratio. If the thesaurus is introduced into current network information retrieval tool, the search technology would be definitely improved with the aid of rich semantic relationship of the thesaurus. This paper proposes an idea of calculating the similarity based on the relationship among the terms in the thesaurus. Utilizing query extension, this paper designs a semantic model with thesaurus for forestry information retrieval (SMTFIR). Finally, this paper compares SMTFIR, Baidu and the method used in agricultural thesaurus with two category realms in forestry thesaurus. The results show that SMTFIR can improve keyword-based retrieval method more effectively using thesaurus. In addition, SMTFIR is also suitable to other domains and provides a new thought for applying thesaurus in network information system.
Key words: forestry thesaurus; semantic retrieval; similarity computation; query extension; webpage grabbing

摘 要:随着互联网的快速发展,基于关键词字面匹配的信息检索方式已不能满足人们的需求。叙词表中所包含的语义关系是提高查全率和查准率的重要途径,如果将叙词表控制机制引入当前网络信息检索工具中,

* The National Natural Science Foundation of China under Grant No. 61170268 (国家自然科学基金); the Fundamental Research Funds for the Central Universities of China under Grant Nos. TD2014-02, xs2014024 (中央高校基本科研业务费专项资金).
Received 2015-02, Accepted 2015-05.
CNKI网络优先出版:2015-05-06, http://www.cnki.net/kcms/detail/11.5602.TP.20150506.1608.001.html

必然能在一定程度上提高信息检索的效率。利用叙词表中的词间关系，提出了一种计算叙词间语义相似度的方法，借助查询扩展的思想，设计了一种基于叙词表的林业信息语义检索模型。最后，以林业汉英拉叙词表中两个类目范畴作为实验对象，分别同百度搜索引擎、农业叙词表中所使用的检索方法进行了比较，实验结果表明，提出的检索模型可以更好地利用叙词表来改进传统的基于关键字的检索方式，此外，所提模型是通用的，为叙词表在网络信息系统中的应用提供了一种新的思路。

关键词：林业叙词表；语义检索；相似度计算；查询扩展；网页抓取

文献标志码：A　**中图分类号：**TP274

1　引言

在当前信息大爆炸的时代，网络上的信息和数据已经变得非常庞大，如何在海量级的数据中进行高效、准确的信息检索得到了越来越多的学者和专家的关注。搜索引擎是目前人们获取网络信息的主要工具。但是，由于目前主流的搜索引擎采用的都是基于关键词的字面匹配模式，即仅以孤立的关键词对信息内容进行标引和检索，人们在搜索内容上想要表达的语义内涵无法被机器所充分理解，进而导致信息检索查全率和查准率下降，在当前多样化的网络信息环境下其不足之处就显而易见了。由于基于关键字匹配的检索方法无法准确地表达出词语的语义内涵，近些年一些新的检索理念被提出，例如概念检索[1-2]和语义检索[3-4]等。本体是实现语义检索的一种较为有效的工具[5-6]，但本体的构建和维护需要大量的工作，与之相对的是，目前很多行业领域都有自己较成熟的叙词表。

叙词表是一个相对完善并且发展成熟的概念知识体系，自其从20世纪50年代诞生以来，经过不断发展和完善，已成为主题法中重要的信息组织工具，并在传统文献标引和检索中发挥过重要作用[7]。如能将叙词表引入到网络信息检索工具中，通过利用叙词表这一语义逻辑，必然能够在一定程度上提高传统信息检索的查全率和查准率。目前，基于叙词表的信息检索方法在医学领域已有较为深入的研究[8]。文献[9]利用随机游动（random walk）的方法借助医学叙词表对用户所输入的检索信息进行语义扩展，进而改善搜索结果。文献[10]对用户搜索语句进行语法分析，根据分析结果利用医学叙词表进行查询扩展。但是以上两种方法均没有对叙词之间的关系类型进行量化分析。文献[11]给出一种基于农业叙词表的检索方法，但该方法在查询扩展时只考虑与核心检索词直接相关的单级扩展，没有考虑其他叙词的影响，而且同样也没有对叙词之间的关系类型进行量化分析。本文在文献[11]的基础上，参考了Li等人的混合相似度算法[12]，以及Liu等人的基于相关概念节点密度的概念向量模型[13]，并结合林业汉英拉叙词表的相应特点，提出了一种综合叙词间多种关系的相似度计算方法，借助查询扩展和加权检索的思想，设计了一种基于叙词表的林业信息语义检索模型（semantic model with thesaurus for forestry information retrieval，SMTFIR）。最后通过实验验证了该模型的有效性。

2　基于叙词间关系的相似度计算方法

2.1　相关定义

定义1（叙词表概念树）在叙词表中，以族首词 O 为根节点，由族首词为 O 的所有叙词的上位叙词和下位叙词构成的树状结构 T 称为叙词表概念树。树结构中的节点 C 称为叙词节点，C_i 为对 T 进行层次遍历的第 i 个节点，根节点 O 记为 C_0。C 的所有祖先节点构成的集合称为 C 的祖先叙词节点 $A(C)$；C 的所有孩子节点构成的集合称为 C 的孩子叙词节点 $L(C)$。若至少存在一个词 W 与 C 所对应的叙词为相关关系，则称 C 所对应的叙词为 W 的相关关联叙词。根节点 O 的深度记为1；树中路径上分支数目为1的两个节点间的距离记为1。

定义2（最短路径长度）在 T 中，两个叙词节点之间分支数目最少的树中路径称为两个节点的最短路径，最短路径所拥有的分支数目称为最短路径长度。

定义3（最近根节点）在 T 中，如果叙词节点 R 是 A 和 B 共同的祖先节点，并且是符合此条件的所有节点中距离根节点最远的一个，则称 R 为 A 和 B 的最近根节点，记为 $R(A,B)$ 或 R 。

定义4（语义范围）在 T 中，以 C 为根的子树所包含的叶子节点数目称为 C 的语义范围，记为 $SCover(C)$ 。

定义5（基于叙词的语义向量）在一个包含 n 个叙词节点的 T 中，节点 C_i 表示成向量 $\boldsymbol{C}_i=(V_{i,1}, V_{i,2}, \cdots, V_{i,n})$, $V_{i,j}(i=1,2,\cdots,n, j=1,2,\cdots,n)$, 该向量称为基于叙词的语义向量。其中维度值定义为：

$$V_{i,j}=\begin{cases}1, C_j\subset\{C_i, L(C_i)\}\\ \dfrac{SCover(C_i)}{SCover(C_j)}, C_j\subset\{A(C_i)\}\\ 0, \text{else}\end{cases} \tag{1}$$

2.2 相关计算公式

本文规定：所有相似度的值均在[0,1]内。即如果权值为0，认为两个叙词之间没有任何关系；如果权值为1，认为两个叙词是等价的。同时规定，如果所求的两个叙词分别位于不同的概念树中，则认为其相似度为0。

设要判断相似度的词为 C_1 和 C_2, 根据 C_1 与 C_2 的关系类型的不同将相似度公式分为3类：等同相似度为 $SimD(C_1, C_2)$, 属分相似度为 $SimF(C_1, C_2)$, 相关相似度为 $SimW(C_1, C_2)$ 。

（1）等同相似度 $SimD(C_1, C_2)$

在叙词表中，等同词即等价关系，即两个词之间可以相互替换使用，故

$$SimD(C_1, C_2)=1 \tag{2}$$

（2）属分相似度 $SimF(C_1, C_2)$

$$SimF(C_1, C_2)=f_1\times f_2\times f_3 \tag{3}$$

其中，f_1 为基于最短路径的相似度，$f_1=e^{-\alpha d}$（d 为 T 中由 C_1 到 C_2 的最短路径长度，α 为调节因子）；f_2 为基于最近根深度的相似度，$f_2=1-e^{-\beta h}$（h 为 $R(C_1, C_2)$ 的深度，β 为调节因子）；f_3 为基于语义向量的相似度，$f_3=\dfrac{\boldsymbol{C}_1\cdot\boldsymbol{C}_2}{|\boldsymbol{C}_1\|\boldsymbol{C}_2|}$（$\boldsymbol{C}_1$、$\boldsymbol{C}_2$ 为根据定义5求得的 C_1、C_2 的语义向量）。

（3）相关相似度 $SimW(C_1, C_2)$

$$SimW(C_1, C_2)=g_1\times g_2 \tag{4}$$

其中，C_1 为 C_2 的相关关联叙词；g_1 为基于相关关联叙词深度的相似度，$g_1=\dfrac{e^{\varepsilon h}-e^{-\varepsilon h}}{e^{\varepsilon h}+e^{-\varepsilon h}}$（$h$ 为 C_1 的深度，ε 为调节因子）；g_2 为基于相关关联叙词密度的相似度，$g_2=1-e^{-\gamma l}$（l 为以 C_1 为根节点的直接子节点数，γ 为调节因子）。

2.3 相似度计算算法步骤

利用2.2节给出的相似度计算公式，相似度计算算法的具体步骤如下：

步骤1 根据叙词表对 K 进行扩展，得到关于 K 的初始查询扩展集合为 $U=\{D, F, W, Y\}$, 其中 D 表示等同词，F 表示 K 的所有上位/下位词（即叙词表概念树 T 的所有节点），W 表示 K 的相关词，Y 表示 F 的等同词和相关词。

步骤2 找到 K 的族首词 O, 以 O 为根节点建立叙词表概念树 T 。

步骤3 根据叙词表概念树 T, 利用式（2）得到 K 与 U 中 D 的相似度 $SimD(K, D)$ ；利用式（3）得到 K 与 U 中 F 的相似度 $SimF(K, F)$ ；利用式（4）得到 K 与 U 中 W 的相似度 $SimW(K, W)$ 。

步骤4 判断 Y 中每一个词 J 与其相对应的 F 中叙词 I 的关系。若 J 与 I 为相等关系，则利用式（2）、式（3）得到 K 与 J 的相似度 $SimF(K, I)\times SimD(I, J)$ ；若 J 与 I 为相关关系，则利用式（3）、式（4）得到 K 与 J 的相似度 $SimF(K, I)\times SimW(I, J)$ 。

步骤5 设置阈值 Q, 判断 U 中每一个词与 K 的相似度是否大于 Q 。若大于，则将该词加入到查询扩展集合 N 中；若小于，则跳过。

相似度计算流程如图1所示。

3 基于叙词表的林业信息语义检索模型

3.1 模型框架

本模型包含叙词标准化、查询扩展、网页抓取及加权排序4个模块。首先，利用林业汉英拉叙词表对用户输入的检索词进行叙词标准化，得到检索词 K ；其次，抓取与 K 相关的网页信息；之后，利用计算叙

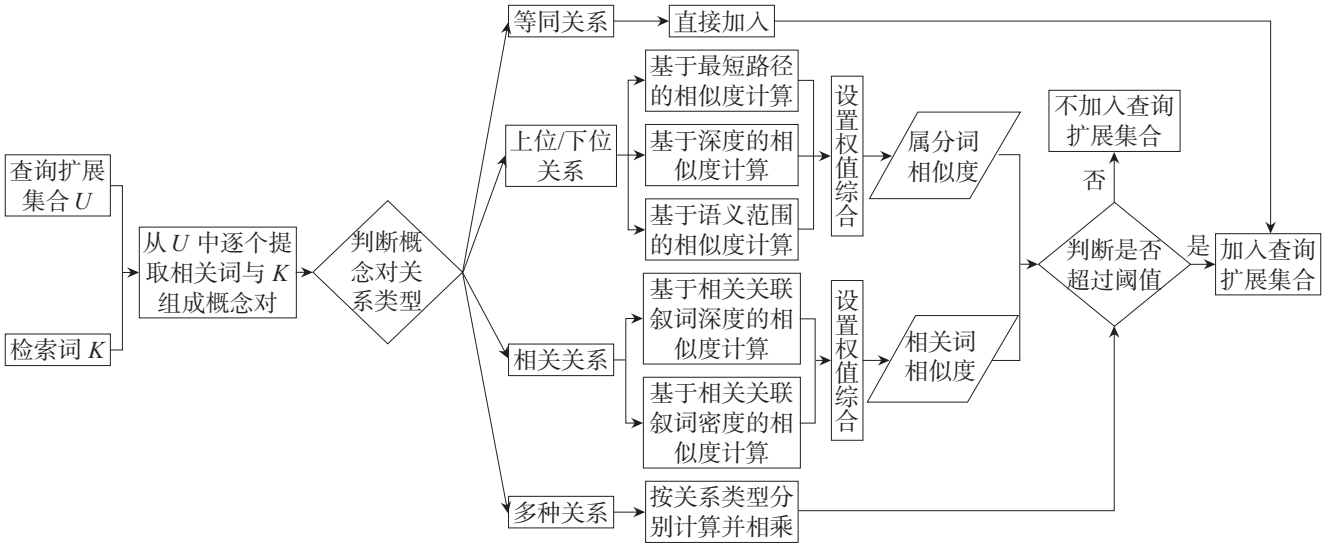

Fig.1 Procedure of similarity calculation

图1 相似度计算流程图

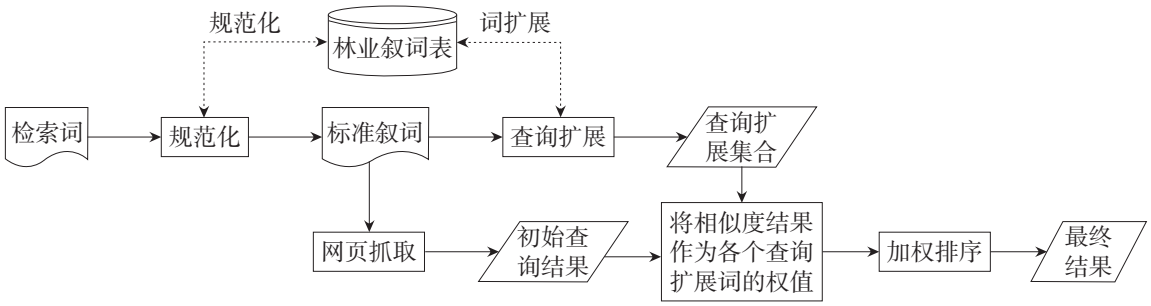

Fig.2 Structure of semantic model with thesaurus for forestry information retrieval

图2 基于叙词表的信息检索模型框架

词间语义相似度的算法得到用于查询扩展的相关词集合及相应权值；最后，根据查询扩展词及其相应权值对抓取的网页信息进行量化分析并排序。

该模型框架如图2所示。

3.2 叙词标准化

首先提取用户输入的检索词，根据叙词表判断是否需要对其进行标准化处理。由于用户检索需求和输入习惯的不同，此处可能遇到4种不同情况：若检索词是叙词，则不必标准化，可直接使用；若检索词为叙词表中的非叙词，则通过叙词表的相等关系将其转化为相应的叙词；若检索词可与叙词表中的叙词部分匹配，则将可匹配的所有叙词返回，供用户从中选择新的检索词；其他情况则保留原检索词，不对其进行查询扩展。

3.3 网页抓取

令由叙词标准化得到的检索词为 K，使用通用搜索引擎以 K 为检索词进行检索，取 s 个结果的URL。利用开源网页分析工具Htmlparser分析这 s 个网址所对应的网页，提取出网页中的标题、摘要、正文等信息。

3.4 查询扩展

利用2.3节所提到的相似度计算方法求出叙词表中所有与 K 相关的词的相似度，通过设置阈值的方式选取符合条件的相关词加入到查询扩展集合 N 中。

3.5 加权排序

在加权计算时，将 N 中相关词与 K 的相似度结果作为相关词的权值，加权排序方法的具体步骤如下：

步骤1 统计查询扩展集合中的每一个相关词在

网页标题中出现的频率 T 以及在网页正文中出现的频率 P。

步骤2 将每个网页的权值求和计算，其公式为：

$$TW_n=\frac{\sum_{i=1}^{m}W_i\times(\omega\times T_i+P_i)}{WN_n} \tag{5}$$

其中，TW_n 为第 n 个网页的总权值；WN_n 为第 n 个网页的字数；m 为查询扩展集合 N 中相关词的数目；W_i 为 N 中第 i 个相关词与检索词 K 的相似度；T_i 和 P_i 分别为该叙词在第 i 个网页的标题和正文中出现的频率；ω 为标题正文比，用于调节标题对于最终结果的重要性，ω 越大，标题对该网页权值的影响越大。

步骤3 将网页按权值由大到小排序并返回给用户。

4 实验及结果分析

4.1 实验数据

本文综合考虑了叙词表词汇量、关系数、实验需要等因素，采用 http://www.lknet.ac.cn/提供的林业汉英拉叙词表的两个词量适中类目范畴中的叙词及词间关系作为叙词表实验数据，分别用于测定相关参数的最优权值和评价相关性排序的效果。

4.2 实验数据检索效果评价指标的选择

检索效果是指利用检索系统进行信息检索产生的有效结果，它是检索系统性能的直接反映。一般来说，基于检索结果相关性的查全率和查准率是传统搜索引擎评价的主要指标。而国外有些学者发现：80%的用户只查看搜索结果的第一页，即对用户而言，其所需要的信息出现在检索结果的前几页比查全率和查准率更重要[14-15]。基于此又有学者提出了搜索长度的概念[16-18]，即指用户发现 n 个相关网页之前需要查看的不相关网页的数目，用来评估搜索引擎是否能够将最相关的网页排列在检索结果集的最前端。本文选择检索结果的相关性和搜索长度这两种指标来评价 SMTFIR 检索的有效性。

考虑到大多数用户检索时只会看返回的第一页结果，本文在进行评价时选择评价前 10 个结果的相关性，用 $P@10$ 表示。计算方法如下所示：

$$P@10=\frac{a}{a+b} \tag{6}$$

其中，a 表示前 10 项结果中与用户检索词相关的结果数量；b 表示前 10 项结果中与用户检索词无关的结果数量。从而可得出前 10 项的平均相关性公式：

$$\overline{P@10}=\frac{1}{n}\sum_{i=1}^{n}P_i \tag{7}$$

其中，P_1 至 P_n 为 n 次独立的实验所求得的 $P@10$。

而搜索长度设定为找到前 5 篇相关结果所需要查看的不相关结果的数量，搜索长度用 L 表示。同理，可以得出平均搜索长度公式：

$$\overline{L}=\frac{1}{n}\sum_{i=1}^{n}L_i \tag{8}$$

其中，L_1 至 L_n 为 n 次独立的实验所求得的 L。

4.3 相关参数权值的测定

通过实验测定两个重要的参数：用于查询扩展模块的阈值 Q 及加权排序模块中的标题正文比 ω。其他相似度算法的参数人工设定为 $\alpha=0.2$，$\beta=0.6$，$\varepsilon=0.6$，$\gamma=0.3$。

为使权值测定尽可能准确，从实验数据中随机选取 10 个叙词进行测试。在实验中，网页抓取模块选择百度搜索结果的前 100 条作为通用搜索引擎的结果进行抓取，将标题正文比先设定为 1。由相关林业方面人员确认返回结果是否与检索词相关。利用最终结果做折线图，如图 3 所示。

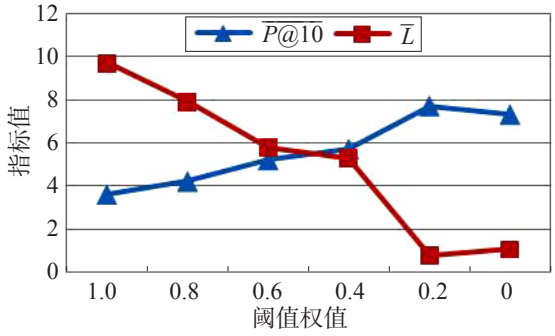

Fig.3 Determination data of threshold value

图3 阈值权值的测定数据

由图 3 可以看出：当阈值 Q 为 0.2 时，$\overline{P@10}$ 的数据值最高，即前 10 条结果的相关度最高；$\overline{L}$ 最低，即找到前 5 条相关结果所需要浏览的无关结果最少。因此，阈值确定为 0.2。

利用确定好的阈值，可以从叙词表中选择与检索词最为接近的词汇用于查询扩展。以检索词为夏绿林为例，通过确定好的阈值可以得到如下相关词汇：落叶阔叶林(0.817 9)，栎林(0.670 3)，桤林(0.670 3)，阔叶林(0.668 3)，常绿阔叶林(0.547 9)，照叶林(0.547 9)，常绿竹林(0.547 7)，硬叶常绿林(0.448 9)，其中括号内数值为其与检索词的相似度。

在得到阈值结果后，将阈值调整为0.2，继续用这10个叙词进行标题正文比的测试。同样，利用最终实验结果分别做折线图，如图4所示。

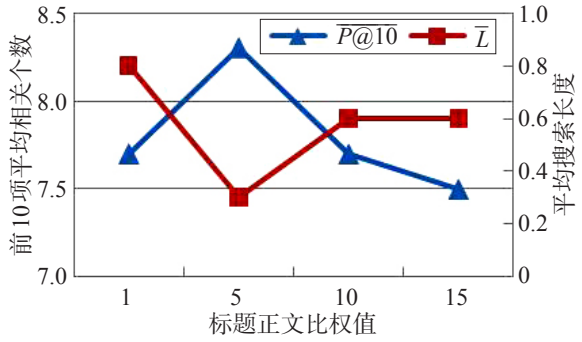

Fig.4 Determination data of title-text rate

图4 标题正文比的权值测定

由图4可以看出，当标题正文比 ω 为5时，此时的 $\overline{P@10}$ 值最高，而 $\overline{L}$ 最小。因此综合两个数据，将标题正文比 ω 确定为5。

4.4 实验结果分析

根据4.3节测定的最优权值，从实验数据中随机选择15个词分别利用百度搜索引擎、文献[11]的方法以及SMTFIR进行搜索，并分别测量在不同情况下返回结果的 $P@10$ 和 L 指标，将实验结果绘制为表1。

根据表1的结果做折线图，如图5和图6所示。从图中可以看出，SMTFIR和文献[11]的方法相较于百度的结果来说均有不同程度的改进，这说明叙词表确实可以提高搜索结果的准确性。与此同时，SMTFIR也要比文献[11]的方法更加准确，说明了本文提出的检索模型可以更好地利用叙词表来改进传统基于关键字的检索方式。

4.5 模型通用性分析

经过几十年的发展，叙词表的编制方法得到不断改善，最终形成了一系列的国际标准。国际标准有1974年发布的ISO 2788和1985年发布的ISO 5964，我国目前的现行标准为1991年发布的GB/T 13190。

Table 1 Results comparison between SMTFIR and other methods

表1 SMTFIR与其他检索方法的对比

序号	检索词	$P@10$			L		
		百度搜索引擎	文献[11]	SMTFIR	百度搜索引擎	文献[11]	SMTFIR
1	夏绿林	5	7	8	5	0	0
2	雨林	5	5	7	3	5	1
3	疏林	5	5	5	5	4	5
4	红杉	2	4	9	14	8	1
5	红云杉	6	4	9	3	8	0
6	黑云杉	4	4	8	6	7	0
7	白冷杉	6	5	5	3	2	0
8	日本铁杉	5	6	5	3	0	0
9	冷杉林	8	7	8	1	1	0
10	照叶林	7	8	9	1	1	0
11	落叶松林	6	6	7	4	1	3
12	常绿落叶阔叶混交林	9	10	10	0	0	0
13	种子林	6	6	9	3	2	0
14	一般法正林	6	8	9	2	0	0
15	池杉	8	9	9	0	0	0

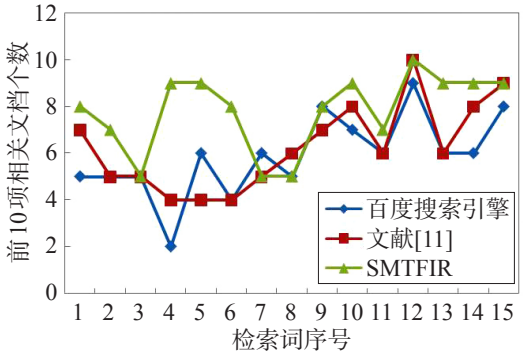

Fig.5 Results comparison between SMTFIR and other methods (*P*@10)

图5 SMTFIR与其他检索方法的对比 (*P*@10)

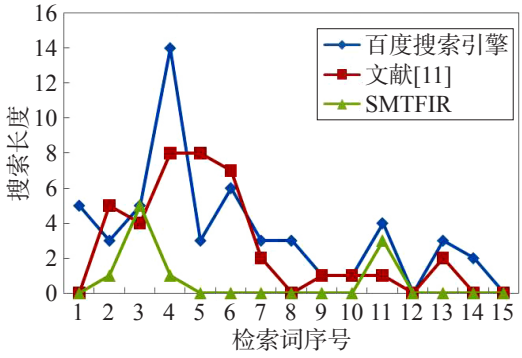

Fig.6 Results comparison between SMTFIR and other methods (*L*)

图6 SMTFIR与其他检索方法的对比 (*L*)

在这些标准中均明确规定了叙词表中的词间关系有3种,分别是本文所提及的等同关系、等级关系和相关关系。2.2节所利用的3种关系在现行任何符合国际标准的叙词表中均是存在的,因此本文所提出的模型具有较强的通用性。

5 结束语

由于基于关键词的传统信息检索方法不能充分表达语义信息,本文利用叙词表的词间关系,提出了一种计算叙词间语义相似度的方法,设计了一种基于叙词表的林业信息语义检索模型,显著提高了查询效果。本文模型同样适合其他的行业领域,这种检索方式为在当前大数据时代如何合理利用叙词表提供了一个新的研究思路。在今后的研究中可以从检索结果相关性评价等方面进行改进和完善。

References:

[1] Qian Xueming, Guo Danping, Hou Xingsong, et al. HWVP: hierarchical wavelet packet descriptors and their applications in scene categorization and semantic concept retrieval[J]. Multimedia Tools and Applications, 2014, 69(3): 897-920.

[2] Aly R, Doherty A, Hiemstra D, et al. The uncertain representation ranking framework for concept-based video retrieval[J]. Information Retrieval, 2013, 16(5): 557-583.

[3] Alghamdi N S, Rahayu W, Pardede E. Semantic-based structural and content indexing for the efficient retrieval of queries over large XML data repositories[J]. Future Generation Computer Systems, 2014, 37: 212-231.

[4] Bergmann R, Gil Y. Similarity assessment and efficient retrieval of semantic workflows[J]. Information Systems, 2014, 40: 115-127.

[5] Rodríguez-García M Á, Valencia-García R, García-Sánchez F, et al. Ontology-based annotation and retrieval of services in the cloud[J]. Knowledge-Based Systems, 2014, 56: 15-25.

[6] Xi Lei, Zheng Guang, Wang Qiang, et al. Intelligent service system of pollution-free agricultural products catalog based on personalized features[J]. Transactions of the Chinese Society of Agricultural Engineering, 2013, 29(20): 142-150.

[7] Agichtein E, Gabrilovich E. Information organization and retrieval with collaboratively generated content[C]//Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, Beijing, China, Jul 25-29, 2011. New York, USA: ACM, 2011: 1307-1308.

[8] Ivanović M, Budimac Z. An overview of ontologies and data resources in medical domains[J]. Expert Systems with Applications, 2014, 41(11): 5158-5166.

[9] Martinez D, Otegi A, Soroa A, et al. Improving search over electronic health records using UMLS-based query expansion through random walks[J]. Journal of Biomedical Informatics, 2014, 51: 100-106.

[10] Azcárate M C, Vázquez J M, López M M. Improving image retrieval effectiveness via query expansion using MeSH hierarchical structure[J]. Journal of the American Medical Informatics Association, 2013, 20(6): 1014-1020.

[11] Xiong Xia. Domain information retrieval based on term rela-

tionships of thesaurus[D]. Beijing: Chinese Academy of Agricultural Sciences, 2011.

[12] Li Yuhua, Bandar Z A, McLean D A. An approach for measuring semantic similarity between words using multiple information sources[J]. IEEE Transactions on Knowledge and Data Engineering, 2003, 15(4): 871-882.

[13] Liu Hongzhe, Bao Hong, Xu De. Concept vector for similarity measurement based on hierarchical domain structure[J]. Computing and Informatics, 2012, 30(5): 881-900.

[14] Jansen B J. An investigation into the use of simple queries on Web IR systems[J]. Information Research: An Electronic Journal, 2000, 6(1): 1-10.

[15] Ali R, Beg M M S. An overview of Web search evaluation methods[J]. Computers & Electrical Engineering, 2011, 37(6): 835-848.

[16] Chignell M H, Gwizdka J, Bodner R C. Discriminating meta-search: a framework for evaluation[J]. Information Processing & Management, 1999, 35(3): 337-362.

[17] Dwivedi S K, Goutam R K. Evaluation of search engines using search length[C]//Proceedings of the International Conference of Computer Modeling and Simulation, 2011: 502-505.

[18] Scaiella U, Ferragina P, Marino A, et al. Topical clustering of search results[C]//Proceedings of the 5th ACM International Conference on Web Search and Data Mining, Seattle, USA, Feb 8-12, 2012. New York, USA: ACM, 2012: 223-232.

附中文参考文献：

[6] 席磊, 郑光, 汪强, 等. 基于个性化特征的无公害农产品目录智能服务系统[J]. 农业工程学报, 2013, 29(20): 142-150.

[11] 熊霞. 基于叙词表词间关系的领域信息检索[D]. 北京: 中国农业科学院, 2011.

HAN Qichen was born in 1992. He is an M.S. candidate at School of Engineering Science, University of Chinese Academy of Sciences. His research interests include information retrieval and personalized recommendation.

韩其琛(1992—),男,山西太原人,中国科学院大学工程科学学院硕士研究生,主要研究领域为信息检索,个性化推荐。

LI Dongmei was born in 1972. She received the Ph.D. degree in artificial intelligence from Beijing Jiaotong University in 2014. Now she is an associate professor at Beijing Forestry University. Her research interests include artificial intelligent, knowledge engineering and semantic Web.

李冬梅(1972—),女,黑龙江大庆人,2014年于北京交通大学获得博士学位,现为北京林业大学信息学院副教授,主要研究领域为人工智能,知识工程,语义Web。