

doi: 10.11830/ISSN.1000-5013.201703022



# 采用相关反馈和文档相似度的 维吾尔语检索词加权方法

于丽, 亚森·艾则孜

(新疆警察学院 信息安全工程系, 新疆 乌鲁木齐 830011)

**摘要:** 针对维吾尔语 Web 文档的有效检索问题, 提出一种基于相关反馈和文档相似度的检索词加权方法。首先, 对维吾尔语文档进行预处理, 获得相应的词干集。然后, 当用户输入多个检索词时, 执行初始检索, 并基于局部相关反馈思想提取出排名靠前的  $N$  个文档。接着, 利用 TF-IDF 算法计算检索词与反馈文档之间的词频相似度, 通过余弦距离计算文档之间的相似度, 并以此对检索词进行两次加权。最后, 根据加权后的检索词进行文档检索。实验结果表明: 该方法能够准确地检索出用户所需的文档, 并将其靠前排序。

**关键词:** 维吾尔语; 文档检索; 检索词加权; 相关反馈; 文档相似度

**中图分类号:** TP 391      **文献标志码:** A      **文章编号:** 1000-5013(2017)03-0408-06

## Uyghur Retrieval Word Weighting Scheme Using Relevance Feedback and Document Similarity

YU Li, YASEN · AIZEZI

(Department of Information Security Engineering, Xinjiang Police College, Urumqi 830011, China)

**Abstract:** For the issue that the effective retrieval of Uyghur web documents, a Uyghur retrieval word weighting scheme based on the relevance feedback and document similarity is proposed. First of all, the Uyghur documents are pre-processed to obtain the corresponding stem set. Then, the initial search is executed when the user input a number of retrieval words, and it extracts the top  $N$  documents based on local relevance feedback. Follow, the TF-IDF algorithm is used to compute the frequency similarity between retrieval word and feedback documents. At the same time, the cosine distance is used to compute the similarity between documents, so as to make twice weighted for retrieval words. Finally, it performs document retrieval according to the weight of retrieval words. Experimental results show that the proposed method can accurately retrieve the documents required by the user, and can sort them in the front.

**Keywords:** Uyghur; document retrieval; weighted retrieval words; relevance feedback; document similarity

随着国家对新疆地区投入的增加, 大量维吾尔语等少数民族语种的文字信息开始以数字化形式呈现<sup>[1]</sup>。若能迅速准确地从维吾尔语书写的 Web 文档数据中获得所需文档, 将会为新疆地区的文化发展作出一定的贡献。在文档检索中, 用户经常会输入多个检索词。在传统检索中, 各检索词的权重一致。但由于用户初始查询时, 其信息需求通常比较模糊, 从而无法准确地反馈实际需求的文档<sup>[2]</sup>。目前, 提高检索系统性能的方法主要包括概率和语言模型方法<sup>[3]</sup>、检索分类方法<sup>[4]</sup>及检索修正方法<sup>[5]</sup>。其中, 检索修正技术是通过改善用户的初始检索词来提高检索性能。检索修正技术<sup>[6]</sup>主要分为检索词重新加权

(retrieval word re-weighting)和查询扩展(query expansion). 检索词重新加权方法是根据初步检索反馈信息,对用户输入的多个检索词进行重要性加权. 针对检索词重新加权技术,学者提出了多种方法. Sembok 等<sup>[7]</sup>基于 N-gram 的向量空间模型和语境分析对检索词进行重新加权,但其没有通过反馈机制获得顶级文档,而是计算所有文档,计算量较大. Zhou 等<sup>[8]</sup>提出了基于加权信息增益(weighted information gain, WIG)检索词重新加权方法,利用单一检索词和检索词组的距离特征估计顶级检索文档的质量. 然而,其没有考虑不同反馈文档对检索词加权具有不同的重要性. 本文主要研究维吾尔语文档检索中的检索词重新加权技术.

### 1 维吾尔语文档预处理

维吾尔语文档预处理主要包括文档过滤和词干提取. 其中,文档过滤用于过滤文档中非维吾尔语文字和停用词,通过和事先准备好的停用词表进行比对,过滤掉停用词. 停用词为对文档主题没有贡献,不包含文章类别信息的词,如介词、副词、代词等. 去掉停留词能够实现特征降维,提高检索精度<sup>[9]</sup>.

词干提取过程中,根据维吾尔语单词与单词间的空格符进行分词. 由于维吾尔语单词是由字母拼写而成的,通过将不同的词缀粘贴到单词的头部实现语法功能,所以提取文档中代表真实含义的词汇是困难的. 在维吾尔语中,同一词干可以演变为很多不同含义的词语,虽然这些词语的词形不同,但词义却不会有很大的区别<sup>[10]</sup>. 一个典型的例子,如表 1 所示. 为了提取单词的词义,考虑特征的数量,以词干(学校)作为特征词,从文档中提取出词干集.

表 1 维吾尔词语变体

Tab. 1 Uighur word variations

词干	词组	变形(维文)	变形(中文)
مەكتەپ (学校)	ئى	مەكتەپنى	把学校
	تە	مەكتەپتە	在学校
	كە	مەكتەپكە	到学校
	ئىن	مەكتەپتىن	从学校
	بىز	مەكتەپىمىز	我们的学校
	بىزدە	مەكتەپىمىزدە	在我们的学校
	ئىكى	مەكتەپتىكى	在学校的
	ئىگىز	مەكتەپىڭىز	你的学校
	ئىزلار	مەكتەپىڭىزلار	你们的学校

### 2 检索词重新加权方法

由于用户输入的检索词具有多个单词,若按照均等权重搜索文档,不能精确地搜索到所需文档,所以,需要对用户检索词进行重新加权. 文中基于局部相关反馈思想,首先,根据初始检索词进行检索,并获得初始检索文档;然后,在初始检索文档中选出排名靠前的  $N$  个顶级检索文档,并利用相关性模型计算各检索词与顶级文档的相关性,以此对检索词进行重要性重新加权.

利用  $Q$  表示用户输入的初始检索词集,  $Q = \{q_1, q_2, \dots\}$ , 其中,  $q_i (i = 1, 2, \dots)$  表示第  $i$  个检索词. 将每个检索词  $q_i$  的最终权重大小定义为

$$W_{q_i} = \sum_{j=1}^N W_{q_i}^{d_j}.$$

(1)

式(1)中:  $W_{q_i}$  为  $q_i$  的最终权重大小;  $N$  为从初始检索文档中选取的顶级文档个数;  $W_{q_i}^{d_j}$  为顶级文档  $d_j$  对检索词  $q_i$  的权重贡献值;  $d_j$  为第  $j$  个顶级文档.

假设  $d_j$  与用户所需的信息相关,那么,  $W_{q_i}^{d_j}$  就是有价值的. 虽然文中的检索引擎将  $d_j$  检索为一个相关文档,  $d_j$  还是有可能为一个不相关的文档. 为了处理检索文档集中文档重要性的问题,将  $W_{q_i}^{d_j}$  定义为文档  $d_j$  中检索词  $q_i$  的二次权重值,表示为

$$W_{q_i}^{d_j} = \omega_{q_i}^{d_j} \times v_{d_j}.$$

(2)

式(2)中:  $\omega_{q_i}^{d_j}$  为文档  $d_j$  中检索词  $q_i$  的基础权重值,即第一次加权项;  $v_{d_j}$  为  $d_j$  与用户所需信息之间的相关性,即第二次加权项.

#### 2.1 基于 TF-IDF 的第一次加权

文档  $d_j$  中检索词  $q_i$  的基础权重  $\omega_{q_i}^{d_j}$  可由标准 TF-IDF 权重模型<sup>[11]</sup>计算,具体表示为

$$\omega_{q_i}^{d_j} = F_{q_i}^{d_j} \times \text{IDF}_{q_i}.$$

(3)

式(3)中: $F_{q_i}^{d_j}$ 为文档 $d_j$ 中检索词 $q_i$ 出现的频率, $F_{q_i}^{d_j} = \frac{f_{q_i}}{n_{d_j}}$ , $f_{q_i}$ 为文档 $d_j$ 中检索词 $q_i$ 出现的次数, $n_{d_j}$ 为文档 $d_j$ 中术语的总数; $IDF_{q_i}$ 为整个集合中包含检索词 $q_i$ 的文档数量倒数的 $\log$ 值,即 $IDF_{q_i} = \log \frac{N_D}{N_{q_i}}$ , $N_D$ 为语料集 $D$ 中文档总数, $N_{q_i}$ 为语料集中包含检索词 $q_i$ 的文档数.

2.2 基于文档相似度的第二次加权

上述过程通过反馈文档中检索词的 TF-IDF 信息对检索词进行加权,只考虑了词频信息.不同的文档与用户实际需要的相关性不同,相关性较大的文档给出的词频信息的重要性也应该较大.为此,需要根据文档与用户的相关程度对该文档给出的词频进行进一步加权.

为了计算一个文档在初始检索中的重要性,通过评估每个文档与检索集合中其他文档的距离,测量每个文档与用户所需信息之间的相关性.通过余弦相似度函数  $\text{Sim}()$ ,计算每个顶级文档与集合中其他顶级文档之间的相似度,并以此给出第二次加权项,即

$$v_{d_j} = \frac{\sum_{k=1, k \neq j}^N \text{Sim}(d_k, d_j)}{N - 1} \tag{4}$$

式(4)中: $N$ 为从初始检索文档中选取的顶级文档个数; $d_k$ 和 $d_j$ 分别表示文档集合中的第 $k$ 个和第 $j$ 个文档的欧式向量,文中利用向量空间模型(vector space modal, VSM)对文档进行表示.将每个文档 $d_j$ 作为一个向量 $d_j = (\omega_{1,j}, \omega_{2,j}, \omega_{m,j}, \cdots, \omega_{M,j})$ .其中,第 $m$ 个组件 $\omega_{m,j}$ 为该文档中第 $m$ 个词干的权重.第 $m$ 个词干的 $\omega_{m,j}$ 根据 TF-IDF 算法,通过计算该词干在文档 $d_j$ 中的词频和在反馈文档中的逆文档频率获得.  $\text{Sim}()$ 用于评估 $d_k$ 和 $d_j$ 间相似度的余弦函数  $\cos()$ ,典型的余弦相似度表达式为

$$\cos(d_i, d_j) = \frac{d_i \cdot d_j}{|d_i| |d_j|} = \frac{\sum_{m=1}^M (\omega_{m,i} \times \omega_{m,j})}{\sqrt{\sum_{m=1}^M \omega_{m,i}^2 \times \sum_{n=1}^N \omega_{m,n}^2}} \tag{5}$$

余弦相似度衡量的是向量空间 2 个向量的夹角,更加体现了方向上的差异<sup>[12-15]</sup>.向量夹角的余弦值越大,表明两者之间的相似度越大.余弦相似度的另一个重要特征是其独立于文档的长度.因此,余弦相似度在向量空间模型中是一种较为有效的相似性度量.

那么,利用式(2),(3)可得

$$W_{q_i} = \sum_{j=1}^N F_{q_i}^{d_j} \times IDF_{q_i} \times v_{d_j} \tag{6}$$

最后,利用  $\log$  标准化计算值,即

$$W_{q_i} = \log(1 + \sum_{j=1}^N F_{q_i}^{d_j} \times IDF_{q_i} \times v_{d_j}) \tag{7}$$

式(7)中:添加了一个常量值,这样可以避免算法获取零值.利用式(7)对检索词进行了二次加权,其中, $W_{q_i}$ 的大小与顶级文档中检索词的频率和整个集合中检索词的 IDF 成正比.

式(4)中,当一个顶级文档与所有其他顶级文档的距离都较小时,该顶级文档的权重较大.即一个文档到聚类中心的距离越近,则这个文档的权重值越大.但这是一个简化的通用假设,即不考虑检索的实际行为.当模糊检索结果中包含的内容不止一个时,其结果集中可能包含多个文档聚类.

为了解决这个问题,将与用户初始检索具有较高相似度的文档分配一个较大的影响权重值,即该文档对检索词加权的影响程度较大.因此,只要文档的主题与初始检索的主题足够贴近,那么,不同聚类中的文档可以获取较大的权重.在这种直观判断基础之上,可以将式(4)改写为

$$v_{d_j} = (K \times \frac{\sum_{k=1, k \neq j}^N \text{Sim}(d_k, d_j)}{N - 1} + (1 - K) \times \text{Sim}(d_j, I))^L \tag{8}$$

式(8)中: $I$ 为初始检索 $Q$ 的欧式向量;变量 $K$ 和 $L$ 为常数,通过大量实验经验,设定 $K=0.7, L=3$ .

将式(8)代替式(7)中的 $v_{d_j}$ ,可得

$$W_{q_i} = \log(1 + \text{IDF}_{q_i} \times \sum_{j=1}^N F_{q_i}^{d_j} \times (K \times \frac{\sum_{k=1, k \neq j}^N \text{Sim}(d_k, d_j)}{N-1} + (1-K) \times \text{Sim}(d_j, I))^L).$$

(9)

式(9)中: $q_i$  和  $d_j$  之间的关系计算了 2 次,即在计算  $F_{q_i}^{d_j} \times \text{IDF}_{q_i}$  和  $\text{Sim}(d_j, I)$  时.

为了减弱  $q_i$  与  $d_j$  关系对权重的影响,定义  $I_i$  为

$$I_i = I - \{q_i\}.$$

(10)

因此,若将  $I$  中的  $q_i$  省略,则可以获取  $I_i$ . 那么,在式(9)中利用  $I_i$  代替  $I$ ,可以减少这种关系计算的次数,得到

$$W_{q_i} = \log(1 + \text{IDF}_{q_i} \times \sum_{j=1}^N F_{q_i}^{d_j} \times (K \times \frac{\sum_{k=1, k \neq j}^N \text{Sim}(d_k, d_j)}{N-1} + (1-K) \times \text{Sim}(d_j, I_i))^L).$$

(11)

最后,利用  $W_{\max}$  对每个检索中的最终权重项归一化到 $[0,1]$ , $W_{\max}$ 表示这次检索中最大的权重项.

3 实验及分析

3.1 实验设置

为了评估文中方法的性能,构建一个实验平台,配置有 Intel 酷睿 i5 CPU,主频为 2.4 GHz,应用 Windows 7 系统,利用 Matlab 2011 进行实验,对检索方法进行性能评估.

对于维吾尔语的检索应用,目前还没有可使用的标准文档集.文中从人民网维吾尔语版和天山网等主流维吾尔语网站上收集了包含政治、经济、体育、旅游、教育和文化等领域的 2 000 篇文档.首先,对维吾尔语文档集进行预处理,为了方便后续处理,把文档转换成 UTF-8 二进制编码格式.然后,过滤掉文档中的非维吾尔语字符和停用词.预处理结束后,获得一个初始词集.最后,进行词干提取,将同一词根演变而来的单词进行聚合,获得文档的词干集合.

根据所收集的文档内容,自行拟定了 10 组检索词组,分别如表 2 所示.表 2 中:每个检索词组至少有 30 个文档的内容与其相关.另外,为了评估检索性能,通过人工方式,将与该 10 组检索词较为相关的 30 个文档进行相关性标注,并标注出最相关的 10 个文档.

表 2 10 种检索词组  
Tab.2 Ten kinds of search team groups

检索序号	查询词组	检索序号	查询词组
1	يۈرەك كېسەللىكى، ئاماك چىكىش (心脏疾病、抽烟)	6	2008-يىللىق بېيجىڭ ئولىمپىك تەنھەركەت مۇسابىقىسى (2008 年北京奥运会)
2	ئىسلام دىنىنىڭ تەبرىك پائالىيەتلىرى (伊斯兰教的庆典)	7	يەر سارى قۇرغاقچىلىق كىۈزۈسى (全球干旱危机)
3	نەشىر ھوقۇقى قانۇنىنىڭ ئۈنۈمى (版权法律的效益)	8	شىنجاڭلىقلارنىڭ نەنەنىۋى تەبرىك پائالىيەتلىرى (新疆人的传统庆典)
4	شىنجاڭدىكى يېزىقنىڭ جايلىشىشى (新疆的石油分布)	9	گېلەم توقۇش (纺织毛毯)
5	شەرىئەت دوختۇرنىڭ نەسلىمىسى (沙里阿提医生的回忆)	10	تېلېۋىزۇر ۋە مەنىۋى مەدەنىيەت (电视和精神文明)

3.2 实验指标

为了评估检索词重新加权方法的有效性,将文中方法与平均权重的传统检索方法、文献[11]提出的基于 WIG 检索词重新加权的检索方法进行比较.

实验最终检索出 30 个相关文档.以召回率和准确率为评估标准,召回率为系统正确检索的文档占用户所需文档的比例,准确度为系统正确检索的文档占系统所检索出的文档总数的比例.同时,统计了平均准确率均值(mean average precision, MAP)和前 10 个检索结果的准确率(P@10). MAP 为所有查询的平均准确率的算术平均,用来衡量检索系统对多个查询的平均检索质量;P@10 为检索结果的前 10 个位置中,出现标注为最相关文档的数量比例,最能反映实际应用情况.另外,实验中的数据都为 10

次实验的平均值.

3.3 顶级文档数量参数的确定

在初始检索中,初始检索反馈的顶级文档数量  $N$  对检索词加权性能具有至关重要的影响. 为此,首先进行实验确认最优的  $N$  值. 设置  $N$  的范围为 $[10, 100]$ ,测试检索性能,并记录 MAP 值,结果如图 1 所示.

由图 1 可知:不同顶级文档数量下的检索性能不同. 这是因为当反馈文档数量较少时,检索词加权过程中可用的相关文档数量太少,不能反映整个文档集的属性. 当反馈文档数量较多时,会存在一些无关或噪声文档,使用这些文档对检索词进行加权会降低加权质量. 综合考虑加权质量和计算量,设定顶级文档数量  $N$  取值为 50.

3.4 实验比较

首先,在查询文档集中,输入 10 种检索词组,利用文中方法,通过相关反馈和文档相似度对检索词进行重新加权,加权结果如表 3 所示. 需要注意的是,各种检索词所分配的权重与文档集内容息息相关.

表 3 文中方法中检索词的加权值

Tab. 3 Weighted value of term in proposed method

检索序号	检索词权重	检索序号	检索词权重
1	(心脏[0.87]疾病[0.83]、抽烟[1])	6	(2008 年[0.25]北京[0.82]奥运会[0.49])
2	(伊斯兰教[0.76]的庆典[0.31])	7	(全球[0.21]干旱[1]危机[0.73])
3	(版权[0.98]法律[0.39]的效益[0.01])	8	(新疆人[0.86]传统的[0.47]庆典[1])
4	(新疆[0.62]的石油[1]分布[0.42])	9	(纺织[0.88]毛毯[0.48])
5	(沙里阿提[1]医生[0.77]的回忆[0.46])	10	(电视[0.19]和精神[1]健康[0.96])

在实验文档集中,执行 3 种检索方法,记录各种检索词组下的检索准确率,结果如表 4 所示. 由表 4 可知:不同的检索词具有不同的检索性能. 这是因为不同的检索词组中,能够反映用户意图的检索词信息清晰度不同. 其中,文中方法在各个检索词组下的准确率比 WIG 和传统方法都高. 这是因为文中根据检索词和反馈文档的相似度对检索词进行了合理的重新加权,这有效提高了检索精度.

传统检索、WIG 方法和文中方法在文档集上检索的准确率( $\eta_a$ )-召回率( $\eta_r$ )曲线,如图 2 所示. 由图 2 可知:文中方法在相同召回率情况下,检索精确率都高于其他方法. 这说明对检索词合理加权能够有效地提高检索性能.

表 4 3 种方法在 10 种检索词组下的准确率

Tab. 4 Accuracy of three methods in 10 kinds of retrieval words

检索序号	$\eta_a/\%$		
	传统检索	WIG	文中方法
1	52.6	70.7	73.4
2	50.5	52.3	75.9
3	45.3	47.3	56.7
4	25.9	26.7	28.5
5	45.5	55.3	64.4
6	27.5	32.2	36.7
7	50.3	55.3	67.3
8	30.4	35.9	49.8
9	25.5	35.5	43.8
10	17.5	27.5	30.5

3 种检索方法的 MAP 和 P@10 性能比较,如表 5 所示. 由表 5 可知:文中方法可以有效地提升检索性能. 其中,文中方法的 MAP 达到了 52.7%,比传统检索和 WIG 加权方法分别提高了 42%和 20%. 在

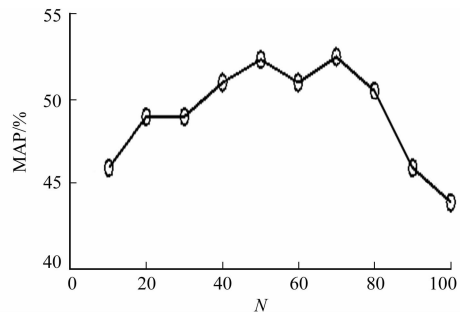


图 1 不同顶级文档数量下的检索性能

Fig. 1 Retrieval performance under different number of top documents

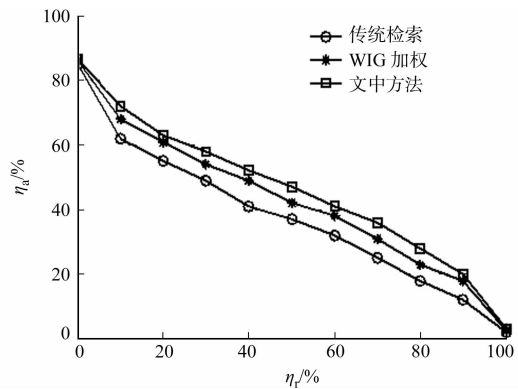


图 2 3 种方法在文档集上检索的准确率-召回率曲线

Fig. 2 Retrieval accuracy-recall rate curve of three methods on document collection

P@10 指标上,文中方法达到了 60.8%,比传统检索和 WIG 加权方法分别提高了 31.6%和 9.3%。这说明文中方法能够更好地为用户提供所需的查询文档,并将其排序到靠前位置。

4 结论

提出一种用于维吾尔语 Web 文档检索的检索词重新加权方法。基于局部相关反馈思想,对文档执行初始检索,并反馈  $N$  个顶级文档。采用 TF-IDF 算法计算检索词与反馈文档之间的词频相似度,利用余弦距离计算文档之间的相似度,并以此对检索词进行两次加权,使检索词更能反映用户的实际需求。与现有重新加权方法相比,文中方法能合理地检索词进行加权,使其能够精确地检索用户所需文档。

另外,文中提出的检索方法应用于维吾尔语文档,但其主体思想也可用于其他语言文档的检索,其不同点在于文档的预处理。文中方法在维吾尔语文档预处理中,考虑了维吾尔语的特征,提取了词干。若将文中方法应用到其他语言中,则需要根据具体语言特征提取有效词干。

文中认为更加精致的加权算法将有助于获取更高的性能。因此,在以后的研究中,将尝试采用一些概率框架进一步加权检索词。

参考文献:

[1] 阿丽亚·艾尔肯,哈力旦·阿布都热依木. KNN 和 SVM 分类器对维吾尔文文本分类性能的比较研究[J]. 新疆大学学报(自然科学维文版),2015,32(2):59-65.

[2] 亚力青·阿里玛斯,哈力旦·阿布都热依木,陈洋. 基于向量空间模型的维吾尔文文本过滤方法[J]. 新疆大学学报(自然科学版),2015,32(2):221-226.

[3] HAN Tiantan,WANG Wendong,GONG Xiangyang, *et al.* Personal multimedia data retrieval query expansion and similarity algorithm improvement based wordNet[J]. International Proceedings of Computer Science and Information Tech,2012,42(3):51-62.

[4] 陈雅兰,胡小华,涂新辉,等. 基于位置语言模型的中文信息检索系统的研究[J]. 计算机科学,2015,42(7):265-269.

[5] HAHM G J,YI M Y,LEE J H, *et al.* A personalized query expansion approach for engineering document retrieval [J]. Advanced Engineering Informatics,2014,28(4):344-359.

[6] ATSUSHI F. Enhancing web document retrieval by the anchor text model and query classification[J]. Ipsj Journal, 2010,51(3):2330-2342.

[7] 李卫疆,赵铁军,王宪刚. 基于上下文的查询扩展[J]. 计算机研究与发展,2010,47(2):300-304.

[8] 陈志敏,姜艺,赵耀. 基于用户查询扩展的自动摘要技术[J]. 计算机应用研究,2011,28(6):2188-2190.

[9] DANG E K F,LUK R W P,ALLAN J. Fast forward index methods for pseudo-relevance feedback retrieval[J]. Acm Transactions on Information Systems,2015,33(4):1-33.

[10] SEMBOK T M T,BAKAR Z A. Characteristics and retrieval effectiveness of n-gram string similarity matching on Malay documents[C]// Proceedings of the 10th WSEAS International Conference on Applied Computer and Applied Computational Science. Stevens Point:ACM Press,2011:165-170.

[11] ZHOU Yun,CROFT W B. Weighted information gain and user clicks on web search results[C]// International ACM SIGIR Conference on Research and Development in Information Retrieval. Amsterdam:ACM Press,2012: 543-550.

[12] 年梅,张兰芳. 维吾尔文网络查询扩展词的构建研究[J]. 计算机工程,2015,41(4):187-189.

[13] 麦热哈巴·艾力,姜文斌,王志洋,等. 维吾尔语词法分析的有向图模型[J]. 软件学报,2012,23(12):94-100.

[14] LEGOWO N,ROJALI S. Design of thesis topic search engine with information retrieval and vector space model of TF-IDF weighting[J]. Australian Journal of Basic and Applied Sciences,2013,42(4):264-273.

[15] 彭凯,汪伟,杨煜普. 基于余弦距离度量学习的伪 K 近邻文本分类算法[J]. 计算机工程与设计,2013,34(6):2200-2203.

表 5 各种检索方法的性能比较  
Tab. 5 Performance comparison  
of various retrieval methods

检索方法	性能指标/%	
	MAP	P@10
传统检索	37.1	46.2
WIG 加权	43.9	55.6
文中方法	52.7	60.8