

基于 U-P2P 元组空间知识协作模式的 中越跨语言查询扩展*

胡积宝¹, 周佳骏¹, 黄名选², 汪婷婷³

(1. 安庆师范学院, 安徽 安庆 246003; 2. 广西教育学院, 南宁 530023; 3. 拉筹伯大学, 澳大利亚 墨尔本 3139)

摘要: 针对中国与东盟国家在知识共享过程中的语言障碍问题, 详细阐述跨语言查询扩展元组空间的逻辑结构, 提出 U-P2P 网络模式下元组空间知识协作的分布式中、越跨语言查询扩展模型。在此基础上, 增加查询扩展代理 peer 的动作语言规则的定义和元组元操作指令集, 依据强化学习算法提出一种基于加权矩阵的自适应路由选择/查询机制。在 Gnutella 环境下的实验表明, 相对于传统的集中式查询而言, 该模型对中越两种语言之间的快速翻译和扩展查询具有较好的准确率和召回率, 在不同知识领域的运用中具有通用性和可扩充性。

关键词: 元组空间; 知识协作; 跨语言; 扩展查询; 对等网络

中图分类号: TP393 **文献标志码:** A **文章编号:** 1001-3695(2013)06-1742-05

doi:10.3969/j.issn.1001-3695.2013.06.037

Cross-language query expansion for knowledge collaboration based on U-P2P tuple space between Chinese and Vietnamese

HU Ji-bao¹, ZHOU Jia-jun¹, HUANG Ming-xuan², WANG Ting-ting³

(1. Anqing Normal University, Anqing Anhui 246003, China; 2. Guangxi College of Education, Nanning 530023, China; 3. Latrobe University, Melbourne 3139, Australia)

Abstract: Aiming to the problem of language barrier about knowledge sharing between China and ASEAN countries, this paper expounded the logical structure of tuple space on cross-language query expansion, and proposed a distribute Chinese-Vietnamese cross-language query expansion model for knowledge collaboration on tuple space based on U-P2P network model. It added the definition of the action language rules and instruction set of tuple operations about query expansion proxy, gave an adaptive routing/query mechanism with weighted matrix according to reinforcement learning algorithm. Relative to centralized query, the experiments in Gnutella network show that the model has a better recall and precision in rapid translation and extends query between Chinese and Vietnamese, and possesses versatility and scalability in the use of the different fields of knowledge.

Key words: tuple space; knowledge collaboration; cross-language; query expansion; P2P network

0 引言

中国—东盟自由贸易区正式成立顺应了经济全球化、区域经济一体化的浪潮, 中国—东盟博览会的召开则更加增进了中国—东盟成员各国之间经济、文化、知识的融合。一方面, 互联网作为中国和东盟各国合作和交流的一个重要的经济、文化和社交平台, 有其自身的优越性。中国—东盟各国公共可以访问的 Web 站点的数量急剧增长, 越来越多的中国—东盟组织和企业拥有自己的网站, 涵盖了丰富的资源信息内容。为了实现互利和双赢的共同目标, 需要各国之间能够通过 Web 网络互通科研、经贸等各类信息, 传承和分享知识, 促进社会的交流。P2P 网络环境下的 Web 社区的出现为组织、企业之间的合作提供了便捷的环境。在这个环境下, 信息共享变得更加简单明确。另一方面, 由于互联网具有多语种、跨文化的特征, 而中国和东盟各国是具有不同语种的

不同区域, 有汉语、英语、越语、老挝语、泰语、印尼语、柬埔寨语、缅甸语等, 因此 Web 的信息分享和交流存在着较为严重的语言沟通障碍问题。跨语言信息检索技术的应用为该问题的解决提供了一种有效的途径。

1 研究现状及问题的提出

跨语言信息检索 (CLIR) 指的是用户以自己熟悉的语言来构建和提交查询, 系统检索出符合用户需求的包含多个语种的相关信息^[1,2]。跨语言信息检索相关问题的研究起源于 20 世纪 70 年代, 直到 90 年代随着互联网的出现, 其研究才得以迅速发展, 取得了相关的研究成果, 得到了众多研究机构和国际会议的重视和讨论。1994 年在台湾国立大学诞生了第一篇关于 CLIR 的博士论文^[3]。1996 年在瑞士苏黎世召开的国际信息检索大会 (SIGIR) 上举行了第一届 CLIR 国际研讨会, 各种跨语言信息检索方法在此次研讨会上被提出^[4]。2000 年以

收稿日期: 2012-10-10; **修回日期:** 2012-11-21 **基金项目:** 国家自然科学基金资助项目 (61262028); 广西自然科学基金资助项目 (2012GXNSFAA053235); 安徽省安庆师范学院青年科研基金资助项目 (KJ201212)

作者简介: 胡积宝 (1979-), 男, 讲师, 硕士, 主要研究方向为网络通信; 周佳骏 (1974-), 男, 副研究员, 硕士, 主要研究方向为对等网络、信息对抗和信息检索 (jiren_191@sina.com); 黄名选 (1966-), 男, 教授, 硕士, 主要研究方向为数据挖掘、信息检索和查询扩展; 汪婷婷 (1980-), 女, 讲师, 博士研究生, 主要研究方向为智能网络。

后,CLIR的研究处于快速发展时期,至今已形成了相对完整的研究体系,跨语言搜索引擎技术得到飞速的发展。

跨语言查询扩展是改善和提高跨语言信息检索系统检索性能的核心技术之一,它指的是利用多种计算机技术,把与原查询相关的词或者词组添加到原查询,得到比原查询更长的新查询,再次检索,以弥补原查询信息的不足。目前跨语言查询扩展主要有以下三种方法:

a) 基于伪相关反馈的跨语言查询扩展。微软亚洲研究院提出的两步伪相关反馈法^[5]是目前跨语言信息检索中使用最广泛的查询扩展方法。美国 Johns Hopkins 大学的计算机科学家 McNamee 等人^[6]以及武汉大学信息管理学院的吴丹等人^[7]在文献^[5]基础上对相关反馈的跨语言查询扩展进行了深入研究,在实验中比较了翻译前查询扩展、翻译后查询扩展以及两者结合的不同效果,取得很好的效果。韩国信息与通信大学的 Seo 等人^[8]研究了英语—韩语跨语言的相关反馈的查询扩展。美国 George Washington 大学的 Bellaachia 等人^[9]提出基于伪相关反馈和 WordNet 结合的英语—阿拉伯语跨语言查询扩展方法,通过伪相关反馈方法获得候选扩展词后,还要与 WordNet 的词作共现检查,以达到消歧的目的。

b) 基于潜在语义的跨语言查询扩展。在跨语言信息检索中应用潜在语义索引技术^[10]也取得了令人满意的效果。广西大学闭剑婷等人^[11]提出一种基于潜在语义分析的跨语言信息查询扩展方法,利用聚类提高扩展文本集合精度,并用潜在语义分析实现无须翻译的查询扩展,减轻翻译歧义带来的影响,获得了较好的检索性能。厦门大学郭文等人^[12]也提出基于语义词典和马尔可夫随机场的潜在语义扩展相结合的跨语言信息查询扩展方法,对提高检索性能有帮助。

c) 基于本体的跨语言查询扩展。吴芳等人^[13]提出了一种基于双语本体(ontology)的跨语言查询扩展模型,利用领域本体在知识表示和知识描述方面的优势,实现基于概念的智能查询扩展,效果良好。

由此可见,国内外学者主要关注跨语言查询扩展的语法和词法以及查询翻译等方面的研究^[14-17],对于跨语言查询扩展在具体网络环境下的应用关注较少。

P2P网络是建立在Internet上的覆盖网。在资源汇集方面,P2P将网络上所有节点的计算能力和存储空间以及各个节点的文件汇合到一起,形成分布式的系统。网络上的各个节点既是客户端又是服务器,能较好地解决待检索资源单点失效问题,具有较强的可扩展性和可靠性,因此被广泛应用在资源共享和协同工作研究领域。

本文主要关注P2P网络下的基于Web环境的跨语言查询扩展的应用,提出一种非结构化P2P(U-P2P)元组空间知识协作的中、越跨语言查询扩展模型CV-QECLIRM,作为相应的东盟博览会跨语言信息检索平台原型,为中国—东盟博览会提供有效的跨语言信息服务。

2 中越跨语言查询扩展系统结构

2.1 U-P2P 模式下的查询扩展宏观思路

U-P2P模式下的查询扩展方法类似于Freebase,核心部分是具有链接指针的包含属性的元组(资源)。用UML图例来描述资源和端点(链接点)的属性,如图1所示。

文档document1和document2之间依靠包含属性值的端点(endpoint1)和相连接的边(edge)发生联系。由图1可以得到

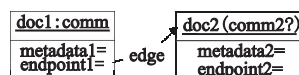


图1 查询扩展模式的元组描述

查询扩展的一般逻辑形式如式(1)。

$$\left\{ \begin{array}{l} x \mid \exists (Y_0, \dots, Y_n), \text{edge}_{b_0}(X, \text{comm}_0, \text{endpoint}_0, Y_0) \\ \wedge \bigwedge_{i=1}^{n-1} \text{edge}_{b_i}(Y_{i-1}, \text{comm}_i, \text{endpoint}_i, Y_i) \\ \wedge \text{edge}_{b_n}(Y_n, \text{comm}_n, \text{endpoint}_n, \text{doc}) \end{array} \right\} \quad (1)$$

其中: $(b_0, \dots, b_n)$ 是布尔型参数,表示边的形式(l型或r型); $(\text{comm}_0, \dots, \text{comm}_n)$ 表示查询词集合,可能是多次出现的特定的查询词; $(\text{endpoint}_0, \dots, \text{endpoint}_n)$ 则为约束端点属性; doc 为文档名。

本文的做法是分解查询 query 的路径 n 为 n 个子查询,分别对应 n 个相应的边(edges),同时生成 n 个代理 peer 去执行调用相关子查询任务。

根据式(1)得到查询词“跨语言信息检索”的查询扩展逻辑描述如式(2)。

$$\left\{ \begin{array}{l} x \mid \exists (Y_0, Y_1), \text{edge}_r(X, \text{title}, \text{expandtitle}, Y_0) \\ \wedge \text{edge}_l(Y_0, \text{title}, \text{expandtitle}, Y_1) \\ \wedge Y_1 = (y_1, \dots, y_n) \\ \wedge R_{\text{doc}}(y_1, \dots, y_n) \\ \wedge y_{i_{\text{title}}} = (\text{"查询扩展"}) \end{array} \right\} \quad (2)$$

其中: title 表示需要查询的题名关键词; expandtitle 为题名扩展关键词; R_{doc} 表示符合查询 query 的一组 doc 文档; i_{title} 表示 R_{doc} 中 title 属性的值。

式(2)的图形结构描述如图2所示。

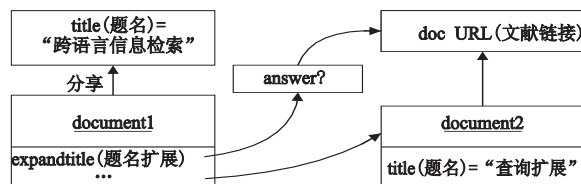


图2 扩展查询图形结构描述

用户首先输入查询条件 query 为文档题名“跨语言信息检索”,查询扩展代理 peer 收到查询,一方面直接用该查询中使用的关系名和属性名检索实例级索引获得 query 的结果。另一方面,peer 按照元组 endpoint 指针的指向找到与 query 语义相似的改写后的查询词,此后 peer 再以改写后的查询检索实例级索引获得进一步的数据。

2.2 跨语言查询扩展模型(CV-QECLIRM)

本节以汉语和越南语之间的查询扩展为例,利用知识协作和共享元组空间形式,在U-P2P模式下构造跨语言查询扩展模型CV-QECLIRM。CV-QECLIRM的系统结构如图3所示。

用户接口模块负责接收来自用户的查询输入以及系统返回的查询结果。CV-QECLIRM采用翻译前扩展策略,当用户输入中文的查询 query,查询扩展模块将用户的查询词先行扩展并分发给翻译模块,翻译为若干越南语扩展查询词,交付本地搜索模块进行匹配搜索。

本地搜索模块负责处理本地共享文件、书签、本地网络爬虫爬过的页面索引。如果查询来自用户,则首先将查询搜索结果送到组合模块,与来自其他 peer 节点查询点击回复合并,同时发送给邻居节点信息模块,以协助评估邻居节点对此查询的响应。如果查询来自其他 peer 节点,则搜索结果会被邻居节点信息模块和通信模块送回到此 peer 节点。

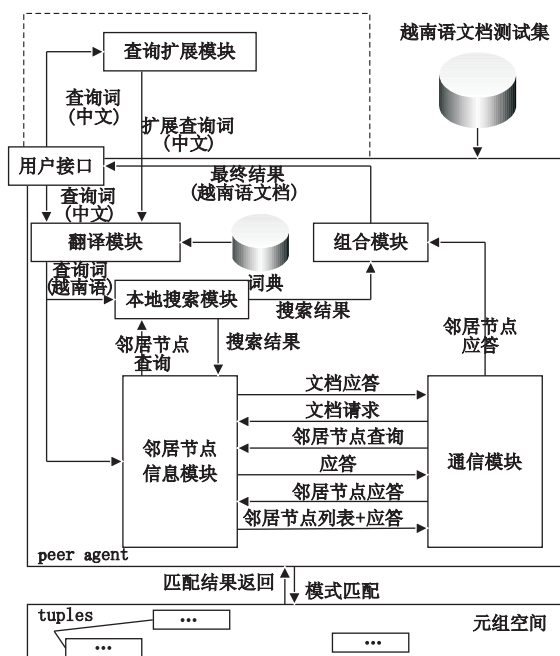


图3 CV-QECLIRM的系统结构

邻居节点信息模块负责处理如何应答邻居节点的查询请求,评判当前 peer 节点到每一个邻居节点的路由路径,判断哪一个是最优化的邻居节点。该模块具有能存储已知节点信息并实时动态更新的数据库,能够处理大量邻居节点查询请求的响应信息。无论接收到来自用户还是其他 peer 节点的查询请求,都将从已知 peer 节点信息数据库中选择一组邻居节点进行转发。

组合模块则应用启发式教育策略,负责将从本地搜索模块获取的点击数(hits)与邻居节点获得的进行合并重新排序,同时删除具有重复信息的 URL 页面。

通信模块是 peer 节点应用层和网络层的接口,其任务主要是本地以及外部 peer 之间的查询转发、答复及结果传递。

2.3 元组空间结构及元组操作算法

共享 P2P 文件的基本操作是借助不同的代理 peer 进行元组交换实现的。每个元组编码的首个字段是动词(verb),用来识别具体的请求动作。随后的几个字段是适当参数。执行元组空间的文件共享操作命令及其对应描述如表 1 所示。

表 1 元组空间的文件共享操作命令及描述

操作	元组描述	输出接口
publish	('publish', communityld, document)	user interface
remove	('remove', communityld, documentld)	user interface
search request	('search', communityld, expr, documentld)	user interface network adapter
search response	('searchresponse', communityld, documentld, peerld, document *, queryld)	repository network adapter
download	('retrieve', peerld, communityld, documentld)	user interface network adapter
download response	('retrieveresponse', communityld, documentld, document)	repository network adapter

一方面,网络适配器传送查询请求到其他符合条件的 peer,处理来自其他 peer 的查询请求,同时对本地存储库进行评估。另一方面,网络适配器也执行文档的下载任务(查询存储在本地 peer 的文档)。

用户接口收集来自用户的输入,并将查询请求提交到本地存储或网络适配器处理。当用户完成查询时,用户接口收集和存储接收到的查询响应到一个临时集合供用户查询。

存储本地文档的本地信息库则对 peer 提出的查询请求进

行响应(包括查询文件或文件属性值)。

每一个 peer 从用户接口或网络接收查询请求并且转发到另一个 peer,其工作的目标是使用文件共享操作指令(如发布、删除、搜索、下载)完成信息查询扩展的功能。

基于寻径模式的查询扩展响应算法如下:

```

function ANSWERQUERY0( dir, comm, endpoint, doc)
if dir = I then
    DEREFERENCE( doc ) /* ensure that doc is stored locally */
    uri-list ← comm. LOOKUP( doc, endpoint )
    for all uri ∈ uri-list do
        results ← results ∪ DEREFERENCE( uri )
    end for
    return results
else if dir = r then
    if p not member-of comm then
        { ( Pi, comm ) }i=1...m ← Community. SEARCH( communityId =
comm )

        j = select in [ 1...m ] /* select a peer */
        Community. DOWNLOAD( pj, comm )
    end if
    results ← comm. SEARCH( endpoint = doc )
    return results
end if
end function

```

Algorithm

```

function ANSWERPATHQUERY( [ b0, ..., bn ], [ comm0, ..., commn ], [ endpoint0, ..., endpointn ], doc )
Ylist ← ANSWERQUERY0( bn, commn, endpointn, doc )
if n = 0 then
    return Ylist
else
    for all Y ∈ Ylist do
        answerlist ← answerlist ∪
        ANSWERPATHQUERY( [ b0, ..., bn ], [ comm0, ..., commn-1 ], [ endpoint0, ..., endpointn-1 ], Y )
    end for
    return answerlist
end if
end function

```

2.4 基于规则的 peer 代理语言

查询代理 peer 的动作语言规则可以用指定的一组向量链表描述。向量链表的开始是规则匹配模板的规则头(head),链表尾部(tail)存储的是规则被激活时执行的元组指令序列(input/output)。元组指令集由元组的元操作(in/out/read)组成。元组只包含字符串字段,并且可以包括变量。元操作指令由模板定义,一个单一的 peer 可以具有多个规则,共享相同的一组变量。

为了 peer 能有效执行 U-P2P 文件共享查询,peer 元组空间文件共享命令的匹配需要固定标准 U-P2P 模板,如表 1 所示。可以从搜索链表的反馈中检索到查询的元数据值。

查询代理 peer 寻径模式的语言结构描述如图 4 所示。

查询词长度为 n 的查询具有 $n+2$ 条规则,具体查询语言规则的定义如下:

```

rule 0:
head: read( PathQuery, ?doc, id0 )

```

```

tail;out(SubQuery,dir_n,$doc,comm_n,endpoint_n,id_i)
rule 1:
head;in(SubQueryAnswer,?doc,id_i)
tail;out(SubQuery,dir_{n-1},$doc,comm_{n-1},endpoint_{n-1},id_2)
...
rule i:
head;in(SubQueryAnswer,?doc,id_i)
tail;out(SubQuery,dir_{n-i},$doc,comm_{n-i},endpoint_{n-i},id_i)
...
rule n+1:
head;in(SubQueryAnswer,?doc,id_{i+1})
tail;out(PathQueryAnswer,?doc,id_0)

```

执行查询的过程如下:a)创建和激活查询代理 peer;b)初始化查询请求,元组(PathQuery,doc,id₀)输出到元组空间 Tuple;c)最终的查询结果和模板(PathQueryAnswer,?doc,id₀)进行匹配。

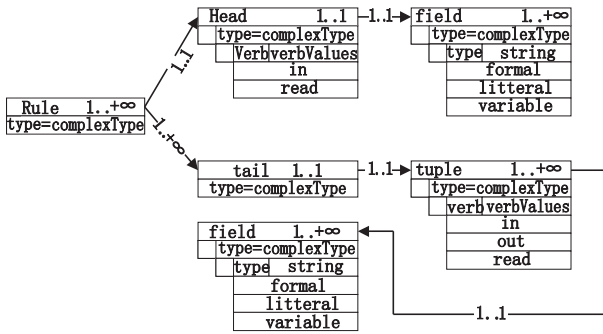


图4 查询代理peer寻径模式的语言结构

3 节点更新/查询的自适应路由机制

CV-QECLIRM中的查询代理 peer 包含查询矩阵 w^f 和查询扩展矩阵 w^e ,分别存储查询和查询扩展信息,且两个矩阵具有相同的结构。开始时,每个 peer 的邻居节点查询矩阵为空,一旦参与查询和转发过程,将会相应更新矩阵的值。

1) 查询矩阵 w^f_{ip} 权值最初更新于 peer 对邻居节点请求的响应,即对查询词 i 的查询请求被提交或被转发。当 peer 接收到查询 query 的响应后,会比较接收到的查询词点击率和本地的查询词点击率,从而对 peer 的查询词 $i \in \text{query}$ 作出一个估计。

2) 查询扩展矩阵 w^e_{ip} 权值通过查询扩展响应的交互进行更新。查询扩展矩阵是实现查询扩展的重要环节。

在元组的向量空间中,向量空间模型通过分配权重(weight)给查询的路由路径,从而将路径表示为权重的向量。本文依据强化学习规则对邻居节点查询矩阵中查询路径权重进行更新,更新式为

$$w_{ip}(t+1) = (1-\gamma)w_{ip}(t) + \gamma \frac{S_p + 1}{S_l + 1} \quad (3)$$

其中: t 为时间步长; S_p 和 S_l 分别为节点peer和本地查询词点击率(hits)平均值对于查询query的响应; γ ($0 < \gamma < 1$)为学习的效率。查询词 i 取决于依据查询query和查询矩阵共同作用的查询学习规则。

邻居节点的更新算法描述为:

- 计算从邻居节点获得的查询词点击率平均值 S_p 和本地接收到的查询词点击率平均值 S_l 。
- 依据式(1)更新 query 中的查询词。
- 若 $S_l < S_p$,则依据式(1)更新扩展 query 中的词,这些词

是邻居节点中点击率超过本地查询词点击率的词语。

d) 发送“成功”信号给邻居节点。

e) 添加能够响应“成功”信号的新节点到已知节点列表。

新的查询路径的路由原则是系统按照式(4)计算出 peer 和 query 之间的查询词相似度 δ 自组织进行。

$$\delta(p, Q) = \sum_{i \in Q} [\alpha \cdot w_{ip}^f + (1-\alpha) \cdot w_{ip}^e] \quad (4)$$

其中: α 是可靠性参数,控制及调节查询和查询扩展矩阵的贡献值。经过多次实验数据分析,可以得出 α ($0 < \alpha < 1$)能够比较真实地反映 w^f 和 w^e 的置信区间。

基于强化学习算法的 peer 节点更新和选择的动态过程描述如图5所示。

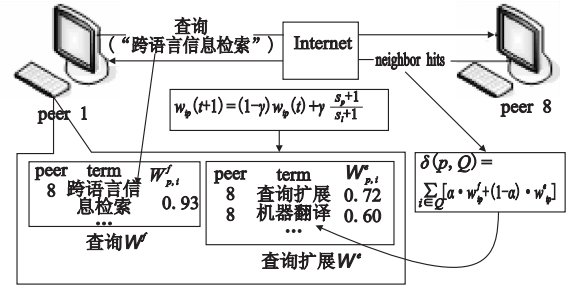


图5 强化学习算法的peer节点更新和选择

图5给出了两个 peer 节点之间查询路径选择以及相应查询词的搜索计算过程,进而得到查询词=“跨语言信息检索”/查询扩展词=“查询扩展”和“机器翻译”的相应矩阵的权值。

4 实验及数据分析

4.1 CV-QECLIRM 的效率评价标准

传统信息查询评价度量精确率 P 和召回率 R 无法捕捉到 U-P2P 网络环境下的分布式搜索过程所耗费的有效时间。以下几个因素将影响到 P 和 R 的有效度:a)没有考虑到用户在查询检索过程中等待的时间(延迟);b)网络带宽的限制;c)每个查询代理 peer 有可能接收到来自不同邻居节点的相同返回结果(重复冗余信息的反馈)。基于以上问题,本文给出一种改进的 U-P2P 网络的分布式查询扩展性能评价度量(分布式精确率 Π 和分布式召回率 Θ)。

分布式精确率和召回率分别定义为

$$\Pi = \frac{\sum_{u \in U} f(u)^{-1}}{|H|} \quad (5)$$

$$\Theta = \frac{\sum_{u \in U} f(u)^{-1}}{|R|} \quad (6)$$

其中: H 是满足查询条件的已经查询到的所有文档集合(推荐集); U 是 H 的子集,是 H 中和已查询到的文档扩展相关的文档集合; R 是整个 U-P2P 网络中所有相关的文档集合(测试集); $f(u)$ 是原始查询和 peer 返回的查询文档 $u \in U$ 之间的最大路由距离跳数。

4.2 CV-QECLIRM 的实验及数据分析

本文以广西壮族自治区图书馆《东南亚研究论文库》以及广西民族大学图书馆东盟文献信息中心馆藏的越南语文献等相关文献资料为依托,自建形成越南语的文档测试集 CV-TSet,测试集长度 32M(包含 6 339 个文档),包含 25 个主题,每个主题包括标题域、描述域和叙述域,且具备相应的查询翻译功能。实验主要采用标题域的内容作为处理基础。实验在 Gnutella 环境下得到 50 个查询的平均准确率和召回率数据,如

图 6 所示。

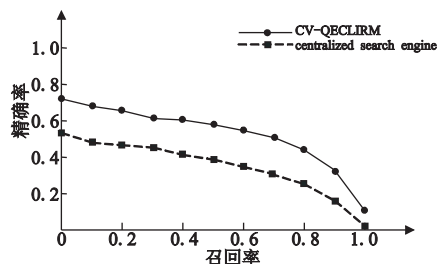


图6 查询的平均召回率—准确率折线图

实验结果显示,由于查询扩展代理 peer 采取了自适应的路径选择算法,CV-QECLIRM 的系统性能达到 71.3%,而 centralized search engine 采用的集中式查询路由算法的系统性能只有 64.8%。此外由图 6 可以看出,CV-QECLIRM 召回率—准确率曲线要明显优于 centralized search engine 曲线。

由于测试集 CV-TSet 文档中包含自然科学,尤其是信息检索方面的文献文档较多,所以在这些方面的 query 中表现的结果较好,而在其他专业领域方面取得的测试效果则较为一般,这些领域的 query 对系统整体的精度性能有一定影响。在后继的实验中,测试文档集应该尽可能囊括中国—东盟各国之间人文、社会、经济以及其他各学科知识,便于加强中国—东盟各国之间的知识协作和交流共享。

5 结束语

本文从社会化搜索以及文件共享的角度出发,融合 U-P2P 网络节点分布式存储信息特征,在传统跨语言查询扩展研究的基础上提出了一种基于 U-P2P 元组空间知识协作的分布式中、越跨言查询扩展模型,重点讨论了 U-P2P 模式下知识查询扩展的元操作过程、查询扩展代理 peer 的语言结构和规则,以及 peer 节点的自适应选择/查询路由算法。实验表明,模型克服了传统集中式方法在效率、知识表示方面的不足,有效地提高了知识的获取率,并且可以应用到各个不同的专业及社会知识领域,具有较好的通用性和扩充性。

参考文献:

- [1] TALVENSAARI T, LAURIKALA J, JARVELIN K, *et al.* Creating and exploiting a comparable corpus in cross-language information retrieval[J]. *ACM Trans on Information Systems*, 2007, 25(1): 40-45.
- [2] TALVENSAARI T, PIRKOLA A, JARVELIN K, *et al.* Focused Web crawling in the acquisition of comparable corpora[J]. *Information Retrieval*, 2008, 11(5): 427-445.
- [3] 陈信希. 跨语言资讯检索导论[D]. 台北: 台湾国立大学, 1994.
- [4] GREFENSTETTE G. Cross-language information retrieval[M]//The Kluwer International Series on Information Retrieval. Boston: Kluwer Academic Publishers, 1996.
- [5] GAO Jian-feng, NIE Jian-yun, ZHANG Jian, *et al.* TREC-9 CLIR experiments at MSRCN[C]//Proc of the 9th Text Retrieval Evaluation Conference. 2001: 343-353.
- [6] McNAMEE P, MAYFIELD J. Comparing cross-language query expansion techniques by degrading translation resources[C]//Proc of the 25th Annual International Conference on Research and Development in Information Retrieval. New York: ACM Press, 2002: 159-166.
- [7] 吴丹, 何大庆, 王惠临. 基于伪相关反馈的跨语言查询扩展[J]. *情报学报*, 2010, 29(2): 232-239.
- [8] SEO H C, KIM S B, RIM H C, *et al.* Improving query translation in English-Korean cross-language information retrieval[J]. *Information Processing and Management*, 2005, 41(3): 507-522.
- [9] BELLAACHIA A, AMORTIJANI G. Enhanced query expansion in English-Arabic CLIR[C]//Proc of the 19th IEEE International Conference on Database and Expert Systems Application. Washington DC: IEEE Computer Society, 2008.
- [10] LITTMAN M L, DUMAIS S T, LANDAUER T K, *et al.* Automatic cross-language retrieval using latent semantic indexing[C]//Proc of AAAI-97 Spring Symposium Series. 1997: 18-24.
- [11] 闭剑婷, 苏一丹. 基于潜在语义分析的跨语言查询扩展方法[J]. *计算机工程*, 2009, 35(10): 49-53.
- [12] 郭文, 陈毅东, 赵欣. 跨语言信息检索中的查询扩展[J]. *心智与计算*, 2009, 3(1): 1-8.
- [13] 吴芳, 丁玲, 张杰, 等. 跨语言信息检索中基于本体的查询扩展模型研究[J]. *计算机教育*, 2009(17): 122-124.
- [14] ARCHANA M. Conventional ontology for cross language information system[C]//Proc of IEEE International Conference on Computing, Electronics and Electrical Technologies. [S. l.]: IEEE Computer Society, 2012: 991-993.
- [15] KUMAR B A. Profound survey on cross language information retrieval methods (CLIR) [C]//Proc of the 2nd IEEE International Conference on Advanced Computing & Communication Technologies. [S. l.]: IEEE Computer Society, 2012: 64-68.
- [16] MANIKANDAN B. A novel approach for cross language information retrieval[C]//Proc of the 3rd IEEE International Conference on Electronics Computer Technology. [S. l.]: IEEE Computer Society, 2011: 34-38.
- [17] ZHUHADAR L, NASRAOUI O, WYATT R. Multi-language ontology-based search engine[C]//Proc of the 3rd IEEE International Conference on Advances in Computer-Human Interactions. [S. l.]: IEEE Computer Society, 2010: 13-18.

下期要目

- ❖ 无线传感器网络中网络层故障容错技术研究进展
- ❖ 武器装备故障预测建模方法选择研究
- ❖ “教”与“学”优化算法研究综述
- ❖ 基于区域—频道访问度的民意热点信息挖掘算法
- ❖ 基于聚类划分的两阶段离群点检测算法
- ❖ 面向个性化推荐系统的二分网络协同过滤算法研究
- ❖ 基于仿射的有向网络聚类算法
- ❖ 基于主题和链接分析的微博社区发现算法
- ❖ 情感分析中极性副词的自动扩展

- ❖ 分数阶线性系统的稳定性证明
- ❖ 基于误差传递建模的仿真可信性控制
- ❖ 基于加权因子的 PID 型模糊控制器分析及设计
- ❖ 一类离散非线性时延系统的滑模变结构控制研究
- ❖ 需求和回收努力扰动下闭环供应链定价与协调
- ❖ 基于系统动力学的自然灾害应急物流逆向回收决策研究
- ❖ 基于制造特征的产品族零件工时定额制定方法
- ❖ 多车型回程车辆调度问题的 ADP 算法研究
- ❖ 基于混合信息滤波的粒子滤波 SLAM 算法
- ❖ 基于改进蛙跳策略的 Map-Reduce 作业调度算法