●刘伟成　孙吉红

跨语言信息检索进展研究*

摘　要　根据研究对象的变迁,国外关于跨语言信息检索的历程主要分为三个阶段。跨语言信息检索目前的主要解决方法是在单语言信息检索系统上增加一个语言转换机制。解决查询条件与查询文档集间的语言障碍有五种不同的技术路线。跨语言信息检索主要研究热点有翻译歧义研究、翻译资源构建、专有名词识别与音译研究等五个领域。图 2。参考文献 25。

关键词　跨语言信息检索　查询语言　语言搜索引擎　语料库

分类号　G354

ABSTRACT　According to the evolution of research objects, there have been three stages in the development of cross-language information retrieval. The main solution for cross-language information retrieval is to add a language transformation mechanism to single-language information retrieval systems. There are five technical approaches to solve the language barriers between retrieval conditions and retrieval file sets. The authors also point some hot issues in the research of cross-language information retrieval, such as the studies on ambiguous meanings, the construction of translation resources, the recognition of proper names and transliteration. 2 figs. 25 refs.

KEY WORDS　Cross-language information retrieval.　Search language.　Language search engine.　Language base.

CLASS NUMBER　G354

传统的信息检索系统主要是针对单一语种的文档集实现,一般是使用用户最为熟悉的语种作为查询语言。跨语言信息检索(cross language information retrieval, CLIR)就是以某种语言检索另外一种语言表达的文献资源的方法和技术。随着信息在全球的自由流动,跨语言信息检索已成为世界范围内一个亟待解决的问题。

1　跨语言信息检索研究的发展历程

根据研究对象的变迁,国外关于跨语言信息检索的研究历程主要分为如下三个阶段。

1.1　萌芽阶段:针对国际联机检索的跨语言检索研究

有关跨语言信息检索效率的最早试验结果,是在 1969 年由 Cornell 大学的 Salton 所记录[1],Salton 通过翻译已有英语概念列表中的一些单词为德语,来构建一个多语概念列表,而后利用该表扩充其 SMART 信息检索系统。在 1973 年的研究中,他实现了英法多语概念列表,并通过在建立一个共同的概念集之后,单独开发针对每种语言的相应部分,来达到更为完整的覆盖范围。在此项研究中,他采取包含 52 篇摘要的法英并行语料库并使用包含 16 个已翻译查询的集合。1972 年 Pevzner 对英俄跨语言信息检索进行了研究,表明受控叙词表对于查询翻译是非常有效的。1978 年国际标准组织颁布了关于多语言叙词表(multilingual thesauri)的国际标准 ISO5964,该标准在 1985 年进行了修改。

上述研究主要是针对国际联机检索进行的,而当时联机检索系统并不普及,国际互联网尚不为人们所知,人们对网络信息的需求并不强烈,研究工作也没有取得重大的进展。

1.2　发展阶段:基于互联网的跨语言信息检索实验系统的出现

跨语言信息检索研究真正活跃起来并取得成果,是在互联网迅猛发展的 20 世纪 90 年代后期,国际上先后有许多相关论文发表,一些实验性跨语言信息检索技术相继问世。1997 年由德国语言技术研究室人工智能研究中心研究开发的 Mulinex 系统是世界上第一个成功地运用跨语言自动翻译技术,使人们利用本国语言就能有效获取网上其他语言信息的跨语言网络信息检索系统。根据 Doug Oard 的不完全统计,到 2002 年 4 月共有 17 个跨语言信息检索系统问世[2]

在这一阶段,跨语言信息检索的理论和技术也得到飞速发展,学者们提出了多种语言转换策略和匹配机制。1990 年,潜语义索引(Latent Semantic Indexing, LSI)技术被应用于跨语言信息检索。1994 年诞生了第一篇关于跨语言信息检索的博士论文(Khaled Radwan)[3]。1996 年同义词表应用于 CLIR (ETH Zurich)。1997 年,卡内基梅隆大学语言技术研究所在跨语言信息检索的理论与实践中首次采用广义向量空间模型(Generalized Vector Space Model, GVSM)[4]。

随着各种试验系统的不断出现,各国学者希望通过评价各种不同的 CLIR 系统,从而进行比较和改进,在这一阶段出现了几个著名的测试平台和测评会议。文本检索会议(TREC)在 1997 年开始将跨语言检索的测评作为中心议题之一。NTCIR(NACSIS Test Collection for IR Systems)成立于 1998 年,NTCIR 第一次工作组会议于 1999 年 8 月在日本的东京举行,该会议主要侧重于亚洲语(如中文、日语、朝鲜语)的跨语言信息检索问题的研究[5]。CLEF[6](Cross Lan-

* 本文系国家自然科学基金资助项目(70473067)研究成果之一。

guage Evaluation Forum)的第一次会议于2000年9月在葡萄牙首都里斯本举行,每年举办一次。该论坛侧重于欧洲范围内跨语言检索问题的评价。

1.3 大型商用阶段:跨语言搜索引擎技术的飞速发展

多语性是互联网世界的特色之一,依据1996年ETHNOLOGUE目录上的统计,全世界语言数目高达6703种(Grimes,1996)[7]。另据其他学者的研究,目前网络上有160种语言的信息。从搜索引擎诞生的那一天起,寻求对多语种的支持就是在激烈的竞争中制胜的法宝之一。2001年的文献[8]显示,当时的Google支持的语言有14种,Altavista支持的语言有25种,而雅虎则推出了数十种本地化的搜索引擎。这些宣称支持多语搜索的搜索引擎其实只是多个单语模式搜索的融合,即用户只能以一种语言提问,返回同一种语言的信息。用户如果需要在多种语言中查找信息,就必须同时使用多种语言提问。

最近两年,真正的跨语言搜索引擎得到飞速发展。目前,通过Google的"使用偏好"选项可以进行跨语言信息检索,Google支持的查询语言有115种,可以检索用35种语言所写成的网页。

2 跨语言信息检索的基本框架

跨语言信息检索就是以单一语言描述的用户查询来检索多语种的信息资源,实质就是单语言的用户查询与多语言的信息(文档)表示之间的匹配。目前主要的解决方法就是在单语言信息检索系统的基础上增加一个语言转换机制(查询翻译,文献翻译或不翻译)。作为传统信息检索的一种扩展,跨语言网络信息检索综合了多种信息处理成果,在进行语言转换之前还要进行一些前期的文本预处理,比如语言识别、信息抽取、分词、信息标引、文本分类等等。跨语言信息检索的基本框架如图1所示。

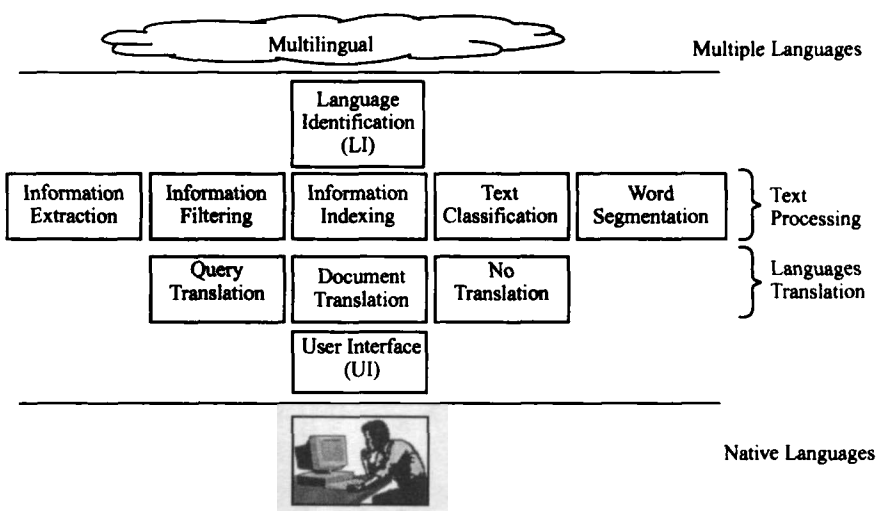

图1 跨语言信息检索的基本框架

3 跨语言信息检索的类型和技术

跨语言信息检索的方法分类如图2所示[9]。一般来说,解决查询条件与查询文档集之间的语言障碍有5种不同的技术路线:同源匹配(cognate matching),查询翻译(query translation),文献翻译(document translation),中间语言技术(interlingual technique),不翻译(no translation)。

3.1 同源匹配

同源匹配根据两种语言的语词拼写形式或读音相似度来判断其中一种语言语词的意义,不进行任何翻译。例如,康奈尔大学的Buckley等人[10]开发了一个英语—法语匹配程序,它将英语单词视为可能拼错的法语单词,以此来实现英语提问式与法语文献的匹配。这种简单方法的效果不错,检索效率可以达到单语检索的60%,而且几乎不需要任何词典知识。然而这种方法只适用于具有相同词源的语言,比如英语和法语,但对于中英文来说则不适用。同源匹配可以单独使用,而多数情况下是与其他策略结合使用,比如在中英文跨语言信息检索中可以用于外来语的音译或反向音译。

3.2 查询翻译

查询翻译将用户输入的提问式(源语言)翻译为系统支持的语言(目标语言),然后再将目标语言的提问式提交给匹配模块,进行单语言信息检索。它是目前最为常用的策略,Kwok[11]认为这种方法简单而有效,所以本文主要围绕基于查询的跨语言信息检索开展研究。其优点是能够在线快速执行,主要缺点是提问式通常很短,语境信息很少,难以消除歧义。每个提问词被其所有可能的译法所替代,翻译模糊性问题严重,故控制翻译的模糊性是设计有效的提问式翻译技术的一个关键问题。一种办法是只翻译短语,因为短语翻译通常表现出较少的模糊性。研究表明,短语识别策略能够大幅度提高检索效率。微软研究院的Jianfeng Gao等人[12]介绍了一种使用统计模型识别名词性短语以提高提问式翻译质量的方法。另一种办法是通过用户的介入(利用交互式用户界面)也可以有效控制翻译的模糊性。Davis和Ogden开发的QUILT系统能够将英语提问词的西班牙语翻译显示给用户,具有西班牙语知识的用户可以对翻译进行识别和判断。Mark Davis开发了一个交互式搜索引擎ARCTOS,通过选择性用户界面,用户可选择将英语、法语、德语或意大利语的提问词翻译为系统支持的其他语言,然后对候

选翻译进行选择，修改提问式并发送给某个特定语言的检索模块。

3.3 文献翻译

文献翻译与查询翻译正好相反，是指先将多语言的文献信息集转换成与查询相同的语言，再进行单语言信息检索。其主要优点是：①由于具有完整的文献语境，故可以提高翻译质量；②可以离线执行。缺点是速度很慢，且需要将文献库中的文献翻译为系统支持的每一种语言，这使得文献库的规模不可能很大。

相对于查询翻译方法，采用文献翻译方法的CLIR系统要少得多。“欧共体远程通信和信息处理技术”(EU Telematics)计划下的Twenty-One项目组开发的Twenty-One系统使用的主要跨语言方法就采用了文献翻译方法，并以查询翻译作为辅助。Gachot, Lange和Yang (1996)使用Systran翻译系统来产生文件对应双语语料库，供区域性回馈使用。对于文献翻译还有另外一种方案，即对每个文献所对应的向量进行翻译来代替文献翻译，目前还没有看到这方面的研究和成果[13]。

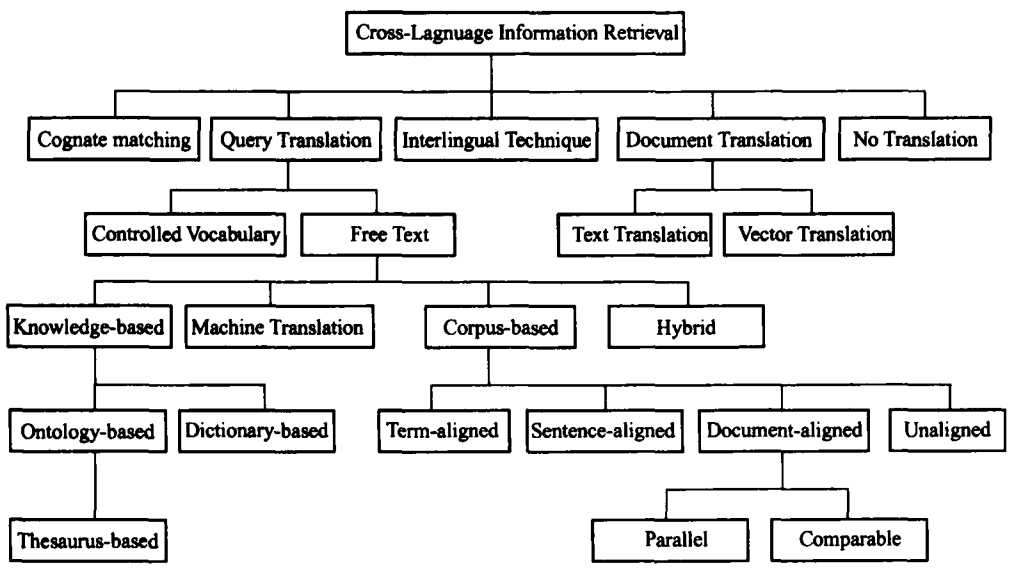

图2 跨语言信息检索方法分类

3.4 中间语言方法

在跨语言信息检索中，解决语言障碍的基本方法是两种语言之间的翻译，然而所有的翻译方法都离不开机器翻译、双语词典、语料库等作为翻译的语言基础。但是，在跨语言信息检索中可能会碰到这样的情形：两种语言直接翻译的语言资源不存在，例如在TREC中很难找到德语和意大利语之间直接对等的语言资源。为此研究人员提出了一种利用中间语言或中枢语言进行翻译的方法：将源语言翻译成中间语言(可以是一种或多种)，然后再将中间语言翻译成目标语言(利用多种中间语言时需要合并)。

前面提到的中间语言是自然语言，此外，人工语言在跨语言信息检索中也经常作为中间语言出现。其典型代表是基于多语种词表的CLIR技术。它将文献和提问式都翻译为受控词表中的语词。MNIS-TextWise实验室的“概念中间语言文献检索”(Conceptual Interlingua Document Retrieval)项目小组开发的CINDOR系统使用了较为独特的语间转换技术来实现CLIR。该系统以WordNet的同义词群“synsets”为基础，通过将几种语言的同义词都链接到表示对应概念的“synset号”上，建立了一个名为“概念中间语言”的概念表示知识库。这样，系统就可以将文献标引词和提问词都转换为“synset号”，从而跨越了语言障碍。Ruiz和他的同事[14]将“概念中间语言”应用于中英文跨语言信息检索，他们通过现有的双语词典和平行语料自动生成汉语概念与英语概念(WordNet)之间的连接。欧洲的MACS (Multilingual Access to Subjects)计划使用多语言词表技术将各种主题词语言按照概念联系起来，该系统允许用户使用其中一种语言(英语，法语，德语)来同时检索多种语言的文献。概念中间语言既可以用于表示文献也可以用于表示提问。这种方法的缺点是新的概念或术语无法及时地补充到概念词表中。

3.5 不翻译

目前不通过翻译进行跨语言信息检索的技术有潜语义索引(Latent Semantic Indexing, LSI)和广义向量空间模型等方法。

Deerwester等人于1990年在单语信息检索中提出了LSI方法。Landauer和Littman同年提出了跨语言潜语义索引(CLLSI: Cross-Language Latent Semantic Indexing)[15]的信息检索技术。它的基本思想是首先通过将有代表性的文档与其对应的翻译文档联系起来形成训练文档集，然后利用奇异值分解技术(SVD: Singular Value Decomposition)对双语检索词—文档关联矩阵进行奇异值分解，获得双语文档集的特征信息以及检索词用法上的映射关系，即构造出不同语种的潜在语义空间，最后根据平行文档中语词的用法特征检索出另一种语种的相关信息。Dumais等人于1997年进行了英法跨语言检索的实验，在其训练过程，英法双语文件、英语词汇、法语词汇都被映射到一个向量空间中，随后加入不同语言的文件。沿用LSI的基本想法，尽管这些术语是不同语言描述的，但是可以进行语义上的匹配比较，不需要翻译转换。过去有

多人在不同的语言匹配上做过实验,比如 Berry&Young(1995)以希腊文—英文、Oard (1996)以西班牙文—英文等进行了试验,验证了这种方法比较有效。

广义向量空间模型的基本思想是根据双语训练文档集分别建立源语与目标语的检索词—文档关联矩阵,在计算查询条件和文档的相似度时,考虑将经典的向量空间模型与两个关联矩阵相结合在源语言与目标语言之间实现映射关系。

卡内基梅隆大学语言技术研究所的 Carbonell 等人[16]曾对语料库导向的翻译方法 (TMT)、伪相关反馈(Pseudo Relevance Feedback,PRF),广义向量空间模型(GVSM)和 LSI 等四种方法,在相同的条件下,做了一系列的实验,结果显示 GVSM 比 LSI 稍微好一点,这两种方法又都比 TMT 和 PRF 好。

4 跨语言信息检索主要研究热点与领域

4.1 跨语言信息检索中的翻译歧义研究

翻译的歧义性是跨语言信息检索的关键问题,对检索效率有重要影响,这也是国外学者广泛关注的研究领域,所依赖的语言资源主要有词典、主题词表、本体、语料库等。

Davis[17]尝试以词性进行消歧,效果不错,平均准确率提高了 37%,达到了单语检索的67.3%。Chen 等人[18]以共现模型(co-occurrence)分析翻译歧义,以虚拟语境(pseudo context)模型分析目标多义,在 TREC-6 的评测中,与仅处理翻译歧义性相比,使检索效率提高了 10.11%。平行语料常用于翻译的词义消歧,但是平行语料加工困难,不易获取。Akira 等人以 Web 文献为语料,利用词汇间的共现信息实现了同样的检索效率,在其试验中检索的平均准确率达到了手工翻译的 97%。Myung-Gil Jang 等人利用从目标文献中获得的互信息(mutual information)统计进行查询翻译的消歧,在这里,互信息不仅用于选择翻译词汇而且对翻译后的查询词赋以权重,通过 TREC-6 标准测试集实验证明,检索效率分别达到了单语检索的 85% 和手工消歧的 96%。

4.2 跨语言信息检索中的翻译资源构建研究

翻译资源的优劣对于跨语言信息检索的性能有着重要影响,所以在跨语言信息检索研究中,国外学者对于翻译资源的构建以及相互之间的比较进行了深入研究。在 CLIR 中,常用的翻译资源有机器翻译系统、双语词典、本体和语料库等。

平行语料(parallel corpora)已经成为 CLIR 应用的一种重要的翻译资源,但是目前所能得到的平行语料库非常有限。国外许多学者探讨通过互联网来自动获取平行语料。双语词典的获取和编撰一直依靠手工方式,更新的周期长而且代价高,这促使许多学者从事基于双语语料库的翻译词典自动获取研究,并取得了丰硕的成果。Jiangping Chen[19]采用汉语词典、英语词典、双语词典等语言学资源构建词典知识库(lexical knowledge base,LKB),试验结果证明 LKB 方法非常有效,能够显著提高查询翻译的正确率,减少未登录词的百分比。Christopher 等人进行了自动构建汉英双语主题词表的研究。WordNet 已成为事实上的语义词典国际标准,为了开发跨语言应用,各国竞相在 WordNet 的基础上开发跨语言本体,其中最著名的就是欧盟资助的 EuroWordNet,目前支持二十几种欧洲语言。Chen 等人[20]在 WordNet、双语词典、平行语料库等翻译资源的基础上自动构建汉英双语本体,并在 TREC 测试集上进行跨语言信息检索试验,结果表明检索的平均准确率提高了 10.02%。

4.3 跨语言信息检索中的专有名词识别与音译研究

由于翻译词典的覆盖度,未登录词(out of vocabulary, OOV)一直是机器翻译和跨语言信息检索的重要问题,专有名词的翻译更是挑战。Chen 等人(1998, 2000), Knight 和 Graehl (1998), Wan 和 Verspoor (1998)都相继提出机器音译的方法来处理这个问题。Yan Qu 等人[21]提出了由英语到日语片假名的音译方法,利用英语语音词典和概率规则来获取候选音译,并通过日语单语语料库自动确认,最终将英语—日语音译词对添加到双语词典中,在 CLEF 测试集上进行日英跨语言信息检索实验,使检索的平均准确率提高了 2.5% ~64.8%。Nasreen 提出一种统计模型(selected n-gram model)进行英语—阿拉伯语间的音译,在 TREC2002 的 CLIR 评测中,无论是对专有名词音译还是对所有未登录词音译都证明是有效的。S. Y. Jung 等人采用隐马尔可夫模型(HMM)进行英语到韩语的音译,召回率达到了 87.5%。Paola 等人[22]将专有名词的音译用于跨语言声音文件的检索,效果也是显著的。

4.4 跨语言信息检索中的翻译技术研究

跨语言信息检索涉及查询语种和检索语种两个基本的概念。查询语种是用户查询请求所属语种,检索语种是检索目标对象所属语种,如何能够在这两者之间建立起沟通的桥梁是目前跨语言信息检索技术研究最核心和关键的问题。所有的这些工作分别从以下 3 个方面来展开:机器翻译系统、基于语料库的方法以及基于字典的方法。20 世纪 90 年代,Brown 等人提出了一种基于机器翻译的统计方法;随后 Nie、BBN 等也相继提出了各自的概率论翻译模型,采用基于平行文本的方法来解决 CLIR 的问题;Kwok、Hedlund 等着重研究了翻译过程中的字典查找模式:Ballesteros 和微软的研究人员在自己的工作中都使用到了基于共现的统计方法,等等[23]。

4.5 跨语言信息检索中的系统评价研究

除了理论和技术外,评估也是跨语言信息检索系统发展过程的重要一环。跨语言信息检索主要有 3 个测试平台:TREC、NTCIR、CLEF。TREC 的跨语言测试项目始于 1997 年,以英语为主,并搭配一些战略语言,如 2001 年的英语—阿拉伯语跨语言信息检索;NTCIR 始于 1999 年,以亚洲语言(中、日、韩)为主,目前已经举办了五届;CLEF 则以欧洲语言为主,从 2000 至今已经举办了六届。这些评测项目和会议对于推动跨语言信息检索的研究和应用起到了良好作用,每次会议都吸引了众多的研究机构和企业,参赛单位的论文和评比结果会在网站上公布。

5 跨语言信息检索的发展趋势与面临的挑战

近年来 CLEF、NTCIR 等评测会议上,很多语言之间(特别是欧洲语言)的检索效率已经达到了单语检索的 80% ~ 90%。但是对其他语言而言跨语言信息检索仍然不是一个已经解决了的问题。闵金明等人[24]对跨语言信息检索的研究现状进行了概括性总结,明确指出该研究中需要解决的关键问题是:WSD(词义消歧)和 OOV。根据我们所掌握的资料和对研究现状的观察,跨语言信息检索还存在许多具有挑战性的研究方向和领域。

(1)利用网络资源进行OOV的翻译是一个研究热点。目前为止,这个问题并没有得到有效解决。现有的方法往往利用特定语言间的网络共现特征来寻找翻译,开发与特定语言无关的OOV翻译方法是值得关注的方向。从机器翻译的角度看,网络可以看成一个大型的可比语料库,通过一些相似特征收集两种语言的网络文本,借助已有的通用双语词典,然后运用机器翻译中的文本对齐技术来实现OOV的翻译应该是一个有效的方法。

(2)本体在词义消歧中的应用研究。词义消歧是一个传统的AI难题,除了传统的WSD方法,越来越多的学者借助WordNet、FrameNet等本体资源来实现跨语言检索中的消歧。多语言本体的构建及其在CLIR中的消歧应用是一个研究热点,也面临许多挑战。

(3)跨语言信息检索和分布式信息检索在很多方面存在着共性,分布式信息检索通常的研究领域包括资源表示、资源选择、结果合并,而跨语言信息检索在这些方面的研究不多,特别是前两者,基本处于空白状态,可以开展的工作还有很多。

(4)未来的发展趋势是多种方法的融合。该领域一个很明显的趋势就是越来越多的研究人员开始考虑结合本文提到的跨语言信息检索实现方法中的几种来进一步改进查询翻译的精度。笔者在博士论文中的案例分析和实证研究也证明了这一点[25]。

(5)跨语言信息检索的检索结果处理上仍然有大量的工作需要完成,即如何将结果文档以用户能够接受的形式表现出来。交互式跨语言信息检索和跨语言摘要技术是有前景的研究方向。

(6)跨语言检索系统的评价一直都是沿用传统检索系统的评价方法。在这种情况下,同样的方法如果因为翻译资源的不同往往会得到不同的结果,这对于有效地评价是不公平的,所以需要加入跨语言检索特有的评价参数,以区分不同的翻译资源带来的不同检索效率。

(7)跨语言信息检索内容不应当局限于文档检索,可以扩展到跨语言图像检索、跨语言语音检索、跨语言交互式检索、跨语言问答系统、跨语言新话题发现和跟踪等。目前CLEF已经在相关领域做了有益的探索。

参考文献

1 Salton G. Automatic processing of foreign language documents. Journal of the American Society for Information Science. 1970,21(3)

2 [2006-01]. http://www.ee.umd.edu/medlab/mlir/systems.html

3 陈信希. 跨语言资讯检索导论. [2005-06-20]. http://nlg3.csie.ntu.edu.tw/courses/IR/slides/Chapter_10_Cross_Language_Information_Retrieval.ppt

4,16 Carbonell, J., et al. (1997). A Realistic Evaluation of Translingual Information Retrieval Methods. Personal communication, LTI, CMU

5 [2006-01]. http://research.nii.ac.jp/ntcir/index-en.html

6 [2006-01]. http://clef.iei.pi.cnr.it

7 Grimes, B. F. (1996). Ethnologue Language Name Index. http://www.sil.org/ethnologue/names

8 黄国才. 跨语言综合搜索引擎设计. 现代图书情报技术, 2001 (4)

9,25 刘伟成. 基于查询翻译的跨语言信息检索研究. [博士学位论文]:武汉大学,2006

10 Buchley, et al. Using clustering and superconcepts within smart: trec6. In Proceedings of the Sixth Retrieval Conference, 1997

11 Kwok, K. L. (2001). NTCIR-2 Chinese, cross language retrieval experiments using PIRCS. Proceedings of NTCIR-2 Workshop, section 5, (pp. 14-20)

12 Gao, J., et al. TREC9 CLIR experiments at MSRCN. Paper presented at the TREC-9, 2000

13 王妙娅,赖茂生. 跨语言信息检索中的询问翻译方法及其进展研究. 现代图书情报技术,2005 (4)

14 Ruiz, M. E., et al. (2001). CINDOR TREC-9 English-Chinese evaluation. Paper presented at the TREC-9. NIST Special Publication 500-249, Gaithersburg: National Institute of Standards and Technology (pp. 379-399)

15 Landauer T. K., Littman M. L. Fully automatic cross-language document retrieval using latent semantic indexing. In: Proc. of the 6th Annual Conf. of the UW Center for the New Oxford English Dictionary and Text Re search, 1990. 31 - 38

17 Davis, M. W. New Experiments in Cross-Language Text Retrieval at NMSU's Computing Research Lab. Proceedings of TREC 5, 1997:391 - 407

18 Hsin-His Chen, Guo-Wei Bian, Wen-Cheng Lin. Resolving translation ambiguity and target polysemy in cross-language information retrieval. Proceedings of the 37th annual meeting of the Association for Computational Linguistics, Jun. 1999, College Park, Maryland, pp. 215 - 222

19 Jiangping Chen. The construction, use, and evaluation of a lexical knowledge base for English-Chinesecross-language information retrieval. PhD Dissertation of Graduate School of Syracuse University. December,2003

20 Hsin-Hsi Chen, Chi-Ching Lin, and Wen-Cheng Lin. Building a Chinese-English WordNet for Translingual Applications. ACM Transactions on Asian Language Information Processing, Vol. 1, No. 2, June 2002, pp. 103 - 122

21 Yan Qu, et al. Automatic Transliteration for Japanese-to-English Text Retrieval. SIGIR'03, July 28-August 1, 2003, Toronto, Canada. pp. 353 - 360

22 Paola Virga, Sanjeev Khudanpur. Transliteration of Proper Names in Cross-Language Applications. SIGIR'03, July 28-August 1, 2003, Toronto, Canada

23 王进,陈恩红,张振亚. 基于本体的跨语言信息检索模型. 中文信息学报,2004,18(3)

24 闵金明,孙乐,张俊林. 重新审视跨语言信息检索. 中文信息学报,2006,20(4)

刘伟成 武汉科技大学管理学院讲师,博士。通讯地址:武汉科技大学(本部)215信箱。邮编430081。

孙吉红 副教授,武汉大学信息管理学院2005级博士研究生。通讯地址:武汉大学信息管理学院。邮编430072。

(来稿时间:2007-03-12)