DOI:10.13530/j.cnki.jlis.150022

近10年来国外多语言信息组织与检索研究进展与启示*

司 莉 庄晓喆 贾 欢

摘 要 近年来国外在多语言信息组织与检索研究领域取得了显著进展。本文以WoS、ACM、Emerald、Elsevier、ProQuest、Springer等数据库收录的文献为基础，对近10年来该领域的研究进行述评。国外研究重点关注以下问题：多语言本体构建与协调，基于关联数据的多语言语义网建设，跨语种语言资源和知识组织系统互操作，多语言文本分类与聚类，多语言环境下的用户信息行为，多语言信息检索模型，多语言信息检索方法与技术，多语言信息检索系统开发及评估，特定领域的多语言信息检索，交互式多语言信息检索。对我国的启示主要体现在：加强实证研究方法的应用，开发面向实用的多语言信息检索系统，注重基于语义的信息组织与检索研究，拓展特定学科领域应用研究。图1。参考文献68。

关键词 多语言信息组织 跨语言信息组织 多语言信息检索 跨语言信息检索

分类号 G250

A Review of Multilingual Information Organization and Retrieval Research Abroad in the Last Ten Years

SI Li, ZHUANG Xiaozhe & JIA Huan

ABSTRACT

The multilinguality of online information has become the major obstacle to the sharing of online information resources across nations. To overcome this obstacle, there is an urgent need for developing methods and technologies of multilingual information organization and retrieval. Recent years witnessed significant progress in the research on both cross lingual and multilingual information organization and retrieval, which dates back to the 1960s but remains relatively inactive for almost 30 years. The purpose of this study is to conduct a systematic review on the status quo of multilingual information organization and retrieval research at abroad based on literature collected from foreign academic databases.

We performed a thorough search in the following ten databases: ACM Digital Library, EBSCO Academic Search Premier, Elsevier ScienceDirect, Emerald, IEEE/IET Electronic Library, ProQuest Digital Dissertations & Theses, Sage Premier, Springer Link, Wiley InterScience, and Web of Science. "Multilingual information retrieval", "crosslingual information retrieval", "cross language information retrieval", "bilingual information re-

* 本文系教育部人文社会科学重点研究基地重大项目"基于内容的多语言信息组织与检索研究"(编号:14JJD870001)研究成果之一。(This article is an outcome of the project "Research on Content-based Multilingual Information Organization and Retrieval" (No. 14JJD870001) supported by Humanities and Social Science Foundation, Ministry of Education of the People's Republic of China.)

通信作者:庄晓喆,Email:978130697@qq.com,ORCID:0000-0002-3304-1867(Correspondence should be addressed to ZHUANG Xiaozhe, Email: 978130697@qq.com, ORCID: 0000-0002-3304-1867)

trieval", "translingual information retrieval", "multilingual information access", "MLIR", "CLIR", "multilingual ontology", "multilingual interoperability" were selected as search terms in title and keyword search. The literature published time is limited to the years after 2004. References in the literature were utilized for retrospective search. The retrieval dead line is March 1st, 2015. The number of related full-text literature is 1 028 in total, and articles written by Chinese scholars in English are excluded.

The literature review consists of three parts: the progress of multilingual information organization research at abroad; the progress of multilingual information retrieval research at abroad; the summarization and enlightenments of multilingual information organization and retrieval research at abroad. Through literature review, it is shown that multilingual information organization research mainly focused on the construction and the alignment of multilingual ontologies, the utilization of linked data in building multilingual semantic network; the interoperation among multilingual language resources and multilingual knowledge organization systems, and multilingual text categorization and clustering. As for multilingual information retrieval research, hotspots include user information behavior under the multilingual environment, multilingual information retrieval models, multilingual information retrieval methods and techniques, multilingual information retrieval tools and their evaluation, multilingual information retrieval in specific domains, and interactive multilingual information retrieval. We find that the research covers various research topics, highlights abundant empirical studies, evolves towards semantic-based information organization and retrieval, and orients to specific domains. Inspired by foreign counterparts, Chinese scholars in this field should strengthen the employment of empirical research methods, make more efforts to develop application-oriented multilingual information retrieval systems, pay more attention to the study of semantic-based multilingual information organization and retrieval, and broaden the study of domain-specific multilingual information retrieval.

However, due to the insufficient background knowledge in linguistics and computer science, a minority of the articles could not be fully understood. Besides, non-full-text literature is not in the scope of our literature review, which may lead to an incomplete reflection of the research status.

Our study attempts to explore and summarize the progress of multilingual information organization and retrieval research outside China since 2005. We identified the research hotspots in the domain of multilingual information organization and retrieval, introduced both theoretical and practical progress, and gave suggestions for the further study of multilingual information organization and retrieval in China. 1 fig. 68 refs.

KEY WORDS

Multilingual information organization. Cross-language information organization. Multilingual information retrieval. Cross-language information retrieval.

0 引言

以自身熟悉的语言构造查询请求，一站式获取其他语种的信息无疑是信息用户所乐见的。为实现这一愿景，学界就多语言信息组织与检索开展长期探索。1969年，Salton利用基于英、德双语概念列表构建的SMART检索系统进行了首次多语言文本检索实验，开多语言信息检索研究之先河[1]。20世纪90年代起，信息检索迈入网络信息检索阶段。随着互联网的日益普及，网络用户的地域分布趋于广泛。截至2014年6月，亚、非各国的网络用户已占全球网络用户总数的55.5%，是欧洲、北美地

区网络用户数的近两倍[2]。与此同时，网络信息的语种分布亦呈现多元化趋势。2003 年，72%的网站用英语表述站内信息[3]，但 2014 年，这一比例已降至 55%[4]。网络信息的多语种化与用户所掌握语言的差异性、有限性之间的矛盾进一步凸显。这不仅降低了信息检索的检全率、检准率，使置身信息海洋的人们更加难以捕获满足需求的信息，而且极大地制约了国家和地区间的信息交流与共享，阻碍数字鸿沟的弥合，也不利于小语种信息资源的挖掘与利用。

因此，多语言信息组织与检索愈加受到不同领域研究者的关注。近 20 年来，相关研究项目层出不穷，如欧盟资助的欧洲多语言信息检索项目（European Multilingual Information Retrieval，EMIR）和 i2010 数字图书馆项目（Digital Libraries Initiative，DLI）、美国资助的跨语言信息检测、抽取和总结项目（Translingual Information Detection，Extraction and Summarization，TIDES）等。美国的文本检索会议（Text Retrieval Conference，TREC）和信息检索特殊兴趣小组会议（Special Interest Group on Information Retrieval，SIGIR）、欧盟的跨语言评价论坛（Cross Language Evaluation Forum，CLEF）、日本的信息检索系统测试集会议（National Center for Science Information Systems Test Collections for Information Retrieval，NTCIR）、印度的信息检索评估论坛（Forum for Information Retrieval Evaluation，FIRE）等国际会议也均聚焦于该领域研究。本文对 2005 年以来国外有关多语言信息组织与检索的研究进行总结、梳理，以期为国内研究者在该领域的探究与实践提供借鉴。

对于跨语言检索与多语言检索的概念，Peters 等指出，跨语言信息检索旨在以某一语种构建的查询条件从多语种信息集合中检出另一语种的信息，多语言信息检索则旨在以任意语种的查询条件从多语种信息集合中发现并获取任何语种的信息，前者是后者的组成部分，后者是前者的积累效应[5]。在文献调研中，笔者发现部分研究者并未严格区分这两种概念，甚至将两者视为同一概念，故本文也将跨语言检索纳入研究范畴。

本文以 Web of Science 数据库为主要信息源，以 ACM、EBSCO、Emerald、Elsevier、IEEE/IET、ProQuest、Sage、Springer、Wiley 等全文数据库为其他信息源，选取“multilingual information retrieval”“crosslingual information retrieval”“cross language information retrieval”“bilingual information retrieval”“translingual information retrieval”“multilingual information access”“MLIR”“CLIR”“multilingual ontology”“multilingual interoperability”等为关键词，在题名和关键词字段中进行检索，并将文献的发表时间限定为 2005 年以后。对检索结果进行去重并剔除无关结果后，共得到全文文献 1 028 篇。检索截止时间为 2015 年 3 月 1 日。按其发表年份进行统计，结果如图 1 所示。

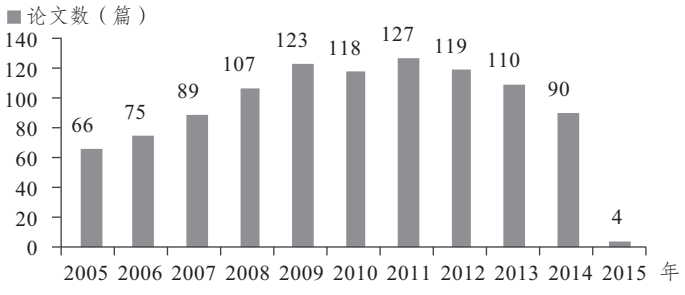

图 1　国外多语言信息组织与检索领域研究论文年度分布

由图1可见，2005年至2011年间，国外多语言信息组织与检索领域的研究论文数总体呈现显著上升趋势，至2011年达到高峰。2012年以后发文量有所下降。2010年前，相关研究主要集中于多语言本体的构建、多语言文本分类、多语言信息检索模型、多语言信息检索的方法与技术等方面。2010年后，研究主题进一步拓展，关联数据应用、多语言文本聚类、多语言信息检索技术与系统评估、多语言信息检索在特定领域的应用、交互式多语言信息检索也成为研究热点。

1　多语言信息组织研究进展

多语言信息组织的研究与实践可追溯到多语种叙词表的编制。随着语义网、关联数据、互操作等技术工具被广泛应用于信息组织中，多语言信息的组织也正向语义化、集成化方向发展。

1.1　多语言本体的构建与协调

多语言本体以概念为基本构成单位，通过概念的属性、概念间的关系及相关约束表达概念的语义，有效缓解了语义的曲解问题。Dragoni等探讨了有机农业领域多语言本体的演化，认为本体演化是由实体弃用（entity deprecation）、本体映射、本体扩充、实体特殊化、实体泛化和实体翻译等情境组成的工作流程，强调领域专家、知识工程专家、语言专家的密切协作。另外，他们还开发了多语言领域本体演化的支持工具MoKi[6]。Salim等指出多语言本体的价值在于可实现对多语言文本的语义标注，提出多语言本体的两种构建方法（基于对网页内容及概念间关系的抽取，基于对传统知识组织系统的复用），并将该本体应用于网络文本的标注，进而实现多语言门户网站的检索功能[7]。

本体协调（ontology alignment）核心是本体映射，可支持不同语言本体间的互操作，以达成对多语种信息的语义描述。Trojahn等提出了一个面向多语言描述型逻辑本体的映射框架。该框架由翻译代理模块（用于将源语言本体译为目标语言本体）、协调代理模块（用于交换不同代理产生的映射结果）、论证模块（用于计算合适的映射集）组成，可实现单一词型概念和复合词型概念的跨语言映射。他还通过从源本体（葡萄牙语）到目标本体（英语）的映射验证该框架的可行性[8]。Fu等将机器翻译技术应用于英汉领域本体的映射，并根据实验结果分析基于机器翻译技术的多语言本体映射的优势与不足，认为引入语义相似度计算方法可有效提升映射的准确度[9]。

1.2　基于关联数据的多语言语义网建设

关联数据可用于标识、发布和连接各类数据、信息、知识。将其应用于多语言信息组织中，可建立不同语种概念间的关系，并指引用户访问资源，是建立多语言语义网的又一重要手段。Pérez等认为RDF数据集的多语言性要求关联数据向多语言化方向发展，并就数据源的特征分析、URI和IRI的设置、RDF数据建模、RDF数据集生成中的语言识别与编码、RDF数据集的互联与发布等问题展开阐述[10]。Gayo等指出多语言关联开放数据的设计需要着重考虑实体命名、参引（dereferencing）、资源标识、资源描述、资源链接（尤其是语言不同但内容相同的资源）、关联数据重用等问题，并分别提出了相应解决方案[11]。Caracciolo等探讨多语言叙词表的关联数据化问题，他们将农业科学多语言叙词表AGROVOC改造为以SKOS-XL模型表示的概念框架，利用SKOS映射机制进行AGROVOC到LCSH、Eurovoc、RAMEAU、DDC以及DBpedia等同样以关联开放数据形式发布的外部资源的映射，以可参引的URI标识词表中的概念，通过URI链接建立这些概念与其他叙词表、分类法和知识库中概念间的关联[12]。Ehrmann等尝试整合不同语种的语义知识库，建成“语言学领域的关联开放

数据云”（linguistic linked open data cloud）。该库使用词典模型 lemon 表述概念的含义，以 SKOS 的概念类表示同义词集合，再以关联数据形式发布于网络，集成了维基百科、Omegawiki、WordNet、Open Multilingual WordNet 等资源，可提供对涉及 50 余种语言、930 多万个概念的详细说明，并揭示其间的词义关系[13]。

1.3 跨语种语言资源与知识组织系统的互操作

互操作指两个及以上系统间相互使用已被交换的信息的能力[14]。跨语言知识组织系统的互操作则是实现不同语种信息跨库检索的基础，也有助于充实概念间的语义关系。在跨语种语言资源的互操作方面，Witt 等提出实现跨语种语言资源互操作的两种途径：基于转换与基于中间语言的途径。前者以映射为手段，实现两种语言间的互译；后者通常选用相对泛化、抽象的描述语言或本体连接两者，更强调对语言结构的分析[15]。Soria 等利用中间语言索引（interlingual index）实现分布式语义词库 ItalWordNet（意大利语）与 Sinica BOW（英语、汉语）的互操作，可实现扩充现有语义资源、验证语义关联、创建新的语义资源等目标[16]。

在跨语言知识组织系统的互操作方面，Nicholson 和 McCulloch 回顾了英国高层叙词表项目 HILT 的第二阶段工作，探讨不同语种知识组织系统的互操作，以实现一站式检索服务。该项目以 DDC 为映射转换中心，将多部分类法、叙词表、标题词表与 DDC 类号建立映射关系。可将用户输入的关键词反馈为可能与其需求相关的 DDC 概念，再使用与概念相关联的 DDC 类号执行查询。作者还设计了基于 SRW 协议的互操作系统架构[17]。Ma 等通过补充对象属性、数据类型属性的方式扩展了 SKOS 模型，使用该模型完成地质年代多语言叙词表的本体化描述，并将改造后的叙词表应用于在线地质图集成服务系统中，使用户可通过一种语言检索涉及 7 种语言的多语种地质图资源[18]。

1.4 多语言文本分类与聚类

多语言文本分类旨在将不同语言的文本/文档自动归入事先建立的分类体系中。Gliozzo 与 Strapparava 提出了一种通过从可比语料库中获取多语言域模型（multilingual domain model，即由多语种词汇组成的词串的集合）来实现多语言文本分类的方法，该方法无需人工干预，也无需引入双语/多语词典等外部跨语种语言资源[19]。Amine 与 Mimoun 利用在线语义词库 WordNet 的 2.1 版进行英语和西班牙语文献的文本分类，首先利用机器翻译将西班牙语文献译为英语，同时将 WordNet 改造为本体，再为 WordNet 中的概念构建同义词集合，并建立集合间的上下位关系，形成概念范畴体系，然后生成待分类文献的概念矢量并赋予权重，最后计算概念矢量与类目轮廓间的相似度[20]。

多语言文本聚类的目标是完成不同语言文本/文档的归类，但分类体系并非人工事先建立，而是自动形成。Kiran 等以维基百科作为外部多语言知识库开展多语言文献聚类，使用由关键词向量、类别向量、外链向量、信息框向量构成的向量空间模型描述不同语种文献的特征，并计算向量空间的相似度，最终实现多语言文本聚类[21]。Cobo 和 Rocha 设计了一种针对经济领域多语言文献的聚类算法，且测试了其对于英语和西班牙语研究论文聚类的有效性。该算法借鉴蚂蚁算法的思想，使用四个向量表示文献特征，通过计算特征向量角间距的凸线性组合，得到文献的相似度，再辅以专有名词识别和词汇抽取技术完成多语言文本聚类[22]。

2 多语言信息检索研究进展

多语言信息检索研究虽然已历经 40 余年，但早期进展较为缓慢。网络环境下，语言隔阂取代空间距离成为信息交流与获取的首要障碍。受此驱动，该领域研究呈现不断上升的态势，相关成果较为丰硕。

2.1 多语言环境下的用户信息行为

了解不同用户群体的检索行为特点是开发多语言信息检索的前提。Ruiz 与 Chin 采用现场观察、访谈、网上调查等方法，探讨用户搜索以非母语语种标注的图像资源的行为，发现基于大众标注的图像检索系统由于缺乏对不同语种标签的规范机制，用户查找时困难较大[23]。Ghorab 等通过分析欧洲图书馆（The European Library，EL）检索系统的日志记录，探究拥有不同语言和文化背景用户的检索行为特征，得出的结论是，用户的检索行为因其语言和文化背景的区别而呈现较明显差异，提出应针对各国/地区用户设计、开展个性化的多语言信息检索服务[24]。Hong 通过问卷调查、访谈、检索实验等方式深入探究双语用户的网络信息搜索行为，她发现该类用户倾向于根据自身需求选择一种语言进行检索，但大多希望搜索引擎提供双语检索界面[25]。Petrelli 与 Clough 分析了具有双语表达能力的高校师生跨语言图像检索时提交的查询请求，发现用户倾向于使用在线免费机器翻译工具，查询请求不合乎语法结构是制约机器翻译准确度的重要因素[26]。

2.2 多语言信息检索模型

多语言信息检索模型是多语言信息检索系统的形式化表示。Lilleng 和 Tomassen 构建了基于特征向量的多语言信息检索模型。针对用户输入检索词，通过预先建成的目标语言本体和对目标文档的语词统计分析构建特征向量空间，以丰富用户查询的语义信息。基于该模型的检索系统能较好地解决用户查询的消歧问题，从而改善用户查询的翻译质量[27]。Ko 等提出基于独立预测模型和联合预测模型的两种跨语言问答系统输出结果的排序模型，并通过基于英—汉和英—日跨语言问答系统的实验进行验证，两种方法均可明显优化由语词组成的答案的排序结果[28]。

研究者还提出了一些专门面向多语言信息检索的模型。Ghorab 等设计了一套自适应多语言信息检索模型，该模型由流程控制器、查询适应与翻译模块、多语言信息检索模块、结果列表适应与翻译模块四部分组成，并应用作者提出的用户兴趣建模算法以及检索结果合并、重排与翻译算法。实验表明，基于该模型的多语言信息检索系统具有较好的个性化检索功能[29]。Jan 等提出了面向跨语言信息检索的统一音译检索模型（unified transliteration retrieval model），将音译相似度测量与相关度评分融为一体，同时提出了一种基于隐马尔科夫模型和统计机器翻译框架的新型音译相似度测量方法[30]。Potthast 等构建了基于维基百科的多语言信息检索模型 CL-ESA。该模型充分利用维基百科中丰富的跨语言语义表达方式，并引入显式语义分析（explicit semantic analysis）法，根据用户查询和目标文献中的语词在维基中出现的位置获取其语义信息，为其逐一添加概念向量，通过计算概念向量间的余弦相似度得到用户查询与文献主题的相似度，从而实现多语言信息检索结果的合理排序[31]。

2.3 多语言信息检索方法与技术

国外研究者围绕跨语言和多语言信息检索的方法与技术开展大量研究，主要包括机器翻译、双语/多语语料库和词典、多语言词汇与信息抽取、命名实体识别、词义消歧、查询扩展等。

2.3.1 机器翻译

机器翻译是较早运用于跨语言和多语言信息检索的技术之一，可分为基于规则的机器翻译与统计机器翻译两类[32]。Abu Shquier 等设计了一种基于规则的新型机器翻译方法，并应用于英语—阿拉伯语机器翻译系统中，此方法可更好地处理性数一致（agreement）和词汇重排列问题，且可扩展性较强[33]。Riesa 提出了针对统计机器翻译的基于句法的词对齐模型，该模型将词对齐系统与基于句法的机器翻译解码器相整合，依托分层翻译算法，使机器翻译系统能持续地自动获得并熟悉翻译规则，他通

过阿拉伯语—英语，汉语—英语信息互译实验，证实了该模型的可行性和可扩展性[34]。此外，Tufis 探讨了针对翻译资源不足的小语种信息的机器翻译策略，指出利用可比语料库的信息抽取技术可获取大量平行语句对，从而有效解决这一问题[35]。Kumaran 论述了复合机器音译系统的设计方法，该系统将单个音译模块串行式或并行式地连接起来，翻译效率显著提高，且无需构建平行名称语料库即可使用[36]。

2.3.2 双语/多语语料库和词典

双语/多语平行语料库和可比（对比）语料库因有利于改善翻译质量而被广泛应用于跨语言、多语言信息检索中。Talvensaari 等建设了一个瑞典语/英语新闻信息语料库，并以该对齐可比语料库为基础开发基于用户查询翻译的跨语言检索系统。他们认为，语料库实质上是跨语言的相似性叙词表，可支持用户查询扩展[37]。研究者还尝试将以维基百科为代表的开放式多语言网络百科当作可免费获取的语料库素材。Otero 和 López 选取考古学领域同时包含英语、西班牙语、葡萄牙语版本的维基百科词条，通过转换源文本为 XML 结构、篇章对齐、关键词抽取等步骤，把半结构化、分类体系的维基百科成功改造为三个特定领域的可比语料库[38]。

利用机读双语/多语词典，将用户查询请求翻译为目标语种后再执行检索是实现跨语言与多语言信息检索的又一策略。Levow 等指出，双语词典乃至双语词汇列表虽结构简单，但在跨语言信息检索中（尤其是在词汇抽取、用户查询翻译方面）却发挥着不可替代的作用。他们分别将权重映射（weight mapping）算法、证据映射（evidence mapping）算法、翻译后语词再分隔（post-translation resegmentation）方法、翻译后查询扩展（post-translation expansion）方法运用于基于双语词典的查询翻译中。他们所进行的四项跨语言检索实验（英—德、英—法、英—汉、英—阿拉伯）表明，采用上述跨语言信息检索策略可进一步优化检索效果[39]。

2.3.3 多语言词汇与信息抽取

多语言词汇与信息抽取有助于使多语言信息检索系统更准确理解目标语言文献的内容。Valderrábanos 等介绍了欧盟委员会资助的跨语言文本信息检索项目 LIQUID 实施中采用的自动化术语抽取方法：首先开发词汇的术语抽取工具，该词汇能用于准确描述特定文献的内容主题，接着将抽取的术语与先行构建的领域本体中的概念建立语义连接关系[40]。Lefever 等讨论如何从平行语料库中抽取双语词汇，他们设计了基于次句（sub-sentential，通常指短语）对齐系统的词汇抽取功能模块，并通过基于统计方法的过滤器对抽取的词汇进行精选[41]。Hecking 等采取浅层句法分析（shallow parsing）和深层、浅层句法分析相结合的方法进行多语言军事领域文献的文本分析与信息抽取，发现后者的效果优于前者，他们还据此开发了文献内容分析原型系统 ZENON，并为之增添基于语义等价表征的逻辑推理机，以进一步优化信息抽取的水平[42]。

2.3.4 命名实体识别

命名实体识别（named entity recognition）又称专名识别，是指识别文本中具有特定意义的实体。多语言环境下，命名实体识别的作用在于捕获不同语言中意义相对应的命名实体。Klementiev 和 Roth 提出了一种从双语可比语料库中自动发现命名实体的算法，基于该算法的命名实体识别省去了冗长的机器学习阶段，较为快捷，且适用于发现小语种文本中的命名实体[43]。Richman 和 Schone 则构建了基于维基百科的多语言命名实体识别系统，他们将维基百科词条的文本视为已经过人工标引处理的多语语料库，以此作为多语言命名实体识别的机器学习训练集，通过维基百科自有的分类结构确定命名实体的类型，进而依托英文版词条与其他语种词条的语义连接查找英文命名实体在其他语言中对应的命名实体[44]。

2.3.5 词义消歧

词义消歧是理解自然语言的必经步骤，可

更好地理解用户的查询请求，同时提高检索精度。Pinto 等提出了一种基于朴素贝叶斯模型的跨语言词义消歧方法，他们使用双语统计词典计算源词汇可被翻译为目标词汇的概率，并在词义消歧过程中自动实现词汇的替换[45]。Guyot 等设计了三种词义消歧算法，其中两种为基于 WordNet 的消歧算法，另一种算法则始终以判别出的第一种含义为词汇的正确含义。作者将其应用于英语—西班牙语跨语言信息检索系统中，发现词义消歧算法仅在某些情况（如用户查询语句较短）下能提高用户检准率，而基于 WordNet 的算法效果较佳[46]。

2.3.6 查询扩展

查询扩展即在用户输入初始的查询请求后，自动根据查询的语义增加新的查询语句，有利于更加完整、规范地反映用户的真实信息需求，使之获取更多相关信息。Jnedie 将查询扩展分为基于用户反馈的交互式查询扩展和自动查询扩展，并指出多语种叙词表和以 WordNet 为代表的语义词库均是多语言环境下用户查询扩展的重要工具。作者还介绍了词汇共现分析、基于上下文的词汇共现分析、局部上下文分析、用户日志挖掘等查询扩展技术[47]。Gavel 和 Andersson 利用 MeSH 叙词表实现了涵盖挪威、瑞典、芬兰三国的医学期刊书目数据库 SveMed+的用户查询自动化扩展功能，同时借鉴 PubMed 数据库的自动术语映射技术进行用户输入的自然语词与叙词的映射，以完成查询的翻译与规范化处理[48]。

2.4 多语言信息检索系统开发及评估

2.4.1 多语言信息检索系统开发

多语言信息检索系统是相关方法与技术的应用载体，是不同地区的用户查找、获取不同语种信息的必经渠道。Stanković等从多语言资源的整合、多语言元数据的创建、软件系统的选择与改造、多语言检索界面的设计等方面，介绍了可实现用户查询的语义扩展和语词形态扩展的多语言期刊论文全文检索系统 Bibliša 的开发流程与要点[49]。Brodeala 等构建了提供多语言检索功能的语义检索系统 SemanQuery，该系统以英语、西班牙语版本的 WordNet 为语义资源，建立两种语言中概念间的等级关系和等同关系，从而能根据用户的查询请求推断其关注领域，将跨语言的相关检索词和检索结果推荐给用户[50]。

跨语言问答系统（cross-language question answering system）是跨语言信息检索技术在自动问答系统中的应用。Dolores 等通过实例调查，比较语料库、自动翻译工具、维基、领域本体、多语词典作为多语言问答系统语言资源的优势与不足，认为多语言领域本体和维基是多语言问答系统语言资源的较好选择[51]。Ferrández 等将维基百科和多语言语义词库 Eurowordnet 嵌入跨语言问答系统 BRILIW 中，实验显示这一基于多语言本体的跨语言问答系统的检准率显著高于基于机器翻译的跨语言问答系统[52]。Cimiano 等设计了基于关联数据的跨语言问答系统模型，该模型选择以关联数据形式发布的多语言知识库本体 DBpedia 为内容资源，将用户查询转换为三元组形式并实例化，生成本体三元组，并构建相应的 SPAQUL 查询语句，获取答案，同时辅以 WordNet、MaltParser 等工具进行语义知识的抽取和用户查询的处理[53]。

2.4.2 多语言信息检索系统评估

多语言信息检索系统的评估对于验证系统的绩效，比较相关方法和技术优劣，改进现有系统，开发新型系统等方面均有重大意义。Sujatha 总结了普通信息检索系统的各种评估方法，其中包括基于上下文相似度（context resemblance）的方法，基于概率相关反馈（probabilistic relevance feedback）的方法，基于归一化折损累积增益的方法（normalized discounted cumulative gain）等，认为上述方法及评估指标大多适用于多语言信息检索系统的评估[54]。Chandra 等提出了一套多语言信息检索系统绩效的评估指标体系，主要指标包括平均检准率、单个主题平均检准率、平均倒数排名（av-

erage mean reciprocal rank)、平均折损累积增益(average discounted cumulative gain)等,并对其设计的基于多语词典的多语言检索系统开展评估[55]。Petrelli 分析在跨语言信息检索系统的各开发阶段开展用户评估的意义,指出需要注重用户评估在交互式跨语言信息检索系统建设中的应用[56]。Shiri 等组织来自加拿大阿尔伯塔大学的 15 名师生对基于叙词表的跨语言检索系统 Searchling 的界面进行评价,设计了三项检索任务,并使用录音和录像软件记录用户的检索行为,用户普遍认为系统能有效帮助其构建规范化的查询请求,同时建议在用户帮助中增加对叙词表的有关说明[57]。

2.5 特定领域的多语言信息检索

由于生物医学、地球科学、法学等学科领域开放性较强,信息交流较为频繁,研究人员对多语种信息的需求较旺盛,相关研究成果较丰富。Hanbury 等开发了生物医学领域的多语言信息检索系统 KHRESMOI,该系统可查询图书、期刊、数据、图像、网站等多种类型的信息,且可通过移动设备访问[58]。跨语言评价论坛(CLEF)自 2005 年起连续多年举办 GeoCLEF 跨语言地理信息检索大会,聚焦地球科学领域的跨语言信息检索研究。Larson 和 Gey 设计了基于逻辑回归和盲相关反馈(blind relevance feedback)算法以及基于隐含地名的地理信息标引方法,并将上述方法应用于英—德跨语言地理信息检索系统,实现了对用户查询地名关键词的自动扩展[59]。Peruginelli 和 Francesconi 则本着通过强化语义标引以实现有效的词义消歧的思路,设计多语种法律信息的检索系统模型,该模型运用基于文本分类的自动化词义消歧策略,可实现多语种法律文献的元数据检索、关键词检索和分类浏览等功能[60]。

此外,多语言信息检索的对象也逐渐由文本信息扩展至多媒体信息。其中针对图像的多语言信息检索尤其受到关注。Ménard 先后以三种方式(同时使用英语、法语受控词汇和非受控词汇;只使用双语受控词汇;只使用双语非受控词汇)标引一个图片库中的每张图片,并随机抽取其中 30 张,将 60 名用户平均分为三组,要求其逐一浏览图片后在图片库中找出这些图片(图片的标引分别采取上述方式之一),研究认为构建基于受控词表的协作式大众标注机制是多语言图像信息检索的最佳策略[61]。Tungkasthan 等介绍了 Yahoo 图像搜索引擎所使用的基于多线程控制的多语言检索框架,该框架由多语言翻译模块和多语言爬虫模块构成,前者用于处理、翻译用户查询请求,后者用来执行图像检索任务,合并检索结果并反馈给用户[62]。

2.6 交互式多语言信息检索

交互式多语言信息检索关注用户在检索过程中产生的反馈信息,是一种通过建立用户与检索系统间的有机联系完成检索任务的新型多语言信息检索模式。Oard 等利用其开发的交互式跨语言信息检索系统 MIRACLE,开展关于用户辅助查询翻译及用户辅助文档选择的研究,主要分析用户在系统的帮助下进行查询翻译和文档选择时表现出的行为特征[63]。Ahmed 和 Nurnberger 对 Mulinex、Keizai、UCLIR 等交互式跨语言检索系统进行评介,认为上述系统在用户查询的词义自动消歧、查询翻译与扩展以及检索结果的可视化呈现等方面存在不足,改善交互式跨语言信息检索系统应充分运用命名实体识别、信息检索可视化、转换生成翻译(transitive translation)等技术[64]。Zazo 等将免费的在线机器翻译工具 Google Linguistic Tools 与 Systran Online 分别嵌入英—西班牙和法—西班牙交互式跨语言问答系统中,用于完成用户查询请求与查询结果(答案)的翻译,查询结果的翻译效果与机器翻译工具的上下文语境分析能力有关[65]。Ruecker 等讨论了交互式多语言检索系统中可视化用户界面的建设问题,他们总结了可视化用户界面在交互式多语言信息

检索中的作用，包括协助用户选择检索词，保留用户已使用的检索策略，提供多样化的检索结果显示方式等，并以基于英—法双语叙词表的跨语言检索系统 Searchling 为例，介绍其用户界面的设计流程[66]。

3　总结及启示

3.1　总结

由上可见，近10年来国外在多语言信息组织与检索方面的研究呈现以下特点。

（1）研究主题广泛。研究涵盖多语言信息组织的技术手段、跨语言知识组织系统的互操作、多语言文本分类和聚类、用户的多语言信息检索行为、多语言信息检索模型、多语言信息检索方法与技术、多语言信息检索系统及其评估、特定领域的多语言信息检索等10多个领域。研究不但围绕全球主要语种的信息检索展开，而且逐步涉足不同小语种网络信息间的检索；不但针对文本信息的检索，还探讨图像等多媒体信息的检索；既运用已有的信息检索模型和技术，也积极开发新的多语言信息检索模型与技术；既面向单种方法与技术的应用，也注重分析多种技术、工具相结合的可行性，如综合运用领域本体、双语统计词典等工具改善词义消歧效果。

（2）注重实证研究方法。国外研究者始终关注多语言环境下的用户信息行为，并围绕该主题开展大量实验研究，了解用户在查找和获取多语言信息过程中所表现的行为特征。此外，在多语言信息检索方法和技术、多语言信息检索系统的评估、交互式多语言信息检索等方面，国外研究者多根据研究目的精心设计用户参与的多语言信息检索任务，开展实验，用以评估和比较不同多语言信息组织与检索方法、技术及系统的效度，并依据实验结果确定后续研究方向和系统优化策略。如跨语言评价论坛（CLEF）自2001年起先后设立交互式跨语言检索（iCLEF）、跨语言检索日志分析（LogCLEF）等主题，鼓励使用检索日志分析、用户实验、现场观察、用户访谈等多种实证方法，开展用户参与的查询翻译与检索界面优化研究。

（3）关注面向特定学科领域的应用研究。多语言领域本体已成为不同领域的多语种信息组织与检索中不可或缺的工具。以实现医学、地球科学、法学等某一学科领域信息的跨语言获取为目标的跨语言与多语言信息检索系统不断问世。跨语言评价论坛近年来亦相继增设CLEF-IP、CLEF eHealth、LifeCLEF、PAN 等议题，分别探索知识产权领域、电子健康学领域、生命科学领域、学术不端检测与作者识别领域的跨语言信息检索方法与工具的评估问题[67]。以上研究显示，国外多语言信息组织与检索研究呈现深化与细化的发展趋势，也从另一角度体现了国外研究者对实证研究的重视。

（4）语义化发展趋势明显。从海量多语言信息中获取真正契合用户需求的知识，无疑有赖于信息组织的语义化。旨在充分揭示概念间关联的多语言本体、关联数据等语义化信息组织工具已得到逐渐运用。潜式和显式语义分析、语义词库、维基百科等方法和工具已被广泛应用于多语言信息检索模型和检索系统的设计中，在文本分析、命名实体识别、词义消歧、用户查询扩展等方面取得系列成果。这使多语言信息检索向精准化、智能化（如查询扩展、联想式检索）方向迈进，以适应用户更为专深化、动态化的信息需求。

3.2　启示

多语言信息组织与检索的实现能消除信息查找过程中的语言屏障，更广泛地满足机构和个人的信息需求，促进国家、民族间的信息交流与共享，同时也是深度挖掘小语种信息资源的前提。国外研究对我国的启示主要体现在四个方面。

（1）加强实证研究方法的应用

国内该领域研究者应更加注重实证研究，

尤其是多语言环境下用户信息需求及行为研究方面。但国内相关研究很少，且研究对象局限于学术用户，目前仅有吴丹发表了两篇论文。首先，可以借鉴国外同类实证研究的流程和策略，积极探索多语言环境下不同用户群体对文本、图像、视频、语音等各类资源的信息需求及检索行为特征。其次，可结合问卷调查、用户信息行为测试、出声思考、深入访谈、人机交互以及用户日志分析等多种研究方法，剖析用户使用数字图书馆、搜索引擎、信息检索系统、专业数据库、电子商务、专利查新、移动信息服务等工具和服务时的多语言信息需求与行为，使多语言信息组织与检索研究和上述领域研究有机融合。再次，多语言信息组织与检索平台的建设也应以用户为中心，深入了解用户对检索系统/平台的具体功能需求与使用偏好，为平台的设计提供支持。

（2）开发面向实用的多语言信息检索系统

国外非常重视多语言信息检索系统的设计与开发，已开发出众多的实验系统，其中有一批系统已投入使用。如，2010 年 6 月正式上线的 WorldWideScience 平台支持用户使用英、汉、德、阿拉伯等 10 种语言，一站式检索来自 70 余个国家的 100 个数据库、机构库与门户网站的学术信息[68]。而迄今为止，国内开发的多语言信息检索系统多为实验系统，可公开访问者非常鲜见。加紧研发可供公众使用的综合性、专业性检索系统，是实现多语种信息检索的迫切需求。

此外，系统设计者与真实用户之间存在知识和使用习惯等方面的差异，可能影响系统的可用性。因此，在开发过程中，应尤为重视系统的可用性评估，注重用户测评工作，将其作为多语言信息检索系统的主要评估手段。可将用户测试作为主要方法，辅以专家测试。测试前，应围绕系统界面的友好性及检索的有效性、效率性与满意度，根据评价的目的设计检索任务，既可将任务限定在特定领域，也可以受试者感兴趣或正在研究的领域为测试内容，或两者兼而有之。通过反复测试与评估，不断优化系统性能。

（3）注重基于语义的信息组织与检索研究

基于语义的多语言信息组织与检索是实现知识挖掘及聚合的关键。大数据环境下，实现基于语义的信息组织与检索更为迫切。国内研究者要以基于内容的信息组织与检索理论为指导，充分利用本体协调、关联数据、维基百科等技术与工具，改善词义消歧、查询扩展与知识挖掘的效果。在此基础上，通过开发和完善多语言信息检索系统的语义扩展检索、概念联想检索、资源智能推荐等功能，揭示知识元之间的联系和脉络，实现多语种、多类型信息检索过程的交互化以及检索结果的高度整合，从而进一步优化跨语言信息检索系统的性能，推动多语言信息检索迈入基于语义的知识检索阶段。

（4）拓展特定学科领域的多语言信息检索研究

目前，国内对多语言信息组织与检索的研究较少针对特定领域。随着多语言信息组织与检索方法及技术的丰富与成熟，该领域研究应该深入各学科领域，解决具体问题。当前，随着全球化的发展和学术研究的细化，各学科领域的开放性、交叉性显著增强，研究人员的跨领域和跨地域交流与合作日趋紧密，对多语种信息的需求愈加迫切。现有的多语言信息组织与检索技术虽具有一定通用性，但由于各学科特点及各领域信息的特质不同，其在特定学科领域的应用模式并非千篇一律，往往需要加以适度改造。加速此方面研究，既有益于检验和完善多语言信息组织与检索理论、方法、技术，也有助于获取学科发展所需的多语种信息资源，从而推进各学科本身研究的发展。

总之，笔者认为，国内相关研究应遵循“理论与模型的创新→技术和方法的一般应用→技术和方法的具体应用→检索工具的开发与应用（实际应用）”这一发展路径，以构建基于内容的多语言信息检索机制为导向，面向实际问题，不断细化现有研究领域，并更加注重研究成果的转化。

参考文献

[1] Oard D W, Dorr B J. A survey of multilingual text retrieval[EB/OL]. [2014-12-29]. http://drum.lib.umd.edu/bitstream/1903/807/2/CS-TR-3615.pdf.

[2] Internet users in the world[EB/OL]. [2014-12-31]. http://www.internetworldstats.com/stats.htm.

[3] Country and language statistics[EB/OL]. [2014-12-31]. http://www.oclc.org/research/activities/wcp/stats/intnl.html?urlm=159859.

[4] Usage of content languages for websites[EB/OL]. [2014-12-31]. http://w3techs.com/technologies/overview/content_language/all.

[5] Peters C, Braschler M, Clough P. Multilingual information retrieval: from research to practice[M]. Berlin: Springer-Verlag, 2012: 5,58.

[6] Dragoni M, Francescomarino C, Ghidini C, et al. Guiding the evolution of a multilingual ontology in a concrete setting[C]//The semantic web: semantics and big data. Berlin: Springer-Verlag, 2013: 608-622.

[7] Salim J, Hashim S, Aris A. A framework for building multilingual ontologies for islamic portal[C]// Proceedings of 2010 international symposium on information technology (Vol.3). New York: Institute of Electrical and Electronics Engineers, 2010: 1302-1307.

[8] Trojahn C, Quaresma P, Vieira R. A framework for multilingual ontology mapping[C]//Proceedings of the international conference on language resources and evaluation. Paris: The European Language Resources Association, 2008: 1034-1037.

[9] Fu B, Brennan R, O'Sullivan D. Cross-lingual ontology mapping-an investigation of the impact of machine translation[C]//The Semantic Web. Berlin: Springer-Verlag, 2009: 1-15.

[10] Pérez A, Suero D, Ponsoda E, et al. Guidelines for multilingual linked data[EB/OL]. [2015-03-14]. http://oa.upm.es/29824/1/INVE_MEM_2013_167952.pdf.

[11] Gayo J, Kontokostas D, Auer S. Multilingual linked open data patterns[EB/OL]. [2015-03-08]. http://www.semantic-web-journal.net/system/files/swj406.pdf.

[12] Caracciolo C, Stellato A, Rajbahndari S, et al. Thesaurus maintenance, alignment and publication as linked data: the AGROVOC use case[J]. International Journal of Metadata, Semantics and Ontologies, 2012(1): 65-75.

[13] Ehrmann M, Cecconi F, Vannella D, et al. Representing multilingual data as linked data: the case of BabelNet 2.0[C]//Proceedings of the ninth international conference on language resources and evaluation. Reykjavik: The European Language Resources Association, 2014: 401-408.

[14] IEEE Standards Board. IEEE standard glossary of software engineering terminology[M]. New York: Institute of Electrical and Electronics Engineers, 1990: 42.

[15] Witt A, Heid U, Sasaki F, et al. Multilingual language resources and interoperability[J]. Language Resources and Evaluation, 2009(1): 1-14.

[16] Soria C, Tesconi M, Marchetti A, et al. Towards agent-based cross-lingual interoperability of distributed lexical resources[C]//Proceedings of the workshop on multilingual language resources and interoperability. Sydney: Association for Computational Linguistics, 2006: 17-24.

[17] Nicholson D, McCulloch E. Investigating the feasibility of a distributed, mapping-based, approach to solving subject interoperability problems in a multi-scheme, cross-service, retrieval environment[EB/OL]. [2015-01-26]. https://pure.strath.ac.uk/portal/files/176991/strathprints002875.pdf.

[18] Ma X, Carranza E, Wu C, et al. A SKOS-based multilingual thesaurus of geological time scale for interoperability of online geological maps[J]. Computers & Geosciences, 2011(10): 1602-1615.

[19] Gliozzo A, Strapparava C. Cross language text categorization by acquiring multilingual domain models from compa-

rable corpora [C]//Proceedings of the ACL workshop on building and using parallel texts. Ann Arbor: Association for Computational Linguistics, 2005: 9-16.

[20] Amine B, Mimoun M. WordNet based multilingual text categorization[C]//Proceedings of ACS/IEEE international conference on computer systems and applications. Amman: Arab Computing Society, 2007: 848-855.

[21] Kiran K N, Santosh G S K, Vasudeva V. Multilingual document clustering using wikipedia as external knowledge [C]//Multidisciplinary Information Retrieval. Berlin: Springer-Verlag, 2011: 108-117.

[22] Cobo Á, Rocha R. Identification of related multilingual documents using ant clustering algorithms[J]. Revista chilena de ingeniería, 2011(3): 351-358.

[23] Ruiz M, Chin P. Users' image seeking behavior in a multilingual tag environment[C]//Multilingual information access evaluation ii: multimedia experiments. Berlin: Springer-Verlag, 2010: 37-44.

[24] Ghorab M, Leveling J, Zhou D, et al. Identifying common user behaviour in multilingual search logs[C]//Multilingual information access evaluation i: text retrieval experiments. Berlin: Springer-Verlag, 2010: 518-525.

[25] Hong W. A descriptive user study of bilingual information seekers searching for online information to complete four tasks[D]. Pittsburgh: University of Pittsburgh, 2011.

[26] Petrelli D, Clough P. Analysing user's queries for cross-language image retrieval from digital library collections [J]. The Electronic Library, 2012(2): 197-219.

[27] Lilleng J, Tomassen S. Cross-lingual information retrieval by feature vectors[C]//Natural language processing and information systems. Berlin: Springer-Verlag, 2007: 229-239.

[28] Ko J, Si L, Nyberg E, et al. Probabilistic models for answer-ranking in multilingual question-answering[J]. ACM Transactions on Information Systems, 2010(3): 16:1-37.

[29] Ghorab M, Leveling J, Lawless S, et al. Multilingual adaptive search for digital libraries[C]//Research and advanced technology for digital libraries. Berlin: Springer-Verlag, 2011: 244-251.

[30] Jan E, Lin S, Chen B. Transliteration retrieval model for cross lingual information retrieval[C]//Information retrieval technology. Berlin: Springer-Verlag, 2010: 183-192.

[31] Potthast M, Stein B, Anderka M. A wikipedia-based multilingual retrieval model[C]//Advances in Information Retrieval. Berlin: Springer-Verlag, 2008: 522-530.

[32] Olive J, Christianson C, McCary J. Handbook of natural language processing and machine translation[M]. Berlin, Germany: Springer-Verlag, 2011: 133.

[33] Abu Shquier M, Al Nabhan M, Sembok T. Adopting new rules in rule-based machine translation[C]//Proceedings of the 12th international conference on computer modelling and simulation. Cambridge: Institute of Electrical and Electronics Engineers, 2010: 62-67.

[34] Riesa J. Syntactic alignment models for large-scale statistical machine translation[D]. Los Angeles: University of South California, 2012.

[35] Tufis D. Finding translation examples for under-resourced language pairs or for narrow domains: the case for machine translation[J]. Computer Science Journal of Moldova, 2012(2): 227-245.

[36] Kumaran A. Compositional machine transliteration[EB/OL]. [2015-01-24]. http://www.cse.iitb.ac.in/~pb/papers/TALIP-CompositionalTransliteration-CRC.pdf.

[37] Talvensaari T, Laurikkala J, Järvelin K, et al. Creating and exploiting a comparable corpus in cross-language information retrieval[J]. ACM Transactions on Information Systems, 2007(1): 1-47.

[38] Otero P, López I. Wikipedia as multilingual source of comparable corpora[C]//Proceedings of the 3rd workshop on building and using comparable corpora. Malta: European Language Resources Association, 2010: 21-25.

[39] Levow G, Oard D, Resnik P. Dictionary-based techniques for cross-language information retrieval[J]. Informa-

tion Processing and Management, 2005(3): 523-547.

[40] Valderrábanos A, Belskis A, Moreno L. Multilingual terminology extraction and validation[EB/OL]. [2014-12-27]. http://www.bitext.com/prensa/ART_EN_LREC_camera_ready__Valderrabanos_Belskis_Iraola_amended.pdf.

[41] Lefever E, Macken L, Hoste V. Language-independent bilingual terminology extraction from a multilingual parallel corpus[C]//Proceedings of the 12th Conference of the European Chapter of the ACL. Athens: Association for Computational Linguistics, 2009: 496-504.

[42] Hecking M, Wotzlaw A, Coote R. Multilingual content extraction extended with background knowledge for military intelligence [EB/OL]. [2015-02-02]. http://www.dodccrp.org/events/16th_iccrts_2011/papers/018.pdf.

[43] Klementiev A, Roth D. Named entity transliteration and discovery from multilingual comparable corpora[C]//Proceedings of the human language technology conference of the north american chapter of the ACL. New York: Association for Computational Linguistics, 2006: 82-88.

[44] Richman A, Schone P. Mining wiki resources for multilingual named entity recognition[C]//Proceedings of the 46th annual meeting of the association for computational linguistics. Columbus: The Association for Computational Linguistics, 2008: 1-9.

[45] Pinto D, Vilariño D, Balderas C, et al. A naive bayes approach to cross-lingual word sense disambiguation and lexical substitution[C]//Advances in pattern recognition. Berlin: Springer-Verlag, 2010: 352-361.

[46] Guyot J, Falquet G, Radhouani S, et al. UNIGE experiments on robust word sense disambiguation[EB/OL]. [2015-01-23]. http://clef.isti.cnr.it/2008/working_notes/guyot_paperCLEF2008.pdf.

[47] Jnedie R. Query expansion seminar cross lingual and multilingual text retrieval[EB/OL]. [2014-12-24]. http://www.iti.cs.uni-magdeburg.de/~fahmed/Paper_Example__Query%20Expansion1.pdf.

[48] Gavel Y, Andersson P. Multilingual query expansion in the SveMed+ bibliographic database: a case study[J]. Journal of Information Science, 2014(3): 269-280.

[49] Stanković R, Krstev C, Obradović I, et al. A tool for enhanced search of multilingual digital libraries of e-journals[C]//Proceedings of the eighth conference on language resources and evaluation. Istanbul: European Language Resources Association, 2012: 1710-1717.

[50] Brodeala L, Martin-Bautista M, Gil R. Combining semantic and multilingual search to databases with recommender systems[C]//Proceedings of the 22nd international workshop on database and expert systems applications. Toulouse, Institute of Electrical and Electronics Engineers, 2011: 544-548.

[51] Dolores M, Lobo O, Artacho J. Language resources used in multi-lingual question-answering systems[J]. Online Information Review, 2011(4): 543-557.

[52] Ferrández S, Toral A, Ferrández Ó, et al. Exploiting Wikipedia and EuroWordNet to solve cross-lingual question answering[J]. Information Sciences, 2009(20): 3473-3488.

[53] Cimiano P, Lopez V, Unger C, et al. Multilingual question answering over linked data (QALD-3): lab overview[G]//Information access evaluation: multilinguality, multimodality, and visualization. Berlin: Springer-Verlag, 2013: 321-332.

[54] Sujatha P. A review on performance evaluation measures of multilingual information retrieval systems[J]. International Journal of Advanced Research in Computer Science and Software Engineering, 2012(8): 440-446.

[55] Chandra M, Sadanandam M, Raju K. Software metric framework for Multilingual Information Retrieval (MLIR) system performance assessment[J]. International Journal of Emerging Trends & Technology in Computer Science, 2013(4): 38-46.

[56] Petrelli D. On the role of user-centred evaluation in the advancement of interactive information retrieval[J]. Information Processing &Management, 2008(1): 22-38.

[57] Shiri A, Ruecker S, Bouchard M, et al. User evaluation of searchling: a visual interface for bilingual digital libraries[J]. The Electronic Library, 2011(1): 71-89.

[58] Hanbury A, Boyer C, Gschwandtner M, et al. KHRESMOI: towards a multi-lingual search and access system for biomedical information [EB/OL]. [2015-01-26]. http://publications.hevs.ch/index.php/attachments/single/321.

[59] Larson R, Gey F. GeoCLEF text retrieval and manual expansion approaches[C]//Evaluation of multilingual and multi-modal information retrieval. Berlin: Springer-Verlag, 2007: 970-977.

[60] Peruginelli G, Francesconi E. Multilingual access modalities to legal resources based on semantic disambiguation [EB/OL]. [2014-12-17]. http://ceur-ws.org/Vol-465/paper9.pdf.

[61] Ménard E. Ordinary image retrieval in a multilingual context: a comparison of two indexing vocabularies[J]. Aslib Proceedings, 2010(4/5): 428-437.

[62] Tungkasthan A, Intarasema S, Premchaisawadi W. A multi-language search scheme using a multithread processing for Yahoo image search[C]//Proceedings of the eighth international symposium on natural language processing. Bangkok: Institute of Electrical and Electronics Engineers, 2009: 30-34.

[63] Oard D W, He D, Wang J. User-assisted query translation for interactive cross-language information retrieval [J]. Information Processing and Management, 2008(1): 181-211.

[64] Ahmed F, Nurnberger A. Literature review of interactive cross language information retrieval tools[J]. The International Arab Journal of Information and Technology, 2012(5): 479-486.

[65] Zazo A, Figuerola C, Berrocal J, et al. Use of free on-line machine translation for interactive cross-language question answering[C]//Accessing Multilingual Information Repositories. Berlin: Springer-Verlag, 2006: 263-272.

[66] Ruecker S, Shiri A, Fiorentino C. Interactive visualization for multilingual search[J]. Bulletin of the American Society for Information Science and Technology, 2012(4): 36-40.

[67] Ferro N. CLEF 15th birthday: past, present, and future[EB/OL]. [2015-03-21]. http://sigir.org/files/forum/2014D/p031.pdf.

[68] WorldWideScience.org[EB/OL]. [2015-05-13]. http://worldwidescience.org/index.html.

司 莉 武汉大学信息资源研究中心教授，博士生导师，图书馆学系主任。湖北 武汉 430072。

庄晓喆 武汉大学信息管理学院博士研究生。湖北 武汉 430072。

贾 欢 武汉大学信息管理学院博士研究生。湖北 武汉 430072。

（收稿日期：2015-04-07；修回日期：2015-05-25）