



跨语言信息检索

——翻译优化理论与技术

吴 丹 著



科学技术文献出版社
SCIENTIFIC AND TECHNICAL DOCUMENTATION PRESS

国家科学技术学术著作出版基金资助出版

跨语言信息检索

——翻译优化理论与技术

Cross-Language Information Retrieval:

Translation Enhancement Theory and Technology

吴 丹 著



科学技术文献出版社
SCIENTIFIC AND TECHNICAL DOCUMENTATION PRESS

图书在版编目(CIP)数据

跨语言信息检索:翻译优化理论与技术/吴丹著. —北京:科学技术文献出版社, 2011. 9

ISBN 978-7-5023-6963-7

I. ①跨… II. ①吴… III. ①情报检索-研究 IV. ①G252.7

中国版本图书馆 CIP 数据核字(2011)第 133034 号

跨语言信息检索

策划编辑:周国臻 责任编辑:马 帅 责任校对:赵文珍 责任出版:王杰馨

出 版 者 科学技术文献出版社
地 址 北京市复兴路 15 号 邮编 100038
编 务 部 (010)58882938, 58882087(传真)
发 行 部 (010)58882868, 58882866(传真)
邮 购 部 (010)58882873
网 址 <http://www.stdp.com.cn>
发 行 者 科学技术文献出版社发行 全国各地新华书店经销
印 刷 者 北京雁林吉兆印刷有限公司
版 次 2011 年 9 月第 1 版 2011 年 9 月第 1 次印刷
开 本 787×1092 16 开
字 数 382 千
印 张 17
书 号 ISBN 978-7-5023-6963-7
定 价 65.00 元



版权所有 违法必究

购买本社图书,凡字迹不清、缺页、倒页、脱页者,本社发行部负责调换

序

语言在信息交流中起着举足轻重的作用。人们有效地使用信息的前提是对信息的正确理解,而语言则在理解信息的过程中扮演着媒介的重要角色。随着互联网深入到全球化经济和多元化社会的方方面面,人们在享受获取信息便利的同时,也日益感受到获取的信息尤其是相关的感兴趣的信息,在很多情况下是用他们不能理解的外语,而不是他们的母语来表达的。为了帮助人们获取这类无法理解的信息,多语言技术应运而生。多语言技术被定义为一系列让信息传播不局限于用户能够理解和使用的语言——从而超越了语言的界限的信息处理能力和工具。过去近 20 年的发展表明,信息化的经济和社会生活的每一个方面都越来越需要多语言技术的帮助,因此研究、开发多语言技术和资源以帮助人们从可能相互关联的多语言信息对象中获取、认识和使用相关信息,是一个很重要也很有意义的课题。

同机器翻译系统、多语言语料库和词库等一样,跨语言信息检索是多语言技术中的一个重要成员。它研究应用一个语言的查询式(如中文)来检索另一个语言的文章(如英文文档)所需要的技术和资源。同时跨语言信息检索还是检索和自然语言处理相结合的产物。随着检索逐步成为人们信息交流的核心手段之一,跨语言信息检索也因日益受人瞩目而成为热门研究课题。

尽管跨语言信息检索技术最早起始于 20 世纪 60 年代,但它真正的发展是在 20 世纪 90 年代中期。经过 10 多年的耕耘,跨语言信息检索技术已经得到了长足的发展。这首先表现在跨语言信息检索技术的多样化上。例如翻译在跨语言信息检索中发挥着核心作用,由于翻译对象的不同,跨语言信息检索可分为基于查询式翻译的方法、基于文档翻译的方法和基于中间表达的方法;或是由于使用翻译资源的不同,跨语言信息检索又可以分为应用机器翻译的方法、应用词典的方法和應用平行语料库的方法。其次表现在检索的成功率上。在研究的初期,跨语言信息检索的成功率一般只有单语言检索的一半,而现在跨语言信息检索的成功率通常可以很接近单语言检索的成功率,或是有时超过一点。第三表现在在互联网上提供跨语言信息检索系统和服务上。有些商业公司已经提供系统和服务,两

个典型的例子是 Google 的 Google Translate 系统和 Yahoo! 的 Babel Fish 系统。

但是,跨语言信息检索中的相关反馈技术的研究还不完善,而且用户参与的交互式跨语言信息检索技术也有待提高。吴丹博士的这本专著很好地弥补了这两个重要方面的不足。

《跨语言信息检索——翻译优化理论与技术》是一部专门论述跨语言信息检索技术的专著。吴丹博士基于她多年在跨语言检索上的研究成果,并综合了相关领域在国内外的最新发展,在这本书中对跨语言信息检索的理论、技术方法和研究进展进行了全面介绍。此书既有作者对跨语言检索理论的深刻理解,也有作者在跨语言检索实验中获得的第一手实践心得。此书对跨语言检索领域的专家们有借鉴比较的作用,对刚刚踏入此领域的学者和学生有介绍指导的作用。

这里尤其想要指出此书的两个重要创新点。第一点是对翻译优化技术的提出和检验。相关反馈技术在单语言信息检索中被广泛使用,在跨语言检索里也有很多研究。而翻译优化则是一种很特殊的相关反馈机制,它与跨语言检索中的翻译机制和资源紧密结合,是一种有效的和专对跨语言检索的相关反馈技术。翻译优化机制的建立及其与查询式扩展的结合还明显地稳定了相关反馈技术在跨语言检索中应用的有效性,使得相关反馈技术更能适应不同种类的查询需求。第二点是对交互式跨语言检索的研究。跨语言检索系统最终是要为用户服务的,要帮助用户找到相关的多语言信息。因此用户直接参与系统的检验是一个很重要的和很必要的环节。但由于用户实验的不确定性、复杂性和巨大开销,用户参与的实验并不多。此书中对翻译优化的检验有自动化的指标实验,更重要的是吴丹博士开发了 ICE-TEA 跨语言检索系统,实现了翻译优化技术,然后在此基础上进行了较大规模的用户实验,所得到的结果就很有说服力。

我有幸在吴丹博士 2006 年到美国匹兹堡大学访问时和她认识,并与她建立了长期和稳定的合作关系。我很欣慰地看到当年许多想法和设计都逐步实现并得到验证。这本书是对这些优异成果的一个总结。吴丹博士能这么快地把想法转化成为经检验是成功的研究成果,令我十分敬佩。对能从一开始就参与这些成果的研究,并有幸在此作序把这些成果介绍给广大读者,我感到十分荣幸。

何大庆

2011 年 2 月 14 日

于美国匹兹堡

前 言

多语言性是网络社会交流的重要特征之一,跨语言信息检索旨在通过以某种语言来检索表达为另外一种语言的信息,达到消除信息检索中语言差异的目的,从而成为近年来信息检索领域很活跃的研究方向。我们对国内外相关研究进行了全面调研,发现目前跨语言信息检索在查询翻译消歧问题上还有待进一步改善,而有关跨语言信息检索的中文研究则缺乏局部的技术实现以及整体的理论框架。鉴于此,本书的研究重点是引入相关反馈技术来提高英汉跨语言信息检索的精准度,尝试对基于相关反馈的翻译优化技术进行一些开创性的研究。

本书全面深入地介绍了跨语言信息检索技术的理论、重要方法 and 研究进展,重点探讨了基于相关反馈的翻译优化理论与技术,对交互式跨语言信息检索进行了研究。在此基础上,通过将翻译优化技术与查询扩展技术进行比较研究,并结合系统相关性与用户相关性理论,构建了一个跨语言检索相关反馈综合模型理论框架(RFIM 模型),开发并实现了一个基于相关反馈的交互式英汉跨语言信息检索系统。最后,进行了一系列实验研究,其中包括:(1)通过自动相关反馈与用户相关反馈两组实验,验证了翻译优化技术的有效性与 RFIM 模型的合理性;(2)将翻译优化技术与其他技术如命名实体翻译、机器翻译、数据融合等技术进行比较与结合,验证了翻译优化技术在众多优化技术中的效果。相信翻译优化这种新的跨语言信息检索技术可以帮助用户跨越语言障碍从而无缝获取信息。

本书的创新之处主要体现在以下 3 个方面:一是提出了一种崭新的跨语言相关反馈技术——翻译优化技术。基于相关反馈的翻译优化技术完全不同于单语言检索中传统的相关反馈思想,它是利用检索结果中相关文献对来优化跨语言检索的查询翻译,解决了部分查询翻译的歧义性问题。二是建立了一个全新的跨语言检索相关反馈综合模型理论框架(RFIM 模型)。具有三层结构的跨语言相关反馈综合模型结合了系统与用户相关性理论、各项相关反馈技术、影响相关反馈的各种因素以及用户行为因素。RFIM 模型的建立合理地解释了在跨语言信息检索中各种相关反馈方法之间的关系及其应用环境。三是开发了一个新颖的具有相关反馈功能的交互式英汉跨语言信息检索系统 ICE-TEA,并首次进行了跨语

言检索的自动相关反馈与用户相关反馈比较实验研究,同时将翻译优化技术与其他技术进行了比较与结合研究。ICE-TEA 系统的构建和一系列实验的结论为翻译优化技术在跨语言信息检索中的实际应用提供了参考。

本书是一部论述跨语言信息检索技术的专著。书中既有对跨语言信息检索技术的理论方法研究的全面、深入的综述,又有针对其中翻译优化理论与技术的专门论述,还结合了大量系统实现与实验研究。因此,本书是一部理论与实践相结合的著作。本书适合高等院校信息管理专业、计算机科学技术专业以及相关专业的教学科研人员、高年级学生和研究生阅读参考,也适合工作范围涉及该领域的工程技术人员参阅。

全书共分 11 章。第 1 章“导论”介绍了跨语言信息检索的意义和国内外研究现状,提出了翻译优化技术,并介绍了全书的内容、技术路线、研究方法与结构。第 2 章“跨语言信息检索基础理论”和第 3 章“跨语言信息检索关键技术”分别概述了跨语言信息检索领域的基础理论与关键技术问题。第 4 章“跨语言信息检索的相关反馈技术”描述了相关性理论与相关反馈技术及其在不同检索模型中的应用,提出了跨语言查询扩展。第 5 章“基于相关反馈的翻译优化技术”重点论述了翻译优化技术的理论基础、处理过程、相关算法与数学模型。第 6 章“跨语言检索相关反馈综合模型构建”将翻译优化与查询扩展进行了对比与结合,提出了跨语言检索相关反馈综合模型理论框架(RFIM),并对模型的各层次进行了要素分析。第 7 章“基于相关反馈的英汉跨语言检索系统实现”详细介绍了所构建系统的功能、结构、设计与评价方法。第 8 章“基于自动相关反馈的翻译优化实验”和第 9 章“基于用户相关反馈的翻译优化实验”,分别采用自动相关反馈方法和用户相关反馈方法对翻译优化技术进行了实验检验。第 10 章“其他优化技术的跨语言信息检索实验”是将翻译优化技术与命名实体翻译、机器翻译、数据融合等技术进行比较与结合,并给出了实验检验结果。第 11 章“结论与展望”总结了全书的研究内容,指出了未来的研究方向。

本书是在我的北京大学博士毕业论文的基础上,补充了近几年的最新研究成果写成的。在此,首先衷心感谢我的博士导师——中国科学技术信息研究所的王惠临研究员,他的悉心教导和不断鼓励,使我在读博士期间收获了很多知识和经验。本书涉及的研究是在美国匹兹堡大学何大庆教授的直接指导与长期合作下完成的,他点明了方向,给了我许多新的思考,并帮助我在美国完成了部分实验,在此,对他表示衷心的感谢。同时,本书内容中的后续研究是我在武汉大学从事博士后工作期间完成的,在此期间得到了博士后合作导师——武汉大学陈传夫教授的中肯指导与大力帮助,他对本书内容提出了许多切合实际的建议,并为我提供了良好的工作条件,在此,表示衷心的感谢。另外,武汉大学信息管理学院信息

检索实验室、中国科学技术信息研究所语言技术实验室、美国匹兹堡大学信息科学学院信息检索实验室都为本书的撰写提供了大力帮助,为本书部分实验内容的完成提供了条件,在此对所有提供过帮助的人深表感谢。本书在撰写和出版过程中得到了科学技术文献出版社周国臻编辑的大力支持,武汉大学信息管理学院的领导和同事们对本书的出版给予了热情鼓励和指导,北京大学信息管理系的老师与同学们给予了我支持与帮助,我的家人给了我极大的关心和爱护,以及其他诸多有关人士付出了辛勤劳动,特此一并致以衷心谢忱!

此外,本书的出版还得到了4个项目的资助,分别是:2009年度国家科学技术学术著作出版基金项目“跨语言信息检索:翻译优化理论与技术”、2009年度国家社会科学基金项目“网络学术信息的多语言表示与获取模式研究(项目编号:09CTQ026)”、2009年度教育部人文社会科学研究项目“多语言信息获取中的用户相关反馈研究(项目编号:09YJC870022)”、2010年武汉大学人文社会科学“70后”学者学术发展计划项目“跨语言环境下的细粒度信息检索及可视化研究”。本书的出版是上述项目的研究成果之一。

吴 丹

2011年2月于武汉珞珈山

作者简介

吴丹,女,1978年生,博士,现为武汉大学信息管理学院副教授,珞珈青年学者。2008年毕业于北京大学情报学专业,获管理学博士学位。2006年至2007年公派赴美国匹兹堡大学从事跨语言信息检索研究。近年先后在国内发表学术论文60余篇,其中被SCI和SSCI检索4篇,参编著作8部,主持国家级、省部级、校级科研课题10项。主要研究领域为跨语言信息检索、语言处理技术、数字图书馆、知识组织等。

目 录

第1章 导论	1
1.1 跨语言信息检索的意义	1
1.1.1 网络资源呈现多语言特征	1
1.1.2 用户对跨语言信息检索的需求	4
1.2 国内外研究现状述评	5
1.2.1 国外跨语言信息检索研究历程	6
1.2.2 国内跨语言信息检索研究现状	7
1.3 翻译优化技术的提出	11
1.3.1 翻译优化的依据	11
1.3.2 翻译优化的意义	13
1.4 本书的结构	14
1.4.1 研究内容	14
1.4.2 技术路线	15
1.4.3 研究方法	16
1.4.4 结构安排	16
第2章 跨语言信息检索基础理论	19
2.1 跨语言信息检索基础	19
2.1.1 跨语言信息检索的概念	19
2.1.2 跨语言信息检索的应用	20
2.2 跨语言信息检索的语言预处理	32
2.2.1 中文信息处理	32
2.2.2 欧洲语言信息处理	35
2.2.3 其他语言信息处理	36
2.3 跨语言信息检索模型	37
2.3.1 检索模型及其分类	37
2.3.2 检索模型在跨语言信息检索中的应用	46

2.4 跨语言信息检索相关资源	53
2.4.1 相关会议	53
2.4.2 相关语料	55
2.5 跨语言信息检索研究展望	56
2.6 本章小结	58
第3章 跨语言信息检索关键技术	59
3.1 跨语言信息检索中的翻译策略	59
3.1.1 翻译策略的选择	59
3.1.2 翻译资源的构建	63
3.1.3 翻译歧义性问题	68
3.2 查询翻译消歧方法	70
3.2.1 利用结构化查询消歧	70
3.2.2 通过语言分析消歧	71
3.2.3 借助语言资源消歧	73
3.2.4 基于用户交互消歧	74
3.3 命名实体识别与翻译	77
3.3.1 命名实体的识别方法	78
3.3.2 命名实体的翻译方法	79
3.4 机器翻译与跨语言信息检索	79
3.4.1 机器翻译在跨语言信息检索应用中的基本问题	80
3.4.2 统计机器翻译与跨语言信息检索	81
3.5 数据融合	82
3.5.1 查询表达式融合	82
3.5.2 检索结果融合	83
3.6 本章小结	84
第4章 跨语言信息检索的相关反馈技术	85
4.1 相关性理论与相关反馈技术	85
4.1.1 相关性与相关性模型	85
4.1.2 相关性判断	88
4.1.3 相关反馈技术	89
4.2 相关反馈技术在不同检索模型中的应用	93
4.2.1 基于经典模型的相关反馈	93
4.2.2 基于语言模型的相关反馈	96
4.2.3 基于语义模型的相关反馈	97
4.3 跨语言查询扩展	98

4.3.1 不同翻译资源下的查询扩展	100
4.3.2 不同语言转换中的查询扩展	105
4.3.3 翻译资源质量对查询扩展的影响	105
4.4 基于语言模型的跨语言相关反馈	107
4.5 跨语言相关反馈中存在的问题	108
4.6 本章小结	109
第5章 基于相关反馈的翻译优化技术	110
5.1 翻译优化技术的基础理论	110
5.1.1 翻译优化的前提假设	110
5.1.2 翻译优化的基本思想	112
5.1.3 翻译优化的理论依据	113
5.2 翻译优化技术的处理过程	115
5.2.1 从相关文献对中抽取检索词及其翻译关系	116
5.2.2 利用检索词及其翻译关系优化查询翻译	121
5.3 相关文献集合中检索词的翻译概率估计算法	123
5.3.1 基于词对齐的翻译算法 TWA	123
5.3.2 保留所有翻译的算法 KAT	124
5.3.3 保留最好翻译的算法 K1T	125
5.3.4 保留最高频率翻译的算法 KFT	126
5.4 翻译优化的数学建模	128
5.4.1 对齐模型	128
5.4.2 翻译概率重新估计模型	130
5.5 本章小结	131
第6章 跨语言检索相关反馈综合模型构建	132
6.1 翻译优化与查询扩展的比较研究	132
6.1.1 翻译优化与查询扩展的对比	132
6.1.2 翻译优化与查询扩展的结合	134
6.2 跨语言检索相关反馈综合模型(RFIM)的构建	135
6.2.1 模型要解决的问题	135
6.2.2 跨语言检索中已有的相关反馈建模分析	135
6.2.3 层次分析法	136
6.2.4 具有层次结构的跨语言相关反馈综合模型总体框架	137
6.3 跨语言相关反馈综合模型各层次要素分析	137
6.3.1 用户行为层	137
6.3.2 方法技术层	139

6.3.3 影响因子层	140
6.4 模型的检验	141
6.5 本章小结	141
第7章 基于相关反馈的英汉跨语言检索系统实现	143
7.1 系统功能与资源	143
7.1.1 系统功能设计	144
7.1.2 系统实现的流程与环境	145
7.1.3 系统的语料	146
7.2 系统体系结构设计	148
7.2.1 查询翻译模块	150
7.2.2 文本预处理模块	151
7.2.3 检索模块	151
7.2.4 结果显示模块	154
7.2.5 相关反馈模块	154
7.3 人机交互接口设计	157
7.3.1 交互性设计	157
7.3.2 多级相关性判断	159
7.4 系统评价体系	159
7.4.1 跨语言信息检索评价模型	159
7.4.2 跨语言信息检索评价指标	160
7.5 本章小结	162
第8章 基于自动相关反馈的翻译优化实验	163
8.1 实验设计	163
8.1.1 实验目标	163
8.1.2 实验内容	163
8.1.3 实验数据收集	164
8.2 基准(Baseline)实验结果分析	166
8.2.1 单语言检索基准实验结果	166
8.2.2 跨语言检索基准实验结果	166
8.3 翻译优化实验结果分析	167
8.3.1 基于词对齐的翻译方法 TWA 的实验结果	168
8.3.2 保留所有翻译的方法 KAT 的实验结果	169
8.3.3 保留最好翻译的方法 K1T 的实验结果	169
8.3.4 保留最高频率翻译的方法 KFT 的实验结果	170
8.3.5 四种翻译优化方法的实验结果比较	171

8.4 查询扩展实验结果分析	172
8.4.1 翻译前查询扩展方法 Pre-QE 的实验结果	172
8.4.2 翻译后查询扩展方法 Post-QE 的实验结果	173
8.4.3 混合式查询扩展方法 Comb-QE 的实验结果	174
8.4.4 3 种查询扩展方法的实验结果比较	175
8.5 翻译优化与查询扩展相结合的实验结果分析	177
8.5.1 翻译优化与查询扩展的实验结果对比	177
8.5.2 翻译优化与查询扩展相结合的实验结果分析	178
8.6 实验结论	179
8.7 本章小结	180
第 9 章 基于用户相关反馈的翻译优化实验	181
9.1 实验设计	181
9.1.1 实验目标	181
9.1.2 实验内容	181
9.2 英汉跨语言信息检索用户相关性标注实验	182
9.2.1 实验数据收集	182
9.2.2 实验步骤	182
9.2.3 用户个体差异与相关性判断结果分析	185
9.2.4 翻译优化的实验结果分析	187
9.2.5 用户对系统的评价分析	190
9.3 用户全程参与的英汉跨语言信息检索相关反馈实验	191
9.3.1 实验资源准备	192
9.3.2 实验步骤	193
9.3.3 实验结果分析	196
9.3.4 用户的评价分析	201
9.4 实验结论	205
9.5 本章小结	206
第 10 章 其他优化技术的跨语言信息检索实验	207
10.1 基于命名实体翻译的跨语言信息检索实验	207
10.1.1 实验概述	207
10.1.2 基于信息抽取的命名实体识别与翻译	208
10.1.3 命名实体翻译在跨语言信息检索中的应用	209
10.1.4 实验结果分析	210
10.2 基于机器翻译的跨语言信息检索实验	214
10.2.1 机器翻译用于查询翻译	214

10.2.2 实验概述.....	215
10.2.3 实验结果分析.....	216
10.3 基于数据融合的跨语言信息检索实验.....	219
10.3.1 实验概述.....	219
10.3.2 实验结果分析.....	220
10.4 实验结论.....	221
10.5 本章小结.....	221
第 11 章 结论与展望	223
11.1 研究结论.....	223
11.2 研究展望.....	224
附录 1 相关反馈实验的英文检索主题(标题)	226
附录 2 相关反馈实验的中文检索主题(标题)	227
附录 3 四种翻译优化方法的平均查准率均值 MAP	228
附录 4 ICE-TEA 系统用户研究调查问卷	234
参考文献.....	241

第 1 章 导 论

1.1 跨语言信息检索的意义

作为一个正式的学术概念,信息检索(Information Retrieval)于 1948 年由美国学者 Mooers¹ 在其硕士学位论文中首次提出。经过半个多世纪的研究,信息检索的发展演变可以看作是不不断消除一道道信息存取障碍的过程。首先,从脱机批处理到计算机检索系统的产生,是超越了信息存取的速度障碍;其次,从单机到网络平台,从集中式网络到分布式、异构性、动态 Web 环境的迁移,是打破了信息存取的空间障碍;再次,从分类法、主题词表到本体(Ontology)的出现和应用,是跨越了信息存取的语义理解障碍。即便如此,在全球信息共享的迫切要求下,依然还有一道障碍未被攻克,即信息存取的语言障碍。

1.1.1 网络资源呈现多语言特征

进入 21 世纪以来,网络技术的发展日新月异促使多语种网络信息资源以惊人的速度丰富起来。据全球标准互联网用户调查和分析权威机构 Nielsen//NetRatings² 统计,以及在 Internet World Stats³ 上的发布显示,从 2000 年到 2010 年,在短短 10 年时间内,全世界各种语言的网络使用增长率达到 444.8%,其中各大洲的统计数字分别为:中东地区 1 825.3%,非洲 2 357.3%,拉丁美洲/加勒比海 1 032.8%,亚洲 621.8%,欧洲 352.0%,大洋洲/澳大利亚 179.0%,北美洲 146.3%(如图 1-1 所示)。由此可见,第三世界国家如非洲、中东和拉美语言的网络使用率增长最快,亚洲和欧洲也保持着高速增长,网络信息资源不再是英语一统天下的局面,而是体现出越来越明显的多语言化特征。目前,网络用户语言的前十位如图 1-2 所示,英语仍是网络世界中使用人数最多的语言,到 2010 年用户数达到 5.36 亿;中文用户位居第二,人数为 4.44 亿。

尽管目前英语仍是网络世界中使用人数最多的语言,但非英语语言的比例也呈现出逐年上升的趋势。图 1-3 反映了进入 21 世纪以来,近 10 年内十大网络语言使用的变化趋势。如图 1-3(e),在 2001 年,英语在网络语言中占近一半的比例;而在图 1-3(d)中,到 2004 年,英语所占比例已下降至 35.2%;在图 1-3(c)中,2007 年英语所占比例为 31.2%;如图 1-3(b),到 2008 年,非英语语言占网络语言的比例超过 70%;如图 1-3(a),在 2010 年,英语比例继续下降,直降至 27.3%,中文及其他语言的比例继续上升,这种语言多元化的趋势还在进一步深化。比较 a、b、c、d、e 这 5 幅图,十大网络语言也发生了变化。21 世纪前期排进前 10 位的荷兰

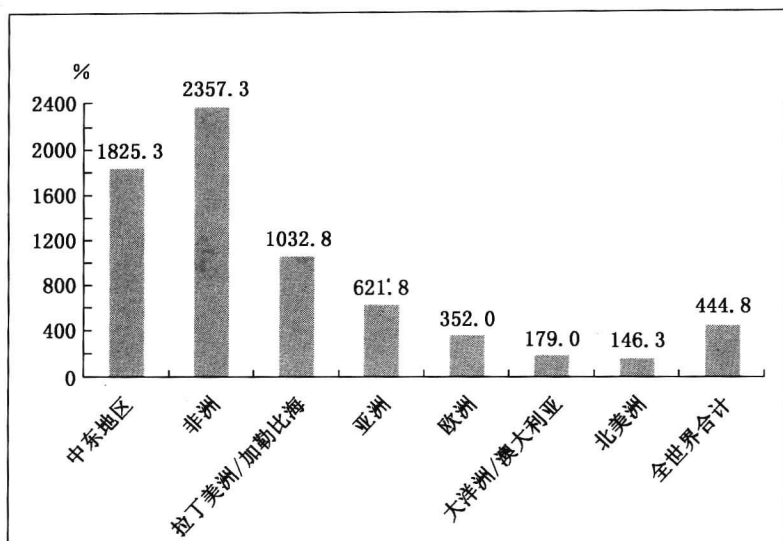


图 1-1 2000—2010 年世界各大洲网络语言使用增长率
(数据更新时间:2011 年 1 月 30 日)

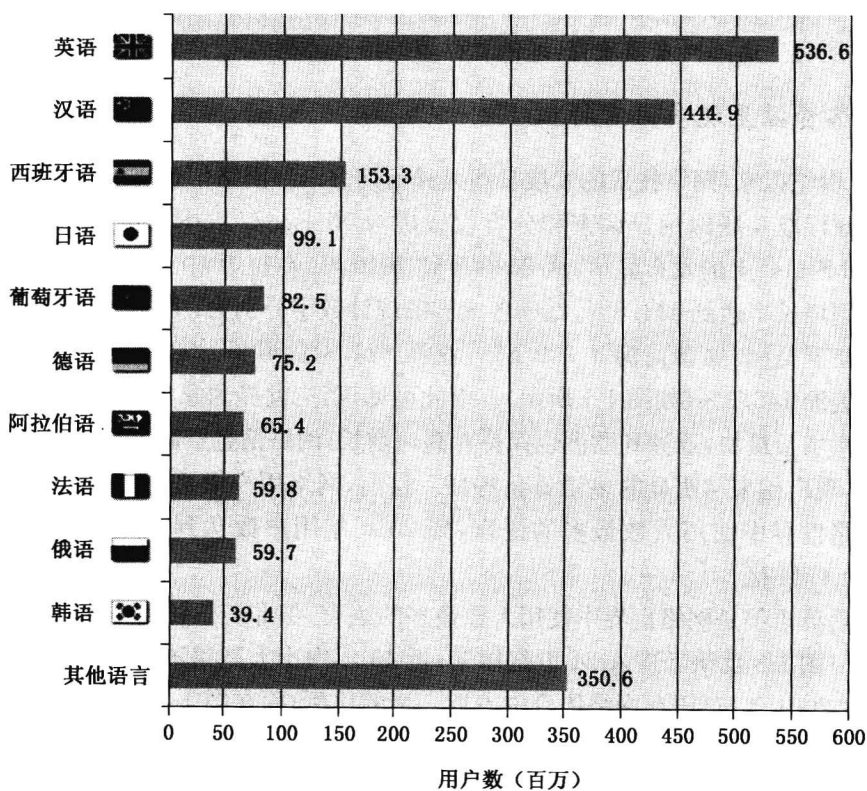


图 1-2 2010 年世界前十位网络语言用户数(数据更新时间:2011 年 1 月 30 日)