上海外国语大学博士后流动站
战略支援部队信息工程大学

基于知识本体的俄汉可比语料库建设与应用研究

JIYU ZHISHI BENTI DE E-HAN KEBI YULIAOKU JIANSHE YU YINGYONG YANJIU

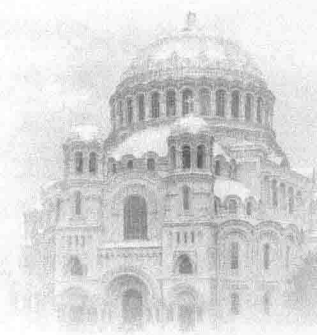
原伟 著

世界图书出版公司

上海外国语大学博士后流动站
战略支援部队信息工程大学

基于知识本体的俄汉可比语料库建设与应用研究

原伟 著



世界图书出版公司
广州·上海·西安·北京

图书在版编目（CIP）数据

基于知识本体的俄汉可比语料库建设与应用研究 / 原伟著. —广州：世界图书出版广东有限公司，2019.4
ISBN 978-7-5192-6145-0

Ⅰ. ①基… Ⅱ. ①原… Ⅲ. ①语料库－建设－研究 Ⅳ. ①H0

中国版本图书馆 CIP 数据核字（2019）第 067857 号

书　　名	基于知识本体的俄汉可比语料库建设与应用研究 JIYU ZHISHI BENTI DE E-HAN KEBI YULIAOKU JIANSHE YU YINGYONG YANJIU
著　　者	原　伟
策划编辑	刘正武
责任编辑	张东文
出版发行	世界图书出版广东有限公司
地　　址	广州市海珠区新港西路大江冲 25 号
邮　　编	510300
发行电话	020-84451969　84459539
网　　址	http://www.gdst.com.cn/
邮　　箱	wpc_gdst@163.com
经　　销	新华书店
印　　刷	虎彩印艺股份有限公司
开　　本	787 mm × 1092 mm　1/16
印　　张	16
字　　数	265 千字
版　　次	2019 年 4 月第 1 版　2019 年 4 月第 1 次印刷
国际书号	ISBN 978-7-5192-6145-0
定　　价	48.00 元

序　言

可比语料库作为近年来语料库研究的热点方向之一，可广泛应用于语言学研究和自然语言处理领域。本体是一种基于语义网技术的知识表示方法，它与可比语料库的融合式研究是对可比语料获取、组织和应用方法的一次革新，将更好地发挥可比语料库的效能并扩展其应用领域。经过前期调查分析，国内外鲜有俄汉可比语料库相关研究，未见基于本体的俄汉可比语料库相关成果。本研究首先在理论层面分析了可比语料库研究的现存问题，提出将本体引入可比语料库研究的思路，并以此为基础提出了基于本体的可比语料库理论构想。随后在实践层面将该理论构想运用到了面向俄汉可比语料库的乌克兰事件多语言复合型本体构建、基于该本体的俄汉新闻及维基百科可比语料获取、语料库构建和语料库应用等核心问题的研究中。

本研究所取得的成果：第一，本研究所提出的基于本体的可比语料库理论构想由“一对关系、三个模型和四个问题”组成，阐释了本体与语料的层次关系，建立了基于本体的单语料、语料库和语料可比关系模型，在理论上讨论了面向可比语料库的本体构建、基于本体的可比语料获取、语料库构建和语料库应用问题。第二，本研究所构建的面向俄汉可比语料库的乌克兰事件多语言复合型本体（МОПКУС）

由表征领域知识和描述存储语料的 2 大类及 8 个子类组成，包含数据属性 46 个，对象属性 9 个，实例 60733 个（领域知识类实例 4525 个，语料描述类实例 56207 个）。第三，本研究所构建的基于МОПКУС 的俄汉可比语料库包含新闻原始语料 3554 篇，维基百科原始语料 1670 篇，其中包括中文语料 163 万字，俄文语料 132 万词；以领域相同、发布时间相似对齐俄汉新闻可比语料文本 50148 对，以俄汉维基语言链接对齐维基百科可比语料文本 835 对，并对俄汉语料各层面的可比程度进行了评估计算，为语料库的应用奠定了坚实的数据基础。第四，在所构建的基于本体的俄汉可比语料库基础上探讨了三个应用问题：复杂语料调用与语义查询、基于多维度特征的可比度评估以及跨语言文本推荐与信息整合，并研制了相关软件系统。

本研究的创新之处在于：第一，通过对现有可比语料库研究理论的阐述与分析，对可比语料的定义和分类问题进行了重新整理与论证，分析了可比语料库研究现存问题，提出了基于本体的可比语料库研究思路。第二，提出并建立了基于本体的可比语料库理论构想，并将该理论在基于本体的俄汉可比语料库构建和应用中进行了验证。第三，尝试提出一种基于多维特征的语料可比度评估方法，并通过软件系统进行了验证。第四，尝试提出一种使用非翻译手段获取跨语言文本信息的方法——基于可比语料的跨语言文本推荐和信息整合，并通过软件系统进行了验证。

在本书选题、撰写和修订过程中，我的老师王松亭教授给予了我悉心的指导，启迪了我的学术思想，倾注了大量的精力和心血，激励我在学术道路上不断探索，对他的帮助我感恩不敢忘怀，希望在今后成长的道路上不辜负老师的谆谆教诲和殷切希望。

感谢我的老师易绵竹教授多年以来对我的悉心指导和无私帮助，他广博的知识和深刻的学术洞察力对我的学习和科研产生了巨大影

响，在困惑时总能为我提出宝贵的意见与建议，此刻心中更多的是感谢与感恩。

感谢我的家人多年以来给予我的宽容、理解、支持与帮助，任何时候他们都是我最坚实的后盾和力量源泉，希望我的点滴成绩能为他们带来一些欣慰与快乐，祝愿他们永远健康平安。

作 者

2019 年 2 月于洛阳

目　录

图表目录