现代图书情报技术
NEW TECHNOLOGY OF LIBRARY AND INFORMATION SERVICE

| 首 页 | 关于我们 | 投稿指南 | 征订服务 | 诚邀合作 | 留 言 |

知识组织与知识管理

最新目录 ｜ 下期目录 ｜ 过刊浏览 ｜ 高级检索　<< Previous Articles | Next Articles >>

用于双语术语抽取的专业领域中英文可比语料库构建

康小丽[1], 章成志[2]

1. 南昌大学图书馆 南昌 330031;
2. 南京理工大学信息管理系 南京 210094

Kang Xiaoli[1], Zhang Chengzhi[2]

1. Library of Nanchang University, Nanchang 330031, China;
2. Department of Information Management, Nanjing University of Science and Technology, Nanjing 210094, China

- 摘要
- 参考文献
- 相关文章

Download: PDF (540KB)　HTML (1KB)　Export: BibTeX or EndNote (RIS)　Supporting Info

摘要 面向双语术语抽取这一应用目标,提出专业领域可比语料库的构建方案并进行实验验证。针对给定的主题领域分别进行中英文专业语料的采集,从中分别获取中英文关键词,根据词语共现统计获取该主题领域的其他相关关键词;以这些关键词作为查询入口,通过学术搜索引擎从网络获取候选可比语料;对可比语料进行定量评估,以剔除不符合要求的语料,最终得到特定主题领域的可比语料库。

关键词： 可比语料库　语料库构建　双语术语抽取

Abstract： In this paper, the process of building comparable corpus in special domain for bilingual terminology is designed. Firstly, bilingual sample corpus in a special domain is collected, and keywords are extracted from the sample corpus based on word co-occurrence method. Then, these keywords are used to be a query to a scholar search engine, and the searched result is used to be candidate comparable corpus. Finally, the comparable corpus in the special domain is obtained after filtering noise documents by quantitative evaluation.

Keywords： Comparable corpus,　Corpus construction,　Bilingual terminology extraction

Service
- 把本文推荐给朋友
- 加入我的书架
- 加入引用管理器
- Email Alert
- RSS

作者相关文章
- 康小丽
- 章成志

引用本文:

康小丽, 章成志 .用于双语术语抽取的专业领域中英文可比语料库构建[J]　现代图书情报技术, 2012,V28(2): 28-33

Kang Xiaoli, Zhang Chengzhi .Chinese-English Comparable Corpus Construction for Bilingual Terminology Extraction[J] ， 2012,V28(2): 28-33

链接本文:

http://www.infotech.ac.cn/CN/ 　或　 http://www.infotech.ac.cn/CN/Y2012/V28/I2/28

[1] 淘宝 [EB/OL]. [2011-11-01]. http://www.taobao.com/. (TaoBao [EB/OL]. [2011-11-01]. http://www.taobao.com/.)

[2] eBay [EB/OL]. [2011-11-01]. http://www.ebay.com/.

[3] Epinions [EB/OL]. [2011-11-01]. http://www.epinions.com/.

[4] Marsh S P. Formalising Trust as a Computational Concept[D]. Scotland :University of Stirling,1994.

[5] 夏云,李德凤.可比语料量化比较分析与应用文体翻译——一项基于自建小型语料库的研究[C].见: 第18届世界翻译大会论文集 .北京:外文出版社,2008: 561-566. (Xia Yun,Li Defeng. Quatitative Comparative Analysis of Comparable Corpus and Translation of Practical Style-A Research Based on Self-built Small Corpus[C].In: *Proceedings of the 18 FIT World Congress Proceedings.* Beijing:Foreign Languages Press,2008: 561-566.)

[6] Jøsang A, Ismail R. The Beta Reputation System[C]. In: *Proceedings of the 15th Bled Conference on Electronic Commerce.* Slovenia:University of Mariber, 2002:1-14.

[7] Salton G, McGill M J. Introduction to Modern Information Retrieval [M]. New York: McGraw Hill Book Company, 1983:201-203.

[1] 冯冠军, 禹龙, 田生伟.**基于CRFs自动构建维吾尔语情感词语料库**[J]. 现代图书情报技术, 2011,27(3): 17-21

[2] 康小丽,章成志,王惠临.**基于可比语料库的双语术语抽取研究述评***[J]. 现代图书情报技术, 2009,(10): 7-13