



主管单位: 中华人民共和国文化部 主办单位: 国家图书馆 中国图书馆学会 ISSN 1001-8867 CN 11-2746/G2

文章摘要

邓三鸿,王昊,秦嘉杭,苏新宁.基于字角色标注的中文书目关键词标引研究[J].中国图书馆学报,2012,38(2):38~49

基于字角色标注的中文书目关键词标引研究

Research on Keywords Indexing for Chinese Bibliography Based on Word Role Annotation

DOI:

中文关键词: [中文书目](#) [关键词标引](#) [字角色](#) [序列标注](#) [自动标引](#)

英文关键词: [Chinese bibliography](#) [Keywords indexing](#) [Word roles](#) [Sequence annotation](#) [Automatic indexing](#)

基金项目:本文系国家社科基金项目“面向语义网本体的知识管理研究”(编号:09CTQ010)的研究成果之一。

作者	单位
邓三鸿	南京大学信息管理学系 南京 210093
王昊	南京大学信息管理学系 南京 210093
秦嘉杭	南京财经大学图书馆 南京 210046
苏新宁	南京大学信息管理学系 南京 210093

摘要点击次数: 1015

全文下载次数: 225

中文摘要:

中文书目机器自动标引是数字图书馆建设中亟待解决的关键问题之一。本文试图将条件随机场(CRFs)序列标注机器学习算法词抽取中,建立面向图书内容、基于字角色标注的中文书目关键词标引模型。将图书内容转化为字序列,进而提出构建关键词角色?综合利用字序列上下文特征的设计思路。通过实验,从题名和内容提要中分别自动抽取关键词,论证该模型的合理性和实用性。图6 考文献23。

英文摘要:

Automatic indexing by computers for Chinese bibliography has become one of the most critical problems which should be so immediately in digital library construction. This paper tries to introduce Conditional Random Fields (CRFs) algorithm into the extraction of Chinese bibliography, and builds the model which faces book contents based on the word roles annotation. The model converts the book contents into sequences of words. Based on that, an idea which combines word roles space model building with comprehensive utilization of word sequence has been proposed. Moreover, the paper also verifies the rationality and practicality of the model by showing the experiment of automatically extracting keywords from titles and abstracts. 6 figs. 3 tabs. 23 refs.

[查看全文](#) [查看/发表评论](#) [下载PDF阅读器](#)

关闭