

《先秦文献信息处理》

图书基本信息

书名：《先秦文献信息处理》

13位ISBN编号：9787510056611

10位ISBN编号：7510056616

出版时间：2013-1

出版社：世界图书出版公司

页数：323

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu000.com

《先秦文献信息处理》

内容概要

《先秦文献信息处理》研究先秦文献的信息处理，包括词汇和历史知识的处理与智能检索。先秦汉语是汉语可以溯及的源头，对先秦文献做词汇处理，目的在于追本溯源，了解先秦汉语词汇的基本面貌。用计算语言学方法梳理这一时期的历史事件、人名、地名等知识，有助于更好地理解先秦文献。与现当代中文文献相比，先秦传世文献规模很小，但相关的注疏文献十分丰富。因此书中采用了与一般中文信息处理不同的方法——基于注疏文献的方法。对比实验表明，这种方法在先秦文献信息处理中具有显著的优势。

《先秦文献信息处理》

作者简介

陈小荷，江西丰城人。北京大学现代汉语专业文学博士。现任南京师范大学文学院语言科学与技术系教授、博士生导师，兼中国中文信息学会理事、《中文信息学报》编委。主要研究方向为中文信息处理、汉语语法学。代表作有《现代汉语自动分析——ViSual C++实现》《汉语口语里表示“……再说”的语素“着”》等。主持国家社会科学基金项目1项，江苏省高校社科重点研究基地重大项目1项，参与国家社会科学基金重大项目1项。

《先秦文献信息处理》

书籍目录

第一章 绪论 第一节 先秦文献 第二节 先秦文献信息处理 第三节 先秦文献信息处理的特点 第四节 基于注疏文献的处理方法 第二章 注疏文献处理 第一节 注疏文献的选取 第二节 注疏文献的内容分析 第三节 先秦文献与其注疏文献的句子对齐和注释对齐 第三章 词语切分 第一节 先秦汉语分词规范 第二节 基于统计模型的自动分词方法 第三节 利用注疏文献的自动分词方法 第四章 词性标注 第一节 词类体系 第二节 基于统计的分词标注一体化方法 第三节 基于注疏文献的词性校正 第五章 古今字、通假字资源建设 第一节 古今字、通假字研究现状 第二节 古今字、通假字数据库的构建 第三节 通假字标注实验 第四节 实验结果及其分析 第六章 词汇概貌 第一节 语料来源和说明 第二节 先秦文献的汉字统计 第三节 先秦文献的词汇统计 第四节 计算每种文献的特色词 第五节 成语来源统计 第七章 词义消歧 第一节 古代汉语词义特点和消歧难点 第二节 义项区分的颗粒度 第三节 基于CRF模型的消歧实验 第四节 分类器集成的消歧实验 第八章 专名标注 第一节 关于“专名”和“命名实体”的辨析 第二节 先秦时期的人名与地名 第三节 基于统计的专名识别 第四节 基于注疏文献的专名识别 第五节 人名所指歧义消解 第六节 基于专名标注的事件划分 第九章 版本异文发现 第一节 异文和版本异文 第二节 个案和研究方法 第三节 基于相似度计算的句珠配对 第四节 基于同文排除的异文配对 第五节 基于双序列比对的一体化处理 第六节 实验结果及其分析 第十章 《左传》中的春秋社会网络分析 第一节 社会网络分析的基本方法和软件 第二节 社会关系数据的获取和网络的定义 第三节 人物—事件网络分析 第四节 人物关系网络分析 第五节 人物关系网络的深入分析 后记

章节摘录

版权页：插图：例如：没有为今读yo的第一人称代词造字，而是借用同音的“予”、“余”来表示。没有为今读wéi的句首句中语气词造字，而是借用同音的“唯”、“惟”、“维”来表示，这就是通用。在通假字中，通用较少，假借居多。因此，本研究的通假字标注，不再细分二者，主要针对假借字标注。《汉语大字典》在《编写细则·释义·通假现象的处理》中说：“通假是指古汉语书面语言中音同音近而意义原无关联的字通用。‘甲’通‘乙’，的条件必须是：（1）甲乙二字同时并存；（2）甲乙二字音同音近；（3）甲乙二字意义原无联系。”可见，通假字的特点之一是借音不借义，重视的是它们之间的语音联系，如，《管子·问》：“授事以能，则人上功。”中“上”通“尚”。借用甲字的声音表示乙字的意义，甲字的意义并不参与其中，但这并不等于说甲字和乙字意义上毫无联系，不能把这作为识别通假字的标准之一。通假字的另一个特点是，通假字和正字同时并存。通假的产生。必须以通假的两个字同时存在为前提，二者是共时的关系。通假的双方，失去任何一方，也就无所谓通假了。这也是通假字区别于古今字的一个重要特点。古今字的古字和今字在时间上有明显的先后关系。古人在写通假字的时候，并不注意它的形体结构所表示的意义，而只是把它当作一个语音符号。古今字则是借助于汉字的表意功能，通常是在原有的形体上增加或者改变意符，从而减轻原来字形表示的意义太多的负担，以便从视觉上把今字和古字区别开来。也就是说，今字分担了古字的全部或部分意义。此外，文字通假现象是利用现有的文字根据音同音近的原则进行替代，并不意味着汉字的数量的增加，而古今字多数情况下，今字产生，古字不废，汉字的数量有所增加。通假字和古今字的根本区别是共时和历时。但是，共时和历时是相对的，古今字和通假字也不是绝对对立的。当今字产生之后，如果有人再用古字，那就不应该视为古今关系，而应视为通假关系了。但在具体实践中，要弄清每一个字产生的时代，区分通假字和古今字并非易事。即使据现有资料，可以说某个字在某个时代就有了，但是也不敢说，这个字就产生于那个时代。一旦有新材料出土，这个字的产生时间就可能提前。

《先秦文献信息处理》

编辑推荐

《先秦文献信息处理》由世界图书出版公司北京公司出版。

《先秦文献信息处理》

精彩短评

- 1、古典文献信息处理，good
- 2、老板要去Harvard做古汉语处理相关的报告，粗略翻了一遍这本书，挺有意思。最后一章还是不能免俗地分析了一把春秋时期的社会网络...

《先秦文献信息处理》

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：www.tushu000.com