

中文句結構樹資料庫

本院覽號

05T-890901

公告日期

智財權狀態

know-how

摘要

「中文句結構樹資料庫」(Sinica Treebank Version 3.0) 包含了6個檔案，61,087個中文樹圖，361,834個詞，是中央研究院詞庫小組從中央研究院平衡語料庫 (Sinica Corpus) 中抽取句子，經由電腦剖析成結構樹，並加以人工修正、檢驗後所得的成果。在中文句結構樹中，我們標示了中文句語意和語法的訊息。此一「中文句結構樹資料庫」目前開放網上檢索及資料移轉，以供學者專家在中文句法、語意關係研究參考之用。請參考 <http://turing.iis.sinica.edu.tw/treesearch/>

技術優勢

本發明為中文文章語意分析的基本工具。乃中央研究院詞庫小組從中央研究院平衡語料庫 (Sinica Corpus) 中抽取句子，經由電腦剖析成結構樹，並加以人工修正、檢驗後所得的成果。因經語言學專家之審定，故具相當之正確性，為中文句法、語意關係研究建立了一正確的基礎工具。

應用範圍

中文資料檢索、分析。 中文文章語意分析。

創作人

陳克健、黃居仁等



中央研究院
ACADEMIA SINICA