

計畫名稱：中央研究院機構計畫數位典藏與數位學習國家型科技計畫—語言典藏—上古
漢語文獻標記語料庫計畫

主持人：魏培泉

執行期間：2007.01.01-2012.12.31

贊助單位：本院資訊科技創新研究

上古漢語文獻標記語料庫計畫

http://dbo.sinica.edu.tw/Ancient_Chinese_tagged/

計畫簡介

「上古漢語文獻標記語料庫計畫」的主要目標，是要建構一個具有語言信息的上古漢語文獻標記語料庫。自從中央研究院開發漢籍全文資料庫以後，中國古代文獻的數位產業也隨之急速發展，到今日我們已經可以看到許多可供查詢的古籍數位資料庫。然而其中尚存在新舊文獻共構以及語言信息缺乏的問題。因此，本計畫所處理的文獻將包括戰國以至西漢的傳世文獻與若干重要的出土文獻，主要工作是在這些文獻上進行斷詞以及標注詞類。其成果將成為可以讓使用者利用詞彙來搜檢以及進行整理統計的語料庫。