

中文詞庫（八萬目詞）

中文詞知識庫為一包含八萬八千目詞的電子辭典。本詞庫收的詞包含一般用詞、常用專有名詞、成語、慣用語、常用派生詞、異體詞、合併詞以及少數特殊領域用語和古漢語詞語。每個詞項包含的訊息有：

1. 詞條(中文詞)，
2. 詞頻(根據中央研究院五百萬詞語料庫)，
3. 發音(包括注音符號及漢語拼音)，
4. 詞類(根據技術報告 93-05 的分類)，
5. 名詞語義分類(根據技術報告 93-05 名詞概念結構的分類)。

註：資料中的漢語拼音符號原有 ê ü 兩個特殊字元，為處理方便，分別做如下的替換：

ㄌㄩㄢˊ lüan->lywan

ㄋㄩˋ nü->nyu

ㄌㄩˋ lü->lyu

ㄝ ê->eh

CKIP Electronic Dictionary

The CKIP Electronic Dictionary is an electronic lexicon for Mandarin Chinese containing 88,000 entries. Each entry contains:

1. print form (Chinese characters),
2. word frequency (based on a 5 million words corpus),
3. pronunciation (National Phonetic Alphabets, Zhu4yin1fu2hao4 and Chinese Phonetic Alphabet, Han4Yu3Pin1Yin1),
4. syntactic category (based on CKIP classification of 198 categories),
5. semantic feature (base on CKIP classification of 123 concept nodes for nouns).

Note: The information in Chinese Phonetic Alphabet contains ê and ü two special symbols, for the convenience of computer processing, the following replacement is adopted:

ㄌㄩㄢˊ lüan->lywan

ㄋㄩˋ nü->nyu

ㄌㄩˋ lü->lyu

ㄝ ê->eh