

# 中華民國計算語言學學會 中文分詞語料庫 申請說明

中文分詞語料庫為一包含兩百萬詞、不含詞類標記的語料庫，每個文句皆根據分詞原則來斷詞。而此分詞原則，乃是中華民國計算語言學學會在經濟部中央標準局委辦的「資訊處理用中文分詞規範調查研究及草案研擬」計畫中所訂定的。本語料庫來源包括書面語和口語兩部分，其中資訊類佔 21%。

## 申請文件：

- 由服務機關開立身份證明書,內文需言明本資料庫僅用於個人研究，絕不用於商業用途。
- 備妥一式兩份使用協議書（本會與申請人各執乙份）

## 請備妥上列文件寄至：

中華民國計算語言學學會  
台北市南港區研究院路二段128號  
中央研究院資訊所（轉）

## 付款方式：

- 劃撥：帳號19166251，戶名：中華民國計算語言學學會
- 支票：請加註「中華民國計算語言學學會」抬頭
- 信用卡：請填寫信用卡資料

## 工本費用：

個人：

- |                                |         |
|--------------------------------|---------|
| <input type="checkbox"/> 會員    | 免費      |
| <input type="checkbox"/> 國內非會員 | NT\$500 |
| <input type="checkbox"/> 國外非會員 | US\$50  |

團體 (1-10 人使用)

- |                                |           |
|--------------------------------|-----------|
| <input type="checkbox"/> 會員    | 免費        |
| <input type="checkbox"/> 國內非會員 | NT\$5,000 |
| <input type="checkbox"/> 國外非會員 | US\$500   |

# 中華民國計算語言學學會

## 中文分詞語料庫授權使用協議書

立協議書人（以下簡稱本人）現經中華民國計算語言學學會授權，取得中文分詞語料庫之使用權。本人同意切實遵守下列條款：

- 一. 中文分詞語料庫之電子型式，組成內容型式，著作權屬中央研究院詞庫小組。智慧財產權屬中央研究院及經濟部中央標準局共有，並授權中華民國計算語言學學會進行非營利性研究授權。
- 二. 中文分詞語料庫文件內容之著作權屬原始著作人所有。
- 三. 本協議書係針對使用上述第一項之著作內容進行研究之授權。
- 四. 牽涉上述第一項及第二項著作內容之任何引用之商業行為應與著作權所有人另定約規定之。
- 五. 本語料庫內之部份或全部內容，及其直接衍生資料（即詞彙庫，字/詞頻統計）不得包含／引用於任何商業產品中。
- 六. 本人不得將此語料庫之全部或部份移轉給第三人。
- 七. 本人應用此語料庫從事學術研究而獲致成果時，應於本人所發表的論文或其他公開文件中申明誌謝 (1) 中華民國計算語言學學會，(2) 中央研究院詞庫小組，及 (3) 經濟部中央標準局。
- 八. 本人如對此語料庫內容或附屬程式有所改進或擴充，願意無條件回饋中華民國計算語言學學會，以共同促進中文計算語言學之發展。
- 九. 本人如有侵犯此語料庫著作權或智慧財產權之行為，應自行負擔法律責任。

立協議書人： (蓋章)

服務單位： (蓋章)

職稱：

E - m a i l：

中 華 民 國                      年                      月                      日

---

審核單位： (由本會填寫)

# 中華民國計算語言學學會 中文分詞語料庫授權使用協議書

立協議書人（以下簡稱本人）現經中華民國計算語言學學會授權，取得中文分詞語料庫之使用權。本人同意切實遵守下列條款：

- 一. 中文分詞語料庫之電子型式，組成內容型式，著作權屬中央研究院詞庫小組。智慧財產權屬中央研究院及經濟部中央標準局共有，並授權中華民國計算語言學學會進行非營利性研究授權。
- 二. 中文分詞語料庫文件內容之著作權屬原始著作人所有。
- 三. 本協議書係針對使用上述第一項之著作內容進行研究之授權。
- 四. 牽涉上述第一項及第二項著作內容之任何引用之商業行為應與著作權所有人另定約規定之。
- 五. 本語料庫內之部份或全部內容，及其直接衍生資料（即詞彙庫，字/詞頻統計）不得包含／引用於任何商業產品中。
- 六. 本人不得將此語料庫之全部或部份移轉給第三人。
- 七. 本人應用此語料庫從事學術研究而獲致成果時，應於本人所發表的論文或其他公開文件中申明誌謝 (1) 中華民國計算語言學學會，(2) 中央研究院詞庫小組，及 (3) 經濟部中央標準局。
- 八. 本人如對此語料庫內容或附屬程式有所改進或擴充，願意無條件回饋中華民國計算語言學學會，以共同促進中文計算語言學之發展。
- 九. 本人如有侵犯此語料庫著作權或智慧財產權之行為，應自行負擔法律責任。

立協議書人： (蓋章)

服務單位： (蓋章)

職稱：

E - m a i l：

中 華 民 國                      年                      月                      日

---

審核單位： (由本會填寫)