

# iCorpus: 台華語新聞語料庫

本院覽號

公告日期

智財權狀態

05T-1070906

know-how

## 摘要

iCorpus的內容為將擷取民視新聞報導的華文文字稿逐句翻譯為台文。內容收集有新聞報導3,266則，共83,544句。台語與華語各約50萬詞、100萬字。

## 技術優勢

- iCorpus 是少數大量的華台平行語料之一，尤其是內容為新聞報導，可以收集到新產生的台語與華語語彙。

## 應用範圍

- 華、台兩種語言自動翻譯之研究與系統開發。



圖1. iCorpus 台華新聞語料庫目錄



圖2. iCorpus 台華新聞語料庫內容頁

## 創作人

陳孟彰、何澤政、高明達



中央研究院  
ACADEMIA SINICA