自然語言處理實驗室

Natural Language Processing Laboratory

指導老師：林川傑 (cjlin@email.ntou.edu.tw)

實驗室網址： nlp.cse.ntou.edu.tw

研究目標

自然語言處理一直是人工智慧研究重要領域。如何讓電腦分析與處理人類語言、讓電腦理解人類意念表達後提供服務、讓電腦與使用者能以人的語言溝通，可說是電腦科學終極目標，也是未來改變世界的重要資訊科技之一。這個領域的研究近年來極為受到重視，未來所有電腦系統都需要更友善的見面，此時自然語言處理扮演極重要的角色。

本實驗室的研究目標，即為所有與自然語言處理技術相關的課題和應用，包括：

❖ 台灣南島語

本實驗室近年來許多研究都和台灣南島語研究有關，語言涵蓋了泰雅語、太魯閣語、賽德克語、排灣語和布農語。未來會拓向更多語言。

❖ 台灣本土語言處理

本實驗室研究不僅針對英文和中文，也可擴展至台灣本地所使用的各種語言，像是台語、客家語、手語等。

❖ 社群網站及部落格之資訊處理

網際網路上的資訊處理是目前最熱門的課題。除了一般網站外，社群、討論區、電商平台這些大量文字資訊也是重要資源。

❖ 自動問答系統

一個能接受自然語言問句、並從大量自然語言文件或網際網路上找尋答案的系統。回應包括答案本身以及出現的網頁，提供迅速找尋答案的方法。

這些技術的發展出可提供學生研究訓練之外，未來成果均可實際應用在各種資訊系統以及網際網路線上系統之上。

研究計畫 (部份)

- ❖ 泰雅語群新創詞推薦系統 (NSTC 113-2221-E-019-037 -)
- ❖ 使用詞綴資訊之泰雅語群與中文神經網路機器翻譯(NSTC 112-2221-E-019 -049 -)
- ❖ 臺灣南島語與中文之神經網路機器翻譯 (MOST 111-2221-E-019 -070 -)
- ❖ 賽德克語-中文機器翻譯之研究 (MOST 110-2221-E-019-068 -)
- ❖ 賽德克語構詞結構之自動解析及標記工作 (MOST 109-2221-E-019-053 -)
- ❖ 中文歧視性文字之自動偵測與分類 (MOST 108-2221-E-019-045 -)

論文發表 (部份)

Chuan-Jie Lin, Li-May Sung, Chun-Kai Yang, and Bo-Yi Jin (2025) "Formosan Language Morphological Analyzer," Proceedings of the 2025 IRI, EMRITE, pp. 246-249.

Chuan-Jie Lin, Chun-Kai Yang, and Li-May Sung (2024) "Solving Out-of-Vocabulary Problem in Atayalic Languages by Morphological Enumeration," Proceedings of the Educational Data Mining in Writing and Literacy Instruction Workshop (WLIEDM), Educational Data Mining 2024, pp. 18-21.

Ting-Hsuan Chou and Chuan-Jie Lin (2023) “台語—國語神經網路翻譯系統 Taiwanese-Mandarin Neural Machine Translation," Computational Linguistics and Chinese Language Processing, Vol. 28, No.1, pp. 35-52. [In Chinese]

I Chung and Chuan-Jie Lin (2021) "TOCAB: A Dataset for Chinese Abusive Language Processing," Proceedings of the IEEE 22nd IRI, EMRITE, pp. 445-452.

Chuan-Jie Lin, Li-May Sung, Jing-Sheng You, Wei Wang, Cheng-Hsun Lee, and Zih-Cyuan Liao (2020) "Analyzing the Morphological Structures in Seediq Words," International Journal of Computational Linguistics and Chinese Language Processing, Volume 25, Number 2, December, 2020, pp. 1-20. [THCI Core]

Hsu Yang and Chuan-Jie Lin (2020) "TOCP: A Dataset for Chinese Profanity Processing," Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying (TRAC-2), International Conference on Language Resources and Evaluation, pp. 6-12.

研究設備

GPU電腦	3台
個人電腦	10台
印表機	2台