

自然語言實驗室研究現況

- 研究方向：自然語言處理、深度學習、文法改錯、語言搜尋引擎
- Prototypes
 - 詞彙語意資料集 LanguageNet (類似 ImageNet資料集)
 - 在 WordNet 加入語意相容的英中例句 (預計產生標示 1000萬筆語意標示)
 - Linggle 語言搜尋引擎 (英文、中文、跨語言——打中查英)
 - 1兆詞英文 (Google), 7億 + 10億字中文 (中央社+聯合報)
 - 自動英文文法改錯工具、自動中文拼字改錯工具
 - CoNLL2013競賽第2名，目前採用 Open NMT 類神經路線，採用聯合報編輯記錄 5千萬詞，14年新聞共9.7億詞，與師大合作透過 Cool English 網站推廣
 - WriteAhead 寫作提示工具
 - 文法句型提示
 - 語言程度等級提升

英文文法改錯工具、華文拼字改錯工具

- 研究方法：類神經機器翻譯架構，採用Open NMT
- 研究資料集：(1) EF Cambridge Dataset 學習者語料庫 (2) 聯合報編輯記錄 5千萬詞
- 雛形系統舉例：
 - 英文文法改錯（可以改正常見錯誤：冠詞、介詞、單複數、動詞時態、型態）
 - Can we discuss **about** this issue ?
 - **Be Being** punctual is very **improtant** important .
 - Look **forword** forward to **see** **seeing** your progress .
 - We can go **to** for a picnic .
 - 華文拼字改錯（可以改正常見同音、近音、形似、數字規範的錯誤）
 - **描瞄准準**暑假商機
 - 這篇文章很有文學造**旨詣**
 - 熱淚盈**匡眶**

plant

noun

1 (botany) a living organism lacking the power of locomotion 植物

Synonyms flora; plant life;

1.1 植物

To use algae as a replacement for fossil fuels , oil is extracted from the **plant** and developed into a biofuel that can be burned in diesel engines .

若用藻類代替化石燃料，油就會從**植物**中提取出來，並被加工為可以在柴油機中燃燒的生物燃料。

1.2 植株

After root reduction , the cluster bud can be grow into complete **plant** , the breeding cycle were shortened greatly .

叢生芽經過生根誘導後就可以形成完整**植株**，從而極大的縮短了繁殖週期。

2 buildings for carrying on industrial labor; 工廠

Synonyms industrial plant; works;

2.1 裝置

Dcs has been used in 40 kt a sulfur recovery **plant** .

a硫磺回收**裝置**自動控制採用DCS集散控制系統。

2.2 工廠

She and her extended family were wearing surgical masks to protect themselves from a chemical leak at a damaged ammonia **plant** a few miles away .

因為幾公里之外一個氨水**工廠**洩漏，她們一家老小得帶著手術面罩來保護自己。

Linggle 英語版、Linggle 華文版

linggle¹⁰¹²

present a ~method



present a method

44.1% 27,000

Show

present a technique

10.7% 6,500

Show

present a system

10.3% 6,300

Show

present a methodology

8.2% 5,000

Show

le

linggle¹⁰¹²

~網紅



美女

18.8% 12,290

Show

名人

17.8% 11,636

Show

社群

14.0% 9,151

Show

當紅

9.0% 5,891

Show

名媛

3.5% 2,281

Show

15

目前研究重點

- 詞彙向量研究
 - 英文、中文並存的詞彙語意向量 (bilingual word embeddings)
 - 詞彙語意向量 (sense embeddings)
- 分析語料庫內詞典未收錄詞彙的詞性、語意
 - 魯蛇、鐵粉、辣妹
- 自然語言理解
 - 詞彙語意解歧 (e.g., 辨識下面三句中 party 分別是什麼意思 (派對、政黨、當事人))
 - Peter throws really wild parties.
 - They contacted party members to ask for their support.
 - It's often difficult to establish who the guilty party is following a traffic accident.
- 輔助閱讀工具
 - 自選教材、自定等級、提示目標單字 (翻譯、例句) 、單字測驗、維基百科超連結