

NLP Lab: Week 7

Grammatical Error Correction

Automated Error Annotation Tool —ERRANT for Mandarin 作業任務

張俊盛 Jason S. Chang

National Tsing Hua University

10820ISA 562100

自然語言處理實作 Natural Language Processing Lab

2020-0414 Google Hangout 網路直播

有助教在 台達館107室

本次作業任務

- 任務 Level 1 (80分)
 - 輸入漢語編輯資料檔案
 - 輸出 ERRANT m2錯誤標示資料 (編輯動作, 詞性, 位置)
- 任務 Level 2 (滿分)
 - 輸入漢語編輯資料檔案
 - 檢查相關相鄰的 INS 和 DEL 是否為同音近音的拼字錯誤
INS + DEL = REP:SPELL (請助教提供中文單字的注音資料)
 - 輸出 ERRANT m2 錯誤標示資料
- 任務 Level 3 (加分題)
 - 輸入漢語編輯資料檔案
 - 考慮字形, 部首, 偏旁, 適當地合併相鄰 INS 和 DEL 為 REP:SPELL
 - 考慮詞性, 適當地把相關相鄰的 INS 和 DEL 合併為 REP
 - 輸出 ERRANT m2錯誤標示資料

期末專題提示

- 期末專題需要提出 proposal 經過審查通過
- 對執行期末專題，表現良好者，將輔導寫作論文，發表論文，修改英語
- 方向
 - 英語之外的其他語言，如漢語
 - 加入更多語言學考慮 ability of thinking 改成 ability to think 中，不是兩個 replace 的錯誤與編輯，而是一個 PREP +ing 改成 to-INF 的 replace 錯誤與編輯
 - 標示引發錯誤的「問題詞」(problem words)。例如，在 rude with him è rude to him 中，rude 是 [-with-]{+to+} 的問題詞
 - 標示錯誤與編輯，所涉及的文法規則 (grammar pattern) 。例如，在 rude with him è rude to him 涉及，rude: ADJ with n è ADJ to n 。
 - 提供制式的錯誤解釋、例句
- 目標：新一代英語或漢語錯誤標示工具