

中文諧音詞創意生成系統：結合語音相似度與 語境分析

Chinese Homophonic Term Generation System: Integrating Phonetic Similarity and Contextual Analysis

指導教授：陳響亮

專題成員：梁禎庭

開發工具：Python、CKIPTransformer、BERT

測試環境：Windows 11

一、簡介：

本系統旨在自動產生中文諧音梗，達成語意合理又具語音雙關效果的文字替換。

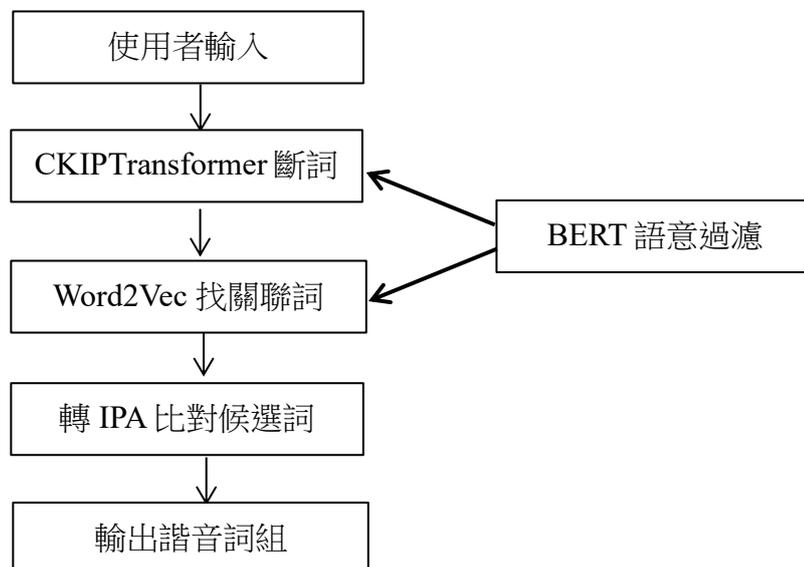
使用者輸入一段文字後，系統先用 CKIPTransformer 進行斷詞，並透過 BERT 模型計算詞與整句的語意關聯性，篩選出語意上關鍵的幾個詞。

接著，使用 Word2Vec 找出與這些詞語意相關性高的詞，再將它們逐字拆開，用 BERT 檢查每個字與原詞的語意相似度，只保留語意相近的字。

然後將這些字與預先建立的候選詞資料庫進行語音比對，轉為 IPA 來計算語音距離，留下發音相近的詞組。

最後從匹配結果中選出語音最接近且語意通順的替代表達，產生具娛樂性與創意性的諧音梗。

以下為系統架構圖：



二、測試結果：

輸入字串：香甜蛋糕，柔軟綿密，多種口味，滿足味蕾。

```
最佳匹配字串：天高地厚 --> 甜糕地厚
匹配分數：2.0000
匹配對應：
  高 → 糕 (相似度：1.0000)
  天 → 甜 (相似度：1.0000)
-----
最佳匹配字串：向高世 --> 香糕世
匹配分數：1.9937
匹配對應：
  高 → 糕 (相似度：1.0000)
  向 → 香 (相似度：0.9937)
-----
最佳匹配字串：高次項 --> 糕次香
匹配分數：1.9937
匹配對應：
  高 → 糕 (相似度：1.0000)
  項 → 香 (相似度：0.9937)
-----
最佳匹配字串：高電位 --> 糕甜位
匹配分數：1.9833
匹配對應：
  高 → 糕 (相似度：1.0000)
  電 → 甜 (相似度：0.9833)
-----
最佳匹配字串：高雄港 --> 糕香港
匹配分數：1.9748
匹配對應：
  高 → 糕 (相似度：1.0000)
  雄 → 香 (相似度：0.9748)
-----
最佳匹配字串：高雄市 --> 糕香市
匹配分數：1.9748
匹配對應：
  高 → 糕 (相似度：1.0000)
  雄 → 香 (相似度：0.9748)
```

可從數個結果中挑選較合適的：

糕香市（高雄市）、甜糕地厚（天高地厚）、香糕世（向高市）