

Named Entity Recognition for Information Extraction

Student: Samantha Tan Swee Yun Supervisor: Hui Siu Cheung



Project Background

Named Entity Recognition (NER) for Information Extraction (IE) has grown in importance due to its capability to streamline processes such as administrative tasks by providing real-time feedback overview. As Singapore is a well-known multicultural country, which consists of unique food, street and location names that may not always be in English, it is thus important for us to investigate NER on Singapore-based datasets. However, as the quality of NER is known to be affected by factors such as data sparsity and data diversity, we propose the use of an Named Entity Matching (NEM) dictionary instead to increase the performance of the IE process.

Project Objectives

The aim of this project is to study and evaluate different NER models for building an NEM dictionary such as a Singapore Food Location NEM Dictionary by:

- Evaluating the performance of FLERT XLM-R, CL-KL and XLNet on a standard CoNLL-2003 benchmark dataset
- Applying the top performing NER models on Singapore-based datasets

Experimental Results

CL-KL model without external context retrieval obtained the highest f1-score on the Singapore Hawker Centre Dataset and was able to better meet our objective of extracting Singapore food locations and addresses based on its performance. Hence, it is the best model to generate our NEM Dictionary.

Best Model: CL-KL Architecture

