

Milestone 1 Report

修課學生：林佳穎 林子翔 蘇柏元 徐彥旻

特別感謝旁聽學生：Hubert Lin 林鑫彤 翁福德

1. Construct the backend databases

- 將[台大醫院掛號系統](#)依病症掛號的資料爬下來存到 division.csv
- 再將[元氣網疾病表列](#)爬下來存到 disease.csv
- 再將 csv 的資料匯入線上資料庫 nosql: mongodb
- 醫生門診時刻表 doctor.csv 會依照使用者查詢內容, 即時去『[台大醫院掛號系統](#)依醫師名稱掛號』爬回來(目前尚未完整實作這個部份)

Disease table(包含：中文病名,英文病名,所屬科別,發生部位,症狀,介紹網站)

中文病名	英文病名	所屬科別	部位	症狀
過敏性鼻炎	['allergic rhinitis']	['耳鼻喉科', '家醫科']	['鼻']	['打噴嚏', '眼睛癢', '耳朵癢', '鼻塞', '流鼻涕', '鼻涕倒流', '頭暈', '注意力無法集中', '黑眼圈']
白內障	['cataract']	['眼科']	['眼']	['視力不穩定', '視力模糊', '遠視', '畏光', '夜間眩光', '物體明暗對比難分辨', '色調改變', '物體顏色變暗', '近視度數逐年大增', '伸手']
青光眼	['Glaucoma']	['眼科']	['眼']	['眼睛痛', '發紅', '同側頭痛', '惡心', '視力模糊', '視覺範圍縮小', '嘔吐']

Division table(包含：科別,中文病名,看診醫生)

科別	症狀	醫師
內科部	['內分泌', '甲(副)狀腺病變', '月經異常', '泌乳異常', '肢端肥大', '突眼', '骨質疏鬆']	['王治元', '李弘元', '呂金盈', '張以承', '吳婉禎', '李']
內科部	['關節炎風濕症', '類風濕性關節炎', '肌肉痠痛', '紅斑性狼瘡', '免疫疾病']	['許秉寧', '李光倫', '陳茂源', '吳政翰', '李克仁', '沈玢好', '李']

Doctor table(包含：網路掛號狀態,醫師,門診時間,院區,地點,科別,門診名稱,診別,電話代碼,可否現場掛號,備註)(尚未完整實作)

掛號	醫師	門診時間	院區	地點	科別	門診名稱	診別	電話代碼	現場掛號	備註
.初診.	林隆君	106.4.14(三)	總院區	西址-1樓	內科部	普通門診	36	100136	可	心臟
.初診.	林隆君	106.4.21(三)	總院區	西址-1樓	內科部	普通門診	36	100136	可	心臟

2. Design the dialogue semantic schema

A. Decide the number of tables/domains to support

- a. disease table(包含：中文病名,英文病名,所屬科別,發生部位,症狀,介紹網站)
- b. division table(包含：科別,中文病名,看診醫生)
- c. doctor table(包含：網路掛號狀態,醫師,門診時間,院區,地點,科別,門診名稱,診別,電話代碼,可否現場掛號,備註) (尚未完整實作)

B. Decide the supported user intents

Five user intents:

- a. 查詢某疾病的症狀
- b. 查詢某疾病的科別
- c. 查詢某疾病或某科別的主治醫生
- d. 查詢門診時間 (尚未完整實作)
- e. 預約掛號 (尚未完整實作)

C. Decide the slots that can be specified

Four slots that can be specified:

- a. disease 疾病名稱
- b. division 門診科別
- c. doctor 醫生名稱
- d. time 門診時間

3. Generate training examples for each intent

Annotate semantic frames (intents and slots)

A. 查詢症狀

- (請問/"") (得/"") +disease+ (會) (怎麼樣/怎樣) ?
- (請問/"") (得/"") +disease+ (會有) (什麼/哪些) 症狀 ?

B. 查詢科別

- (請問/"") +disease+ (是/屬於/要看/要掛/要掛號) (哪/什麼)
科 ?

C. 查詢醫生

- (請問/"") + (time/"") + (有/"") (什麼/哪些) 醫生 (可以/"") 看
+ (disease/division) ?
- (請問/"") + (time/"") + (disease/division) + (要/可以/"") (看/
找/掛/掛號) (什麼/哪些) 醫師 ?

D. 查詢門診

- (請給我/給我/請告訴我/告訴我/請問) + (time/"") +
(doctor/disease/division) +的門診 (時刻表/時間)

E. 預約掛號

- 我 (想/要/想要) (掛/掛號/預約) + (time/"") +
(doctor/disease/division) (""/的門診)

4. Use machine learning method to train your system

Intent Prediction: LSTM for intent classification

前處理

用 jieba 將輸入的句子斷詞，再依照 index，將句子中的每個詞轉為 one-hot 的形式。

模型架構

一層 LSTM 加上 full-connected layer，輸入是整個句子，輸出是六維的 numpy array，每一維代表 "greeting" 或上述其中一種 intent。

Slot Filling:

我們參考的是 Github 上的這個模型： Implements RNNs in Keras
(<https://github.com/chsasank/ATIS.keras>)

前處理

用 jieba 將輸入的句子斷詞 (有建立相關的詞的字典，比如醫生名字、疾病、症狀等等)，再依照 index，將每個詞轉為一個整數。

模型架構

將一個句子的每個詞經過 embedding 之後，整句話經過 CNN (128 個長度為 5 的 filter)，接者再經過 GRU，然後接上 full-connected layer 做每個詞的 classification。整個模型的輸出即為模型為每個詞標記的結果。