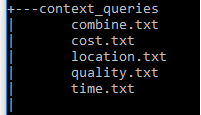


List các file và folder

**MÔ TẢ CÁC FOLDER VÀ FILE QUAN TRỌNG**

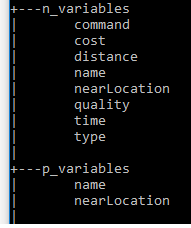
+ Folder context\_queries: chứa các file mẫu câu hạt giống cho từng ngữ cảnh. Các file trong folder context\_queries liệt kê trong hình 1.



Hình 1: các file trong folder context\_queries

Ví dụ: file cost chứa các câu liên quan đến giá. File combine chứa các câu liên quan đến nhiều ngữ cảnh.

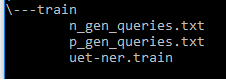
+ Folder p\_variables và n\_variables: chứa các giá trị của các lớp ở file tương ứng. Các file trong hai folder trên liệt kê trong hình 2. p\_variables (primary variables) chứa giá trị cho lớp NAME và LOCATION. n\_variables (normal variables) chứa giá trị cho tất cả các lớp.



Hình 2: Các file trong folder n\_variables và p\_variables

+ Hai file p\_queries.txt và n\_queries.txt: chứa các mẫu câu hạt giống ghép từ các file trong folder context\_queries. Trong đó file p\_queries.txt chứa các mẫu câu cho ngữ cảnh về location bởi vì trong ngữ cảnh location chỉ chứa giá trị các lớp NAME và LOCATION. Và các mẫu câu về hai lớp NAME và LOCATION cần tần suất xuất hiện nhiều hơn các lớp khác. File n\_queries.txt chứa các mẫu câu hạt giống cho các ngữ cảnh còn lại.

+ Folder train: chứa các file được sinh ra cho quá trình huấn luyện. Hình 3 liệt kê các file trong folder train



Hình 3: Các file trong folder train

+ Hai file p\_gen\_queries.txt và n\_gen\_queries.txt: chứa các mẫu câu hạt giống và các mẫu câu được sinh ra từ mẫu câu hạt giống từ hai file file p\_queries.txt và n\_queries.txt với giá trị của các lớp trong folder p\_variables và n\_variables.

+ File uet-ner.train: chứa các mẫu câu đã được gán nhãn trong hai file p\_gen\_queries.txt và n\_gen\_queries.txt.

+ File get\_q.py: script để ghép các file trong folder context\_queries ra hai file p\_queries.txt và n\_queries.txt. Và sinh gaz-list cho một số lớp có ít giá trị chứa trong file uet-search-gaz.txt.

+ File lets\_train.py: script để sinh mẫu câu, tự động gán nhãn và chạy lệnh train.

+ Folder test và các file run\_test\_in.py, run\_test\_nc.py: phục vụ cho việc test.

**CHẠY PROJECT**

Để chạy project yêu cầu python 2.7 trở lên được cài đặt.

Chạy thứ tự hai câu lệnh dưới tại cmd để tiến hành huấn luyện. Kết quả model ở trong file model/uet-search-ner-new.gz

python get\_q.py

python lets\_train.py