Author: Kevin Zen

Makefile : Easy commands to setup local environment.
README.md : How to interact with the repo.
environment.yml : Defines environment dependencies.
ubi_vocab.code-workspace : Ignore - VSCode workspace file.

./data:
all-the-news-2-1.csv : Extracted news article data as csv.
all-the-news-2-1.zip : Zipped news article data as csv
gre_vocab.csv : Cleaned Gre vocab file from magoosh helper.
master_labeled.csv : Final labeled dataset with lesk, pos and bert results for wsd.
labeled_lesk.csv : Ignore: intermediary labeled dataset with lesk results. See master_labeled.csv
labeled_pos_matched.csv : Ignore, contains pos matching algorithm results on labeled data, see master_labeled.csv
labeled_pos_matched_2.csv: Ignore, contains pos matching algorithm results on labeled data, see master_labeled.csv
news_small.csv: A 1k sampled version of all-the-news-2-1.csv
tmp_small_merged_df.csv : Ignore

./ubi_vocab/utils:
__init__.py
__pycache__
api.py : Main interface for replacing words in a body of text with gre vocabulary words.
article_class.py : Defines class to house article data.
constants.py : Defines constants used in project.
data_io.py : Defines reading data methods.
demo.ipynb : Demo interactive notebook shared on binder.
logger_utils.py : Commonly used logging function.
main.py : Non user facing main interactive function, contains eda and main internal functionalities for replacing sentences, and applying WSD techniques.
metrics.py : Calculate metrics like precision,recall, f1 on the results.
pre_process.py : Functions to extract synonyms.
transformer.py : Function to implement baseline BERT model.
wsd.py : LESK and BERT for WSD functions.

./ubi_vocab/utils/__pycache__:
api.cpython-37.pyc
article_class.cpython-37.pyc
constants.cpython-37.pyc
data_io.cpython-37.pyc
logger_utils.cpython-37.pyc
logging.cpython-37.pyc

main.cpython-37.pyc
metrics.cpython-37.pyc
pre_process.cpython-37.pyc
transformer.cpython-37.pyc
wsd.cpython-37.pyc