User Guide

This repo provides the code related to my journal work. It has the Geo Parser, the classifier code and the transfer learning code.

GeoParser
The Geo Parser performs a lookup from 4 gazetters (Open Street Map, Ordanance Survey, GeoText, Geonames) to detect the occurence of a geographic featuretype or a placename from some text. It uses API calls and string matching to detect occurences of the above mentioned entities. The module is different from other placename detection tools in the aspect that it also detects geographic feature types. I have used ADL feature types dictionary for this.

To run: run the main.py file.
Input: all_18828.txt
Output: returns counts of the place names and geographic feature types in a sentence

Dataset:
Contains the data for the classifiers as well as the transfer learning code

MLcode

Classifiers
Standalone classifier

Naive Bayes, Random Forests, SVM's

To run: run the Jupyter notebook (SVM_NB_RF_Classifiers)
Input: WEKA_dataset.csv
Output: predicted class.

Metaclassifier
To run: run the Jupyter notebook (SVM_SVM_Metaclassifier)
Input: WEKA_dataset.csv
Output: predicted class.

Transfer learning
UlmFit (Fast ai)
To run: run the Jupyter notebook (GeoSpatial_ULMFIT)
Input: wordemb_dataset.csv
Output: predicted class.

BERT (hugging face)
To run: run the Jupyter notebook (Geosp_BERT_finalcode)
Input: wordemb_dataset.csv
Output: predicted class.