

NER-validation

June 19, 2022

```
[11]: import spacy
import pandas as pd
# Load English tokenizer, tagger, parser and NER
nlp = spacy.load("en_core_web_sm")

base_path = '/home/manju/Desktop/assign/task_data/'

def readInputFile(csv_file):
    data = pd.read_csv(csv_file)
    transcript = data['transcription']
    return transcript

def main(base_path):
    train_file = base_path + 'train_data.csv'
    valid_file = base_path + 'valid_data.csv'

    transcript = readInputFile(valid_file)
    print("Transcription analysis: Noun Phrases and Verbs are printed, if there are any NER present in the sentence they will be printed ")
    transcript = transcript[225:250] #Select few sentences randomly for testing
    # Process documents one by one
    for text in transcript:
        doc = nlp(text)

        print("\nSentence: \"", text, end = '\"; ')
        # Analyze syntax
        print("Nouns:", [chunk.text for chunk in doc.noun_chunks], end = '; ')
        print("Verbs:", [token.lemma_ for token in doc if token.pos_ == "VERB"], end = '; ')

        # Find named entities, phrases and concepts
        for entity in doc.ents:
            print("NER:", entity.text, entity.label_, end = '; ')

if __name__ == "__main__":
    main(base_path)
```

Transcription analysis: Noun Phrases and Verbs are printed, if there are any NER

present in the sentence they will be printed

Sentence: " Set the language"; Nouns: ['the language']; Verbs: ['set'];
Sentence: " Language settings"; Nouns: ['Language settings']; Verbs: [];
Sentence: " Set language to Chinese"; Nouns: ['Set language', 'Chinese']; Verbs: ['set']; NER: Chinese NORP;
Sentence: " Set language to German"; Nouns: ['Set language']; Verbs: ['set']; NER: German NORP;
Sentence: " Set language to Korean"; Nouns: ['Set language']; Verbs: ['set']; NER: Korean NORP;
Sentence: " Set language to English"; Nouns: ['Set language', 'English']; Verbs: ['set']; NER: English LANGUAGE;
Sentence: " Set my device to Chinese"; Nouns: ['my device', 'Chinese']; Verbs: ['set']; NER: Chinese NORP;
Sentence: " OK now switch the main language to German"; Nouns: ['the main language']; Verbs: ['switch']; NER: German NORP;
Sentence: " OK now switch the main language to Korean"; Nouns: ['the main language']; Verbs: ['switch']; NER: Korean NORP;
Sentence: " OK now switch the main language to English"; Nouns: ['the main language', 'English']; Verbs: ['switch']; NER: English LANGUAGE;
Sentence: " Set my phone's language to Chinese"; Nouns: ["my phone's language", 'Chinese']; Verbs: ['set']; NER: Chinese NORP;
Sentence: " Set my phone's language to German"; Nouns: ["my phone's language"]; Verbs: ['set']; NER: German NORP;
Sentence: " Set my phone's language to Korean"; Nouns: ["my phone's language"]; Verbs: ['set']; NER: Korean NORP;
Sentence: " Set my phone's language to English"; Nouns: ["my phone's language", 'English']; Verbs: ['set']; NER: English LANGUAGE;
Sentence: " I need to practice my Chinese. Switch the language"; Nouns: ['I', 'my Chinese', 'the language']; Verbs: ['need', 'practice', 'switch']; NER: Chinese NORP;
Sentence: " I need to practice my German. Switch the language"; Nouns: ['I', 'my German', 'the language']; Verbs: ['need', 'practice', 'switch']; NER: German NORP; NER: Switch the language PERSON;
Sentence: " I need to practice my Korean. Switch the language"; Nouns: ['I', 'the language']; Verbs: ['need', 'practice', 'switch']; NER: Korean NORP; NER: Switch the language PERSON;
Sentence: " I need to practice my English. Switch the language"; Nouns: ['I', 'my English', 'the language']; Verbs: ['need', 'practice', 'switch']; NER: English LANGUAGE;
Sentence: " Turn the washroom lights off"; Nouns: []; Verbs: ['turn', 'light'];
Sentence: " Pause"; Nouns: ['Pause']; Verbs: [];
Sentence: " Turn on the lights"; Nouns: ['the lights']; Verbs: ['turn'];
Sentence: " Turn off the lights"; Nouns: ['the lights']; Verbs: ['turn'];
Sentence: " Change language"; Nouns: ['Change language']; Verbs: ['change'];
Sentence: " Pause the music"; Nouns: ['the music']; Verbs: ['pause'];
Sentence: " Resume"; Nouns: ['Resume']; Verbs: [];

[]: