
codeswitch Documentation

Release latest

Nov 02, 2020

Contents

1	Supported Code-Mixed Language	3
1.1	Language Code	3
2	Installation	5
2.1	Dependency	5
3	Features & Supported Language	7
4	Language Identification	9
5	POS Tagging	11
6	NER Tagging	13
7	Sentiment Analysis	15
8	Acknowledgement	17

CodeSwitch is a NLP tool, can use for language identification, pos tagging, name entity recognition, sentiment analysis of code mixed data.

Supported Code-Mixed Language

We used [LinCE](#) dataset for training **multilingual BERT** model using huggingface [transformers](#). LinCE has four language mixed data. We took three of it `spanish-english`, `hindi-english` and `nepali-english`. Hope we will train and add other language and task too.

- Spanish-English(`spa-eng`)
- Hindi-English(`hin-eng`)
- Nepali-English(`nep-eng`)

1.1 Language Code

- `spa-eng` for spanish-english
- `hin-eng` for hindi-english
- `nep-eng` for nepali-english

2.1 Dependency

- pytorch >=1.6.0

Features & Supported Language

- Language Identification
 - spanish-english
 - hindi-english
 - nepali-english
- POS
 - spanish-english
 - hindi-english
- NER
 - spanish-english
 - hindi-english
- Sentiment Analysis
 - spanish-english

CHAPTER 4

Language Identification

```
from codeswitch.codeswitch import LanguageIdentification
lid = LanguageIdentification('spa-eng')
# for hindi-english use 'hin-eng',
# for nepali-english use 'nep-eng'
text = "" # your code-mixed sentence
result = lid.identify(text)
print(result)
```


CHAPTER 5

POS Tagging

```
from codeswitch.codeswitch import POS
pos = POS('spa-eng')
# for hindi-english use 'hin-eng'
text = "" # your mixed sentence
result = pos.tag(text)
print(result)
```


CHAPTER 6

NER Tagging

```
from codeswitch.codeswitch import NER
ner = NER('spa-eng')
# for hindi-english use 'hin-eng'
text = "" # your mixed sentence
result = ner.tag(text)
print(result)
```


CHAPTER 7

Sentiment Analysis

```
from codeswitch.codeswitch import SentimentAnalysis
sa = SentimentAnalysis('spa-eng')
sentence = "" # your mixed sentence
result = sa.analyze(sentence)
print(result)
```


CHAPTER 8

Acknowledgement

- LinCE
- BERT
- huggingface