
BlaBla

May 21, 2020

Contents

1	High level overview	1
1.1	Introduction	1
1.2	Installing BlaBla	1
2	BlaBla	3
2.1	DocumentProcessor	3
2.2	Document	3
3	Features Table	5

1.1 Introduction

BlaBla is a package for linguistic feature extraction written in Python. Information and tutorials can be found in our README on GitHub. Please see our paper for more details.

1.2 Installing BlaBla

Installing BlaBla is as easy as installing any other normal Python package

```
git clone https://github.com/novoic/blabla
cd blabla
pip install .
```

BlaBla also uses Stanford CoreNLP for analysing text. To set up CoreNLP, run the following command `./setup_corenlp.sh` after changing `corenlp_dir` and `lang` if required. Please refer to our README file for more details on the installation.

```
./setup_corenlp.sh
```


At the heart of BlaBla is the `DocumentProcessor` and the `Document` class. You have to import the `DocumentProcessor` class to process a piece of input text as shown in the below piece of code.

```
from bla_bla.document_processor import DocumentProcessor
with DocumentProcessor("stanza_config/stanza_config.yaml", "en") as doc_proc:
    content = "The picture shows a boy walking to the kitchen to pick a cookie from_
↳the cookie jar."
    doc = doc_proc.analyze(content, "string")
    res_json = doc.compute_features("noun_rate")
    print(res_json)
```

Under the hood, the `DocumentProcessor` object has an `analyze` method that will return an object of type `Document` class which can be used to compute features

2.1 DocumentProcessor

This page outlines the methods from the `DocumentProcessor` class.

2.2 Document

This page outlines the methods from the `Document` class.

CHAPTER 3

Features Table

- genindex
- search