

TOOL

Library and Functions M2

Below are short examples of functions and methods that are used in Module 2.

Downloading Gutenberg Corpora from NLTK

```
import nltk
nltk.download(['gutenberg'])
nltk.corpus.gutenberg.raw(txt_name)
nltk.corpus.gutenberg.words(txt_name)
nltk.corpus.gutenberg.sents(txt_name)
```

where `txt_name` is string of text name in Gutenberg Library

Tokenization

```
import nltk
nltk.sent_tokenize(s)
nltk.word_tokenize(w)
```

where `s` is sentence, `w` is a word

Normalization

```
import unicodedata
unicodedata.normalize('NFKD', s).encode('ascii', 'ignore').
decode('utf-8', 'ignore')
```

Expanding Contractions

```
import contractions
contractions.fix(t)
```

where `t` is a text



Using TextBlob's Word Object

```
from textblob import Word
Word(w).correct()
Word(w).spellcheck()
```

Stemming and Lemmatization

```
import nltk
pso = nltk.stem.PorterStemmer()
wlo = nltk.stem.WordNetLemmatizer()
pso.stem(s)
wlo.lemmatize(s,tag)
```

where **tag** is a POS tag

