Practical NLTK

David Arroyo Menéndez

December 3, 2018

Practical NLTK: Summary

```
$ git clone https://github.com/davidam/python-examples.git
```

1. Sentiment Analysis
2. Detect Gender
3. Sentence Similarity
4. Text Summary
5. Classify Documents
6. Manage Words

singulars/plurals, dictionary entries, stopwords

1. Gramatical Trees
2. Extract Keywords
3. Disambiguation

Sentiment Analysis

Tweets positives versus negatives

```
$ python3 tweet-sentyment.py
$ python3 tweepy-example.py
```

Detect Gender

Your name is your gender

```
$ python3 sexmachine.py
$ python3 perceval_git_counter_sexmachine.py
$ python3 perceval_mbox_sexmachine.py
```

Sentence Similarity

Your name is your gender

```
$ python3 sentence-similarity.py
$ nosetests3 test/test_sentencesimilarity.py
```

Classify Documents

Classify Documents

```
$ python3 doc-classification-ch06.py
```

Classify Newsgroups

Classify Newsgroups

```
$ python3 nltk-sklearn.py
```

Synonims and Antonims

Synonims and Antonims

```
$ python3 synonims-antonims.py
```

Singulars and Plurals

```
$ python3 stem.py
$ nosetests3 test/test_stem.py
```

Stopwords

For some search engines, these are some of the most common, short function words, such as the, is, at, which, and on. In this case, stop words can cause problems when searching for phrases that include them

Singulars and Plurals

```
$ python3 stopwords.py
$ nosetests3 test/test_stopwords.py
```

Lemmas: dictionary entries

```
$ python3 wordnet-lemmatizer.py
$ nosetests3 test/test_wordnet.py:TddInPythonExample.test_syno
```

Trees

```
I can build a gramatic or semantic tree from a sentence
$ python3 semantic-tree.py
I can generate sentences from a gramatic
$ python3 howtos/generate.py
I can visualize a gramatic
$ python3 parse-tree.py
I can obtain bigrams, trigrams or ngrams
$ python3 bigrams-trigrams.py
$ nosetests3 test/test_bigrams_trigrams.py
I can print a tree from sintactic pairs
$ python3 code-chinker.py
I can build sintactic pairs from a sentence and print a tree
$ python3 tokenizeandtag.py
$ nosetests3 test/test_tokenizeandtag.py
```

Corpus

Corpus

```
$ python3 gutenberg.py
$ nosetests3 test/test_gutenberg.py
$ python3 corpus-howto-new-corpus.py
```

Keywords: rake algorithm

```
$ python3 nltk-rake.py
```

Disambiguation

Disambiguation

```
$ python test_all_words_wsd.py
$ python test_wsd.py

# Remember synset
$ python3 wordnet-example.py
$ nosetests3 test/test_wordnet.py
```

Sharing data models: pickle

```
$ python3 nltk-pickle.py
```

Support Python and Libremanuals

El Tutorial de Python por Guido Van Rossum