Package Index > **liwc** > **0.3.0**

liwc 0.3.0

Linguistic Inquiry and Word Count (LIWC) analyzer (proprietary data not included)   **Downloads** ↓

Linguistic Inquiry and Word Count (LIWC) analyzer.

The LIWC lexicon is proprietary, so it is *not* included in this repository, but this Python package requires it. The lexicon data can be acquired (purchased) from **liwc.net**. This package reads from the `LIWC2007_English100131.dic` (MD5: 2a8c06ee3748218aa89b975574b4e84d) file, which must be available on any system where this package is used.

The LIWC2007 `.dic` format looks like this:

```
%
1    funct
2    pronoun
[...]
%
a    1    10
abdomen*     146  147
about    1    16   17
[...]
```

Setup

Install from **PyPI**:

```
pip install -U liwc
```

Example

```
import re
from collections import Counter

def tokenize(text):
    # you may want to use a smarter tokenizer
    for match in re.finditer(r'\w+', text, re.UNICODE):
        yield match.group(0)

import liwc
parse, category_names = liwc.load_token_parser('LIWC2007_English100131.dic')
```

- `parse` is a function from a token of text (a string) to a list of matching LIWC categories (a list of strings)
- `category_names` is all LIWC categories in the lexicon (a list of strings)

```
gettysburg = '''Four score and seven years ago our fathers brought forth on
  this continent a new nation, conceived in liberty, and dedicated to the
  proposition that all men are created equal. Now we are engaged in a great
  civil war, testing whether that nation, or any nation so conceived and so
  dedicated, can long endure. We are met on a great battlefield of that war.
  We have come to dedicate a portion of that field, as a final resting place
  for those who here gave their lives that that nation might live. It is
  altogether fitting and proper that we should do this.'''
gettysburg_tokens = tokenize(gettysburg)
# now flatmap over all the categories in all of the tokens using a generator:
gettysburg_counts = Counter(category for token in gettysburg_tokens for category in parse(token))
# and print the results:
print(gettysburg_counts)
```

License

Copyright (c) 2012-2017 Christopher Brown. **MIT Licensed**.

File	Type	Py Version	Uploaded on	Size
liwc-0.3.0-py2.py3-none-any.whl (md5)	Python Wheel	py2.py3	2017-10-20	5KB
liwc-0.3.0.tar.gz (md5)	Source		2017-10-20	3KB

Author: Christopher Brown
Home Page: https://github.com/chbrown/liwc-python
License: MIT
Package Index Owner: chbrown
DOAP record: liwc-0.3.0.xml