NLP Basics: What is Natural Language Processing & the Natural Language Toolkit?

+ Code + Text

How to install NLTK on your local machine

Both sets of instructions below assume you already have Python installed. These instructions are taken directly from http://www.nltk.org/install.html.

Mac/Unix

From the terminal:

1. Install NLTK: run `pip install -U nltk`
2. Test installation: run `python` then type `import nltk`

Windows

1. Install NLTK: http://pypi.python.org/pypi/nltk
2. Test installation: `Start>Python35`, then type `import nltk`

Download NLTK data

```
import nltk
nltk.download()
```

```
    showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml
    True
```

```
dir(nltk)
```

```
    ['AbstractLazySequence',
     'AffixTagger',
     'AlignedSent',
     'Alignment',
     'AnnotationTask',
     'ApplicationExpression',
```

```
'Assignment',
'BigramAssocMeasures',
'BigramCollocationFinder',
'BigramTagger',
'BinaryMaxentFeatureEncoding',
'BlanklineTokenizer',
'BllipParser',
'BottomUpChartParser',
'BottomUpLeftCornerChartParser',
'BottomUpProbabilisticChartParser',
'Boxer',
'BrillTagger',
'BrillTaggerTrainer',
'CFG',
'CRFTagger',
'CfgReadingCommand',
'ChartParser',
'ChunkParserI',
'ChunkScore',
'ClassifierBasedPOSTagger',
'ClassifierBasedTagger',
'ClassifierI',
'ConcordanceIndex',
'ConditionalExponentialClassifier',
'ConditionalFreqDist',
'ConditionalProbDist',
'ConditionalProbDistI',
'ConfusionMatrix',
'ContextIndex',
'ContextTagger',
'ContingencyMeasures',
'CoreNLPDependencyParser',
'CoreNLPParser',
'Counter',
'CrossValidationProbDist',
'DRS',
'DecisionTreeClassifier',
'DefaultTagger',
'DependencyEvaluator',
'DependencyGrammar',
'DependencyGraph',
'DependencyProduction',
'DictionaryConditionalProbDist',
'DictionaryProbDist',
'DiscourseTester',
'DrtExpression',
'DrtGlueReadingCommand',
```

```
    'ELEProbDist',
    'EarleyChartParser',
    'Expression',
    'FStructure',
    'FeatDict'
```

What can you do with NLTK?

```
from nltk.corpus import stopwords

stopwords.words('english')[0:500:25]
```

```
['i', 'herself', 'been', 'with', 'here', 'very', 'doesn', 'won']
```