

```
In [1]: import nltk
        from nltk.corpus import names
        from pylab import *
        import random as pyrandom
```

Dialog Act Type Classification

```
In [63]: posts = nltk.corpus.nps_chat.xml_posts()[10000]
```

```
In [64]: def features(post):
        f = {}
        for w in nltk.word_tokenize(post): f[w.lower()] = True
        return f
```

```
In [72]: posts[333].text
```

```
Out[72]: 'wouldnt let her date'
```

```
In [73]: posts[333].get('class')
```

```
Out[73]: 'Emotion'
```

```
In [65]: print set([p.get('class') for p in posts])

set(['Emotion', 'ynQuestion', 'yAnswer', 'Continuer', 'whQuestion', 'System',
    'Accept', 'Clarify', 'Emphasis', 'nAnswer', 'Greet', 'Statement', 'Reject',
    'Bye', 'Other'])
```

```
In [66]: featuresets = [(features(p.text),p.get('class')) for p in posts]
```

```
In [67]: training_set = featuresets[1000:]
        test_set = featuresets[:1000]
```

```
In [68]: classifier = nltk.NaiveBayesClassifier.train(training_set)
        print nltk.classify.accuracy(classifier,test_set)

0.66
```

```
In [ ]:
```