In [1]:

```
import nltk
from nltk.stem import PorterStemmer
stemmerporter=PorterStemmer()
words=['prettier','glasses','carrier','birds','itemization','brighter','chilled','exclaimir
for word in words:
    print(stemmerporter.stem(word))
```

```
prettier
glass
carrier
bird
item
brighter
chill
exclaim
```

In [5]:

```
from nltk.tokenize import RegexpTokenizer
s = "Good muffins cost $3.88\nin New York.  Please buy me\ntwo of them.\n\nThanks."
s1 = "I had a great day today. I am hoping for an even better day tomorrow as it is going t
tokenizer = RegexpTokenizer('\w+|\$[\d\.]+|\S+')
#tokenizer = RegexpTokenizer('[\S]+')
tokenizer.tokenize(s)
tokenizer.tokenize(s1)
```

Out[5]:

```
['I',
 'had',
 'a',
 'great',
 'day',
 'today',
 '.',
 'I',
 'am',
 'hoping',
 'for',
 'an',
 'even',
 'better',
 'day',
 'tomorrow',
 'as',
 'it',
 'is',
 'going',
 'to',
 'be',
 'my',
 'birthday',
 '.']
```

In [6]:

```
from nltk.tokenize import RegexpTokenizer
s = "Good muffins cost $3.88\nin New York.  Please buy me\ntwo of them.\n\nThanks."
tokenizer = RegexpTokenizer('\w+|\$[\d\.]+|\S+')
#tokenizer = RegexpTokenizer('[\S]+')
tokenizer.tokenize(s)
```

Out[6]:

```
['Good',
 'muffins',
 'cost',
 '$3.88',
 'in',
 'New',
 'York',
 '.',
 'Please',
 'buy',
 'me',
 'two',
 'of',
 'them',
 '.',
 'Thanks',
 '.']
```

In [7]:

```
from nltk.stem.snowball import SnowballStemmer
stemmer = SnowballStemmer("english")
words=['sibling','prettier','glasses','queue','carrier','birds','running','angrier','pretti
for word in words:
    print(word,":",stemmer.stem(word))
```

```
sibling : sibl
prettier : prettier
glasses : glass
queue : queue
carrier : carrier
birds : bird
running : run
angrier : angrier
prettiest : prettiest
kilometer : kilomet
```

In [ ]: