In [3]:

```
import nltk
from nltk.corpus import stopwords
stopwords.words("russian")
```

Out[3]:

```
['и',
 'в',
 'во',
 'не',
 'что',
 'он',
 'на',
 'я',
 'с',
 'со',
 'как',
 'а',
 'то',
 'все',
 'она',
 'так',
 'его',
 'но',
```

In [4]:

```
entries = nltk.corpus.cmudict.entries()
len(entries)
for entry in entries:
    print(entry)
```

```
('a', ['AH0'])
('a.', ['EY1'])
('a', ['EY1'])
('a42128', ['EY1', 'F', 'AO1', 'R', 'T', 'UW1', 'W', 'AH1', 'N', 'T', 'UW
1', 'EY1', 'T'])
('aaa', ['T', 'R', 'IH2', 'P', 'AH0', 'L', 'EY1'])
('aaberg', ['AA1', 'B', 'ER0', 'G'])
('aachen', ['AA1', 'K', 'AH0', 'N'])
('aachener', ['AA1', 'K', 'AH0', 'N', 'ER0'])
('aaker', ['AA1', 'K', 'ER0'])
('aalseth', ['AA1', 'L', 'S', 'EH0', 'TH'])
('aamodt', ['AA1', 'M', 'AH0', 'T'])
('aancor', ['AA1', 'N', 'K', 'AO2', 'R'])
('aardema', ['AA0', 'R', 'D', 'EH1', 'M', 'AH0'])
('aardvark', ['AA1', 'R', 'D', 'V', 'AA2', 'R', 'K'])
('aaron', ['EH1', 'R', 'AH0', 'N'])
("aaron's", ['EH1', 'R', 'AH0', 'N', 'Z'])
('aarons', ['EH1', 'R', 'AH0', 'N', 'Z'])
('aaronson', ['EH1', 'R', 'AH0', 'N', 'S', 'AH0', 'N'])
```

In [6]:

```
from nltk.corpus import wordnet as wn
synonym=list()
for wns in wn.synsets('shout'):
    for lemmas in wns.lemmas():
        synonym.append(lemmas)
print(synonym)
```

```
[Lemma('cry.n.01.cry'), Lemma('cry.n.01.outcry'), Lemma('cry.n.01.call'), Le
mma('cry.n.01.yell'), Lemma('cry.n.01.shout'), Lemma('cry.n.01.vociferatio
n'), Lemma('shout.v.01.shout'), Lemma('shout.v.02.shout'), Lemma('shout.v.0
2.shout_out'), Lemma('shout.v.02.cry'), Lemma('shout.v.02.call'), Lemma('sho
ut.v.02.yell'), Lemma('shout.v.02.scream'), Lemma('shout.v.02.holler'), Lemm
a('shout.v.02.hollo'), Lemma('shout.v.02.squall'), Lemma('exclaim.v.01.excla
im'), Lemma('exclaim.v.01.cry'), Lemma('exclaim.v.01.cry_out'), Lemma('excla
im.v.01.outcry'), Lemma('exclaim.v.01.call_out'), Lemma('exclaim.v.01.shou
t'), Lemma('abuse.v.03.abuse'), Lemma('abuse.v.03.clapperclaw'), Lemma('abus
e.v.03.blackguard'), Lemma('abuse.v.03.shout')]
```

In [8]:

```
import nltk
texts=["""In morphology and lexicography, a lemma is the canonical form, dictionary form, o
for text in texts:
    sentences=nltk.sent_tokenize(text)
    for sentence in sentences:
        words=nltk.word_tokenize(sentence)
        print(words)
        tagged_words= nltk.pos_tag(words)
        print(tagged_words)
```

```
['In', 'morphology', 'and', 'lexicography', ',', 'a', 'lemma', 'is', 'the',
'canonical', 'form', ',', 'dictionary', 'form', ',', 'or', 'citation', 'for
m', 'of', 'a', 'set', 'of', 'words', '.']
[('In', 'IN'), ('morphology', 'NN'), ('and', 'CC'), ('lexicography', 'NN'),
(',', ','), ('a', 'DT'), ('lemma', 'NN'), ('is', 'VBZ'), ('the', 'DT'), ('ca
nonical', 'JJ'), ('form', 'NN'), (',', ','), ('dictionary', 'JJ'), ('form',
'NN'), (',', ','), ('or', 'CC'), ('citation', 'NN'), ('form', 'NN'), ('of',
'IN'), ('a', 'DT'), ('set', 'NN'), ('of', 'IN'), ('words', 'NNS'), ('.',
'.')]
['In', 'English', ',', 'for', 'example', ',', 'run', ',', 'runs', ',', 'ra
n', 'and', 'running', 'are', 'forms', 'of', 'the', 'same', 'lexeme', ',', 'w
ith', 'run', 'as', 'the', 'lemma']
[('In', 'IN'), ('English', 'NNP'), (',', ','), ('for', 'IN'), ('example', 'N
N'), (',', ','), ('run', 'VB'), (',', ','), ('runs', 'VB'), (',', ','), ('ra
n', 'NN'), ('and', 'CC'), ('running', 'VBG'), ('are', 'VBP'), ('forms', 'NN
S'), ('of', 'IN'), ('the', 'DT'), ('same', 'JJ'), ('lexeme', 'NN'), (',',
','), ('with', 'IN'), ('run', 'VBN'), ('as', 'IN'), ('the', 'DT'), ('lemma',
'NN')]
```

In [ ]:

```

```