In [2]:
```
import pandas as pd
import sklearn as sk
import math
import nltk
from nltk.corpus import stopwords

set(stopwords.words('english'))

def computeTF(wordDict, doc):
    """
        tf(t,d) = count of t in d / number of words in d

    :param wordDict:
    :param doc:
    :return:
    """
    tfDict = {}
    corpusCount = len(doc)
    for word, count in wordDict.items():
        tfDict[word] = count/float(corpusCount)
    return(tfDict)

def computeIDF(docList):
    """
        idf(t) = log(N/(df + 1))
    :param docList:
    :return:
    """
    idfDict = {}
    N = len(docList)

    idfDict = dict.fromkeys(docList[0].keys(), 0)
    for word, val in idfDict.items():
        idfDict[word] = math.log10(N / (float(val) + 1))

    return (idfDict)

def computeTFIDF(tfBow, idfs):
    """
    tf-idf(t, d) = tf(t, d) * log(N/(df + 1))
    :param tfBow:
    :param idfs:
    :return:
    """
    tfidf = {}
    for word, val in tfBow.items():
        tfidf[word] = val*idfs[word]
    return(tfidf)


def create_word_dict(total, sentence):
    wordDict = dict.fromkeys(total, 0)
    for word in sentence:
        wordDict[word] += 1
    return wordDict

sentence1 = "Go until jurong point, crazy.. Available only in bugi
s n great world la e buffet... Cine there got amore wat..."
```

```
sentence2 = "Free entry in 2 a wkly comp to win FA Cup final tkts
21st May 2005. Text FA to 87121 to receive entry question(std txt
rate)T&C's apply 08452810075over18's"

#split so each word have their own string
sentence1_list = nltk.word_tokenize(sentence1)
sentence2_list = nltk.word_tokenize(sentence2)
total= set(sentence1_list).union(set(sentence2_list))

wordDictA = create_word_dict(total,sentence1_list)
wordDictB = create_word_dict(total,sentence2_list)

tfFirst = computeTF(wordDictA, sentence1_list)
tfSecond = computeTF(wordDictB, sentence2_list)

idfs = computeIDF([wordDictA, wordDictB])

#running our two sentences through the IDF:
idfFirst = computeTFIDF(tfFirst, idfs)
idfSecond = computeTFIDF(tfSecond, idfs)

#putting it in a dataframe
idf = pd.DataFrame([idfFirst, idfSecond])
print(idf)
```

```
        ...        Go   receive      rate         a     final
apply  \
0  0.026177  0.013088  0.000000  0.000000  0.000000  0.000000  0.0
00000
1  0.000000  0.000000  0.008136  0.008136  0.008136  0.008136  0.0
08136

          &     great       wat  ...    jurong     there      wkly
C  \
0  0.000000  0.013088  0.013088  ...  0.013088  0.013088  0.000000
0.000000
1  0.008136  0.000000  0.000000  ...  0.000000  0.000000  0.008136
0.008136

       tkts      Free        's        la     amore     entry
0  0.000000  0.000000  0.000000  0.013088  0.013088  0.000000
1  0.008136  0.008136  0.016272  0.000000  0.000000  0.016272

[2 rows x 53 columns]
```

In [ ]: