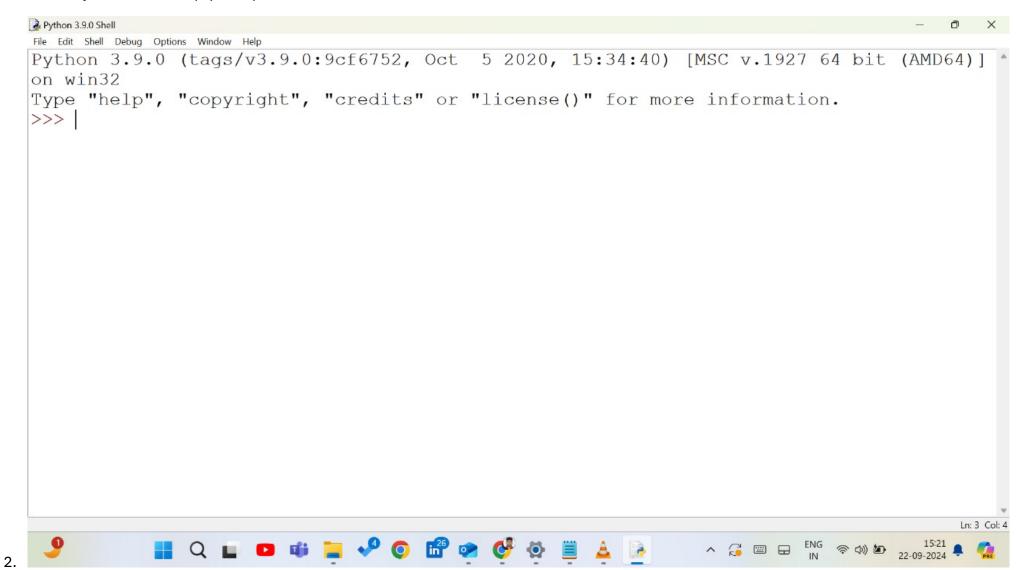
Assignment 1 Information Retrieval

Name: Vikas Reddy Mothe

Net ID: mothev2023

1. Install Python and NLTK (3 points)



```
PS C:\Users\vikas reddy mothe> pip install -U nltk
Requirement already satisfied: nltk in c:\users\vikas reddy mothe\appdata\local\programs\python\python38\lib\site-packages (3.8.1)
Collecting nltk
 Downloading nltk-3.9.1-py3-none-any.whl.metadata (2.9 kB)
Requirement already satisfied: click in c:\users\vikas reddy mothe\appdata\local\programs\python\python38\lib\site-packages (from nlt
k) (8.1.3)
Requirement already satisfied: joblib in c:\users\vikas reddy mothe\appdata\local\programs\python\python38\lib\site-packages (from nl
tk) (1.2.0)
Requirement already satisfied: regex>=2021.8.3 in c:\users\vikas reddy mothe\appdata\local\programs\python\python38\lib\site-packages
 (from nltk) (2022.10.31)
Requirement already satisfied: tqdm in c:\users\vikas reddy mothe\appdata\local\programs\python\python38\lib\site-packages (from nltk
 (4.65.0)
Requirement already satisfied: colorama in c:\users\vikas reddy mothe\appdata\local\programs\python\python38\lib\site-packages (from
click->nltk) (0.4.6)
Downloading nltk-3.9.1-pv3-none-anv.whl (1.5 MB)
                                          - 1.5/1.5 MB 11.4 MB/s eta 0:00:00
Installing collected packages: nltk
 Attempting uninstall: nltk
   Found existing installation: nltk 3.8.1
   Uninstalling nltk-3.8.1:
     Successfully uninstalled nltk-3.8.1
Successfully installed nltk-3.9.1
```

PS C:\Users\vikas reddv mothe>

```
#Importing required libraries and methods
import nltk
nltk.download('punkt')
nltk.download('stopwords')
from nltk.tokenize import sent_tokenize, word_tokenize
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity

Inltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Package stopwords is already up-to-date!
```

```
#Reading the contents of document 1, 100554newsML.txt with open('100554newsML.txt', 'r') as file:
file_contents = file.read()
```

document1 = file_contents

print(document1)

Channel tunnel operator Eurotunnel on Monday announced details of a deal giving bank creditors 45.5 percent of the company in return for wiping out 1.0 billion. The long-awaited but highly complex restructuring of nearly nearly nine billion pounds of debt and unpaid interest throws the company a lifeline which could see The deal, announced simultaneously in Paris and London, brings the company back from the brink of bankruptcy but leaves current shareholders, who have already s "We have fixed and capped the interest payments and arranged only to pay what is available in cash," Eurotunnel co-chairman Alastair Morton told reporters at a Morton said the plan provides the Anglo-French company with the medium term financial stability to consolidate its commercial position and develop its operation Although shareholders will see their holdings diluted, they were offered the prospect of a brighter future and urged to be patient after months of uncertainty we Eurotunnel, which has taken around half of the market in the busiest cross-Channel route from the European ferry companies, said a strong operating performance French co-chairman Patrick Ponsolle told reporters at a Paris news conference that the dividend could come as early as 2004 if the company performed "very well" Eurotunnel and the banks have come up with an ingenious formula to help the company get over the early years of the deal when, despite the swaps of debt for equ If its revenue, after costs and depreciation, is less than 400 million pounds, then the company will issue "Stabilisation notes" to a maximum of 1.85 billion pc Analysts said that under the deal, Eurotunnel's ability to finance its debt would become sustainable, at least for a few years.

"If you lack at the current cash flow of between 150 and 200 million pounds, what they can't find (to meet the bill) they will rely forward into the cash.

"If you look at the current cash flow of between 150 and 200 million pounds a year, what they can't find (to meet the bill) they will roll forward into the stat "So they are here for that time," he added.

The company said in a statement there was still considerable work to be done to finalise and agree the details of the plan before it can be submitted to sharehc Eurotunnel said the debt-for-equity swap would be at 130 pence, or 10.40 francs, per share — considerably below the level of 160 pence widely reported in the r The company said a further 3.7 billion pounds of debt would be converted into new financial instruments and existing shareholders would be able to participate i If they choose not to take up free warrants entitling them to subscribe to this, Eurotunnel said shareholders' interests may be reduced further to just over 39 Eurotunnel's shares, which were suspended last week at 113.5 pence ahead of Monday's announcement, will resume trading on Tuesday.

Shareholders and all 225 creditor banks have to agree the deal. "I'm hopeful but I'm not taking it (approval) for granted," Morton admitted, "Shareholders are packed what would happen if the banks reject the deal, Morton said, "Nobody wants a collapse, nobody wants a doomsday scenario." (\$1=.6393 Pound)

```
[36] #Reading the contents of document 2, 100593newsML.txt
with open('100593newsML.txt', 'r') as file:
    file_contents = file.read()

document2 = file_contents
print(document2)
```

Anglo-French Channel Tunnel operator Eurotunnel Monday announced a deal giving its creditor banks 45.5 percent of the company in return for wiping out one bil The long-awaited restructuring brings to an end months of wrangling between Eurotunnel and the 225 banks to which it owes nearly nine billion pounds (\$14.1 bing The deal, announced simultaneously in Paris and London, brings the company back from the brink of insolvency but leaves shareholders owning only 54.5 percent "The restructuring plan provides Eurotunnel with the medium-term financial stability to allow it to consolidate its substantial commercial achievements to dat The firm was now making a profit before interest, he added.

Although shareholders will see their interests diluted, they were offered the prospect of a brighter future after months of uncertainty while Eurotunnel wrest Eurotunnel, which has taken around half the cross-Channel market from the European ferry companies, said a strong operating performance could allow it to pay French co-chairman Patrick Ponsolle said shareholders would have to be patient before they could reap the benefits of the company's success. He called the debt restructuring plan "an acceptable compromise" for holders of Eurotunnel shares.

The company said there was still considerable work to be done to finalise and agree on the details of the plan before it can be submitted to shareholders and Monday's announcement followed two weeks of highly secretive negotiations between Eurotunnel and its six leading banks. This was extended to the 24 "instructi Eurotunnel said the debt-for-equity swap would be at 130 pence, or 10.40 francs, per share. That is considerably below the level of around 160 pence widely re The company said a further 3.7 billion pounds (\$5.8 billion) of debt would be converted into new financial instruments and existing shareholders would be able If they choose not to take up free warrants entitling them to subscribe to this, Eurotunnel said shareholders' interests may be reduced further to just over Eurotunnel's shares, which were suspended last week at 113.5 pence ahead of Monday's announcement, should resume trading on Tuesday, the company said.





```
#Reading the contents of document 3, 100618newsML.txt
with open('100618newsML.txt', 'r') as file:
    file_contents = file.read()

document3 = file_contents

print(document3)
```

Anglo-French Channel Tunnel operator Eurotunnel on Monday announced a deal giving creditor banks 45.5 percent of the company in return for wiping out one billic The long-awaited restructuring brings to an end months of wrangling between Eurotunnel and the 225 banks to which it owes nearly nine billion pounds. The deal, announced simultaneously in Paris and London, brings the company back from the brink of insolvency but leaves shareholders owning only 54.5 percent of "The restructuring plan provides Eurotunnel with the medium term financial stability to allow it to consolidate its substantial commercial achievements to date The firm was now making a profit before interest, he added.

Although shareholders will see their interests diluted, they were offered the prospect of a brighter future after months of uncertainty while Eurotunnel wrestle Eurotunnel, which has taken around half the cross-Channel market from the European ferry companies, said a strong operating performance could allow it to pay it French co-chairman Patrick Ponsolle said shareholders would have to be patient before they could reap the benefits of the company's success. He called the debt restructuring plan "an acceptable compromise" for holders of Eurotunnel shares.

The company said in a statement there was still considerable work to be done to finalise and agree the details of the plan before it can be submitted to sharehce Monday's announcement followed two weeks of highly secretive negotiations between Eurotunnel and its six leading banks. This was extended to the 24 "instructing Eurotunnel said the debt-for-equity swap would be at 130 pence, or 10.40 francs, per share. That is considerably below the level of around 160 pence widely report to company said a further 3.7 billion pounds of debt would be converted into new financial instruments and existing shareholders would be able to participate in the subscribe to this, Eurotunnel said shareholders' interests may be reduced further to just over 39 Eurotunnel's shares, which were suspended last week at 113.5 pence ahead of Monday's announcement, should resume trading on Tuesday, the company said. (\$1=.6393)

3. Tokenize the documents into words, remove stop words, and conduct stemming (5 points)

```
#tokenizing three documents
    tokenized_document1 = []
    sentences1 = sent tokenize(document1)
    for sentence in sentences1:
      tokenized_document1.extend(word_tokenize(sentence))
    print("Tokenized Document 1\n", tokenized_document1, end='\n\n')
    tokenized_document2 = []
    sentences2 = sent_tokenize(document2)
    for sentence in sentences2:
      tokenized document2.extend(word tokenize(sentence))
    print("Tokenized Document 2\n", tokenized_document2, end='\n\n')
    tokenized_document3 = []
    sentences3 = sent tokenize(document3)
    for sentence in sentences3:
      tokenized_document3.extend(word_tokenize(sentence))
    print("Tokenized Document 3\n", tokenized document3, end='\n\n')
→ Tokenized Document 1
     ['Channel', 'tunnel', 'operator', 'Eurotunnel', 'on', 'Monday', 'announced', 'details', 'of', 'a', 'deal', 'giving', 'bank', 'creditors', '45.5', 'percent', '
   Tokenized Document 2
     ['Anglo-French', 'Channel', 'Tunnel', 'operator', 'Eurotunnel', 'Monday', 'announced', 'a', 'deal', 'giving', 'its', 'creditor', 'banks', '45.5', 'percent', '
    Tokenized Document 3
     ['Anglo-French', 'Channel', 'Tunnel', 'operator', 'Eurotunnel', 'on', 'Monday', 'announced', 'a', 'deal', 'giving', 'creditor', 'banks', '45.5', 'percent', 'c
```

Tokenized Document 1

['Channel', 'tunnel', 'operator', 'Eurotunnel', 'on', 'Monday', 'announced', 'details', 'of', 'a', 'deal', 'giving', 'bank', 'creditors', '45.5', 'percent', 'of', 'the', 'company', 'in', 'return', 'for', 'wiping', 'out', '1.0', 'billion', 'pounds', '(', '\$', '1.6', 'billion', ')', 'of', 'its', 'massive', 'debts', '!, 'The', 'long-awaited', 'but', 'highly', 'complex', 'restructuring', 'of', 'nearly', 'nearly', 'nine', 'billion', 'pounds', 'of', 'debt', 'and', 'unpaid', 'interest', 'throws', 'the', 'company', 'a', 'lifeline', 'which', 'could', 'secure', 'what', 'is', 'still', 'likely', 'to', 'be', 'a', 'difficult', 'future', '.', 'The', 'deal', ', 'announced', 'simultaneously', 'in', 'Paris', 'and', 'London', ',', 'brings', 'the', 'company', 'back', 'from', 'the', 'brink', 'of', 'bankruptcy', 'but', 'leaves', 'current', 'shareholders', ',', 'who', 'have', 'already', 'seen', 'their', 'investment', 'dwindle', ',, 'owning', 'only', '54.5', 'percent', 'of', 'the', 'company', ', '``', 'We', 'have', 'fixed', 'and', 'capped', 'the', 'interest', 'payments', 'and', 'arranged', 'only', 'to', 'pay', 'what', 'is', 'available', 'in', 'cash', ';, """, 'Eurotunnel', 'co-chairman', 'Alastair', 'Morton', 'told', 'reporters', 'at', 'a', 'news', 'conference', '.', '```, 'Avoiding', 'having', 'to', 'do', 'this', 'again', 'is', 'the', 'name', 'of', 'the', 'game', '!, "''", 'Morton', 'said', 'the', 'plan', 'provides', 'the', 'Anglo-French', 'company', 'with', 'the', 'medium', 'term', 'financial', 'stability', 'to', 'consolidate', 'its', 'commercial', 'position', 'and', 'develop', 'its', 'operations', '', 'adding', 'that', 'the', 'firm', 'was', 'now', 'making', 'a', 'profit', 'before', 'interest', '.', 'Although', 'shareholders', 'will', 'see', 'their', 'holdings', 'diluted', ',', 'they', 'were', 'offered', 'the', 'prospect', 'of', 'a', 'brighter', 'future', 'and', 'urged', 'to', 'be', 'patient', 'after', 'months', 'of', 'uncertainty', 'while', 'Eurotunnel', 'wrestled', 'to', 'reduce', 'the', 'crippling', 'interest', 'payments', 'negotiated', 'during', 'the', 'tunnel', "'s", 'construction', '., 'Eurotunnel', '', 'which', 'has', 'taken', 'around', 'half', 'of', 'the', 'market', 'in', 'the', 'busiest', 'cross-Channel', 'route', 'from', 'the', 'European', 'ferry', 'companies', ',, 'said', 'a', 'strong', 'operating', 'performance', 'could', 'allow', 'it', 'to', 'pay', 'its', 'first', 'dividend', 'within', 'the', 'next', '10', 'years', '.', 'French', 'co-chairman', 'Patrick', 'Ponsolle', 'told', 'reporters', 'at', 'a', 'Paris', 'news', 'conference', 'that', 'the', 'dividend', 'could', 'come', 'as', 'early', 'as', '2004', 'if', 'the', 'company', 'performed', '``', 'very', 'well', "'", '., 'Eurotunnel', 'and', 'the', 'banks', 'have', 'come', 'up', 'with', 'an', 'ingenious', 'formula', 'to', 'help', 'the', 'company', 'get', 'over', 'the', 'early', 'years', 'of', 'the', 'deal', 'when', ',', 'despite', 'the', 'swaps', 'of', 'debt', 'for', 'equity', 'and', 'bonds', ',, 'it', 'will', 'still', 'not', 'be', 'able', 'to', 'afford', 'the', 'annual', 'interest', 'bill', 'of', '400', 'million', 'pounds', '., 'If', 'its', 'revenue', ',', 'after', 'costs', 'and', 'depreciation', ',, 'is', 'less', 'than', '400', 'million', 'pounds', ',', 'then', 'the', 'company', 'will', 'issue', '``', 'Stabilisation', 'notes', """, 'to', 'a', 'maximum', 'of', '1.85', 'billion', 'pounds', 'to', 'the', 'banks', '.; 'Eurotunnel', 'would', 'not', 'pay', 'interest', 'on', 'these', 'notes', '(', 'which', 'would', 'constitute', 'a', 'debt', 'issue', ')', 'for', 'ten', 'years', '., 'Analysts', 'said', 'that', 'under', 'the', 'deal', ',', 'Eurotunnel', "'s", 'ability', 'to', 'finance', 'its', 'debt', 'would', 'become', 'sustainable', ',', 'at', 'least', 'for', 'a', 'few', 'years', '.', '``', 'If', 'you', 'look', 'at', 'the', 'current', 'cash', 'flow', 'of', 'between', '150', 'and', '200', 'million', 'pounds', 'a', 'year', ',, 'what', 'they', 'ca', "n't", 'find', '(', 'to', 'meet', 'the', 'bill', ')', 'they', 'will', 'roll', 'forward', 'into', 'the', 'stabilisation', 'notes', ',, 'and', 'they', 'can', 'keep', 'that', 'going', 'for', 'seven', ',, 'eight', ',, 'nine', 'years', ',, "''", 'said', 'an', 'analyst', 'at', 'one', 'major', 'investment', 'bank', '., '``', 'So', 'they', 'are', 'here', 'for', 'that', 'time', ',, """, 'he', 'added', '., 'The', 'company', 'said', 'in', 'a', 'statement', 'there', 'was', 'still', 'considerable', 'work', 'to', 'be', 'done', 'to', 'finalise', 'and', 'agree', 'the', 'details', 'of', 'the', 'plan', 'before', 'it', 'can', 'be', 'submitted', 'to', 'shareholders', 'and', 'the', 'bank', 'group', 'for', 'approval', ',', 'probably', 'early', 'in', 'the', 'Spring', 'of', '1997', '., 'Eurotunnel', 'said', 'the', 'debt-for-equity', 'swap', 'would', 'be', 'at', '130', 'pence', ',, 'or', '10.40', 'francs', ',, 'per', 'share', '--', 'considerably', 'below', 'the', 'level', 'of', '160', 'pence', 'widely', 'reported', 'in', 'the', 'run', 'up', 'to', 'the', 'deal', 'The', 'company', 'said', 'a', 'further', '3.7', 'billion', 'pounds', 'of', 'debt', 'would', 'be', 'converted', 'into', 'new', 'financial', 'instruments', 'and', 'existing', 'shareholders', 'would', 'be', 'able', 'to', 'participate', 'in', 'this', 'issue', '.', 'If', 'they', 'choose', 'not', 'to', 'take', 'up', 'free', 'warrants', 'entitling', 'them', 'to', 'subscribe', 'to', 'this', ',', 'Eurotunnel', 'said', 'shareholders', "'", 'interests', 'may', 'be', 'reduced', 'further', 'to', 'just', 'over', '39', 'percent', 'of', 'the', 'company', 'by', 'the', 'end', 'of', 'December', '2003', '., 'Eurotunnel', "'s", 'shares', '., 'which',

'were', 'suspended', 'last', 'week', 'at', '113.5', 'pence', 'ahead', 'of', 'Monday', "'s", 'announcement', ',', 'will', 'resume', 'trading', 'on', 'Tuesday', '.', 'Shareholders', 'and', 'all', '225', 'creditor', 'banks', 'have', 'to', 'agree', 'the', 'deal', '.', '``', 'I', "'m", 'hopeful', 'but', 'I', "'m", 'not', 'taking', 'it', '(', 'approval', ')', 'for', 'granted', ',', "''", 'Morton', 'admitted', ',', '``', 'Shareholders', 'are', 'pretty', 'angry', 'in', 'France', '.', "''", 'Asked', 'what', 'would', 'happen', 'if', 'the', 'banks', 'reject', 'the', 'deal', ',', 'Morton', 'said', ',', '``', 'Nobody', 'wants', 'a', 'collapse', ',', 'nobody', 'wants', 'a', 'doomsday', 'scenario', '.', "'"", '(', '\$', '1=.6393', 'Pound', ')']

Tokenized Document 2

['Anglo-French', 'Channel', 'Tunnel', 'operator', 'Eurotunnel', 'Monday', 'announced', 'a', 'deal', 'giving', 'its', 'creditor', 'banks', '45.5', 'percent', 'of', 'the', 'company', 'in', 'return', 'for', 'wiping', 'out', 'one', 'billion', 'pounds', '(', '\$', '1.56', 'billion', ')', 'of', 'its', 'debt', '.', 'The', 'long-awaited', 'restructuring', 'brings', 'to', 'an', 'end', 'months', 'of', 'wrangling', 'between', 'Eurotunnel', 'and', 'the', '225', 'banks', 'to', 'which', 'it', 'owes', 'nearly', 'nine', 'billion', 'pounds', '(', '\$', '14.1', 'billion', ')', '.', 'The', 'deal', ',, 'announced', 'simultaneously', 'in', 'Paris', 'and', 'London', ',, 'brings', 'the', 'company', 'back', 'from', 'the', 'brink', 'of', 'insolvency', 'but', 'leaves', 'shareholders', 'owning', 'only', '54.5', 'percent', 'of', 'the', 'company', '', '``', 'The', 'restructuring', 'plan', 'provides', 'Eurotunnel', 'with', 'the', 'medium-term', 'financial', 'stability', 'to', 'allow', 'it', 'to', 'consolidate', 'its', 'substantial', 'commercial', 'achievements', 'to', 'date', 'and', 'to', 'develop', 'its', 'operations', ',, """, 'Eurotunnel', 'co-chairman', 'Alastair', 'Morton', 'said', '., 'The', 'firm', 'was', 'now', 'making', 'a', 'profit', 'before', 'interest', ',, 'he', 'added', '., 'Although', 'shareholders', 'will', 'see', 'their', 'interests', 'diluted', ',, 'they', 'were', 'offered', 'the', 'prospect', 'of', 'a', 'brighter', 'future', 'after', 'months', 'of', 'uncertainty', 'while', 'Eurotunnel', 'wrestled', 'to', 'reduce', 'crippling', 'interest', 'payments', 'negotiated', 'during', 'the', 'tunnel', "'s", 'construction', '.', 'Eurotunnel', ',', 'which', 'has', 'taken', 'around', 'half', 'the', 'cross-Channel', 'market', 'from', 'the', 'European', 'ferry', 'companies', ',, 'said', 'a', 'strong', 'operating', 'performance', 'could', 'allow', 'it', 'to', 'pay', 'its', 'first', 'dividend', 'within', 'the', 'next', '10', 'years', '., 'French', 'co-chairman', 'Patrick', 'Ponsolle', 'said', 'shareholders', 'would', 'have', 'to', 'be', 'patient', 'before', 'they', 'could', 'reap', 'the', 'benefits', 'of', 'the', 'company', "'s", 'success', '.', 'He', 'called', 'the', 'debt', 'restructuring', 'plan', '``', 'an', 'acceptable', 'compromise', "'", 'for', 'holders', 'of', 'Eurotunnel', 'shares', '.', 'The', 'company', 'said', 'there', 'was', 'still', 'considerable', 'work', 'to', 'be', 'done', 'to', 'finalise', 'and', 'agree', 'on', 'the', 'details', 'of', 'the', 'plan', 'before', 'it', 'can', 'be', 'submitted', 'to', 'shareholders', 'and', 'the', 'full', '225', 'bank', 'syndicate', 'for', 'approval', ',', 'probably', 'early', 'in', '1997', '.', 'Monday', "'s", 'announcement', 'followed', 'two', 'weeks', 'of', 'highly', 'secretive', 'negotiations', 'between', 'Eurotunnel', 'and', 'its', 'six', 'leading', 'banks', '.', 'This', 'was', 'extended', 'to', 'the', '24', '``', 'instructing', 'banks', """, 'at', 'a', 'meeting', 'late', 'last', 'week', 'in', 'London', '.', 'Eurotunnel', 'said', 'the', 'debt-for-equity', 'swap', 'would', 'be', 'at', '130', 'pence', ',', 'or', '10.40', 'francs', ',', 'per', 'share', '.', 'That', 'is', 'considerably', 'below', 'the', 'level', 'of', 'around', '160', 'pence', 'widely', 'reported', 'before', 'announcement', 'of', 'the', 'deal', ',', 'and', 'will', 'reduce', 'outstanding', 'debt', 'of', '8.7', 'billion', 'pounds', '(', '\$', '13.6', 'billion', ')', 'by', '1.0', 'billion', '(', '\$', '1.56', 'billion', ')', '.', 'The', 'company', 'said', 'a', 'further', '3.7', 'billion', 'pounds', '(', '\$', '5.8', 'billion', ')', 'of', 'debt', 'would', 'be', 'converted', 'into', 'new', 'financial', 'instruments', 'and', 'existing', 'shareholders', 'would', 'be', 'able', 'to', 'participate', 'in', 'this', 'issue', '!, 'If', 'they', 'choose', 'not', 'to', 'take', 'up', 'free', 'warrants', 'entitling', 'them', 'to', 'subscribe', 'to', 'this', ', 'Eurotunnel', 'said', 'shareholders', "'", 'interests', 'may', 'be', 'reduced', 'further', 'to', 'just', 'over', '39', 'percent', 'of', 'the', 'company', 'by', 'the', 'end', 'of', 'December', '2003', '.', 'Eurotunnel', "'s", 'shares', '.', 'which', 'were', 'suspended', 'last', 'week', 'at', '113.5', 'pence', 'ahead', 'of', 'Monday', "'s", 'announcement', ',', 'should', 'resume', 'trading', 'on', 'Tuesday', ',', 'the', 'company', 'said', '.']

Tokenized Document 3

['Anglo-French', 'Channel', 'Tunnel', 'operator', 'Eurotunnel', 'on', 'Monday', 'announced', 'a', 'deal', 'giving', 'creditor', 'banks', '45.5', 'percent', 'of', 'the', 'company', 'in', 'return', 'for', 'wiping', 'out', 'one', 'billion', 'pounds', '(', '\$', '1.56', 'billion', ')', 'of', 'its', 'debt', 'mountain', '.', 'The', 'long-awaited', 'restructuring', 'brings', 'to', 'an', 'end', 'months', 'of', 'wrangling', 'between', 'Eurotunnel', 'and', 'the', '225', 'banks', 'to', 'which', 'it', 'owes', 'nearly', 'nine', 'billion', 'pounds', '.', 'The', 'deal', ',, 'announced', 'simultaneously', 'in', 'Paris', 'and', 'London', ',, 'brings', 'the', 'company', 'back', 'from', 'the', 'brink', 'of', 'insolvency', 'but', 'leaves', 'shareholders', 'owning', 'only', '54.5', 'percent', 'of', 'the', 'company', '.', '``', 'The', 'restructuring', 'plan', 'provides', 'Eurotunnel', 'with', 'the', 'medium', 'term', 'financial', 'stability', 'to', 'allow', 'it', 'to', 'consolidate', 'its', 'substantial', 'commercial', 'achievements', 'to', 'date', 'and', 'to', 'develop', 'its', 'operations'. '.. """. 'Eurotunnel', 'co-chairman', 'Alastair', 'Morton', 'said', '.', 'The', 'firm', 'was', 'now', 'making', 'a', 'profit', 'before', 'interest', ',, 'he', 'added', '.', 'Although', 'shareholders', 'will', 'see', 'their', 'interests', 'diluted', ',', 'they', 'were', 'offered', 'the', 'prospect', 'of', 'a', 'brighter', 'future', 'after', 'months', 'of', 'uncertainty', 'while', 'Eurotunnel', 'wrestled', 'to', 'reduce', 'crippling', 'interest', 'payments', 'negotiated', 'during', 'the', 'tunnel', "'s", 'construction', '., 'Eurotunnel', '., 'which', 'has', 'taken', 'around', 'half', 'the', 'cross-Channel', 'market', 'from', 'the', 'European', 'ferry', 'companies', ',, 'said', 'a', 'strong', 'operating', 'performance', 'could', 'allow', 'it', 'to', 'pay', 'its', 'first', 'dividend', 'within', 'the', 'next', '10', 'years', '., 'French', 'co-chairman', 'Patrick', 'Ponsolle', 'said', 'shareholders', 'would', 'have', 'to', 'be', 'patient', 'before', 'they', 'could', 'reap', 'the', 'benefits', 'of', 'the', 'company', "'s", 'success', '.', 'He', 'called', 'the', 'debt', 'restructuring', 'plan', '``', 'an', 'acceptable', 'compromise', "''", 'for', 'holders', 'of', 'Eurotunnel', 'shares', '.', 'The', 'company', 'said', 'in', 'a', 'statement', 'there', 'was', 'still', 'considerable', 'work', 'to', 'be', 'done', 'to', 'finalise', 'and', 'agree', 'the', 'details', 'of', 'the', 'plan', 'before', 'it', 'can', 'be', 'submitted', 'to', 'shareholders', 'and', 'agree', 'the', 'details', 'of', 'the', 'plan', 'before', 'it', 'can', 'be', 'submitted', 'to', 'shareholders', 'and', 'agree', 'the', 'details', 'of', 'the', 'plan', 'before', 'it', 'can', 'be', 'submitted', 'to', 'shareholders', 'and', 'agree', 'the', 'details', 'of', 'the', 'plan', 'before', 'it', 'can', 'be', 'submitted', 'to', 'shareholders', 'and', 'agree', 'the', 'details', 'of', 'the', 'plan', 'before', 'it', 'can', 'be', 'submitted', 'to', 'shareholders', 'and', 'agree', 'the', 'details', 'of', 'the', 'plan', 'before', 'the', 'shareholders', 'and', 'agree', 'the', 'glan', 'before', 'to', 'shareholders', 'and', 'agree', 'the', 'glan', 'before', 'the', 'before', 'the', 'glan', 'before', 'the', 'glan', 'before', 'the', 'glan', 'gl 'the', 'full', '225', 'bank', 'syndicate', 'for', 'approval', ',', 'probably', 'early', 'in', '1997', '.', 'Monday', "'s", 'announcement', 'followed', 'two', 'weeks', 'of', 'highly', 'secretive', 'negotiations', 'between', 'Eurotunnel', 'and', 'its', 'six', 'leading', 'banks', '', 'This', 'was', 'extended', 'to', 'the', '24', '``', 'instructing', 'banks', "'"", 'at', 'a', 'meeting', 'late', 'last', 'week', 'in', 'London', '.', 'Eurotunnel', 'said', 'the', 'debt-for-equity', 'swap', 'would', 'be', 'at', '130', 'pence', ',, 'or', '10.40', 'francs', ',, 'per', 'share', '., 'That', 'is', 'considerably', 'below', 'the', 'level', 'of', 'around', '160', 'pence', 'widely', 'reported', 'in', 'the', 'run', 'up', 'to', 'the', 'deal', '., 'and', 'will', 'reduce', 'outstanding', 'debt', 'of', '8.7', 'billion', 'pounds', 'by', '1.0', 'billion', ', 'The', 'company', 'said', 'a', 'further', '3.7', 'billion', 'pounds', 'of', 'debt', 'would', 'be', 'converted', 'into', 'new', 'financial', 'instruments', 'and', 'existing', 'shareholders', 'would', 'be', 'able', 'to', 'participate', 'in', 'this', 'issue', '.', 'If', 'they', 'choose', 'not', 'to', 'take', 'up', 'free', 'warrants', 'entitling', 'them', 'to', 'subscribe', 'to', 'this', ',', 'Eurotunnel', 'said', 'shareholders', "'", 'interests', 'may', 'be', 'reduced', 'further', 'to', 'just', 'over', '39', 'percent', 'of', 'the', 'company', 'by', 'the', 'end', 'of', 'December', '2003', '.', 'Eurotunnel', "'s", 'shares', ',', 'which', 'were', 'suspended', 'last', 'week', 'at', '113.5', 'pence', 'ahead', 'of', 'Monday', "'s", 'announcement', ',', 'should', 'resume', 'trading', 'on', 'Tuesday', ',', 'the', 'company', 'said', '.', '(', '\$', '1=.6393', 'Pound', ')']

```
#removing stopwords from all the three documents
    stop_words = set(stopwords.words('english'))
    filered_document1 = []
    for word in tokenized_document1:
      if word not in stop words:
        filered_document1.append(word)
    print("Filtered Document 1\n", filered_document1, end='\n\n')
    filered_document2 = []
    for word in tokenized_document2:
     if word not in stop_words:
        filered_document2.append(word)
    print("Filtered Document 2\n", filered_document2, end='\n\n')
    filered_document3 = []
    for word in tokenized document3:
     if word not in stop_words:
        filered_document3.append(word)
    print("Filtered Document 3\n", filered_document3, end='\n\n')
→ Filtered Document 1
```

['Channel', 'tunnel', 'operator', 'Eurotunnel', 'Monday', 'announced', 'details', 'deal', 'giving', 'bank', 'creditors', '45.5', 'percent', 'company', 'retur

['Anglo-French', 'Channel', 'Tunnel', 'operator', 'Eurotunnel', 'Monday', 'announced', 'deal', 'giving', 'creditor', 'banks', '45.5', 'percent', 'company', '

['Anglo-French', 'Channel', 'Tunnel', 'operator', 'Eurotunnel', 'Monday', 'announced', 'deal', 'giving', 'creditor', 'banks', '45.5', 'percent', 'company', '

Filtered Document 2

Filtered Document 3

Filtered Document 1

['Channel', 'tunnel', 'operator', 'Eurotunnel', 'Monday', 'announced', 'details', 'deal', 'giving', 'bank', 'creditors', '45.5', 'percent', 'company', 'return', 'wiping', '1.0', 'billion', 'pounds', '(', '\$', '1.6', 'billion', ')', 'massive', 'debts', '', 'The', 'long-awaited', 'highly', 'complex', 'restructuring', 'nearly', 'nearly', 'nine', 'billion', 'pounds', 'debt', 'unpaid', 'interest', 'throws', 'company', 'lifeline', 'could', 'secure', 'still', 'likely', 'difficult', 'future', '., 'The', 'deal', ', 'announced', 'simultaneously', 'Paris', 'London', ',', 'brings', 'company', 'back', 'brink', 'bankruptcy', 'leaves', 'current', 'shareholders', ',, 'already', 'seen', 'investment', 'dwindle', ',', 'owning', '54.5', 'percent', 'company', '.', '``', 'We', 'fixed', 'capped', 'interest', 'payments', 'arranged', 'pay', 'available', 'cash', ',', "''", 'Eurotunnel', 'co-chairman', 'Alastair', 'Morton', 'told', 'reporters', 'news', 'conference', '!, '``', 'Avoiding', 'name', 'game', '!, "''", 'Morton', 'said', 'plan', 'provides', 'Anglo-French', 'company', 'medium', 'term', 'financial', 'stability', 'consolidate', 'commercial', 'position', 'develop', 'operations', ',, 'adding', 'firm', 'making', 'profit', 'interest', ', 'Although', 'shareholders', 'see', 'holdings', 'diluted', ', 'offered', 'prospect', 'brighter', 'future', 'urged', 'patient', 'months', 'uncertainty', 'Eurotunnel', 'wrestled', 'reduce', 'crippling', 'interest', 'payments', 'negotiated', 'tunnel', "'s", 'construction', ". 'Eurotunnel', ", 'taken', 'around', 'half', 'market', 'busiest', 'cross-Channel', 'route', 'European', 'ferry', 'companies', 'j, 'said', 'strong', 'operating', 'performance', 'could', 'allow', 'pay', 'first', 'dividend', 'within', 'next', '10', 'years', '., 'French', 'co-chairman', 'Patrick', 'Ponsolle', 'told', 'reporters', 'Paris', 'news', 'conference', 'dividend', 'could', 'come', 'early', '2004', 'company', 'performed', '``', 'well', "''", '!, 'Eurotunnel', 'banks', 'come', 'ingenious', 'formula', 'help', 'company', 'get', 'early', 'years', 'deal', ',', 'despite', 'swaps', 'debt', 'equity', 'bonds', ',', 'still', 'able', 'afford', 'annual', 'interest', 'bill', '400', 'million', 'pounds', '.', 'If', 'revenue', ',', 'costs', 'depreciation', ',', 'less', '400', 'million', 'pounds', ',, 'company', 'issue', '``', 'Stabilisation', 'notes', "'", 'maximum', '1.85', 'billion', 'pounds', 'banks', ',, 'Eurotunnel', 'would', 'pay', 'interest', 'notes', '(', 'would', 'constitute', 'debt', 'issue', ')', 'ten', 'years', '.', 'Analysts', 'said', 'deal', ',', 'Eurotunnel', "'s", 'ability', 'finance', 'debt', 'would', 'become', 'sustainable', ',' 'least', 'years', '., '``', 'If', 'look', 'current', 'cash', 'flow', '150', '200', 'million', 'pounds', 'year', ', 'ca', "n't", 'find', '(', 'meet', 'bill', ')', 'roll', 'forward', 'stabilisation', 'notes', ',', 'keep', 'going', 'seven', ',', 'eight', ',', 'nine', 'years', ',', """, 'said', 'analyst', 'one', 'major', 'investment', 'bank', '.', '``', 'So', 'time', ',', """, 'added', '.', 'The', 'company', 'said', 'statement', 'still', 'considerable', 'work', 'done', 'finalise', 'agree', 'details', 'plan', 'submitted', 'shareholders', 'bank', 'group', 'approval', ',' 'probably', 'early', 'Spring', '1997', '.', 'Eurotunnel', 'said', 'debt-for-equity', 'swap', 'would', '130', 'pence', ',, '10.40', 'francs', ',, 'per', 'share', '--', 'considerably', 'level', '160', 'pence', 'widely', 'reported', 'run', 'deal', 'The', 'company', 'said', '3.7', 'billion', 'pounds', 'debt', 'would', 'converted', 'new', 'financial', 'instruments', 'existing', 'shareholders', 'would', 'able', 'participate', 'issue', '', 'If', 'choose', 'take', 'free', 'warrants', 'entitling', 'subscribe', '', 'Eurotunnel', 'said', 'shareholders', """, 'interests', 'may', 'reduced', '39', 'percent', 'company', 'end', 'December', '2003', ", 'Eurotunnel', "'s", 'shares', ", 'suspended', 'last', 'week', '113.5', 'pence', 'ahead', 'Monday', "'s", 'announcement', ', 'resume', 'trading', 'Tuesday', ', 'Shareholders', '225', 'creditor', 'banks', 'agree', 'deal', ', '``', 'I', "'m", 'hopeful', 'I', "'m", 'taking', '(', 'approval', ')', 'granted', ',', "''", 'Morton', 'admitted', ',', '``', 'Shareholders', 'pretty', 'angry', 'France', '., "''", 'Asked', 'would', 'happen', 'banks', 'reject', 'deal', ',', 'Morton', 'said', ',', '``', 'Nobody', 'wants', 'collapse', ',', 'nobody', 'wants', 'doomsday', 'scenario', '.', """, '(', '\$', '1=.6393', 'Pound', ')']

Filtered Document 2

['Anglo-French', 'Channel', 'Tunnel', 'operator', 'Eurotunnel', 'Monday', 'announced', 'deal', 'giving', 'creditor', 'banks', '45.5', 'percent', 'company', 'return', 'wiping', 'one', 'billion', 'pounds', '(', '\$', '1.56', 'billion', ')', 'debt', '.', 'The', 'long-awaited', 'restructuring', 'brings', 'end', 'months', 'wrangling', 'Eurotunnel', '225', 'banks', 'owes', 'nearly', 'nine', 'billion', 'pounds', '(', '\$', '14.1', 'billion', ')', '., 'The', 'deal', ', 'announced', 'simultaneously', 'Paris', 'London', ', 'brings', 'company', 'back', 'brink', 'insolvency', 'leaves', 'shareholders', 'owning', '54.5', 'percent', 'company', '., '``', 'The', 'restructuring', 'plan', 'provides', 'Eurotunnel', 'mediumterm', 'financial', 'stability', 'allow', 'consolidate', 'substantial', 'commercial', 'achievements', 'date', 'develop', 'operations', ',, """, 'Eurotunnel', 'co-chairman', 'Alastair', 'Morton', 'said', '., 'The', 'firm', 'making', 'profit', 'interest', ',, 'added', '., 'Although', 'shareholders', 'see', 'interests', 'diluted', ',, 'offered', 'prospect', 'brighter', 'future', 'months', 'uncertainty', 'Eurotunnel', 'wrestled', 'reduce', 'crippling', 'interest', 'payments', 'negotiated', 'tunnel', "'s", 'construction', '.', 'Eurotunnel', ',', 'taken', 'around', 'half', 'cross-Channel', 'market', 'European', 'ferry', 'companies', ',, 'said', 'strong', 'operating', 'performance', 'could', 'allow', 'pay', 'first', 'dividend', 'within', 'next', '10', 'years', '.', 'French', 'co-chairman', 'Patrick', 'Ponsolle', 'said', 'shareholders', 'would', 'patient', 'could', 'reap', 'benefits', 'company', "'s", 'success', '., 'He', 'called', 'debt', 'restructuring', 'plan', '``', 'acceptable', 'compromise', "''", 'holders', 'Eurotunnel', 'shares', '.', 'The', 'company', 'said', 'still', 'considerable', 'work', 'done', 'finalise', 'agree', 'details', 'plan', 'submitted', 'shareholders', 'full', '225', 'bank', 'syndicate', 'approval', ',, 'probably', 'early', '1997', '.', 'Monday', "'s", 'announcement', 'followed', 'two', 'weeks', 'highly', 'secretive', 'negotiations', 'Eurotunnel', 'six', 'leading', 'banks', '.', 'This', 'extended', '24', '``', 'instructing', 'banks', "'", 'meeting', 'late', 'last', 'week', 'London', '.', 'Eurotunnel', 'said', 'debt-for-equity', 'swap', 'would', '130', 'pence', ',', '10.40', 'francs', ',', 'per', 'share', '.', 'That', 'considerably', 'level', 'around', '160', 'pence', 'widely', 'reported', 'announcement', 'deal', ',', 'reduce', 'outstanding', 'debt', '8.7', 'billion', 'pounds', '(', '\$', '13.6', 'billion', ')', '1.0', 'billion', '(', '\$', '1.56', 'billion', ')', '.', 'The', 'company', 'said', '3.7', 'billion', 'pounds', '(', '\$', '1.56', 'billion', ')', '.', 'The', 'company', 'said', '3.7', 'billion', 'pounds', '(', '\$', '1.56', 'billion', ')', '.', 'The', 'company', 'said', '3.7', 'billion', 'pounds', '(', '\$', '13.6', 'billion', ')', '1.0', 'billion', 'l', 'billion', 'pounds', '(', '\$', '13.6', 'billion', 'billion', 'pounds', '(', '\$', '13.6', 'billion', '\$', '5.8', 'billion', ')', 'debt', 'would', 'converted', 'new', 'financial', 'instruments', 'existing', 'shareholders', 'would', 'able', 'participate', 'issue', '.', 'If', 'choose', 'take', 'free', 'warrants', 'entitling', 'subscribe', ',, 'Eurotunnel', 'said', 'shareholders', "'", 'interests', 'may', 'reduced', '39', 'percent', 'company', 'end', 'December', '2003', '., 'Eurotunnel', "'s", 'shares', ',, 'suspended', 'last', 'week', '113.5', 'pence', 'ahead', 'Monday', "'s", 'announcement', ',, 'resume', 'trading', 'Tuesday', ',' 'company', 'said', '!]

Filtered Document 3

['Anglo-French', 'Channel', 'Tunnel', 'operator', 'Eurotunnel', 'Monday', 'announced', 'deal', 'giving', 'creditor', 'banks', '45.5', 'percent', 'company', 'return', 'wiping', 'one', 'billion', 'pounds', '(', '\$', '1.56', 'billion', ')', 'debt', 'mountain', '.', 'The', 'long-awaited', 'restructuring', 'brings', 'end', 'months', 'wrangling', 'Eurotunnel', '225', 'banks', 'owes', 'nearly', 'nine', 'billion', 'pounds', '', 'The', 'deal', '', 'announced', 'simultaneously', 'Paris', 'London', '', 'brings', 'company', 'back', 'brink', 'insolvency', 'leaves', 'shareholders', 'owning', '54.5', 'percent', 'company', '., '``', 'The', 'restructuring', 'plan', 'provides', 'Eurotunnel', 'medium', 'term', 'financial', 'stability', 'allow', 'consolidate', 'substantial', 'commercial', 'achievements', 'date', 'develop', 'operations', ',, """, 'Eurotunnel', 'co-chairman', 'Alastair', 'Morton', 'said', '., 'The', 'firm', 'making', 'profit', 'interest', ',, 'added', '., 'Although', 'shareholders', 'see', 'interests', 'diluted', ',, 'offered', 'prospect', 'brighter', 'future', 'months', 'uncertainty', 'Eurotunnel', 'wrestled', 'reduce', 'crippling', 'interest', 'payments', 'negotiated', 'tunnel', "'s", 'construction', '.', 'Eurotunnel', ',', 'taken', 'around', 'half', 'cross-Channel', 'market', 'European', 'ferry', 'companies', ',, 'said', 'strong', 'operating', 'performance', 'could', 'allow', 'pay', 'first', 'dividend', 'within', 'next', '10', 'years', '.', 'French', 'co-chairman', 'Patrick', 'Ponsolle', 'said', 'shareholders', 'would', 'patient', 'could', 'reap', 'benefits', 'company', "'s", 'success', '., 'He', 'called', 'debt', 'restructuring', 'plan', '``', 'acceptable', 'compromise', "''", 'holders', 'Eurotunnel', 'shares', '.', 'The', 'company', 'said', 'statement', 'still', 'considerable', 'work', 'done', 'finalise', 'agree', 'details', 'plan', 'submitted', 'shareholders', 'full', '225', 'bank', 'syndicate', 'approval', ',', 'probably', 'early', '1997', '.', 'Monday', "'s", 'announcement', 'followed', 'two', 'weeks', 'highly', 'secretive', 'negotiations', 'Eurotunnel', 'six', 'leading', 'banks', '.', 'This', 'extended', '24', '``', 'instructing', 'banks', "'", 'meeting', 'late', 'last', 'week', 'London', '.', 'Eurotunnel', 'said', 'debt-for-equity', 'swap', 'would', '130', 'pence', '., '10.40', 'francs', ', 'per', 'share', '., 'That', 'considerably', 'level', 'around', '160', 'pence', 'widely', 'reported', 'run', 'deal', '., 'reduce', 'outstanding', 'debt', '8.7', 'billion', 'pounds', '1.0', 'billion', '.', 'The', 'company', 'said', '3.7', 'billion', 'pounds', 'debt', 'would', 'converted', 'new', 'financial', 'instruments', 'existing', 'shareholders', 'would', 'able', 'participate', 'issue', '.', 'If', 'choose', 'take', 'free', 'warrants', 'entitling', 'subscribe', '.', 'Eurotunnel', 'said', 'shareholders', "'", 'interests', 'may', 'reduced', '39', 'percent', 'company', 'end', 'December', '2003', '.', 'Eurotunnel', "'s", 'shares', ',, 'suspended', 'last', 'week', '113.5', 'pence', 'ahead', 'Monday', "'s", 'announcement', ';, 'resume', 'trading', 'Tuesday', ';, 'company', 'said', ':, '(', '\$', '1=.6393', 'Pound', ')']

```
#Removing stemwords from all three documents
    stemmer = PorterStemmer()
    stemmed_document1 = []
    for word in filered document1:
      stemmed_document1.append(stemmer.stem(word))
    print("Stemmed Document 1\n", stemmed_document1, end='\n\n')
    stemmed document2 = []
    for word in filered_document2:
     stemmed_document2.append(stemmer.stem(word))
    print("Stemmed Document 2\n", stemmed_document2, end='\n\n')
    stemmed_document3 = []
    for word in filered document3:
      stemmed_document3.append(stemmer.stem(word))
    print("Stemmed Document 3\n", stemmed_document3, end='\n\n')
→ Stemmed Document 1
    ['channel', 'tunnel', 'oper', 'eurotunnel', 'monday', 'announc', 'detail', 'deal', 'give', 'bank', 'creditor', '45.5', 'percent', 'compani', 'return', 'wipe'
```

['anglo-french', 'channel', 'tunnel', 'oper', 'eurotunnel', 'monday', 'announc', 'deal', 'give', 'creditor', 'bank', '45.5', 'percent', 'compani', 'return',

['anglo-french', 'channel', 'tunnel', 'oper', 'eurotunnel', 'monday', 'announc', 'deal', 'give', 'creditor', 'bank', '45.5', 'percent', 'compani', 'return',

Stemmed Document 2

Stemmed Document 3

Stemmed Document 1

I'channel', 'tunnel', 'oper', 'eurotunnel', 'monday', 'announc', 'detail', 'deal', 'give', 'bank', 'creditor', '45.5', 'percent', 'compani', 'return', 'wipe', '1.0', 'billion', 'pound', '(', '\$', '1.6', 'billion', ')', 'massiv', 'debt', '.', 'the', 'long-await', 'highli', 'complex', 'restructur', 'nearli', 'nearli', 'nine', 'billion', 'pound', 'debt', 'unpaid', 'interest', 'throw', 'compani', 'lifelin', 'could', 'secur', 'still', 'like', 'difficult', 'futur', '.', 'the', 'deal', '.', 'announc', 'simultan', 'pari', 'london', '.', 'bring', 'compani', 'back', 'brink', 'bankruptci', 'leav', 'current', 'sharehold', ',, 'alreadi', 'seen', 'invest', 'dwindl', ',, 'own', '54.5', 'percent', 'compani', ', '``', 'we', 'fix', 'cap', 'interest', 'payment', 'arrang', 'pay', 'avail', 'cash', ',, "''", 'eurotunnel', 'co-chairman', 'alastair', 'morton', 'told', 'report', 'news', 'confer', '.', '``', 'avoid', 'name', 'game', '.', "''", 'morton', 'said', 'plan', 'provid', 'anglo-french', 'compani', 'medium', 'term', 'financi', 'stabil', 'consolid', 'commerci', 'posit', 'develop', 'oper', ', 'ad', 'firm', 'make', 'profit', 'interest', '., 'although', 'sharehold', 'see', 'hold', 'dilut', ',, 'offer', 'prospect', 'brighter', 'futur', 'urg', 'patient', 'month', 'uncertainti', 'eurotunnel', 'wrestl', 'reduc', 'crippl', 'interest', 'payment', 'negoti', 'tunnel', "'s", 'construct', '., 'eurotunnel', ',' taken', 'around', 'half', 'market', 'busiest', 'cross-channel', 'rout', 'european', 'ferri', 'compani', '., 'said', 'strong', 'oper', 'perform', 'could', 'allow', 'pay', 'first', 'dividend', 'within', 'next', '10', 'year', '., 'french', 'co-chairman', 'patrick', 'ponsol', 'told', 'report', 'pari', 'news', 'confer', 'dividend', 'could', 'come', 'earli', '2004', 'compani', 'perform', '``', 'well', "'"", '', 'eurotunnel', 'bank', 'come', 'ingeni', 'formula', 'help', 'compani', 'get', 'earli', 'year', 'deal', ', 'despit', 'swap', 'debt', 'equiti', 'bond', ', 'still', 'abl', 'afford', 'annual', 'interest', 'bill', '400', 'million', 'pound', '.', 'if', 'revenu', ',', 'cost', 'depreci', ',', 'less', '400', 'million', 'pound', ',', 'compani', 'issu', '``', 'stabilis', 'note', "''", 'maximum', '1.85', 'billion', 'pound', 'bank', '., 'eurotunnel', 'would', 'pay', 'interest', 'note', '(', 'would', 'constitut', 'debt', 'issu', ')', 'ten', 'year', '., 'analyst', 'said', 'deal', ',, 'eurotunnel', "'s", 'abil', 'financ', 'debt', 'would', 'becom', 'sustain', ',, 'least', 'year', '', '``', 'if', 'look', 'current', 'cash', 'flow', '150', '200', 'million', 'pound', 'year', '', 'ca', "n't", 'find', '(', 'meet', 'bill', ')', 'roll', 'forward', 'stabilis', 'note', ',', 'keep', 'go', 'seven', ',', 'eight', ',', 'nine', 'year', ',', """, 'said', 'analyst', 'one', 'major', 'invest', 'bank', '., '``', 'so', 'time', ',', """, 'ad', '.', 'the', 'compani', 'said', 'statement', 'still', 'consider', 'work', 'done', 'finalis', 'agre', 'detail', 'plan', 'submit', 'sharehold', 'bank', 'group', 'approv', ',', 'probabl', 'earli', 'spring', '1997', '.', 'eurotunnel', 'said', 'debt-for-equ', 'swap', 'would', '130', 'penc', ',', '10.40', 'franc', ',', 'per', 'share', '--', 'consider', 'level', '160', 'penc', 'wide', 'report', 'run', 'deal', 'the', 'compani', 'said', '3.7', 'billion', 'pound', 'debt', 'would', 'convert', 'new', 'financi', 'instrument', 'exist', 'sharehold', 'would', 'abl', 'particip', 'issu', '.', 'if', 'choos', 'take', 'free', 'warrant', 'entitl', 'subscrib', ',' 'eurotunnel', 'said', 'sharehold', "'", 'interest', 'may', 'reduc', '39', 'percent', 'compani', 'end', 'decemb', '2003', '.', 'eurotunnel', "'s", 'share', ',, 'suspend', 'last', 'week', '113.5', 'penc', 'ahead', 'monday', "'s", 'announc', ',, 'resum', 'trade', 'tuesday', '.', 'sharehold', '225', 'creditor', 'bank', 'agre', 'deal', '.', '``', 'i', "'m", 'hope', 'i', "'m", 'take', '(', 'approv', ')', 'grant', ',', "''", 'morton', 'admit', ',' '``', 'sharehold', 'pretti', 'angri', 'franc', '.', "''", 'ask', 'would', 'happen', 'bank', 'reject', 'deal', ',, 'morton', 'said', ',, '``', 'nobodi', 'want', 'collaps', ',, 'nobodi', 'want', 'doomsday', 'scenario', '.', "'", '(', '\$', '1=.6393', 'pound', ')']

Stemmed Document 2

['anglo-french', 'channel', 'tunnel', 'oper', 'eurotunnel', 'monday', 'announc', 'deal', 'give', 'creditor', 'bank', '45.5', 'percent', 'compani', 'return', 'wipe', 'one', 'billion', 'pound', '(', '\$', '1.56', 'billion', ')', 'debt', '.', 'the', 'long-await', 'restructur', 'bring', 'end', 'month', 'wrangl', 'eurotunnel', '225', 'bank', 'owe', 'nearli', 'nine', 'billion', 'pound', '(', '\$', '14.1', 'billion', ')', '.', 'the', 'deal', ', 'announc', 'simultan', 'pari', 'london', ', 'bring', 'compani', 'back', 'brink', 'insolv', 'leav', 'sharehold',

'own', '54.5', 'percent', 'compani', '', '``', 'the', 'restructur', 'plan', 'provid', 'eurotunnel', 'medium-term', 'financi', 'stabil', 'allow', 'consolid', 'substanti', 'commerci', 'achiev', 'date', 'develop', 'oper', '", """, 'eurotunnel', 'co-chairman', 'alastair', 'morton', 'said', '", 'the', 'firm', 'make', 'profit', 'interest', 'yad', '", 'although', 'sharehold', 'see', 'interest', 'dilut', '", 'offer', 'prospect', 'brighter', 'futur', 'month', 'uncertainti', 'eurotunnel', 'wrestl', 'reduc', 'crippl', 'interest', 'payment', 'negoti', 'tunnel', ""s", 'construct, '", 'eurotunnel', '", 'taken', 'around', 'half', 'cross-channel', 'market', 'european', 'ferri', 'compani', '", 'said', 'strong', 'oper', 'perform', 'could', 'allow', 'pay', 'first', 'dividend', 'within', 'next', '10', 'year', '", 'french', 'co-chairman', 'patrick', 'ponsol', 'said', 'sharehold', 'would', 'patient', 'could', 'reap', 'benefit', 'companii, "'s", 'success', '", 'he', 'call', 'debt', 'restructur', 'plan', '``, 'accept', 'compromis', """, 'holder', 'eurotunnel', 'share', '", 'the', 'companii, 'said', 'still', 'consider', 'work', 'done', 'finalis', 'agre', 'detail', 'plan', 'submit', 'sharehold', 'full', '225', 'bank', 'syndic', 'approv', '", 'probabl', 'earli', '1997', '", 'monday', ""s", 'announc', 'follow', 'two', 'week', 'highli', 'secret', 'negoti', 'eurotunnel', 'six', 'lead', 'bank', '", 'thi, 'extend', '24', '```, 'instruct', 'bank', """, 'meet', 'late', 'last', 'week', 'london', '", 'eurotunnel', 'said', 'debt-for-equ', 'swap', 'would', '130', 'penc', '", '10.40', 'franc', '", 'per', 'share', '", 'that', 'consider', 'level', 'around', '160', 'penc', 'wide', 'report', 'announc', 'deal', '", 'reduc', 'outstand', 'debt', '8.7', 'billion', 'pound', '(', '\$', '1.56', 'billion', ')', '", 'the', 'compani', 'said', '3.7', 'billion', 'pound', '(', '\$', '5.8', 'billion', ')', 'debt', 'would', 'convert', 'new', 'financi', 'instrument', 'exist', 'sharehold', 'would', 'abl', 'perticip', 'isau', 'looso', 'lake', 'free', 'warrant', 'entitl', '

Stemmed Document 3

['anglo-french', 'channel', 'tunnel', 'oper', 'eurotunnel', 'monday, 'announc', 'deal', 'give', 'creditor', 'bank', '45.5', 'percent', 'compani', 'return', 'wipe', 'one', 'billion', 'pound', '(, '\$', '1.56', 'billion', ')', 'debt', 'mountain', '', 'the', 'long-await', 'restructur', 'bring', 'end', 'month', 'wrangl', 'eurotunnel', '225', 'bank', 'owe', 'nearli', 'nine', 'billion', 'pound', 'i, 'the', 'deal', 'j, 'announc', 'simultan', 'pari', 'london', 'j, 'bring', 'compani', 'back', 'brink', 'insolv', 'leav', 'sharehold', 'own', '54.5', 'percent', 'compani', '!, '``', 'the', 'restructur', 'plan', 'provid', 'eurotunnel', 'medium', 'term', 'financi', 'stabil', 'allow', 'consolid', 'substanti', 'commerci', 'achiev', 'date', 'develop', 'oper', 'j, """, 'eurotunnel', 'co-chairman', 'alastair', 'morton', 'said', '', 'the', 'firm', 'make', 'profit', 'interest', 'j, 'ad', '!, 'although', 'sharehold', 'see', 'interest', 'dilut', 'j, 'offer', 'prospect', 'brighter', 'futur', 'monthh', 'uncertainti', 'eurotunnel', 'wrestl', 'reduc', 'crippl', 'interest', 'payment', 'negoti', 'tunnel', "s', 'construct', 'j, 'eurotunnel', 'j, 'taken', 'around', 'half', 'cross-channel', 'market', 'european, 'ferri', 'companii, 'j, 'said', 'strong', 'oper', 'perform', 'could', 'allow', 'pay', 'first', 'dividend', 'within', 'next', '10, 'year', 'j, 'french', 'co-chairman', 'patrick', 'ponsol', 'said', 'sharehold', 'would', 'patient', 'could', 'reap', 'benefit', 'companii, "said', 'statement', 'still', 'consider', 'work', 'done', 'finalis', 'agre', 'detail', 'plan', 'submit', 'sharehold', 'full', '225', 'bank', 'syndic', 'approv', 'j, 'probabl', 'earli', '1997', 'j, 'monday', ""s", 'announc', 'follow', 'two', 'week', 'highli', 'secret', 'negoti', 'eurotunnel', 'six', 'lead', 'bank', 'j, 'thi', 'extend', '24', '``', 'instruct', 'bank', """, 'meet', 'late', 'last', 'week', 'london', 'j, 'eurotunnel', 'said', 'debt-for-equ', 'swap', 'would', 'abl', 'particip', 'isau', 'j, 'the', 'compani', 'said', '3-7', 'billion', 'pound', 'debt', 'cony, 'take',

'warrant', 'entitl', 'subscrib', ',', 'eurotunnel', 'said', 'sharehold', "'", 'interest', 'may', 'reduc', '39', 'percent', 'compani', 'end', 'decemb', '2003', '.', 'eurotunnel', "'s", 'share', ',', 'suspend', 'last', 'week', '113.5', 'penc', 'ahead', 'monday', "'s", 'announc', ',' 'resum', 'trade', 'tuesday', ',' 'compani', 'said', '.', '(', '\$', '1=.6393', 'pound', ')']

4. Calculate tf-idf for each word in each document and generate document-word matrix (each element in the matrix is the tf-idf score for a word in a document) (7 points)

```
#Creating new documents i.e after tokenizing the documents, removing stopwords and stemming the document

new_document1 = ' '.join(stemmed_document1)
new_document2 = ' '.join(stemmed_document2)
new_document3 = ' '.join(stemmed_document3)

vectorizer = TfidfVectorizer()
tfidf_matrix = vectorizer.fit_transform([new_document1, new_document2, new_document3])

print("TF-IDF Matrix:\n")
print(tfidf_matrix.toarray())
```

→ TF-IDF Matrix:

```
[[0.05149603 0.02574801 0.
                                   0.02574801 0.
                                                         0.04359518
 0.02574801 0.02574801 0.04359518 0.02574801 0.04359518 0.02574801
 0.
            0.02574801 0.02574801 0.08719035 0.02574801 0.02574801
            0.03315525 0.04359518 0.04359518 0.05149603 0.
            0.05149603 0.04359518 0.04359518 0.05149603 0.02574801
 0.02574801 0.02574801 0.04359518 0.02574801 0.08719035 0.02574801
 0.04359518 0.07724404 0.04359518 0.05149603 0.02574801 0.04359518
 0.04359518 0.04359518 0.04359518 0.02574801 0.02574801 0.18023609
 0.04359518 0.04359518 0.
                                   0.08719035 0.12874006 0.04359518
 0.02574801 0.02574801 0.02574801 0.04359518 0.04359518 0.
 0.04359518 0.08719035 0.05149603 0.05149603 0.02574801 0.05149603
 0.04359518 0.08719035 0.02574801 0.30897616 0.04359518 0.
 0.08719035 0.05149603 0.02574801 0.04359518 0.02574801 0.02574801
  0.04359518 0.07724404 0.05149603 0.02574801 0.02574801 0.08719035
             0.18023609 0.18023609 0.02574801 0.04359518 0.04359518
  0.05149603 0.02574801 0.04359518 0.02574801 0.05149603 0.02574801
  0.04359518 0.04359518 0.07724404 0.04359518 0.02574801 0.02574801
 0.02574801 0.04359518 0.02574801 0.25748013 0.02574801 0.
 0.02574801 0.02574801 0.04359518 0.05149603 0.04359518 0.02574801
 0.02574801 0.04359518 0.04359518 0.
                                              0.02574801 0.04359518
  0.04359518 0.05149603 0.02574801 0.05149603 0.
                                                         0.05149603
```

TF-IDF Matrix:

```
[[0.05149603 0.02574801 0. 0.02574801 0. 0.04359518
 0.02574801 0.02574801 0.04359518 0.02574801 0.04359518 0.02574801
            0.02574801 0.02574801 0.08719035 0.02574801 0.02574801
          0.03315525 0.04359518 0.04359518 0.05149603 0.
            0.05149603 0.04359518 0.04359518 0.05149603 0.02574801
 0.02574801 0.02574801 0.04359518 0.02574801 0.08719035 0.02574801
 0.04359518 0.07724404 0.04359518 0.05149603 0.02574801 0.04359518
 0.04359518 0.04359518 0.04359518 0.02574801 0.02574801 0.18023609
 0.04359518 0.04359518 0. 0.08719035 0.12874006 0.04359518
 0.02574801 0.02574801 0.02574801 0.04359518 0.04359518 0.
 0.04359518 0.08719035 0.05149603 0.05149603 0.02574801 0.05149603
 0.04359518 0.08719035 0.02574801 0.30897616 0.04359518 0.
 0.08719035 0.05149603 0.02574801 0.04359518 0.02574801 0.02574801
 0.04359518 0.07724404 0.05149603 0.02574801 0.02574801 0.08719035
           0.18023609 0.18023609 0.02574801 0.04359518 0.04359518
 0.05149603 0.02574801 0.04359518 0.02574801 0.05149603 0.02574801
 0.04359518 0.04359518 0.07724404 0.04359518 0.02574801 0.02574801
 0.02574801 0.04359518 0.02574801 0.25748013 0.02574801 0.
 0.02574801 0.02574801 0.04359518 0.05149603 0.04359518 0.02574801
 0.02574801 0.04359518 0.04359518 0. 0.02574801 0.04359518
 0.04359518 0.05149603 0.02574801 0.05149603 0. 0.05149603
 0.04359518 0.04359518 0.02574801 0.04359518 0.04359518 0.04359518
 0.02574801 0.04359518 0. 0.04359518 0.02574801 0.04359518
 0. 0.04359518 0.07724404 0.04359518 0. 0.
 0.02574801 0.18023609 0.08719035 0.07724404 0.04359518 0.02574801
     0. 0.04359518 0.02574801 0.04359518 0.02574801
 0.04359518 0.04359518 0.02574801 0.02574801 0.04359518 0.04359518
 0.02574801 0.02574801 0.04359518 0.04359518 0.02574801 0.02574801
 0.02574801 0.13078553 0.05149603 0.02574801 0.10299205 0.
 0.04359518 0.05149603 0.02574801 0.02574801 0.08719035 0.02574801
 0.05149603 0.08719035 0.13078553 0.02574801 0.02574801 0.07724404
          0. 0.02574801 0.05149603 0.02574801 0.02574801
 0.02574801 0.07724404 0.05149603 0.07724404 0.02574801 0.07724404
 0.05149603 0.05149603 0.02574801 0.04359518 0.2059841 0.04359518
 0.02574801 0.02574801 0.02574801 0.02574801 0. 0.05149603
 0.04359518 0.07724404 0.02574801 0.02574801 0.02574801 0.04359518
 0.04359518 0.04359518 0.03315525 0.23173212 0.04359518 0.
 0.04359518 0.02574801 0.04359518 0.04359518 0.05149603 0.18023609
 0.02574801 0. 0.04359518 0.04359518 0.02574801 0.08719035
 0.03315525 0.07724404 0.02574801 0.02574801 0.02574801 0.
 0. 0.02574801 0.04359518 0.05149603 0. 0.05149603
 0.02574801 0.04359518 0.02574801 0. 0.10299205 0.
 0.04359518 0.04359518 0.08719035 0.02574801 0.02574801 0.05149603
```

```
0. 0.02574801 0.04359518 0.04359518 0.08719035 0.02574801
0.04359518 0.02574801 0.04359518 0.02574801 0.02574801 0.02574801
0.02574801 0.18023609 0. 0.02574801 0.154488081
[0.06704793 0.03352397 0.05676101 0.03352397 0.05676101 0.
0.03352397 0.03352397 0. 0.03352397 0. 0.06704793
0.04316821 0.03352397 0.03352397 0. 0.03352397 0.03352397

      0.08633642
      0.
      0.
      0.
      0.03352397
      0.03352397
      0.03352397
      0.04316821

      0.04316821
      0.03352397
      0.
      0.
      0.03352397
      0.03352397
      0.03352397

      0.03352397
      0.06704793
      0.
      0.03352397
      0.03352397
      0.03352397

      0.
      0.16761983
      0.
      0.03352397
      0.06704793
      0.

      0.
      0.
      0.03352397
      0.03352397
      0.16761983

      0.
      0.
      0.03352397
      0.3352397
      0.16761983

      0.
      0.04316821
      0.
      0.33523966
      0.

            0. 0.04316821 0. 0.33523966 0.
0.03352397 0.06704793 0.03352397 0. 0. 0.04316821
0. 0.06704793 0.06704793 0.03352397 0.06704793

      0.
      0.03352397
      0.30171569
      0.
      0.04316821

      0.
      0.06704793
      0.03352397
      0.
      0.03352397
      0.03352397

0. 0.06704793 0.03352397 0.03352397 0.03352397 0.
0.04316821 0.1005719 0.16761983 0.03352397 0. 0.
0.03352397 0.03352397 0. 0.03352397 0.03352397
0. 0.03352397 0. 0.06704793 0.03352397
0.03352397 0. 0.03352397 0.36876362 0.03352397 0.04316821
0.03352397 0.03352397 0. 0.06704793 0. 0.03352397
0.03352397 0. 0. 0.04316821 0.03352397 0.
0. 0.03352397 0.03352397 0.06704793 0.04316821 0.03352397

      0.
      0.
      0.03352397 0.
      0.
      0.
      0.

      0.03352397 0.
      0.04316821 0.
      0.03352397 0.
      0.

      0.04316821 0.
      0.04316821 0.04316821
      0.04316821

0.03352397 0.13409586 0. 0.03352397 0. 0.06704793

      0.04316821
      0.04316821
      0.
      0.03352397
      0.
      0.03352397
      0.

      0.
      0.06704793
      0.03352397
      0.
      0.
      0.

0.03352397 0.03352397 0. 0. 0.03352397 0.03352397
0.03352397 0. 0.1005719 0.06704793 0.03352397 0.
0. 0.03352397 0.06704793 0.03352397 0. 0.03352397
0.03352397 0. 0. 0.03352397 0.03352397 0.1005719
0.04316821 0.04316821 0.03352397 0.03352397 0.03352397 0.03352397
0.03352397 0.03352397 0.03352397 0.1005719 0.03352397 0.1005719
0.03352397 0.1005719 0.03352397 0. 0.13409586 0.
0.03352397 0.03352397 0.03352397 0.03352397 0.04316821 0.1005719
 0. 0.03352397 0.1005719 0.03352397 0.03352397 0.
0. 0. 0.26819172 0. 0.04316821

      0.
      0.03352397 0.
      0.
      0.1005719 0.20114379

      0.03352397 0.04316821 0.
      0.
      0.03352397 0.

 0. 0.03352397 0.03352397 0.03352397 0.03352397 0.04316821
0.04316821 0.03352397 0. 0.03352397 0.04316821 0.03352397
0.03352397 0. 0.03352397 0.04316821 0.20114379 0.04316821
```

```
      0.
      0.
      0.03352397
      0.03352397
      0.06704793

      0.04316821
      0.03352397
      0.
      0.
      0.03352397

0. 0.1005719 0. 0.03352397 0.03352397 0.03352397
0.03352397 0.13409586 0.04316821 0.03352397 0.033523971
[0.06971872 0.03485936 0.0.03485936 0.0.03485936 0.0.03485936 0.03485936 0.0.03485936 0.0.06971872

      0.04488777
      0.03485936
      0.03485936
      0.
      0.03485936
      0.03485936
      0.03485936
      0.03485936
      0.04488777

      0.04488777
      0.03485936
      0.
      0.
      0.03485936
      0.03485936

      0.03485936
      0.06971872
      0.
      0.03485936
      0.
      0.03485936

      0.
      0.13943744
      0.
      0.03485936
      0.06971872
      0.

      0.
      0.03485936
      0.03485936
      0.1742968

      0.
      0.04488777 0.
      0.20915616 0.

      0.03485936 0.06971872 0.03485936 0.
      0.04488777

      0.
      0.06971872
      0.06971872
      0.03485936
      0.06971872

      0.
      0.03485936
      0.31373424
      0.
      0.04488777

0.04488777 0.10457808 0.1742968 0.03485936 0. 0.
0.03485936 0.03485936 0. 0.03485936 0.03485936 0.03485936
0. 0.03485936 0. 0.06971872 0.03485936

      0.03485936
      0.03485936
      0.
      0.06971872
      0.
      0.03485936

      0.03485936
      0.
      0.04488777
      0.03485936
      0.

0. 0.03485936 0.03485936 0.06971872 0.04488777 0.03485936
0. 0. 0.03485936 0. 0. 0.

      0.03485936 0.
      0.04488777 0.
      0.03485936 0.

      0.04488777 0.
      0.03485936 0.
      0.04488777 0.04488777

0.03485936 0.13943744 0. 0.03485936 0. 0.06971872
0.04488777 0.04488777 0. 0.03485936 0. 0.03485936
0. 0.06971872 0.03485936 0. 0.
0.03485936 0.03485936 0. 0. 0.03485936 0.03485936
0.03485936 0. 0.10457808 0.06971872 0.03485936 0.05902203
0. 0.03485936 0.06971872 0.03485936 0. 0.03485936
0.03485936 0.03485936 0.03485936 0.10457808
0.04488777 0.04488777 0.03485936 0.03485936 0.03485936 0.03485936
0.03485936 0.03485936 0.03485936 0.10457808 0.03485936 0.10457808
0.03485936 0.10457808 0.03485936 0. 0.1742968 0.
 0.03485936 0.03485936 0.03485936 0.03485936 0.04488777 0.10457808
0. 0.03485936 0.10457808 0.03485936 0.03485936 0.
0.
          0. 0.04488777 0.27887488 0. 0.04488777
0. 0.03485936 0. 0. 0.10457808 0.20915616
0.03485936 0.04488777 0. 0. 0.03485936 0.
0.04488777 0.03485936 0.03485936 0.03485936 0.03485936 0.04488777
```

```
      0.03485936
      0.
      0.03485936
      0.04488777
      0.20915616
      0.04488777

      0.
      0.
      0.03485936
      0.03485936
      0.06971872

      0.04488777
      0.03485936
      0.
      0.03485936
      0.03485936

      0.
      0.10457808
      0.
      0.03485936
      0.03485936
      0.03485936

      0.03485936
      0.13943744
      0.04488777
      0.03485936
      0.03485936]]
```

5. Calculate pairwise cosine similarity for the documents (5 points)

[0.76020969 0.98134469 1.