Phase 2: Final Search Engine

Course code: CSI 4107

Student 1: Catherine DesOrmeaux, 7746139

Student 2: NamChi Nguyen, 7236760

Module 8a - Bigram Language Model

Functionality

The class BigramModel.java has the functionality of Module 8a. It uses the corpus taken by the DictionaryBuilder.java which are already tokenized, normalized and has stopword removal, but no stemming. Then, using the built-in NGramGenerator from Apache OpenNLP, a list of bigrams was created.

Limitations (cases not handled)

This is dependent on the preprocessing (tokenization, etc) from the DictionaryBuilder. If this is poorly done, then the bigram list will contain words that aren't in pairs.

Problems encountered (if any, as you developed the module)

The bigram list originally included extra whitespace, single words and empty strings which was taken from corpus that was not properly tokenized. So, extra steps had to be taken to remove these from the list.

Module 8b - Query Completion Module

Functionality

The class BigramModel.java also has the functionality of Module 8b. We decided to combine these two modules together into a single class since the query completion is just a single function in BigramModel.

From the list of bigrams, we get the first and second terms individually and calculate the frequency of when the second term appears after the first term is given. We compare the first term with the query word and the second term becomes the suggested word. And then sort the list from highest frequency to lowest. The suggested words for query completion will take the top n from the list.

Limitations (cases not handled)

It only takes a single query word at a time to generate a list of suggested words.
Also, since the bigram list is generated from the Reuters' corpus, the next suggested word will most likely be the same word that appears in the text.

If the user types a word that does not have a following word, then there is nothing in the combo box. For Boolean model, the user must manually add AND, OR, AND_NOT in between the words since query completion won't have suggestions for AND, OR AND_NOT.

Problems encountered (if any, as you developed the module)

N/A

Module 9a - Automatic Thesaurus Construction

Functionality

The class Thesaurus.java has the functionality of Module 9a. It uses the Jaccard coefficient similarity to compare documents as a unit of comparison. From the documents, the words (e.g. tokens) in the document are used as input for Jaccard. The thesaurus uses a HashMap, with the set of two words as the key, and a double as the similarity.

Limitations (cases not handled)

N/A

Problems encountered (if any, as you developed the module)

N/A

Module 9b - Global Query Expansion (in VSM)

Functionality

The class VSM.java has the functionality of Module 9b and Thesaurus.java provides a function to find the max similarity using the Jaccard coefficient. We chose to implement with implicit expansion with only a limit of 1, in which the word is only expanded once. In the case that the query has more than 1 word, we only look at the last word in the query and expand on that word for simplicity.

Limitations (cases not handled)

It doesn’t expand on all words if there are more than one in a query.

Problems encountered (if any, as you developed the module)

N/A

Module 10a - Text categorization with kNN

Functionality

The class MachineLearning.java has the functionality of Module 10a. Since this is also part of preprocessing, we read the file outputted from PreprocessorReuters.java and do tokenization, stopword removal, and normalization on the Reuters' text. All the documents that didn't have a topic assigned are put into a test set and the remaining documents with topics are used as a training set.

Next, we used the Jaccard coefficient similarity measure with k = 1. So, from the training set, the document with the highest similarity is used to assign the topic to the document with no topic. For multiple topic assignment, if the chosen document had multiple topics, we assigned all topics to the test document.

Limitations (cases not handled)

Since we chose k = 1, the documents could easily be misclassified.

Problems encountered (if any, as you developed the module)

Originally, DictionaryBuilder tokenized and found all the documents that had no topics. However, it took a long time to process to pass the documents back and forth between DictionaryBuilder and this module. So we decided to process everything and give DictionaryBuilder a final text file that had topics assigned to all documents.

Module 10b - Topic Restriction

Functionality

The class MainPage.java has the functionality of Module 10b. It reads an external text file *all-topics-strings.lc.txt* to populate a list of possible topics in the UI. From that, the documents will be filtered based on the topics a user selects by using an inverted index except for topics (e.g. topic → [list of docIDs]) that is created by DictionaryBuilder.

If the document collection U of O courses are selected, a user can pick a topic and the option is not disabled, but no topics will be shown in the results since it is not applicable.

Limitations (cases not handled)

N/A

Problems encountered (if any, as you developed the module)

N/A

Optional Module - Visualization of automatic thesaurus

Functionality

The classes Thesaurus.java and VisualizeThesaurus.java have the functionality of an Optional Module.

Note: After asking the professor, we had the option to visualize the thesaurus as a table by sorting the similarity and taking the top n results. We chose this option instead of visualizing the thesaurus as a semantic map/graph.

From the thesaurus, we decided to return the top 15 highest similarities to a given word. If the similarity was 0.0, we removed it from our list or found the next smallest value to substitute. Then the similarities were sorted.

Limitations (cases not handled)

N/A

Problems encountered (if any, as you developed the module)

A pair of similar words (<chosen word, similar word> → similarity) returned a Set, but we needed it as a String to be usable in the UI. We removed the chosen word and kept the similar word and then converted it to a String.

Optional Module - Text categorization with Naive Bayes

Functionality

The class MachineLearning.java has the functionality of an Optional Module. Similar to KNN, the documents are split into a training and testing set respectively. The prior and posterior probabilities are calculated and the topic assigned to the test document is chosen by the max probability.

Limitations (cases not handled)

The training and test set are small, so the topics assigned are sensitive to the biased prior probabilities found.

Problems encountered (if any, as you developed the module)

N/A

Describe how you dealt with the Reuters collection

<u>NOTE</u>: The reuters text files can be found in *Sample Reuters Text File Outputs* folder

1. Were you able to process all the files?
Yes, we were able to process and extract the title, topic and text/body of all the Reuters outputted by PreprocessorReuters.java (see *reuters_output.txt*) into a single text file.

2. If not, why? What caused problems?
But, due to time constraint and testing purposes of our modules, the large collection made it difficult to confirm whether our implementations of the modules were working as intended.

So, we decided to test on 1 Reuters file reut2-000.sgm (see *single_reuters_output.txt*). Unfortunately, one problem we had was building the automatic thesaurus which was too time-consuming with even a single sgm file.

3. How many documents did you end up with?
Therefore, we reduced the number of documents used even more in order to test our implementations (see *test_reuters_output.txt*). This text file contains the first 40 documents from reut2-000.sgm/single_reuters_output.txt.

4. Did you have any execution time issues? (searches being too long for example). If yes, what did you do?
As explained in Questions 2-3 on reducing our document collection, yes, we had execution time issues. When it came to testing our modules like the thesaurus or assigning topics to unassigned documents, the time to execute took at least more than 10 minutes of waiting. Since we couldn’t afford to wait, we had to reduce the collection in order to obtain immediate results.

5. How long does it take to generate the dictionary? Did you set up some constraints to make the dictionary generation faster?
As explained in Questions 2-3, since our collection was limited, the generation of the dictionary did not take a long time to process. However, we wrote all our frequently used modules to an external text file (e.g. dictionary, inverted index, list of tokens) that are read only once in the constructor. We added a parameter to determine the type of document collection used as well as whether KNN or Naive Bayes was chosen to preprocess the documents. These external files are used as input for our other modules.

6. In general, describe how more challenging it is to work with the Reuters collection than with the CSI collection that was used for the vanilla system.

The Reuters collection was a lot more challenging to work with. The preprocessing took time due to inconsistency in the appearance of some tags in the sgm files that had to be handled. The body of the Reuters also increased considerably which would increase the time for building the dictionary and inverted index as well as the thesaurus as it compares the similarity of every pair of words.

Additional Info

How did we split the work?

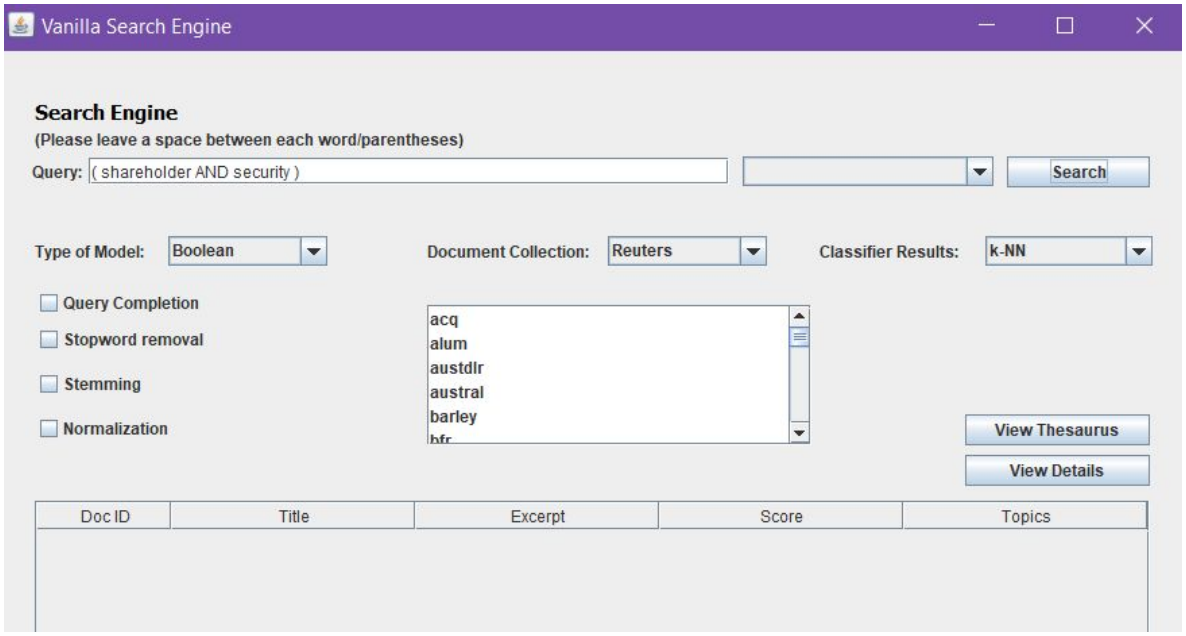
We met each day to work on the project. All members have contributed to each module equally.

Screenshots of the results

NOTE: Due to our chosen subset of the Reuters document collection, we will first present the screenshots of the required queries (red). Then, include our own queries (blue) to showcase the Boolean and VSM models if the required queries didn't return any results.

Boolean Model

(shareholder AND security)



Search Engine

(Please leave a space between each word/parentheses)

Query: (shareholder AND security)

Type of Model: Boolean

Document Collection: Reuters

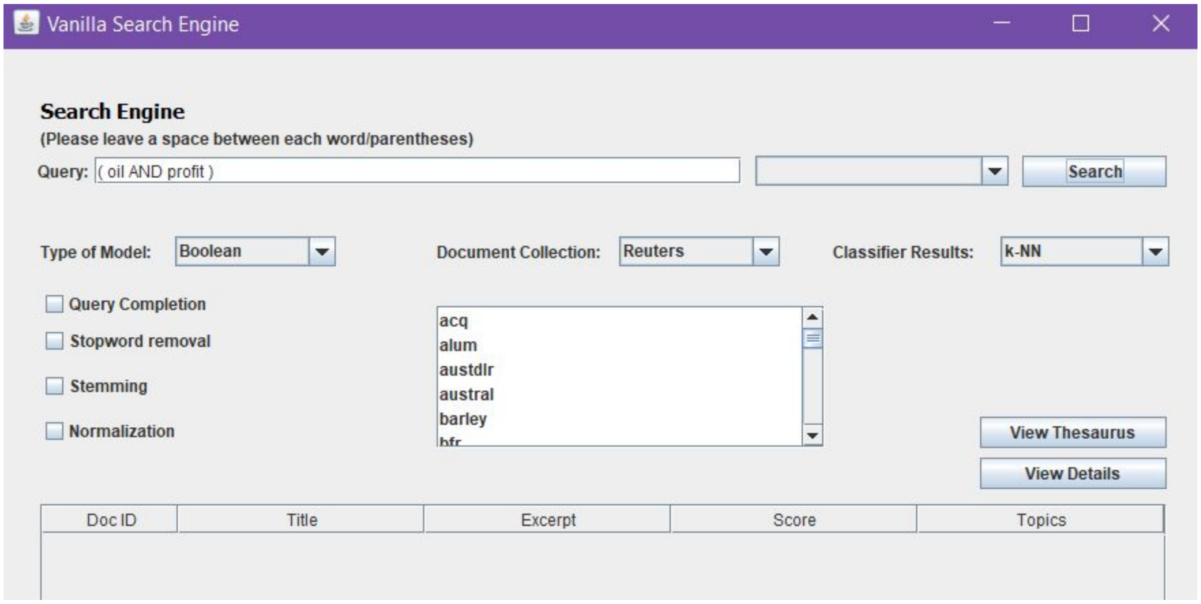
Classifier Results: k-NN

☐ Query Completion
☐ Stopword removal
☐ Stemming
☐ Normalization

acq
alum
austdlr
austral
barley
bfr

Doc ID	Title	Excerpt	Score	Topics

(oil AND profit)



Search Engine

(Please leave a space between each word/parentheses)

Query: (oil AND profit)

Type of Model: Boolean

Document Collection: Reuters

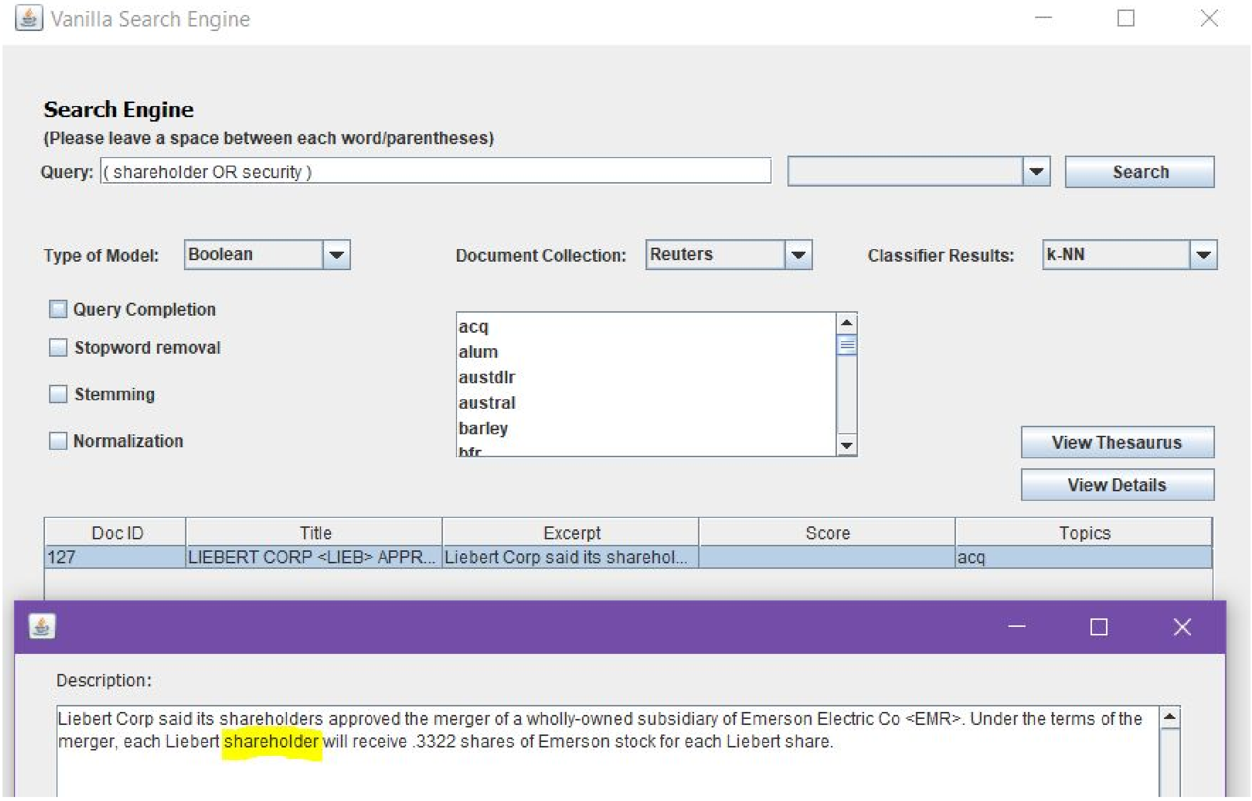
Classifier Results: k-NN

☐ Query Completion
☐ Stopword removal
☐ Stemming
☐ Normalization

acq
alum
austdlr
austral
barley
bfr

Doc ID	Title	Excerpt	Score	Topics

(shareholder OR security)



Search Engine

(Please leave a space between each word/parentheses)

Query: (shareholder OR security)

Type of Model: Boolean

Document Collection: Reuters

Classifier Results: k-NN

□ Query Completion

□ Stopword removal

□ Stemming

□ Normalization

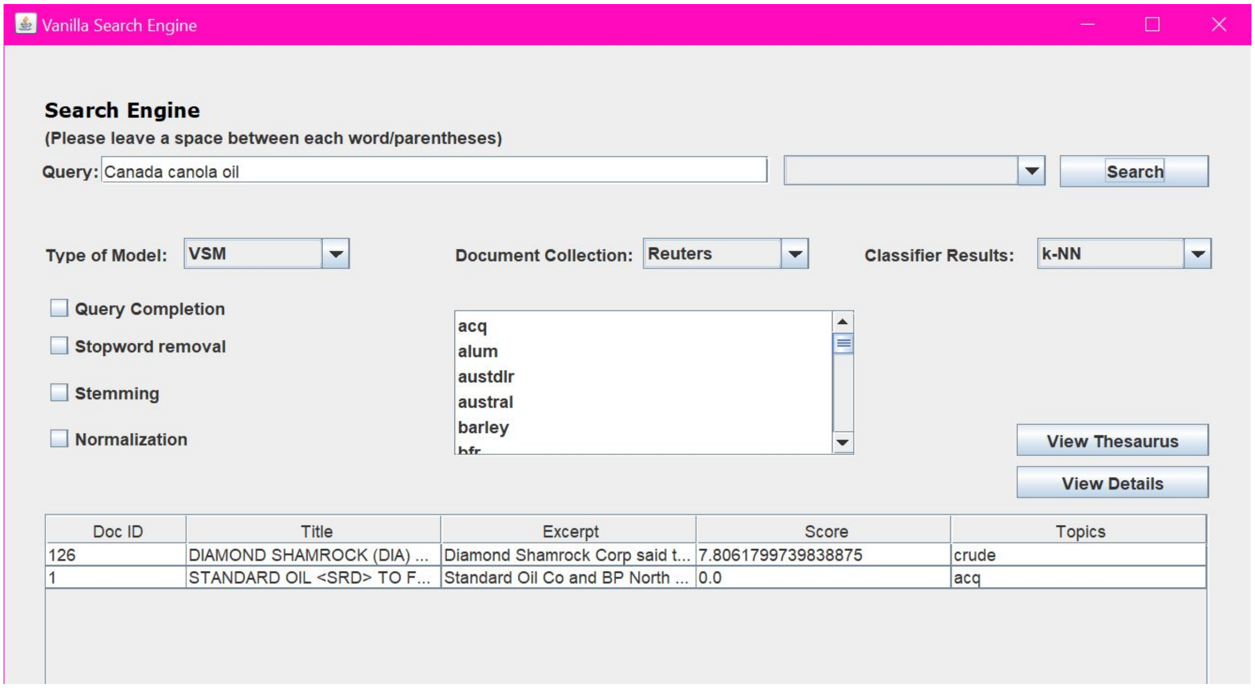
acq
alum
austdlr
austral
barley
bfr

Doc ID	Title	Excerpt	Score	Topics
127	LIEBERT CORP <LIEB> APPR...	Liebert Corp said its sharehol...		acq

Description:

Liebert Corp said its shareholders approved the merger of a wholly-owned subsidiary of Emerson Electric Co <EMR>. Under the terms of the merger, each Liebert shareholder will receive .3322 shares of Emerson stock for each Liebert share.

Canada canola oil (VSM) - kNN



Search Engine

(Please leave a space between each word/parentheses)

Query: Canada canola oil

Type of Model: VSM

Document Collection: Reuters

Classifier Results: k-NN

□ Query Completion

□ Stopword removal

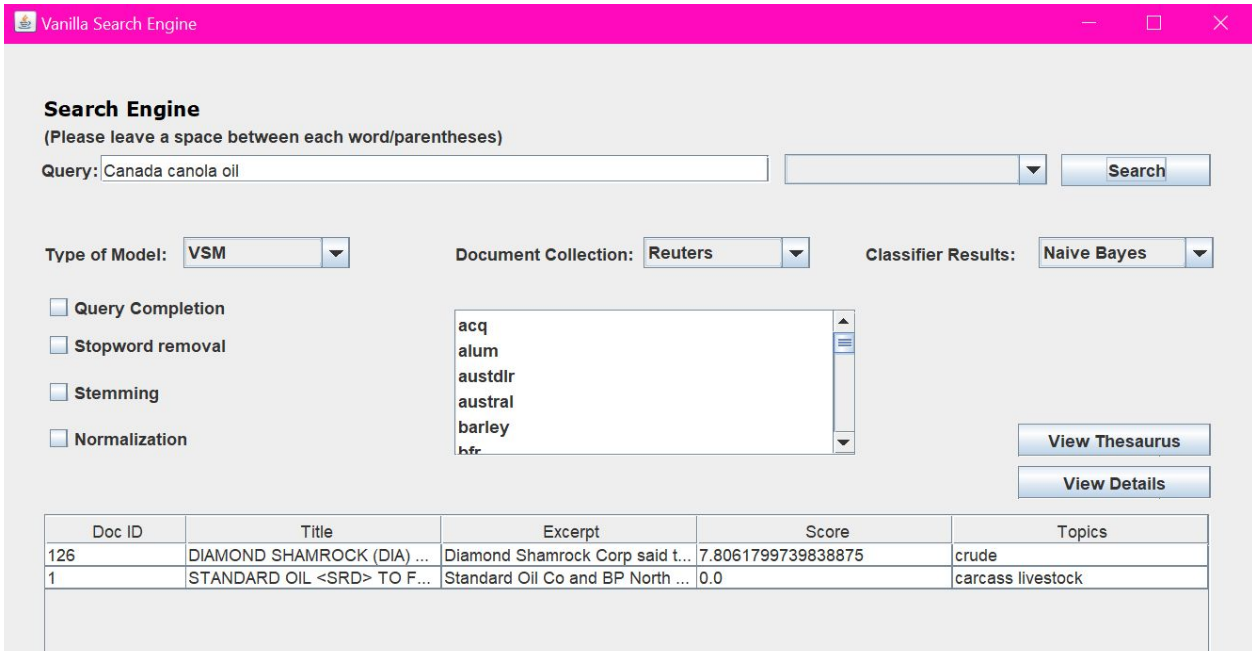
□ Stemming

□ Normalization

acq
alum
austdlr
austral
barley
bfr

Doc ID	Title	Excerpt	Score	Topics
126	DIAMOND SHAMROCK (DIA) ...	Diamond Shamrock Corp said t...	7.8061799739838875	crude
1	STANDARD OIL <SRD> TO F...	Standard Oil Co and BP North ...	0.0	acq

Canada canola oil (VSM) - Naive Bayes



Search Engine

(Please leave a space between each word/parentheses)

Query: Canada canola oil

Type of Model: VSM

Document Collection: Reuters

Classifier Results: Naive Bayes

☐ Query Completion

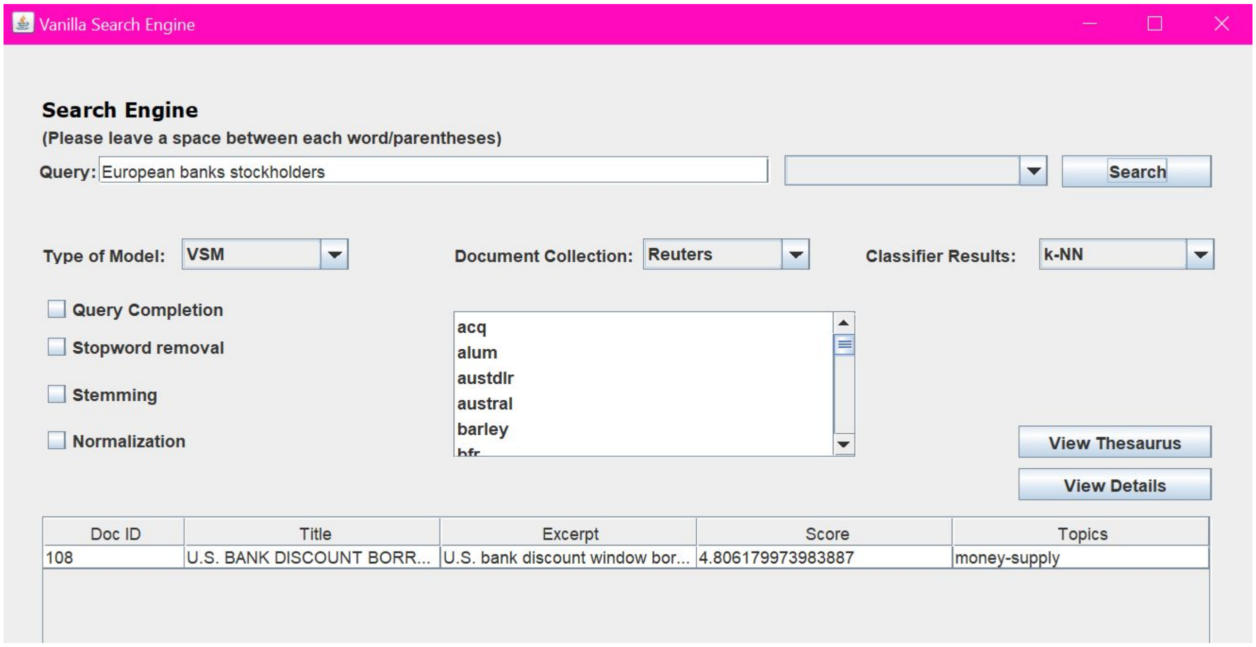
☐ Stopword removal

☐ Stemming

☐ Normalization

Doc ID	Title	Excerpt	Score	Topics
126	DIAMOND SHAMROCK (DIA) ...	Diamond Shamrock Corp said t...	7.8061799739838875	crude
1	STANDARD OIL <SRD> TO F...	Standard Oil Co and BP North ...	0.0	carcass livestock

European banks stockholders (VSM) - kNN



Search Engine

(Please leave a space between each word/parentheses)

Query: European banks stockholders

Type of Model: VSM

Document Collection: Reuters

Classifier Results: k-NN

☐ Query Completion

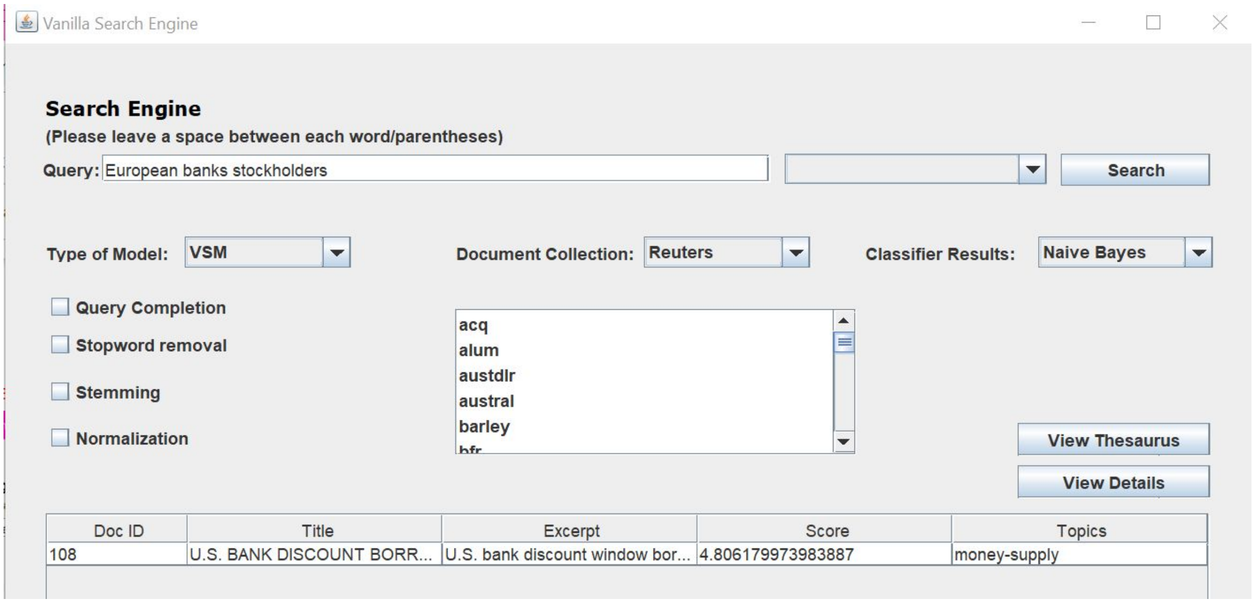
☐ Stopword removal

☐ Stemming

☐ Normalization

Doc ID	Title	Excerpt	Score	Topics
108	U.S. BANK DISCOUNT BORR...	U.S. bank discount window bor...	4.806179973983887	money-supply

European banks stockholders (VSM) - Naive Bayes



Search Engine

(Please leave a space between each word/parentheses)

Query: European banks stockholders

Type of Model: VSM

Document Collection: Reuters

Classifier Results: Naive Bayes

□ Query Completion

□ Stopword removal

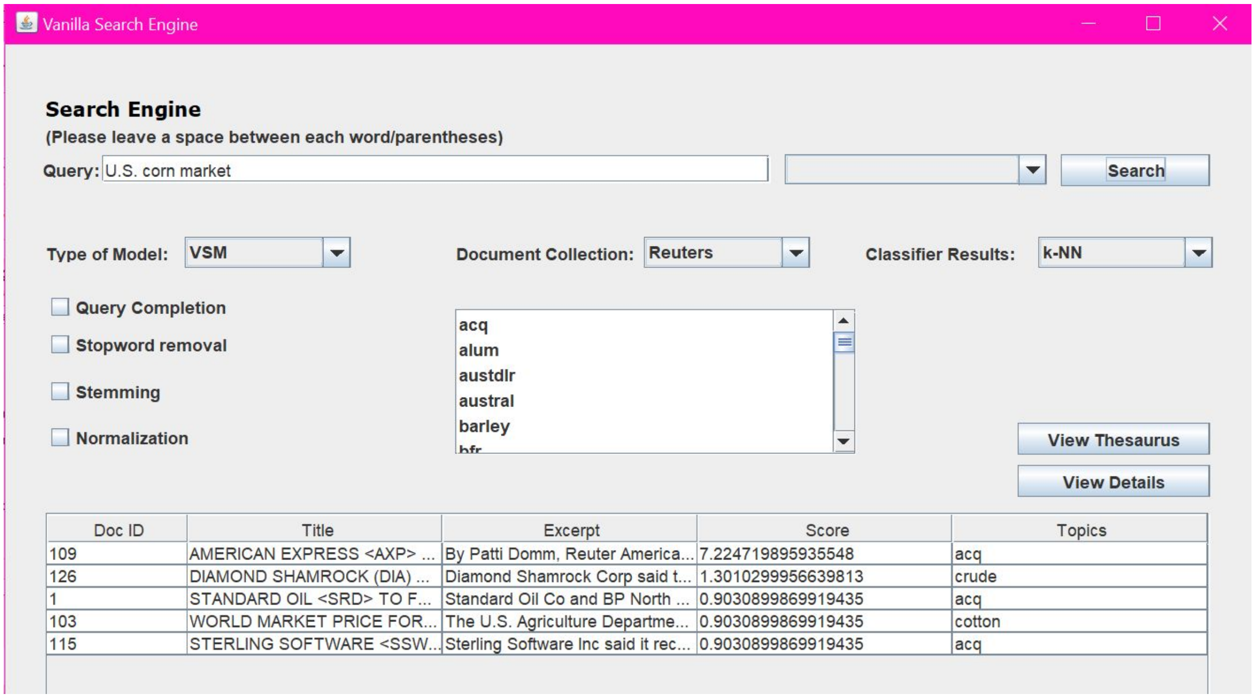
□ Stemming

□ Normalization

acq
alum
austdlr
austral
barley
bfr

Doc ID	Title	Excerpt	Score	Topics
108	U.S. BANK DISCOUNT BORR...	U.S. bank discount window bor...	4.806179973983887	money-supply

U.S. corn market (VSM) - kNN



Search Engine

(Please leave a space between each word/parentheses)

Query: U.S. corn market

Type of Model: VSM

Document Collection: Reuters

Classifier Results: k-NN

□ Query Completion

□ Stopword removal

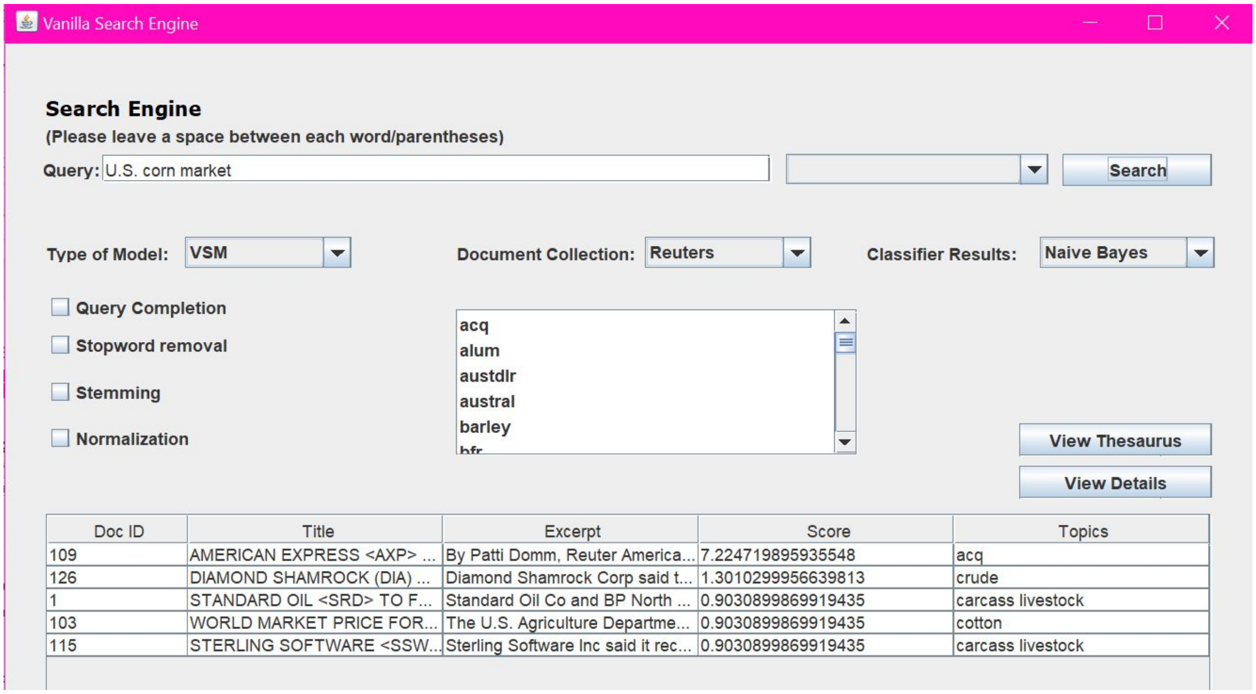
□ Stemming

□ Normalization

acq
alum
austdlr
austral
barley
bfr

Doc ID	Title	Excerpt	Score	Topics
109	AMERICAN EXPRESS <AXP> ...	By Patti Domm, Reuter America...	7.224719895935548	acq
126	DIAMOND SHAMROCK (DIA) ...	Diamond Shamrock Corp said t...	1.3010299956639813	crude
1	STANDARD OIL <SRD> TO F...	Standard Oil Co and BP North ...	0.9030899869919435	acq
103	WORLD MARKET PRICE FOR...	The U.S. Agriculture Departme...	0.9030899869919435	cotton
115	STERLING SOFTWARE <SSW...	Sterling Software Inc said it rec...	0.9030899869919435	acq

U.S. corn market (VSM) - Naive Bayes



Search Engine

(Please leave a space between each word/parentheses)

Query: U.S. corn market

Type of Model: VSM

Document Collection: Reuters

Classifier Results: Naive Bayes

□ **Query Completion**

□ **Stopword removal**

□ **Stemming**

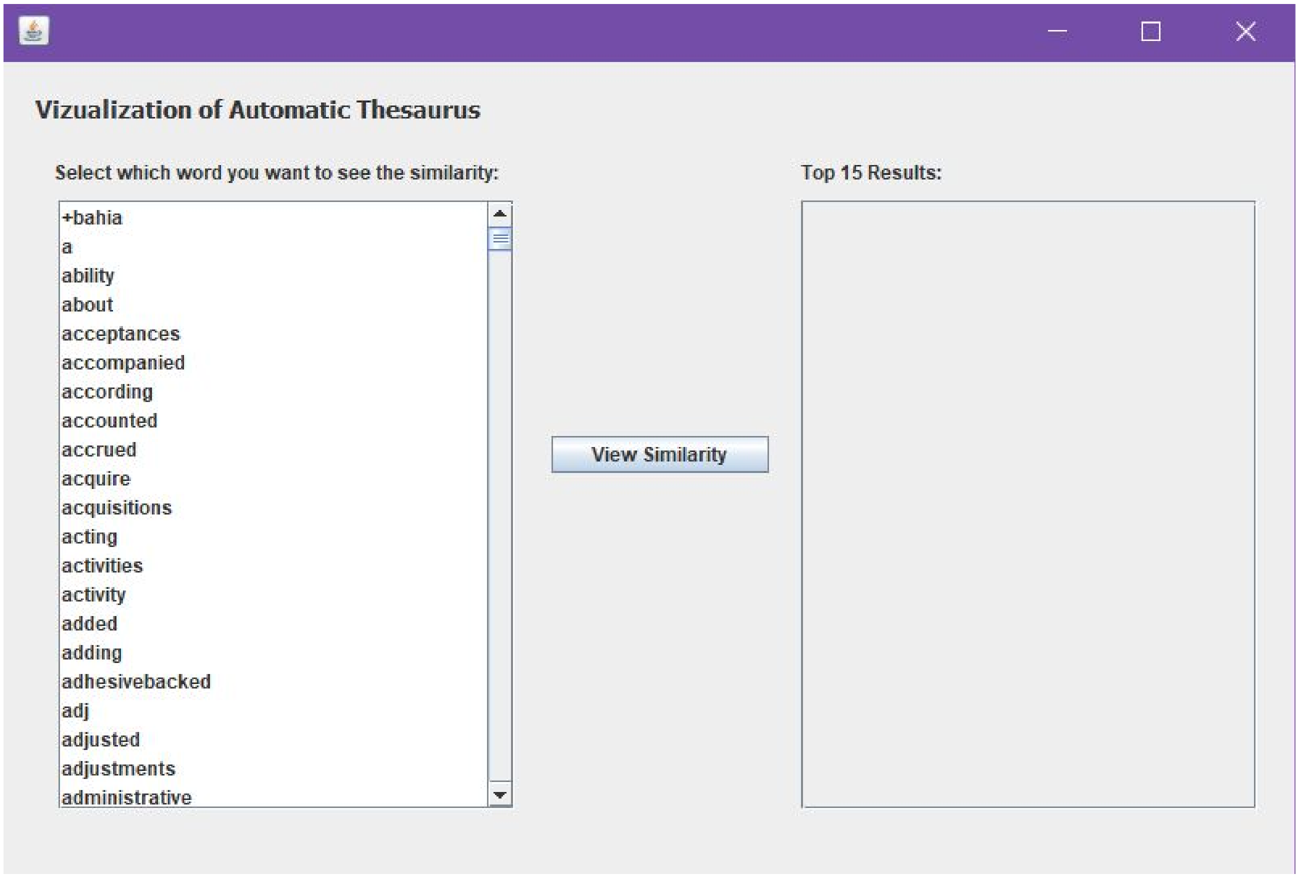
□ **Normalization**

acq
alum
austdlr
austral
barley
bfr

Doc ID	Title	Excerpt	Score	Topics
109	AMERICAN EXPRESS <AXP> ...	By Patti Domm, Reuter America...	7.224719895935548	acq
126	DIAMOND SHAMROCK (DIA) ...	Diamond Shamrock Corp said t...	1.3010299956639813	crude
1	STANDARD OIL <SRD> TO F...	Standard Oil Co and BP North ...	0.9030899869919435	carcass livestock
103	WORLD MARKET PRICE FOR...	The U.S. Agriculture Departme...	0.9030899869919435	cotton
115	STERLING SOFTWARE <SSW...	Sterling Software Inc said it rec...	0.9030899869919435	carcass livestock

NOTE; We assume that the Probabilistic Relevance Retrieval model only works for U of O courses

Visualization of automatic thesaurus



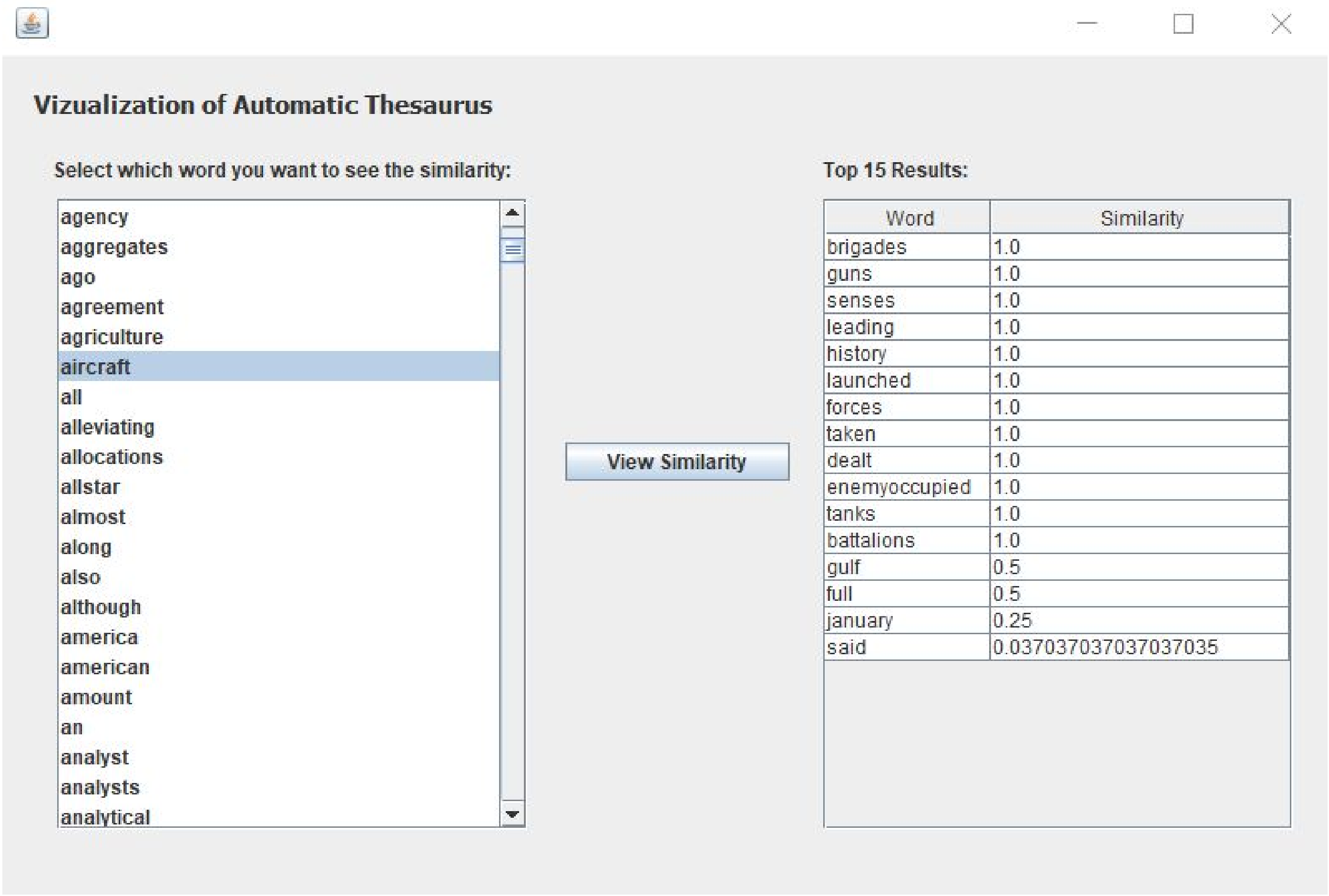
Vizualization of Automatic Thesaurus

Select which word you want to see the similarity:

+bahia
a
ability
about
acceptances
accompanied
according
accounted
accrued
acquire
acquisitions
acting
activities
activity
added
adding
adhesivebacked
adj
adjusted
adjustments
administrative

Top 15 Results:

E.g. Top 15 similar words of *aircraft* based on Jaccard



Vizualization of Automatic Thesaurus

Select which word you want to see the similarity:

agency
aggregates
ago
agreement
agriculture
aircraft
all
alleviating
allocations
allstar
almost
along
also
although
america
american
amount
an
analyst
analysts
analytical

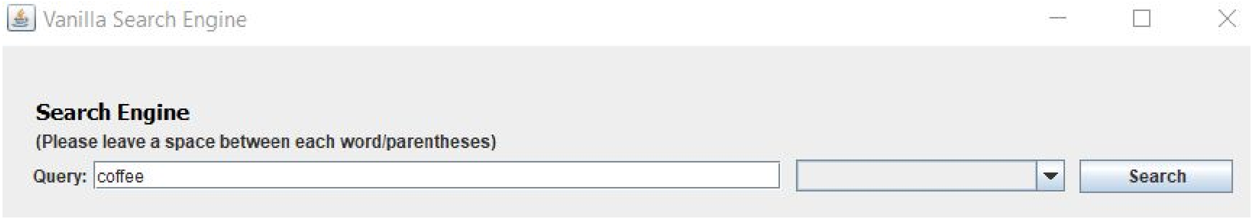
Top 15 Results:

Word	Similarity
brigades	1.0
guns	1.0
senses	1.0
leading	1.0
history	1.0
launched	1.0
forces	1.0
taken	1.0
dealt	1.0
enemyoccupied	1.0
tanks	1.0
battalions	1.0
gulf	0.5
full	0.5
january	0.25
said	0.037037037037037035

Bigram Language Model

NOTE: Functionality will be shown in demo

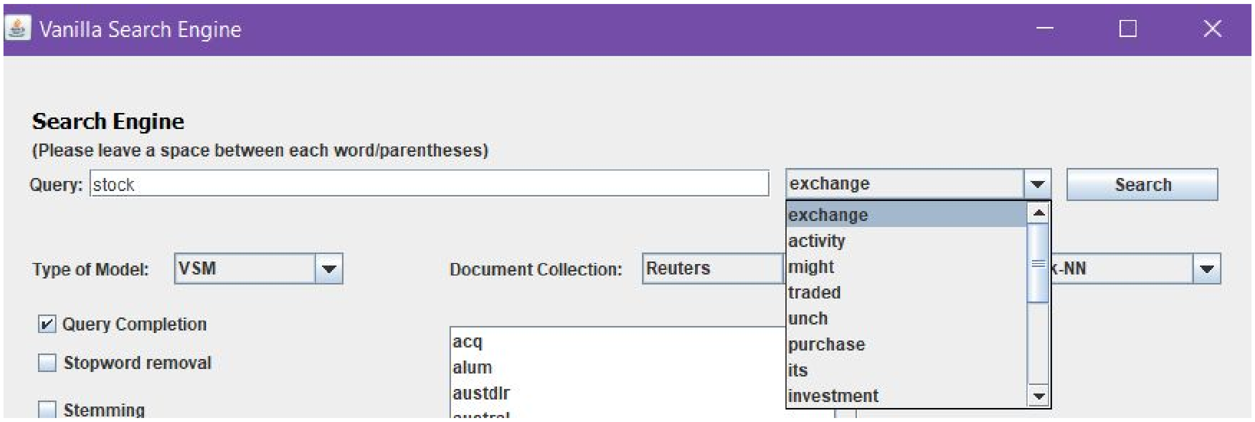
Five test words coffee, stock, oil, product and grain are shown below:



Search Engine

(Please leave a space between each word/parentheses)

Query: coffee



Search Engine

(Please leave a space between each word/parentheses)

Query: stock

exchange

exchange
activity
might
traded
unch
purchase
its
investment

Type of Model: VSM

Document Collection: Reuters

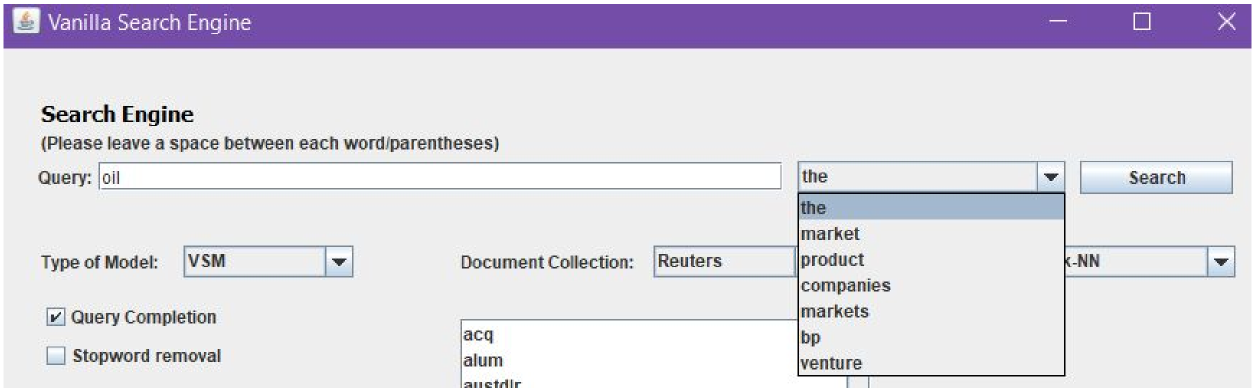
k-NN

☑ Query Completion

□ Stopword removal

□ Stemming

acq
alum
austdlr



Search Engine

(Please leave a space between each word/parentheses)

Query: oil

the

the
market
product
companies
markets
bp
venture

Type of Model: VSM

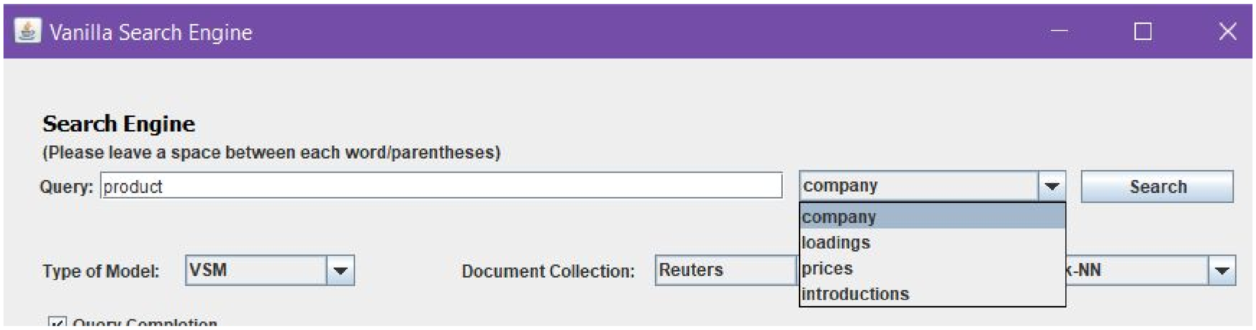
Document Collection: Reuters

k-NN

☑ Query Completion

□ Stopword removal

acq
alum
austdlr



Search Engine

(Please leave a space between each word/parentheses)

Query: product

company

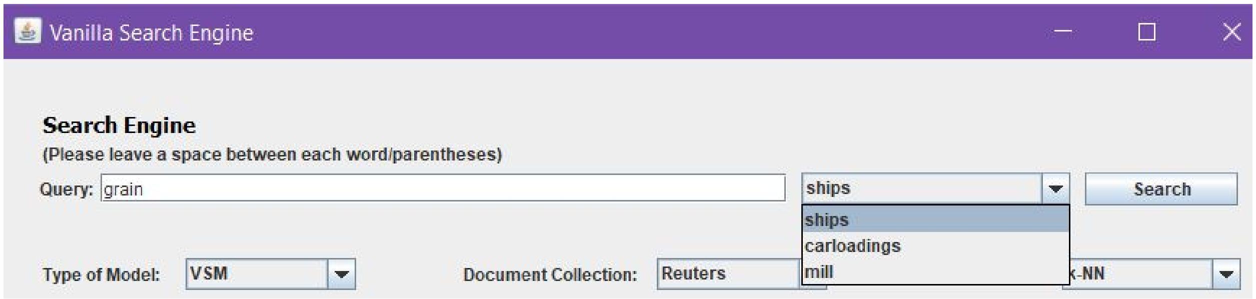
company
loadings
prices
introductions

Type of Model: VSM

Document Collection: Reuters

k-NN

☑ Query Completion



Query: grain

ships

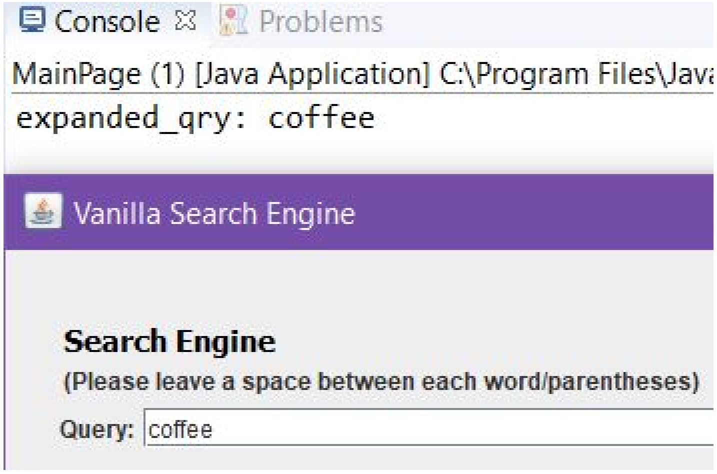
Type of Model: VSM

Document Collection: Reuters

k-NN

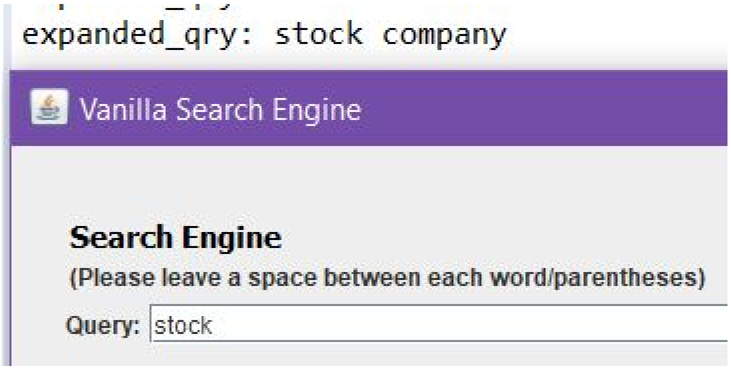
Global Query Expansion (in VSM)

We only expand the query with 1 word.



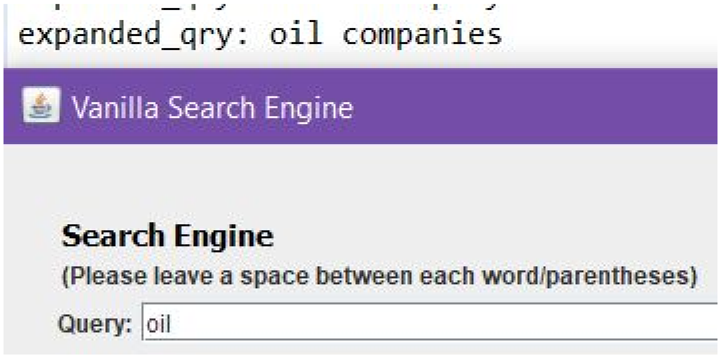
```
MainPage (1) [Java Application] C:\Program Files\Jav
expanded_qry: coffee
```

Query: coffee



```
expanded_qry: stock company
```

Query: stock

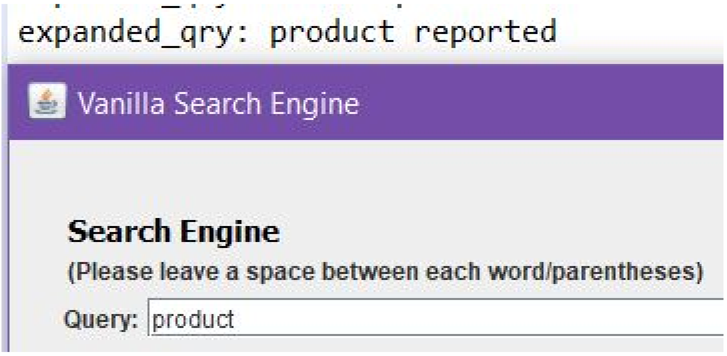


```
expanded_qry: oil companies
```

Search Engine

(Please leave a space between each word/parentheses)

Query: oil

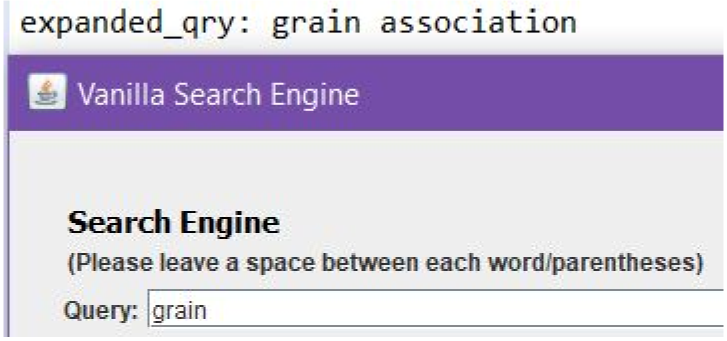


```
expanded_qry: product reported
```

Search Engine

(Please leave a space between each word/parentheses)

Query: product



```
expanded_qry: grain association
```

Search Engine

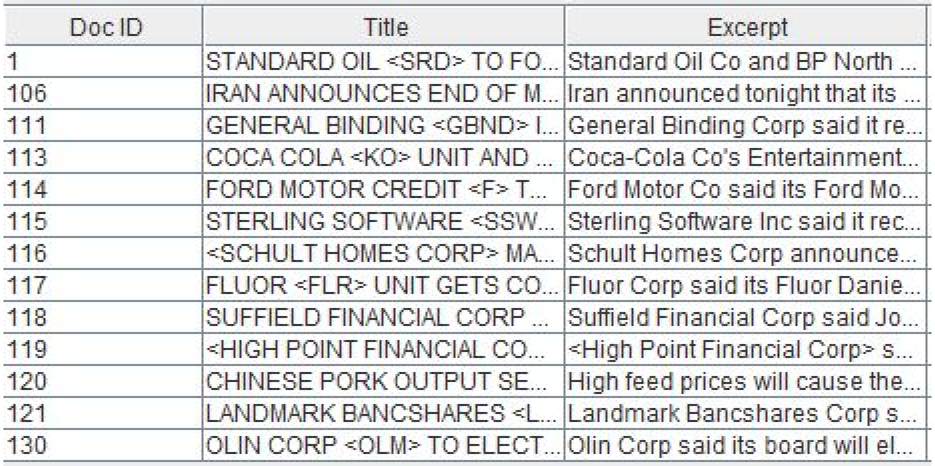
(Please leave a space between each word/parentheses)

Query: grain

Topic Classification

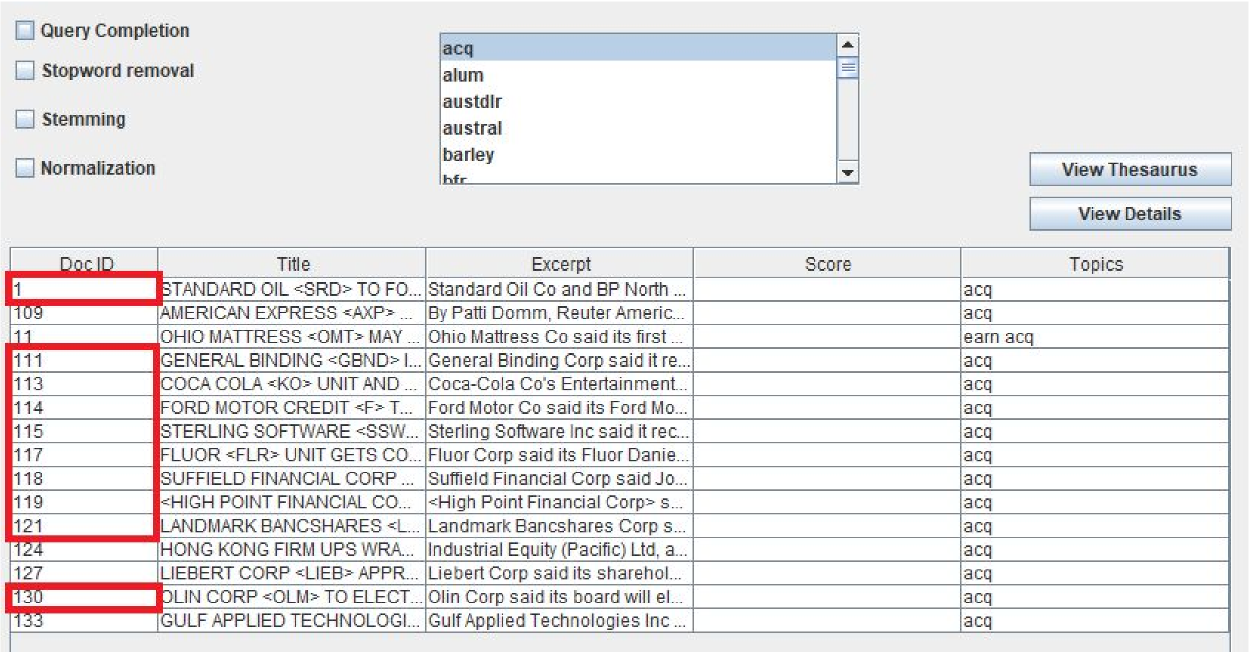
NOTE: Due to our reduced document collection, 5 topics can't be shown for the documents that we classified. With this subset, the documents were classified into 3 different topics.

Our documents with NO topics (by docID):



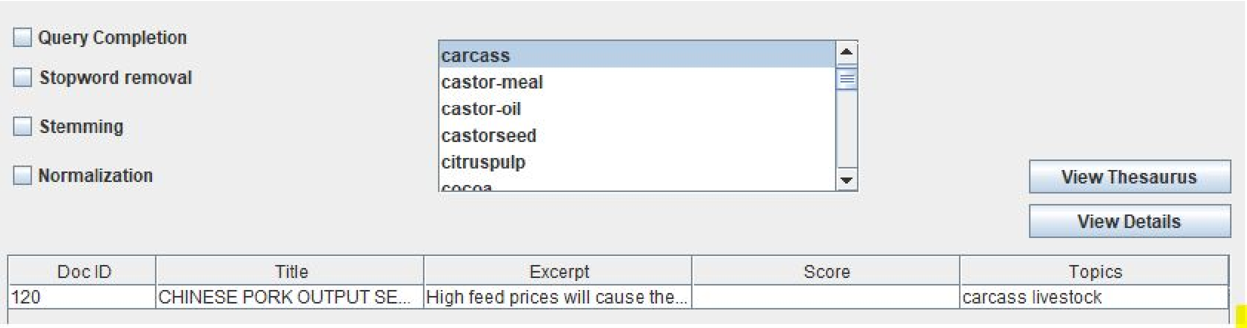
Doc ID	Title	Excerpt
1	STANDARD OIL <SRD> TO FO...	Standard Oil Co and BP North ...
106	IRAN ANNOUNCES END OF M...	Iran announced tonight that its ...
111	GENERAL BINDING <GBND> I...	General Binding Corp said it re...
113	COCA COLA <KO> UNIT AND ...	Coca-Cola Co's Entertainment...
114	FORD MOTOR CREDIT <F> T...	Ford Motor Co said its Ford Mo...
115	STERLING SOFTWARE <SSW...	Sterling Software Inc said it rec...
116	<SCHULT HOMES CORP> MA...	Schult Homes Corp announce...
117	FLUOR <FLR> UNIT GETS CO...	Fluor Corp said its Fluor Danie...
118	SUFFIELD FINANCIAL CORP ...	Suffield Financial Corp said Jo...
119	<HIGH POINT FINANCIAL CO...	<High Point Financial Corp> s...
120	CHINESE PORK OUTPUT SE...	High feed prices will cause the...
121	LANDMARK BANCSHARES <L...	Landmark Bancshares Corp s...
130	OLIN CORP <OLM> TO ELECT...	Olin Corp said its board will el...

Topic: acq



Doc ID	Title	Excerpt	Score	Topics
1	STANDARD OIL <SRD> TO FO...	Standard Oil Co and BP North ...		acq
109	AMERICAN EXPRESS <AXP> ...	By Patti Domm, Reuter Americ...		acq
11	OHIO MATTRESS <OMT> MAY ...	Ohio Mattress Co said its first ...		earn acq
111	GENERAL BINDING <GBND> I...	General Binding Corp said it re...		acq
113	COCA COLA <KO> UNIT AND ...	Coca-Cola Co's Entertainment...		acq
114	FORD MOTOR CREDIT <F> T...	Ford Motor Co said its Ford Mo...		acq
115	STERLING SOFTWARE <SSW...	Sterling Software Inc said it rec...		acq
117	FLUOR <FLR> UNIT GETS CO...	Fluor Corp said its Fluor Danie...		acq
118	SUFFIELD FINANCIAL CORP ...	Suffield Financial Corp said Jo...		acq
119	<HIGH POINT FINANCIAL CO...	<High Point Financial Corp> s...		acq
121	LANDMARK BANCSHARES <L...	Landmark Bancshares Corp s...		acq
124	HONG KONG FIRM UPS WRA...	Industrial Equity (Pacific) Ltd, a...		acq
127	LIEBERT CORP <LIEB> APPR...	Liebert Corp said its sharehol...		acq
130	OLIN CORP <OLM> TO ELECT...	Olin Corp said its board will el...		acq
133	GULF APPLIED TECHNOLOGI...	Gulf Applied Technologies Inc ...		acq

Topics: carcass, livestock



□ Query Completion

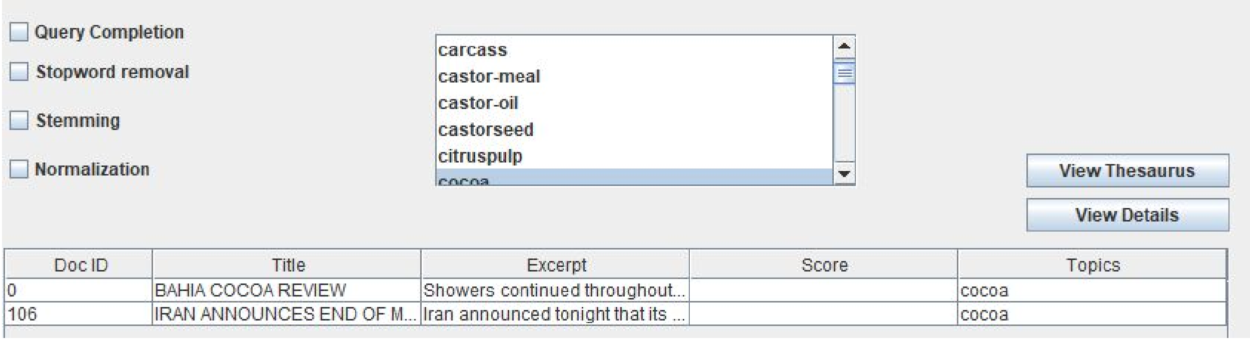
□ Stopword removal

□ Stemming

□ Normalization

Doc ID	Title	Excerpt	Score	Topics
120	CHINESE PORK OUTPUT SE...	High feed prices will cause the...		carcass livestock

Topic: cocoa



□ Query Completion

□ Stopword removal

□ Stemming

□ Normalization

Doc ID	Title	Excerpt	Score	Topics
0	BAHIA COCOA REVIEW	Showers continued throughout...		cocoa
106	IRAN ANNOUNCES END OF M...	Iran announced tonight that its ...		cocoa

Since k=1, it can easily be misclassified (as in this case for docID = 106)

What k did you use for kNN?
We used k = 1.

We are not doing a formal evaluation, but does the classification seem to make sense on the 5 examples you chose?
In the previous screenshot, most results make sense except the last one.

Do you think the kNN is a good approach here?
Yes, it is good approach. However, it is hard to analyze the results as we only used a very small set of the reuters data.

References

**All references have also been inspired by the theory seen in the lectures.
**All references that we copied/modified for the following classes:

BigramModel.java

- createBigram function: https://stackoverflow.com/questions/29656071/java-arraylist-remove-multiple-element-by-index
- totalNumWord function: http://www.java67.com/2016/09/3-ways-to-count-words-in-java-string.html
- getKeyFromValue function: http://www.java2s.com/Code/Java/Collections-Data-Structure/GetakeyfromvaluewithanHashMap.htm
- getDocIDs function: http://www.java2s.com/Code/Java/Collections-Data-Structure/GetakeyfromvaluewithanHashMap.htm
- getDocIDs_UI function: http://www.java2s.com/Code/Java/Collections-Data-Structure/GetakeyfromvaluewithanHashMap.htm

Boolean_Model.java

- infixToPostfix function:
    - http://interactivepython.org/runestone/static/pythonds/BasicDS/InfixPrefixandPostfixExpressions.html
    - https://www.geeksforgeeks.org/stack-set-2-infix-to-postfix/
- Prec function: https://www.geeksforgeeks.org/stack-set-2-infix-to-postfix/
- postfixEval function: http://interactivepython.org/runestone/static/pythonds/BasicDS/InfixPrefixandPostfixExpressions.html
- performBooleanOperation function: http://interactivepython.org/runestone/static/pythonds/BasicDS/InfixPrefixandPostfixExpressions.html
- totalNumWord function: http://www.java67.com/2016/09/3-ways-to-count-words-in-java-string.html
- isWildcard function: https://stackoverflow.com/questions/5238491/check-if-string-contains-only-letters

Description.java

- initialize function: https://stackoverflow.com/questions/1052473/scrollbars-in-jtextarea

DictionaryBuilder.java

- read_reuters function: https://stackoverflow.com/questions/29061782/java-read-txt-file-to-hashmap-split-by
- isStopWord function: https://coderanch.com/t/631347/java/Search-word-text-file

MachineLearning.java

- read_reuters function: https://stackoverflow.com/questions/29061782/java-read-txt-file-to-hashmap-split-by
- Jaccard_Similarity function: https://stackoverflow.com/questions/51113134/union-and-intersection-of-java-sets
- isStopWord function: https://coderanch.com/t/631347/java/Search-word-text-file

MainPage.java

- totalNumWord function: http://www.java67.com/2016/09/3-ways-to-count-words-in-java-string.html

Preprocessor.java

- write function: https://bukkit.org/threads/saving-loading-hashmap.56447/

PreprocessorReuters.java

- getsgmFileName function: https://stackoverflow.com/questions/1384947/java-find-txt-files-in-specified-folder
- write function: https://bukkit.org/threads/saving-loading-hashmap.56447/

Probabilistic.java

- getDocIDs function: http://www.java2s.com/Code/Java/Collections-Data-Structure/GetakeyfromvaluewithanHashMap.htm
- getScore function: http://www.java2s.com/Code/Java/Collections-Data-Structure/GetakeyfromvaluewithanHashMap.htm
- readFile function: https://www.geeksforgeeks.org/different-ways-reading-text-file-java/
- writeToFile function: https://stackoverflow.com/questions/26188532/iterate-through-nested-hashmap
- isqueryMemoryExist function: https://alvinalexander.com/java/java-file-exists-directory-exists

QueryPreprocessor.java

- isStopWord function: https://coderanch.com/t/631347/java/Search-word-text-file

Thesaurus.java

- Jaccard_Similarity function: https://stackoverflow.com/questions/51113134/union-and-intersection-of-java-sets
- getKeyFromValue function: http://www.java2s.com/Code/Java/Collections-Data-Structure/GetakeyfromvaluewithanHashMap.htm
- getDocIDs function: http://www.java2s.com/Code/Java/Collections-Data-Structure/GetakeyfromvaluewithanHashMap.htm
- getScore function: http://www.java2s.com/Code/Java/Collections-Data-Structure/GetakeyfromvaluewithanHashMap.htm

VisualizeThesaurus.java

- N/A

VSM.java

- getKeyFromValue function: http://www.java2s.com/Code/Java/Collections-Data-Structure/GetakeyfromvaluewithanHashMap.htm
- getDocIDs function: http://www.java2s.com/Code/Java/Collections-Data-Structure/GetakeyfromvaluewithanHashMap.htm
- getScore function: http://www.java2s.com/Code/Java/Collections-Data-Structure/GetakeyfromvaluewithanHashMap.htm