

Report File for Assignment 1

Question 1:

Preprocessing is done in 5 steps.

1. Lowercasing: It is simply done using `string.lower()` function
2. Tokenising: It is done using `TweetTokenizer()` function from the `nlTK` library. First we remove all the patterns that we do not want to tokenize and then tokenize the remaining words
3. Stopwords: Stop words are simply removed using a for loop. We get all the stop words from the `nlTK` library using `nlTK.download('stopwords')`.
4. Punctuations: Punctuation marks are removed using `string.punctuation()`.
5. Blank words: Blank words are removed using `strip()` function

Printing the original and the preprocessed text after each step for the first 5 files.

```
pre processing file 1
original text for file 1
Loving these vintage springs on my vintage strat. They have a good tension and great stability. If you are floating your bridge and want the most out of your springs than these are the way to go.
Lowercase text for file 1
loving these vintage springs on my vintage strat. they have a good tension and great stability. if you are floating your bridge and want the most out of your springs than these are the way to go.
Performing tokenization for file 1
['loving', 'these', 'vintage', 'springs', 'on', 'my', 'vintage', 'strat', 'they', 'have', 'a', 'good', 'tension', 'and', 'great', 'stability', 'if', 'you', 'are', 'floating', 'your', 'bridge', 'and', 'want', 'the', 'most', 'out', 'of', 'your', 'springs', 'than', 'these', 'are', 'the', 'way', 'to', 'go']
Removing stopwords for file 1
['loving', 'vintage', 'springs', 'vintage', 'strat', 'good', 'tension', 'great', 'stability', 'floating', 'bridge', 'want', 'springs', 'way', 'go']
Removing punctuations for file 1
['loving', 'vintage', 'springs', 'vintage', 'strat', 'good', 'tension', 'great', 'stability', 'floating', 'bridge', 'want', 'springs', 'way', 'go']
Removing blank spaces for file 1
['loving', 'vintage', 'springs', 'vintage', 'strat', 'good', 'tension', 'great', 'stability', 'floating', 'bridge', 'want', 'springs', 'way', 'go']
pre processing file 2
original text for file 2
Works great as a guitar bench mat. Not rugged enough for abuse but if you take care of it, it will take care of you. Makes organization of workspace much easier because screws won't roll around. Color is good too.
Lowercase text for file 2
works great as a guitar bench mat. not rugged enough for abuse but if you take care of it, it will take care of you. makes organization of workspace much easier because screws won't roll around. color is good too.
Performing tokenization for file 2
['works', 'great', 'as', 'a', 'guitar', 'bench', 'mat', 'not', 'rugged', 'enough', 'for', 'abuse', 'but', 'if', 'you', 'take', 'care', 'of', 'it', 'it', 'will', 'take', 'care', 'of', 'you', 'makes', 'organization', 'of', 'workspace', 'much', 'easier', 'because', 'screws', 'won't', 'roll', 'around', 'color', 'is', 'good', 'too']
Removing stopwords for file 2
['works', 'great', 'guitar', 'bench', 'mat', 'rugged', 'enough', 'abuse', 'take', 'care', 'take', 'care', 'makes', 'organization', 'workspace', 'much', 'easier', 'screws', 'roll', 'around', 'color', 'good']
Removing punctuations for file 2
['works', 'great', 'guitar', 'bench', 'mat', 'rugged', 'enough', 'abuse', 'take', 'care', 'take', 'care', 'makes', 'organization', 'workspace', 'much', 'easier', 'screws', 'roll', 'around', 'color', 'good']
Removing blank spaces for file 2
['works', 'great', 'guitar', 'bench', 'mat', 'rugged', 'enough', 'abuse', 'take', 'care', 'take', 'care', 'makes', 'organization', 'workspace', 'much', 'easier', 'screws', 'roll', 'around', 'color', 'good']
pre processing file 3
original text for file 3
We use these for everything from our acoustic bass down to our ukuleles. I know there is a smaller model available for ukes, violins, etc.; we haven't yet ordered those, but these will work on smaller instruments if one doesn't extend the fret to their maximum width. They're gentle on the instruments, and the grippy material keeps them secure.
The greatest benefit has been when writing music at the computer and needing to set a guitar down to use the keyboard/mouse – just easier for me than a hanging stand.
We have several and gave one to a friend for Christmas as well. I've used mine on stage, and it folds up small enough to fit right in my gig bag.
Lowercase text for file 3
we use these for everything from our acoustic bass down to our ukuleles. i know there is a smaller model available for ukes, violins, etc.; we haven't yet ordered those, but these will work on smaller instruments if one doesn't extend the fret to their maximum width. they're gentle on the instruments, and the grippy material keeps them secure.
the greatest benefit has been when writing music at the computer and needing to set a guitar down to use the keyboard/mouse – just easier for me than a hanging stand.
we have several and gave one to a friend for christmas as well. i've used mine on stage, and it folds up small enough to fit right in my gig bag.
Performing tokenization for file 3
['we', 'use', 'these', 'for', 'everything', 'from', 'our', 'acoustic', 'bass', 'down', 'to', 'our', 'ukuleles', 'i', 'know', 'there', 'is', 'a', 'smaller', 'model', 'available', 'for', 'ukes', 'violins', 'etc', 'we', 'haven't', 'yet', 'ordered', 'those', 'but', 'these', 'will', 'work', 'on', 'smaller', 'instruments', 'if', 'one', 'doesn't', 'extend', 'the', 'fret', 'to', 'their', 'maximum', 'width', 'they're', 'gentle', 'on', 'the', 'instruments', 'and', 'the', 'grippy', 'material', 'keeps', 'them', 'secure', 'the', 'greatest', 'benefit', 'has', 'been', 'when', 'writing', 'music', 'at', 'the', 'computer', 'and', 'needing', 'to', 'set', 'a', 'guitar', 'down', 'to', 'use', 'the', 'keyboard', 'mouse', 'just', 'easier', 'for', 'me', 'than', 'a', 'hanging', 'stand', 'we', 'have', 'several', 'and', 'gave', 'one', 'to', 'a', 'friend', 'for', 'christmas', 'as', 'well', 'i've', 'used', 'mine', 'on', 'stage', 'and', 'it', 'folds', 'up', 'small', 'enough', 'to', 'fit', 'right', 'in', 'my', 'gig', 'bag']
Removing stopwords for file 3
['use', 'everything', 'acoustic', 'bass', 'ukuleles', 'know', 'smaller', 'model', 'available', 'ukes', 'violins', 'etc', 'yet', 'ordered', 'work', 'smaller', 'instruments', 'one', 'extend', 'fret', 'maximum', 'width', 'they're', 'gentle', 'instruments', 'grippy', 'material', 'keeps', 'secure', 'greatest', 'benefit', 'writing', 'music', 'computer', 'needing', 'set', 'guitar', 'use', 'keyboard', 'mouse', 'easier', 'hanging', 'stand', 'several', 'gave', 'one', 'friend', 'christmas', 'well', 'i've', 'used', 'mine', 'stage', 'folds', 'small', 'enough', 'fit', 'right', 'gig', 'bag']
Removing punctuations for file 3
['use', 'everything', 'acoustic', 'bass', 'ukuleles', 'know', 'smaller', 'model', 'available', 'ukes', 'violins', 'etc', 'yet', 'ordered', 'work', 'smaller', 'instruments', 'one', 'extend', 'fret', 'maximum', 'width', 'they're', 'gentle', 'instruments', 'grippy', 'material', 'keeps', 'secure', 'greatest', 'benefit', 'writing', 'music', 'computer', 'needing', 'set', 'guitar', 'use', 'keyboard', 'mouse', 'easier', 'hanging', 'stand', 'several', 'gave', 'one', 'friend', 'christmas', 'well', 'i've', 'used', 'mine', 'stage', 'folds', 'small', 'enough', 'fit', 'right', 'gig', 'bag']
Removing blank spaces for file 3
['use', 'everything', 'acoustic', 'bass', 'ukuleles', 'know', 'smaller', 'model', 'available', 'ukes', 'violins', 'etc', 'yet', 'ordered', 'work', 'smaller', 'instruments', 'one', 'extend', 'fret', 'maximum', 'width', 'they're', 'gentle', 'instruments', 'grippy', 'material', 'keeps', 'secure', 'greatest', 'benefit', 'writing', 'music', 'computer', 'needing', 'set', 'guitar', 'use', 'keyboard', 'mouse', 'easier', 'hanging', 'stand', 'several', 'gave', 'one', 'friend', 'christmas', 'well', 'i've', 'used', 'mine', 'stage', 'folds', 'small', 'enough', 'fit', 'right', 'gig', 'bag']
```

```

pre processing file 4
original text for file 4
Great price and good quality. It didn't quite match the radius of my sound hole but it was close enough.
Lowercase text for file 4
great price and good quality. it didn't quite match the radius of my sound hole but it was close enough.
Performing tokenization for file 4
['great', 'price', 'and', 'good', 'quality', 'it', 'didn't', 'quite', 'match', 'the', 'radius', 'of', 'my', 'sound', 'hole', 'but', 'it', 'was', 'close', 'enough']
Removing stopwords for file 4
['great', 'price', 'good', 'quality', 'quite', 'match', 'radius', 'sound', 'hole', 'close', 'enough']
Removing punctuations for file 4
['great', 'price', 'good', 'quality', 'quite', 'match', 'radius', 'sound', 'hole', 'close', 'enough']
Removing blank spaces for file 4
['great', 'price', 'good', 'quality', 'quite', 'match', 'radius', 'sound', 'hole', 'close', 'enough']
pre processing file 5
original text for file 5
I bought this bass to split time as my primary bass with my Dean Edge. This might be winning me over. The bass boost is outstanding. The active pickups really allow you to adjust to the sound you want. I recommend this for anyone. If you're a beginner like I was not too long ago, it's an excellent bass to start with. If you're on tour and/or music is making you money, this bass will be beautiful on stage. The color is a bit darker than in the picture. But, all around, this is a great buy.
Lowercase text for file 5
i bought this bass to split time as my primary bass with my dean edge. this might be winning me over. the bass boost is outstanding. the active pickups really allow you to adjust to the sound you want. i recommend this for anyone. if you're a beginner like i was not too long ago, it's an excellent bass to start with. if you're on tour and/or music is making you money, this bass will be beautiful on stage. the color is a bit darker than in the picture. but, all around, this is a great buy.
Performing tokenization for file 5
['i', 'bought', 'this', 'bass', 'to', 'split', 'time', 'as', 'my', 'primary', 'bass', 'with', 'my', 'dean', 'edge', 'this', 'might', 'be', 'winning', 'me', 'over', 'the', 'bass', 'boost', 'is', 'outstanding', 'the', 'active', 'pickups', 'really', 'allow', 'you', 'to', 'adjust', 'to', 'the', 'sound', 'you', 'want', 'i', 'recommend', 'this', 'for', 'anyone', 'if', 'you're', 'a', 'beginner', 'like', 'i', 'was', 'not', 'too', 'long', 'ago', 'it's', 'an', 'excellent', 'bass', 'to', 'start', 'with', 'if', 'you're', 'on', 'tour', 'and', 'or', 'music', 'is', 'making', 'you', 'money', 'this', 'bass', 'will', 'be', 'beautiful', 'on', 'stage', 'the', 'color', 'is', 'a', 'bit', 'darker', 'than', 'in', 'the', 'picture', 'but', 'all', 'around', 'this', 'is', 'a', 'great', 'buy']
Removing stopwords for file 5
['bought', 'bass', 'split', 'time', 'primary', 'bass', 'dean', 'edge', 'might', 'winning', 'bass', 'boost', 'outstanding', 'active', 'pickups', 'really', 'allow', 'adjust', 'sound', 'want', 'recommend', 'anyone', 'beginner', 'like', 'long', 'ago', 'excellent', 'bass', 'start', 'tour', 'music', 'making', 'money', 'bass', 'beautiful', 'stage', 'color', 'bit', 'darker', 'picture', 'around', 'great', 'buy']
Removing punctuations for file 5
['bought', 'bass', 'split', 'time', 'primary', 'bass', 'dean', 'edge', 'might', 'winning', 'bass', 'boost', 'outstanding', 'active', 'pickups', 'really', 'allow', 'adjust', 'sound', 'want', 'recommend', 'anyone', 'beginner', 'like', 'long', 'ago', 'excellent', 'bass', 'start', 'tour', 'music', 'making', 'money', 'bass', 'beautiful', 'stage', 'color', 'bit', 'darker', 'picture', 'around', 'great', 'buy']
Removing blank spaces for file 5
['bought', 'bass', 'split', 'time', 'primary', 'bass', 'dean', 'edge', 'might', 'winning', 'bass', 'boost', 'outstanding', 'active', 'pickups', 'really', 'allow', 'adjust', 'sound', 'want', 'recommend', 'anyone', 'beginner', 'like', 'long', 'ago', 'excellent', 'bass', 'start', 'tour', 'music', 'making', 'money', 'bass', 'beautiful', 'stage', 'color', 'bit', 'darker', 'picture', 'around', 'great', 'buy']

```

All the preprocessed files are then stored in a new folder and saved with the same name as the original file

Question 2

1) In question 2 a unigram inverted index is created in the same way as described in the lecture.

To create an inverted index we follow these steps:

- Initialize an empty dictionary
- Iterate over our preprocessed files.
- Then over each preprocessed files, iterate through each word in the document
- If the word is not present in the document then add the word to the dictionary and initialize a list associated with it containing the current file number
- If the word is present in the dictionary then simply append the file number in the list of that particular word

2)Then we use Python's pickle module to save and load the created unigram inverted index.

3) To provide support for AND,OR,AND NOT,OR NOT functions we use the set library provided by python.

5) After taking the query as the input we first preprocess the query in the same way as defined earlier. Then simply iterate through each word in the query, find the documents in which it is stored and perform the operation which is given as input.

Running for sample input:

a. **Input:**

2

Car bag in a canister

OR, AND NOT

Coffee brewing techniques in cookbook

AND, OR NOT, OR

```
2
Enter query: Car bag in a canister
Enter operations: OR,AND NOT
canister
Query 1 : car OR bag AND NOT canister
Number of documents retrieved for query 1 : 31
Names of the documents retrieved for query 1 : file3.txt file264.txt file780.txt file404.txt file665.txt file797.txt file542.txt file930.txt file166.txt file682.txt file174.
txt file686.txt file942.txt file313.txt file698.txt file699.txt file956.txt file573.txt file73.txt file459.txt file466.txt file981.txt file860.txt file863.txt file864.txt fi
le738.txt file746.txt file363.txt file886.txt file118.txt file892.txt
Enter query: Coffee brewing techniques in cookbook
Enter operations: AND,OR NOT,OR
techniques
cookbook
Query 2 : coffee AND brewing OR NOT techniques OR cookbook
Number of documents retrieved for query 2 : 999
Names of the documents retrieved for query 2 : file1.txt file2.txt file3.txt file4.txt file5.txt file6.txt file7.txt file8.txt file9.txt file10.txt file11.txt file12.txt fil
e13.txt file14.txt file15.txt file16.txt file17.txt file18.txt file19.txt file20.txt file21.txt file22.txt file23.txt file24.txt file25.txt file26.txt file27.txt file28.tx
t file29.txt file30.txt file31.txt file32.txt file33.txt file34.txt file35.txt file36.txt file37.txt file38.txt file39.txt file40.txt file41.txt file42.txt file43.txt file44.t
xt file45.txt file46.txt file47.txt file48.txt file49.txt file50.txt file51.txt file52.txt file53.txt file54.txt file55.txt file56.txt file57.txt file58.txt file59.txt file6
0.txt file61.txt file62.txt file63.txt file64.txt file65.txt file66.txt file67.txt file68.txt file69.txt file70.txt file71.txt file72.txt file73.txt file74.txt file75.txt fi
le76.txt file77.txt file78.txt file79.txt file80.txt file81.txt file82.txt file83.txt file84.txt file85.txt file86.txt file87.txt file88.txt file89.txt file90.txt file91.txt
file92.txt file93.txt file94.txt file95.txt file96.txt file97.txt file98.txt file99.txt file100.txt file101.txt file102.txt file103.txt file104.txt file105.txt file106.txt
file107.txt file108.txt file109.txt file110.txt file111.txt file112.txt file113.txt file114.txt file115.txt file116.txt file117.txt file118.txt file119.txt file120.txt file1
21.txt file122.txt file123.txt file124.txt file125.txt file126.txt file127.txt file128.txt file129.txt file130.txt file131.txt file132.txt file133.txt file134.txt file135.tx
t file136.txt file137.txt file138.txt file139.txt file140.txt file141.txt file142.txt file143.txt file144.txt file145.txt file146.txt file147.txt file148.txt file149.txt fil
e150.txt file151.txt file152.txt file153.txt file154.txt file155.txt file156.txt file157.txt file158.txt file159.txt file160.txt file161.txt file162.txt file163.txt file164.
txt file165.txt file166.txt file167.txt file168.txt file169.txt file170.txt file171.txt file172.txt file173.txt file174.txt file175.txt file176.txt file177.txt file178.txt f
ile179.txt file180.txt file181.txt file182.txt file183.txt file184.txt file185.txt file186.txt file187.txt file188.txt file189.txt file190.txt file191.txt file192.txt file19
3.txt file194.txt file195.txt file196.txt file197.txt file198.txt file199.txt file200.txt file201.txt file202.txt file203.txt file204.txt file205.txt file206.txt file207.txt
file208.txt file209.txt file210.txt file211.txt file212.txt file213.txt file214.txt file215.txt file216.txt file217.txt file218.txt file219.txt file220.txt file221.txt file
222.txt file223.txt file224.txt file225.txt file226.txt file227.txt file228.txt file229.txt file230.txt file231.txt file232.txt file233.txt file234.txt file235.txt file236.t
xt file237.txt file238.txt file239.txt file240.txt file241.txt file242.txt file243.txt file244.txt file245.txt file246.txt file247.txt file248.txt file249.txt file250.txt fi
le251.txt file252.txt file253.txt file254.txt file255.txt file256.txt file257.txt file258.txt file259.txt file260.txt file261.txt file262.txt file263.txt file264.txt file265
.txt file266.txt file267.txt file268.txt file269.txt file270.txt file271.txt file272.txt file273.txt file274.txt file275.txt file276.txt file277.txt file278.txt file279.txt
file280.txt file281.txt file282.txt file283.txt file284.txt file285.txt file286.txt file287.txt file288.txt file289.txt file290.txt file291.txt file292.txt file293.txt file2
94.txt file295.txt file296.txt file297.txt file298.txt file299.txt file300.txt file301.txt file302.txt file303.txt file304.txt file305.txt file306.txt file307.txt file308.tx
t file309.txt file310.txt file311.txt file312.txt file313.txt file314.txt file315.txt file316.txt file317.txt file318.txt file319.txt file320.txt file321.txt file322.txt fil
e323.txt file324.txt file325.txt file326.txt file327.txt file328.txt file329.txt file330.txt file331.txt file332.txt file333.txt file334.txt file335.txt file336.txt file337.
txt file338.txt file339.txt file340.txt file341.txt file342.txt file343.txt file344.txt file345.txt file346.txt file347.txt file348.txt file349.txt file350.txt file351.txt f
ile352.txt file353.txt file354.txt file355.txt file356.txt file357.txt file358.txt file359.txt file360.txt file361.txt file362.txt file363.txt file364.txt file365.txt file36
6.txt file367.txt file368.txt file369.txt file370.txt file371.txt file372.txt file373.txt file374.txt file375.txt file376.txt file377.txt file378.txt file379.txt file380.txt
file381.txt file382.txt file383.txt file384.txt file385.txt file386.txt file387.txt file388.txt file389.txt file390.txt file391.txt file392.txt file393.txt file394.txt file
395.txt file396.txt file397.txt file398.txt file399.txt file400.txt file401.txt file402.txt file403.txt file404.txt file405.txt file406.txt file407.txt file408.txt file409.t
xt file410.txt file411.txt file412.txt file413.txt file414.txt file415.txt file416.txt file417.txt file418.txt file419.txt file420.txt file421.txt file422.txt file423.txt fi
le424.txt file425.txt file426.txt file427.txt file428.txt file429.txt file430.txt file431.txt file432.txt file433.txt file434.txt file435.txt file436.txt file437.txt file438
.txt file439.txt file440.txt file441.txt file442.txt file443.txt file444.txt file445.txt file446.txt file447.txt file448.txt file449.txt file450.txt file451.txt file452.txt
file453.txt file454.txt file455.txt file456.txt file457.txt file458.txt file459.txt file460.txt file461.txt file462.txt file463.txt file464.txt file465.txt file466.txt file4
67.txt file468.txt file469.txt file470.txt file471.txt file472.txt file473.txt file474.txt file475.txt file476.txt file477.txt file478.txt file479.txt file480.txt file481.tx
t file482.txt file483.txt file484.txt file485.txt file486.txt file487.txt file488.txt file489.txt file490.txt file491.txt file492.txt file493.txt file494.txt file495.txt fil
```

```
2.txt file713.txt file714.txt file715.txt file716.txt file717.txt file718.txt file719.txt file720.txt file721.txt file722.txt file723.txt file724.txt file725.txt file726.txt
file727.txt file728.txt file729.txt file730.txt file731.txt file732.txt file733.txt file734.txt file735.txt file736.txt file737.txt file738.txt file739.txt file740.txt file
741.txt file742.txt file743.txt file744.txt file745.txt file746.txt file747.txt file748.txt file749.txt file750.txt file751.txt file752.txt file753.txt file754.txt file755.t
xt file756.txt file757.txt file758.txt file759.txt file760.txt file761.txt file762.txt file763.txt file764.txt file765.txt file766.txt file767.txt file768.txt file769.txt fi
le770.txt file771.txt file772.txt file773.txt file774.txt file775.txt file776.txt file777.txt file778.txt file779.txt file780.txt file781.txt file782.txt file783.txt file784
.txt file785.txt file786.txt file787.txt file788.txt file789.txt file790.txt file791.txt file792.txt file793.txt file794.txt file795.txt file796.txt file797.txt file798.txt
file799.txt file800.txt file801.txt file802.txt file803.txt file804.txt file805.txt file806.txt file807.txt file808.txt file809.txt file810.txt file811.txt file812.txt file8
13.txt file814.txt file815.txt file816.txt file817.txt file818.txt file819.txt file820.txt file821.txt file822.txt file823.txt file824.txt file825.txt file826.txt file827.tx
t file828.txt file829.txt file830.txt file831.txt file832.txt file833.txt file834.txt file835.txt file836.txt file837.txt file838.txt file839.txt file840.txt file841.txt fil
e842.txt file843.txt file844.txt file845.txt file846.txt file847.txt file848.txt file849.txt file850.txt file851.txt file852.txt file853.txt file854.txt file855.txt file856.
txt file857.txt file858.txt file859.txt file860.txt file861.txt file862.txt file863.txt file864.txt file865.txt file866.txt file867.txt file868.txt file869.txt file870.txt f
ile871.txt file872.txt file873.txt file874.txt file875.txt file876.txt file877.txt file878.txt file879.txt file880.txt file881.txt file882.txt file883.txt file884.txt file88
5.txt file886.txt file887.txt file888.txt file889.txt file890.txt file891.txt file892.txt file893.txt file894.txt file895.txt file896.txt file897.txt file898.txt file899.txt
file900.txt file901.txt file902.txt file903.txt file904.txt file905.txt file906.txt file907.txt file908.txt file909.txt file910.txt file911.txt file912.txt file913.txt file
914.txt file915.txt file916.txt file917.txt file918.txt file919.txt file920.txt file921.txt file922.txt file923.txt file924.txt file925.txt file926.txt file927.txt file928.t
xt file929.txt file930.txt file931.txt file932.txt file933.txt file934.txt file935.txt file936.txt file937.txt file938.txt file939.txt file940.txt file941.txt file942.txt fi
le943.txt file944.txt file945.txt file946.txt file947.txt file948.txt file949.txt file950.txt file951.txt file952.txt file953.txt file954.txt file955.txt file956.txt file957
.txt file958.txt file959.txt file960.txt file961.txt file962.txt file963.txt file964.txt file965.txt file966.txt file967.txt file968.txt file969.txt file970.txt file971.txt
file972.txt file973.txt file974.txt file975.txt file976.txt file977.txt file978.txt file979.txt file980.txt file981.txt file982.txt file983.txt file984.txt file985.txt file9
86.txt file987.txt file988.txt file989.txt file990.txt file991.txt file992.txt file993.txt file994.txt file995.txt file996.txt file997.txt file998.txt file999.txt
vansh Gandhi | Vanshs-MacBook-Air-4 Assignment1 % █
```

Question 3.

- 1) To create a positional index we follow the same procedure as mentioned in the slides. I have created a dictionary of the same format as given in the slides to store the positional index.

<term, number of docs containing **term**;

doc1: position1, position2 ... ;

doc2: position1, position2 ... ;

etc.>

This is done in the following steps

- Define a dictionary. Dictionary will of this format: {word: [frequency, {fileNo.: [position1, position2...]}]}
 - Iterate over each preprocessed file
 - Iterate over each word of the file, if the word is not present in the dictionary then add the word long with the list [1, {docID: [Position]}] where 1 is the frequency of the word
 - If the word is present in the dictionary then update the frequency by one. If the doc Id is not present in the sub dictionary then create a new docId along with a list consisting of the current position, if it is present in the sub dictionary then just append the position number in the respected fileID.
- 2) Dump the dictionary using pickle module.
 - 3) For query parsing we first preprocess the query. Then over each word in the query we find the potential files, which are the files in which all the words of the query are present. After finding the potential files we check in which all files the words in the query are placed one after the other.

Running the code for sample test case:

Input

2

Car bag in a canister

Coffee brewing techniques in cookbook

```
2
Enter query 1: Car bag in a canister
Number of documents retrieved for query1using positional index: 0
Names of documents retrieved for query1using positional index:
Enter query 2: Coffee brewing techniques in cookbook
Number of documents retrieved for query2using positional index: 0
Names of documents retrieved for query2using positional index:
vanshgandhi@Vanshs-MacBook-Air-4 Assignment1 %
```

Running the code over a few more examples

```
3
Enter query 1: vintage springs
Number of documents retrieved for query1using positional index: 1
Names of documents retrieved for query1using positional index: file1.txt
Enter query 2: youtube videos
Number of documents retrieved for query2using positional index: 2
Names of documents retrieved for query2using positional index: file851.txt file965.txt
Enter query 3: I am youtube have videos
Number of documents retrieved for query3using positional index: 2
Names of documents retrieved for query3using positional index: file851.txt file965.txt
vanshgandhi@Vanshs-MacBook-Air-4 Assignment1 %
```