

To run this java program, the user must have the Java SDK installed on the computer.

To make sure the command prompt can run the java compiler (javac) and program

- First, find where your java SDK is located, an example of where it can be found is **C:\Program Files\Java\jdk1.8.0\_241\bin**  
Keep this in mind for later
- Then go to Control Panel > System and Security > System  
On the left side click Advanced system settings, then Environment Variables.
- If there is a Path variable already there, click on it then edit. Once the new window opens click New and copy/paste the location of the Java SDK location there.
  - If there is no Path variable, you can create it by clicking New, and entering Path for the variable name, and the SDK location for the value.

Once in the command prompt, use 'cd' to change directories to reach the location where the java files are (SearchEngine.java, ItemIndex.java, and Porter.java should be in the same location).

The user should first enter 'javac SearchEngine.java' to compile with the java compiler, then the user can enter the command as shown below. The flags can be used in any order.

The command line used: java **SearchEngine** -CorpusDir **PathOfDir** -PremadeInvertedIndex **NameOfPremadeInvertedIndex** -Queries **QueryFile** -Results **ResultsFile** -Locations **YesOrNo** -InvertedIndex **NameOfIndexFile** -StopList **NameOfStopListFile** -Stemming **YesOrNo** -

- CorpusDir & **PathOfDir** (REQUIRED) - Enter the absolute path to the chosen PathOfDir where the corpus is stored, where all the HTML files and the StopList should be located. The same PathOfDir will hold the InvertedIndex/Results text files after the program has finished running.
- PremadeInvertedIndex & **NameOfPremadeInvertedIndex** (OPTIONAL) - If included, the program will read the text file of the premade inverted index and use it for the queries.
  - This means the program will ignore the InvertedIndex, StopList, and Stemming flags because it will not create an inverted index.
- Queries & **QueryFile** (REQUIRED) - Enter the name of the text file within the same directory. The text file should be formatted like this per line:
  - Query <term>
  - Frequency <term>
- Results & **ResultsFile** (OPTIONAL) - If included, the results of the queries on the InvertedIndex will be saved to the name given, otherwise it will be by default Results.txt
  - Results are in the format:  
<filename> <frequency> (If Frequency is the query) <Locations> (If Locations is YES)

- The locations are an array of the lines where the <term> appears, if the html file was opened in a text editor.
- **Locations & YesOrNo** (OPTIONAL) - If “YES” is entered, then the lines of where each word appears will be printed in the results. If “NO” is entered or the Locations flag is not included, no locations will be printed in the results.
  - If anything other than “YES” is entered, it will act as if “NO” was entered.
- **InvertedIndex & NameOfIndexFile** (OPTIONAL) - If included, the InvertedIndex will be saved to the name given, otherwise it will be by default InvertexIndex.txt.
  - Will only work if PremadeInvertedIndex is not provided, because it does not need to create an inverted index.
- **StopList & NameOfStopListFile** (OPTIONAL) - If included, the named text file will be the stop list, otherwise there will be no stop list to filter out words during parsing.
  - Will only work if PremadeInvertedIndex is not provided, because it does not need to parse the stoplist .
- **Stemming & YesOrNo** (OPTIONAL) - If “YES” is entered, then the inverted index will use Porter’s Stemming Algorithm. If “NO” is entered or the Stemming flag is not included, the stemming algorithm will not be used.
  - Will only work if PremadeInvertedIndex is not provided, because it does not need to create an inverted index.
  - Porter’s Stemming Algorithm will turn every <term> into LowerCase, so “Get” will become “get” in the InvertedIndex, along with applying the stemming algorithm to shorten some words.
- **NOTE:** Each flag can only be used once, if multiple of the same flags are inputted then the program will not run.
- **NOTE:** If NameOfIndexFile & ResultsFile already exists in the CorpusDir, it will be overwritten by the new one.
- **NOTE:** The term(s) in the Queries.txt will be case-sensitive when searching the InvertedIndex. This means that “Get” is different from “get”.
- **Example:** java SearchEngine -CorpusDir  
C:\Users\23617034\eclipse-workspace\IRPhase1\Corpus -StopList StopList.txt -Queries  
Queries.txt
  - InvertedIndex and Results will be saved to InvertedIndex.txt and Results.txt respectively by default because not spoken

## Images

In these example images, only 15 HTML files are being parsed for the InvertedIndex to save time/space.

```
C:\Users\23617034\eclipse-workspace\IRPhase1\src>java SearchEngine -CorpusDir C:\Users\23617034\eclipse-workspace\IRPhase1\Corpus -StopList StopList.txt -Queries Queries.txt -Locations YES
Results was not provided, defaulting to Results.txt
invertedIndexFile was not provided, building an InvertedIndex from the CorpusDir
invertedIndexFileName was not provided, defaulting to InvertedIndex.txt
Porter's Stemming Algorithm will not be implemented.
| Parsing 10\208 |
Parsing finished, 15 HTML files have been parsed
Saved InvertedIndex to C:\Users\23617034\eclipse-workspace\IRPhase1\Corpus\InvertedIndex.txt
Saved Results to C:\Users\23617034\eclipse-workspace\IRPhase1\Corpus\Results.txt
```

**InvertedIndex - Notepad**

```
File Edit Format View Help
10 Baking Tips for Perfect Cakes _ Sally's Baking Addiction.html AFFILIATE 1 [835]
10 Baking Tips for Perfect Cakes _ Sally's Baking Addiction.html Ads 1 [2444]
10 Baking Tips for Perfect Cakes _ Sally's Baking Addiction.html Akismet 1 [1965]
10 Baking Tips for Perfect Cakes _ Sally's Baking Addiction.html Also 1 [1366]
10 Baking Tips for Perfect Cakes _ Sally's Baking Addiction.html Another 1 [1655]
10 Baking Tips for Perfect Cakes _ Sally's Baking Addiction.html Any 1 [1856]
10 Baking Tips for Perfect Cakes _ Sally's Baking Addiction.html April 8 [2014, 2026, 2038, 2050, 2062, 2074, 2086, 2098]
10 Baking Tips for Perfect Cakes _ Sally's Baking Addiction.html Arthur 1 [1410]
10 Baking Tips for Perfect Cakes _ Sally's Baking Addiction.html August 8 [2010, 2022, 2034, 2046, 2058, 2070, 2082, 2094]
10 Baking Tips for Perfect Cakes _ Sally's Baking Addiction.html BEST 1 [1788]
10 Baking Tips for Perfect Cakes _ Sally's Baking Addiction.html Baking 11 [15, 731, 832, 1611, 2166, 2167, 2209, 2314, 238]
10 Baking Tips for Perfect Cakes _ Sally's Baking Addiction.html Basically 1 [1655]
```

**Queries - Notepad**

```
File Edit Format View Help
Query Get
Frequency Get
```

**Results - Notepad**

```
File Edit Format View Help
Query Get
10 Baking Tips for Perfect Cakes _ Sally's Baking Addiction.html [1028]
100 Best Movies of All Time That You Should Watch Immediately.html [1123, 15301, 15622, 16151, 16302]
100 Best Movies on Netflix Right Now (March 2020).html [3310]
25+ Imaginative Gifts Inspired by Your Favorite Famous Artists.html [818]
30 Modern & Contemporary Artists - Artland Magazine.html [234]
312 Famous Artists And Their Studios _ Bored Panda.html [454]
4 Ways to Bake a Cake - wikiHow.html [1503, 1770, 3285, 3453, 3676]
4 Ways to Cure a Fever at Home - wikiHow.html [2913, 2931, 3099]
42 maps that explain World War II - Vox.html [1856, 1949]
50 Best Movies of 2019 - Top 2019 Movies List of the Year.html [1962, 3494]
Frequency Get
10 Baking Tips for Perfect Cakes _ Sally's Baking Addiction.html 1 [1028]
100 Best Movies of All Time That You Should Watch Immediately.html 5 [1123, 15301, 15622, 16151, 16302]
100 Best Movies on Netflix Right Now (March 2020).html 1 [3310]
25+ Imaginative Gifts Inspired by Your Favorite Famous Artists.html 1 [818]
30 Modern & Contemporary Artists - Artland Magazine.html 1 [234]
312 Famous Artists And Their Studios _ Bored Panda.html 1 [454]
4 Ways to Bake a Cake - wikiHow.html 5 [1503, 1770, 3285, 3453, 3676]
4 Ways to Cure a Fever at Home - wikiHow.html 3 [2913, 2931, 3099]
42 maps that explain World War II - Vox.html 2 [1856, 1949]
50 Best Movies of 2019 - Top 2019 Movies List of the Year.html 2 [1962, 3494]
```

## PremadeInvertedIndex

```
C:\Users\23617034\eclipse-workspace\IRPhase1\src>java SearchEngine -CorpusDir C:\Users\23617034\eclipse-workspace\IRPhase1\Corpus -PremadeInvertedIndex InvertedIndex.txt -Queries Queries.txt -Locations YES
Results was not provided, defaulting to Results.txt
| Parsed 10% of PremadeInvertedIndex |Done parsing PremadeInvertedIndex
Saved Results to C:\Users\23617034\eclipse-workspace\IRPhase1\Corpus\Results.txt
```

## PremadeInvertedIndex without Locations in the Results

```
C:\Users\23617034\eclipse-workspace\IRPhase1\src>java SearchEngine -CorpusDir C:\Users\23617034\eclipse-workspace\IRPhase1\Corpus -PremadeInvertedIndex InvertedIndex.txt -Queries Queries.txt
Results was not provided, defaulting to Results.txt
Locations will not be provided.
| Parsed 10% of PremadeInvertedIndex |Done parsing PremadeInvertedIndex
Saved Results to C:\Users\23617034\eclipse-workspace\IRPhase1\Corpus\Results.txt
```

Results - Notepad

File Edit Format View Help

Query Get

10 Baking Tips for Perfect Cakes \_ Sally's Baking Addiction.html  
100 Best Movies of All Time That You Should Watch Immediately.html  
100 Best Movies on Netflix Right Now (March 2020).html  
25+ Imaginative Gifts Inspired by Your Favorite Famous Artists.html  
30 Modern & Contemporary Artists - Artland Magazine.html  
312 Famous Artists And Their Studios \_ Bored Panda.html  
4 Ways to Bake a Cake - wikiHow.html  
4 Ways to Cure a Fever at Home - wikiHow.html  
42 maps that explain World War II - Vox.html  
50 Best Movies of 2019 - Top 2019 Movies List of the Year.html  
Frequency Get  
10 Baking Tips for Perfect Cakes \_ Sally's Baking Addiction.html 1  
100 Best Movies of All Time That You Should Watch Immediately.html 5  
100 Best Movies on Netflix Right Now (March 2020).html 1  
25+ Imaginative Gifts Inspired by Your Favorite Famous Artists.html 1  
30 Modern & Contemporary Artists - Artland Magazine.html 1  
312 Famous Artists And Their Studios \_ Bored Panda.html 1  
4 Ways to Bake a Cake - wikiHow.html 5  
4 Ways to Cure a Fever at Home - wikiHow.html 3  
42 maps that explain World War II - Vox.html 2  
50 Best Movies of 2019 - Top 2019 Movies List of the Year.html 2