

Wynnter Ong
Phase 1 - Search Engine

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 (both SearchEngine.java and ItemIndex.java should be in the same location).

The user can enter the command as shown below

The command line used: java **SearchEngine** -CorpusDir **PathOfDir** -InvertedIndex

NameOfIndexFile -StopList **NameOfStopListFile** -Queries **QueryFile** -Results **ResultsFile**

- 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 in. The same PathOfDir will hold the InvertedIndex/Results text files after the program has finished running.
- InvertedIndex & **NameOfIndexFile** (OPTIONAL) - If included, the InvertedIndex will be saved to the name given, otherwise it will be by default InvertexIndex.txt.
- 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.
- 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
- **NOTE:** Each flag can only be used once, if multiple of the same flags are inputted then the program will not run.
- **NOTE:** NameOfIndexFile & ResultsFile cannot already exist in the PathOfDir, to successfully save the results those files must be deleted
- 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

```
C:\Users\23617034\eclipse-workspace\IRPhase1\src>java SearchEngine -CorpusDir C:\Users\23617034
\eclipse-workspace\IRPhase1\Corpus -StopList StopList.txt -Queries Queries.txt
invertedIndex was not provided, defaulting to InvertedIndex.txt
Results was not provided, defaulting to Results.txt
| Parsing 10\203 | | Parsing 20\203 | | Parsing 30\203 |
| Parsing 40\203 | | Parsing 50\203 | | Parsing 60\203 |
| Parsing 70\203 | | Parsing 80\203 | | Parsing 90\203 |
| Parsing 100\203 | | Parsing 110\203 | | Parsing 120\203 |
| Parsing 130\203 | | Parsing 140\203 | | Parsing 150\203 |
| Parsing 160\203 | | Parsing 170\203 | | Parsing 180\203 |
| Parsing 190\203 | | Parsing 200\203 |
Parsing finished, 200 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
```



Queries - Notepad

File Edit Format View Help

Query France

Frequency France



Results - Notepad

File Edit Format View Help

Query France

France Travel Guide by Rick Steves.html

France travel guide worldtravelguide.html

The 30 best video games of the decade, ranked - CNET.html

VIVE™ _ Discover Virtual Reality Beyond Imagination.html

Virtual Reality - CNET.html

Frequency France

France Travel Guide by Rick Steves.html 1

France travel guide worldtravelguide.html 1

The 30 best video games of the decade, ranked - CNET.html 1

VIVE™ _ Discover Virtual Reality Beyond Imagination.html 1

Virtual Reality - CNET.html 1