**The guide for the user and the system admin**

Information Retrieval final project

Bar Arbel 308276666

Miryam Nudelman 206262438

**Admin Read Me:**

The password for the Admin zone is: 1234

In order to add files to the system you first need to press the "Choose Files" button and select the files you wish to add.

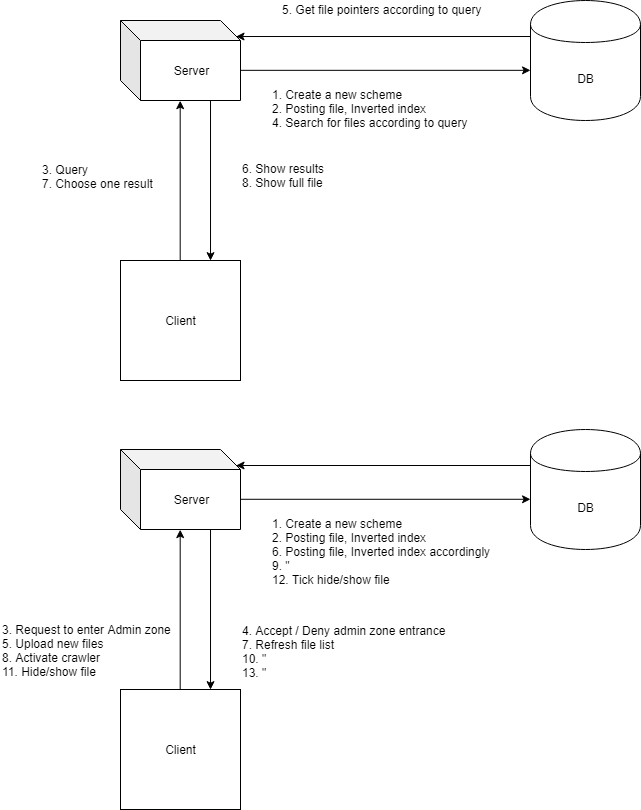
After the selection press "Submit" and the adding and indexing process will begin.

Alternatively, you can enter a number in the field next to the "Run Crawler" button and then press the button to activate the crawler.

The crawler will then extract information from the web and will generate a number of files based on the input, later it will start the adding and indexing process for those files.

In order to enable/disable files you need to select the file you wish to toggle in the show/hide list.

Admin flow:



**The files structure:**

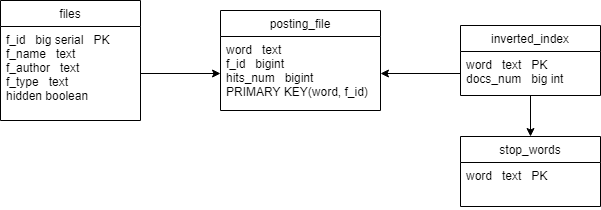
The system indexes text files with the following format:

The file's name holds the name of the song/book/etc.

The first line holds information about the text – first the name of the author/song writer and then the type of the document: song/book/letter/etc. The two are separated by a comma.

And then comes the text itself.

**The database and tables:**



**The program structure:**

The program was made using python and utilizes a database based on PostgresSQL. The GUI was made using HTML, CSS and JS.  
The system works with a web GUI on local files.

The GUI is built from 4 pages:

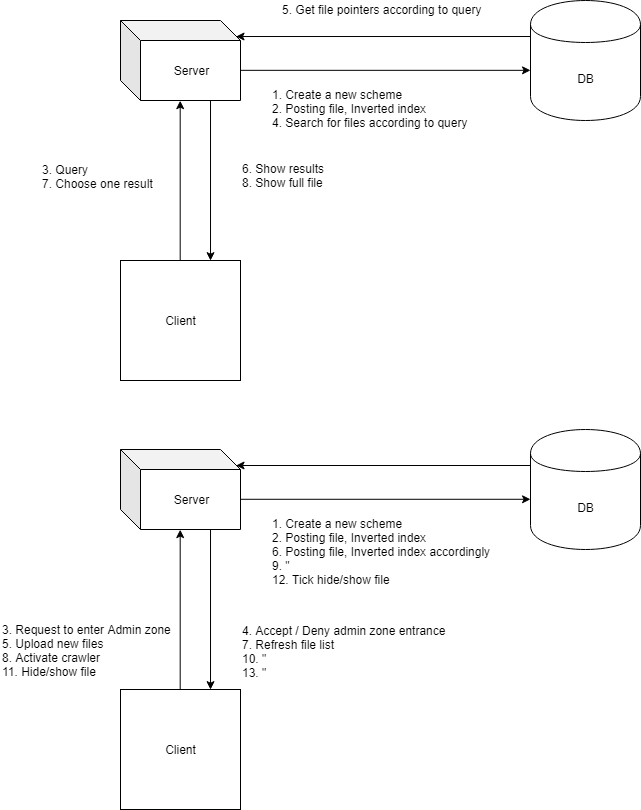
The main page – the user area, allows for searching the DB using a Boolean expression.

The result page – after the Boolean expression is processed by the system the files are shown to the user

The file page – a file can be selected from the list of files that are shown in the result page to be shown in full.

The admin page – allows for adding of new files to the system and enabling/disabling existing files in the system.

User flow:



The main program is using a number of modules:

The add module – this module is responsible for adding new files into the system, updating the DB, and indexing.

The crawler module – this module implements a web crawler which creates new files in the correct structure for the system to index.

The search module – this module deals with processing the user built Boolean expression and creating SQL queries in order to retrieve the correct files from the index.

The dbtools module – initializes the SQL database using python and SQL code.

The app module – using Flask, creates a connection between the functions built in python and the web GUI, JS functions and users input.