

1) How does the program work? (very similar approach like in IR-P01!)

The program uses the Apache Lucene 4.6 library to create an index of a given document folder. The document folder will be analyzed by the `listFileTree` function. That is, the program searches for suitable file endings like “.txt” and “.html” only and it considers also containing subfolders of the document folder. All suitable files are saved to a `ArrayList files` variable so that the indexing process can be started.

To create the index we have to use an `IndexWriter` that requires the location of the index, in this case we save it to memory (RAM), and a configuration object. Last one is set to the standard analyzer which uses the standard tokenizer and stemmer. To add documents to the index we need a document object and have to determine the document's fields. There are different fields, such as `TextFields` or `StoredFields`, with different purposes. First one is used only for analyzing the content and get certain terms, here we don't save the whole content. In our case we use it for the plain content of our documents, because we have the original documents on storage and know their paths so we would be able to open them. `StoredFields` are not be analyzed and therefore are not added to the index, but the content itself is stored. The program uses some file object properties like `fileName` (path) and `lastModTime` (last modified) as `StoredFields`, so that we are able to give these contents back to the user if we want.

Until now, the program creates the index and prints out the amount of indexed files. In the next step the user inputs the search result limit and the search query. This query string will be transformed with help of a query builder. We get the most relevant documents back and print their properties to the list view control.

2) How can I use it?

```
java -jar IR-P02.jar
```

IR_P02 - Information Retrieval System (by Andy Koch)

1. 2. total files indexed 11..

document folder: /media/andy/bigdata/study/DKE-OvGU-Master/Information Retrieval/WS1617/assignments/ap1/data

search results: 3.

search query: 4. 5.

/media/andy/bigdata/study/DKE-OvGU-Master/Information Retrieval/WS1617/assignments/ap1/data/subfolder2/3-s2.0-B9780123704528500268-main.txt
/media/andy/bigdata/study/DKE-OvGU-Master/Information Retrieval/WS1617/assignments/ap1/data/3-s2.0-B9780123704528500256-main.txt
/media/andy/bigdata/study/DKE-OvGU-Master/Information Retrieval/WS1617/assignments/ap1/data/subfolder1/3-s2.0-B978012370452850027X-main.txt
/media/andy/bigdata/study/DKE-OvGU-Master/Information Retrieval/WS1617/assignments/ap1/data/3-s2.0-B9780123704528500281-main.txt
/media/andy/bigdata/study/DKE-OvGU-Master/Information Retrieval/WS1617/assignments/ap1/data/subfolder1/3-s2.0-B978012370452850044X-main.txt
/media/andy/bigdata/study/DKE-OvGU-Master/Information Retrieval/WS1617/assignments/ap1/data/subfolder2/3-s2.0-B978012370452850030X-main.txt

rank: 2 last modified: 28.11.2016 17:15:19
score: 0.5122

1. Press the button to set your document folder, only folders are chooseable (Now, you can see that the document folder text label was set to your choice)
2. Press the button to build the index, only *.txt and *.html files are processed (The amount of indexed files will be shown in the left standing text label)
3. In the search result text field you can determine the limit of search results (the default value is 20)
4. In this text field you have to write your search query
5. Press the button to start the search process, the results are shown in the list view. You can select a result line from the list view and get further information (the rank, score and last modification date) that will be presented under the list view control

Warning! At the moment these steps are not secured by the program w.r.t. the order of execution, so the user have to folllow this execution order strictly!!!

3) Known Issues/Problems

- as mentioned above, the program does not lead the user through the single program steps (e.g. some button could be disabled until other steps are executed and so on)
- a repeating search process (without to close the program) with a lower search result limit is not shown correctly in the list view control (sry it's my first JavaFX program)
- this program requires Java 8 and JavaFX
- the source code has to be refactored strongly and some code snippets have to be commented! (sry I've a lack of time)