

The Tenth International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA 2018), May 20 - 24, 2018 – Nice/France

Tutorial: How to build a Search-Engine with Common Unix-Tools Andreas Schmidt

Exercise III

This exercise continues exercise II, by formulating a query against the previously build index.

1. Formulate a query: Get all documents, containing the terms “anna” and “martinez”. As ranking criteria use the sum of occurrences of the search terms. Show the document path of the matching documents by decreasing relevance. (use `join`, `awk`, `sort`). The result of the query should look like this:

```
d:/data/bbcsport/tennis/084.txt      4
d:/data/bbcsport/tennis/056.txt      2
```

2. Hardstuff: Write a bash-script that takes a number of terms as arguments and returns the documents, which contain all the terms, ordered by the number of matches in the document.

You can use the following script (`query-template.sh`) as a starting point:

```
$ cat query-template.sh
#!/usr/bin/bash

echo " -- $1 --"
for word in ${@:2}; do
    echo " -- $word -- "
done

$ sh query-template.sh anna martinez thailand
-- anna --
-- martinez --
-- thailand --
```