Universität Regensburg

Helge Knüttel
University Library of Regensburg
D-93042 Regensburg, Germany
helge.knuettel@ur.de

Command line tools for the expert searcher: Some applied library carpentry

Introduction

Medical librarians are dealing a lot with text data when developing search strategies, handling search results and documenting the search process. Dedicated software such as reference managers as wells as general word processors are usually employed in these tasks. Yet, a lot of manual work remains and many functions wanted are not or not well supported by these programs. Classic command line tools could be candidates for easier, semi-automated workflows.

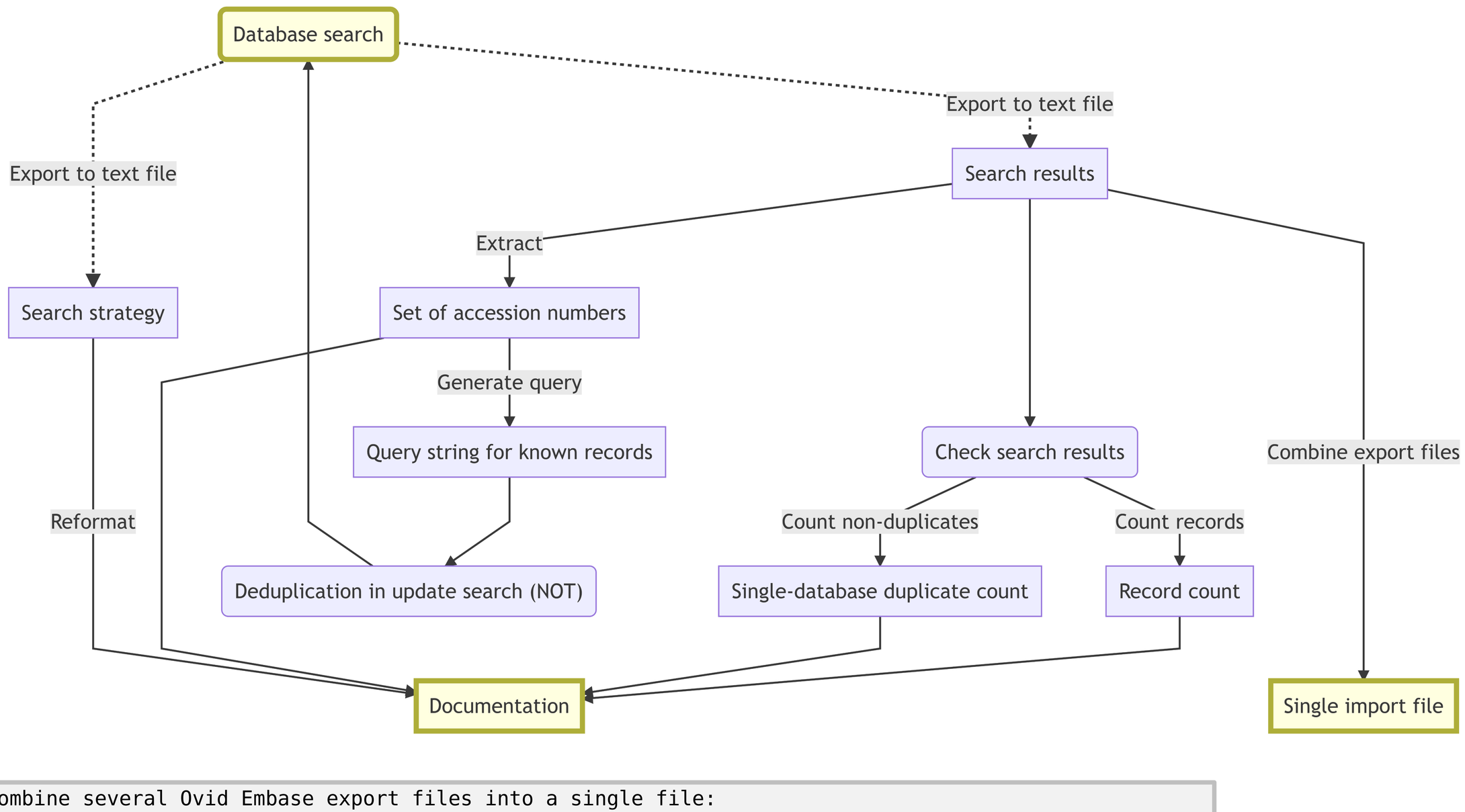
What is the command line?

What do I mean by the "command line" here? Two things, actually: One of the shell programs commonly used with Unix-like operating systems that provide the command line interface used to interact with the computer (e.g. bash) and a basic set of programs expected to exist on every such system. In particular, these are the the GNU core utilities (Coreutils) and the non-interactive text editor sed.

Conclusions

Command line tools are descendants of a decades-long tradition of Unix-like operating systems. The philosophy of these tools is to have simple programs that "should do one thing well" and then combine these programs to accomplish a specific task. Already a limited set of commands and their combination provided a number of practical tools for the expert searcher that helped to identify and avoid errors, to save time and to achieve tasks that would have been impossible otherwise. On the other side, there is some investment in learning to work with the command line.
Besides these practical considerations, there also is a matter of philosophy. Currently, the work of the expert searcher/information specialist often seems to be tied closely to a range of GUI-based software products that shape the way one works and thinks. Getting some basic understanding of command line tools may save from many of the repetitive tasks that computers were built for and concentrate on the things that (currently) only people can do. Generally, this will enhance ones skill set and may introduce a new way of thinking about software. Which may make you happier and more confident and set you on a path that will lead to new ideas (more automation, new analyses).

Where the command line was helpful:




```
# Combine several Ovid Embase export files into a single file:
cat myproject_EMBASE_2018-12-13_r*-*.ovd > myproject_EMBASE_2018-12-13_records-combined.ovd

# Count records in an Ovid Embase export file:
grep --count "^DB  - Embase" myproject_EMBASE_2018-12-13_records-combined.ovd
  3831

# Count unique records in an Ovid Embase export file (there are single-database duplicates):
grep "^UI  - " myproject_EMBASE_2018-12-13_records-combined.ovd | sort | uniq | wc –l
  3813

# Extract Embase accession numbers from Ovid Embase export file into a text file:
grep "^UI  - " myproject_EMBASE_2018-12-13_records-combined.ovd | sed -e 's/^UI  - // -e 's/\r//g' >
     myproject_EMBASE_2018-12-13_records-combined_uid.txt
```

More information with use cases and code examples on GitHub:
https://github.com/knh11545/commandline4expertsearchers

EAHIL 2020
Be Open Act Together