

Progress Report for ManBearPig

spiros thanasoulas st19@illinois.edu

November 28, 2020

Status

As outlined in the project proposal the goal of this project is to document the current status and lay the foundation for providing full text search capabilities to the mandoc system. Mandoc is a modern implementation of a manual page system <https://mandoc.bsd.lv> and part of the project is to identify its current search capabilities (how the command apropos / man -k with a keyword works) and provide a design and parts of an implementation for a full text search option on it.

Progress made so far

I have analyzed the current way that apropos accepts search terms and passes them to the current database implementation. Also i have designed an initial algorithm based on n-grams and a matching database format that could replace (or go side by side) with the current database in order to provide some basic full text search capabilities. I have also written a scraper which can consume the mandoc language and extract the words that are to be fed to the database builder.

Remaining tasks

Actually implement a database format that conforms to the apropos search api. Connect it to the apropos code and issue queries. Write tests to make sure that the words that have been inputed have matching n-grams in the correct format in the binary database. Write the report that describes the work done

Challenges faced

Being a large codebase it requires many hours of familiarizing with a part of it before you can start making sense of how to interface with it. Also being a pure C project , always requires treading carefully with things like memory management, IO etc so the progress done is slower than it would be in a completely dynamic/scripting language working with a smaller framework.

members

st19 / solo project