EZ-Bash: Designing a Metasearch Engine for Bash Command Queries

Arpit Dev Mathur, Arshdeep S Sabharwal, Ken Wang Mentor: Amirhossein Aleyasen

Motivation

Often, a new programmer in CS finds himself faced with a tar file (as an example) and he is asked to unzip it. The command to unzip it is incredibly complicated and not intuitive at all. Linux/Unix based systems have plenty of these commands that a novice would not be aware of. Our motivation is to bridge that information gap with our interface. We plan on adding this functionality into terminal so the user does not have to google for the correct command and has a set of scraped and ranked commands ready to choose from.

Linux console Interface

Project Description

The program will google the phrase on the professional coding websites (e.g. stackoverflow), where somebody probably has already asked the same question, extract the proper answer codes and give the suggested answers to users. We use several methods for validating and ranking the results (e.g. number of votes, programming language detection methods) to make sure they're appropriate results. The project can be also expanded to other programming languages (e.g. Java and Python.)

What We Have Done

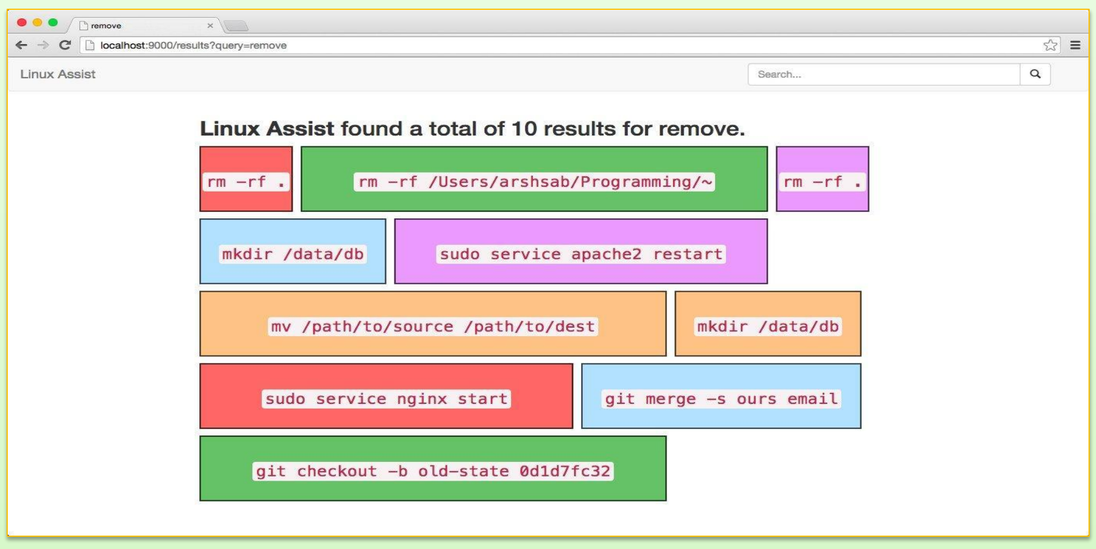
- Implemented a system to retrieve websites with possible solutions.
- Scraped the web pages to retrieve relevant code snippets.
- Filtered relevant code snippets to remove duplicates as well as redundant syntax.
- Developed console-based and web-based user Interfaces

Ranking Results

We used a classifier to rank results. The classifier uses the following features and some manual training to rank results:

- length of command
- votes given to command
- position on page
- number of commands on page
- language of command

Web Interface



- Web interface allows the user to search for command without installing the software
- Uses the same backend logic as the command line interface to generate and rank results.
- Built using the play framework, jQuery, isotope.js, and bootstrap in order to create and format webpages.

Technology Stack

To give our search tool persistency of data beyond system shutdown and restart, we implemented a MongoDB database to store relevant data. The data we stored included:

- Sources : The source websites that include Linux commands like cyberciti.biz, stackoverflow.com
- Code Selectors : The selectors we used to parse webpages.

CS @ ILLINOIS

P.U.R.E. Promoting Undergraduate Research in Engineering

Rockwell Collins

CAT