# Commad Line Research Tool for Linux

Pranava Swaroopa
Clemson University - ICAR
Dept. of Automotive Engineering
Greenville, SC, USA
pswaroo@clemson.edu

Jacob Schechter
Clemson University
School of Computing
Clemson, SC, USA
jcschec@clemson.edu

Jacob Webb
Clemson University
School of Computing
Clemson, SC, USA
jawebb@clemson.edu

Manning Graham
Clemson University
School of Computing
Clemson, SC, USA
mcgraha@clemson.edu

## ABSTRACT

Researching on the internet is the first step to understanding a new topic and also when needing further information regarding one. Researching on the internet involves visiting several websites and collecting information which can take hours. A student needs to gather information about new topics every other day, and the process is repetitive. Wikipedia[4] is an internet source where a basic understanding of every topic can be found. Our tool, which has both a command line and GUI, helps quickly gather information from Wikipedia with additional reference external links with a click of a button. This provides an easy and efficient way to start researching a topic which also allows further understanding through additional links.

## 1 INTRODUCTION

Research predominantly involves going through pre-existing literature, and the internet is the first place in the modern world to search for information. With the vast amount of information available, it is sometimes difficult to get the right information, and Wikipedia is one such source where the information[7] is hosted for free. As noted in [7], topics in science and technology usually contain credible information. Google also uses Wikipedia to serve information about popular topics. The tool we developed gives access to this information from the Linux terminal.

### 1.1 Background

Research can often be tedious and difficult. Due to Wikipedia's community editing nature, the reliability of pages on Wikipedia is often questionable. As a result, we decided to design a program that not only gathers information from Wikipedia's page but also displays and ranks the credible references cited by Wikipedia. Searching from the command line is a popular requirement as we can see from the number of tools available in GitHub. There exist implementations for searching Stack Overflow[2] where questions can be typed in the terminal, and answers can be searched right from the terminal. In[1], we can search GitHub repos from the terminal. These tools increase productivity by creatively using command line tools to improve development speed.

As our research was often needed while working in a Linux environment, we created both a command line tool and a Linux application to handle these searches. By containing our program in the Linux environment, we have streamlined the research process.

## 2 DESIGN

In this project, we designed and developed a command line tool and a GUI(Graphical User Interface). Both were implemented using Python language, as the library support and development ease is high in Python.

### 2.1 Comman Line Tool

The command line tool was implemented using the 'argparse' library. To install the command line tool, a user must clone our GitHub repository[3] and run the command "sudo pip3 install .". They will then have access to all commands available. The user also has several options, like receiving only information regarding a topic, a summary, and a list of references and allowing querying results in a specific language. Additionally, the user may query for content on the reference pages of the chosen Wikipedia page. Depending on the type of search used, the user must enter the corresponding search flag, followed by their chosen topic. If they want to query the reference page content, they must add another flag and their query. The results are displayed on the terminal and in an output text file. Available flags are shown in 2.

### 2.2 GUI

The GUI is developed on top of the code for the command line tool. It implements the same methods in a user-friendly GUI. The different options are implemented as a drop-down menu where users can specify the query type.

In design, we made the GUI as simple as possible so that the users could understand the application's functions easily. The application was implemented using the 'Tkinter' GUI library. The user would enter the desired word into the search bar1 and select the type of results they would receive on the screen. Upon submitting their entry, the application would execute the command line tool with

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
*Fall Class Project'22, November 2022, Clemson,SC, USA*
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

the specified parameters through the' Subprocess' library. These parameters would act within the running executable subprocess in the same manner as previously stated in the command line tool implementation. Finalizing its execution through the export of its findings into a text box within the application, providing the user with a friendly viewing format in which they can review their results.

## 3 ANALYSIS

Our application analyzed the data pulled from the websites by separating this data from actual information found by a searched topic, and the sources used to provide this information. Our CLI options were designed to separate this data. Using option -i (information), all information on a subject would be provided and displayed on the stdout. Using option -q (quick summary), a quick summary of the information and all relevant sources would be provided and displayed on the stdout. Lastly, using -s (similar), we could provide information and sources similar to the input keyword. Overall, the data analysis was done in a very organized and efficient manner within our application.

## 4 FUTURE WORK

We worked to develop a solution that could quickly gather data from the internet for an optimized research strategy. We used Wikipedia as a test website where information regarding a large number of topics is available for free. While Wikipedia tries to require authors to include external references for credibility, it is known in the academic world that the information available lacks credibility[6]. But the information can give a good understanding for the user to delve further into the topic.

Similarly, we can adapt the tool to scrape references from a research article and summarize their content. This would help all researchers as research articles are usually vetted for the information presented by journals, and reading through hundreds of references is time-confusing. New Natural Language Processing (NLP) models that are developed to summarize[5] a research article will benefit from this kind of tool which will make it easy for a researcher to understand the breadth of a topic.

## ACKNOWLEDGMENTS

In this project, all authors contributed equally to conceptualizing the project and deciding the major steps and goals. Jacob Webb and Manning Graham are undergraduates who developed the command line tool. Jacob Shecter and Pranava Swaroopa, graduate students, worked on implementing the GUI as an additional feature for their graduate-level requirement. Everyone was involved in preparing and presenting in front of the class and the final report.

## REFERENCES

[1] feinoujc/gh-search-cli: searches github via command line. supports repo, issues, code and commit searches and listing notifications.
[2] gautamkrishnar/socli: Stack overflow command line client. search and browse stack overflow without leaving the terminal.
[3] ppswaroopa/wikicrawler: Final project for system administration and security class.
[4] Wikipedia.
[5] Rahul, Adhikar, S., and Monika. Nlp based machine learning approaches for text summarization. *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020* (3 2020), 535–538.

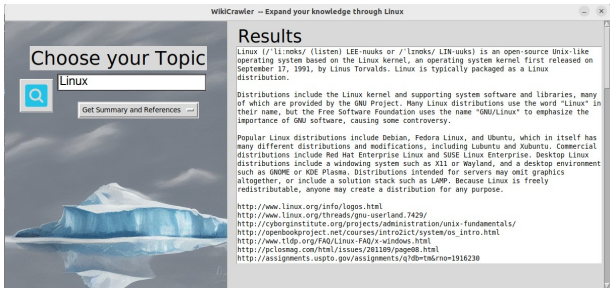


**Figure 1: GUI of WikiCrawler displaying the interface and an example**

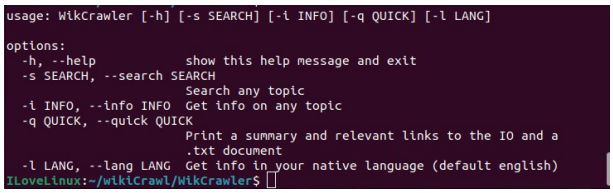


**Figure 2: Commad Line Interface of the tool displaying the syntax and options available to the user.**

[6] Waters, N. L. Why you can't cite wikipedia in my class. *Communications of the ACM 50* (9 2007), 15–17.
[7] Wilkinson, D. M., and Huberman, B. A. Cooperation and quality in wikipedia. *Proceedings of the Conference on Object-Oriented Programming Systems, Languages, and Applications, OOPSLA* (2007), 157–164.