CSC320 Project Final Report

Carson-Newman University

**Cybersecurity News Headline Scraper**

Submitted by

Jack O'Hare

jmohare@cn.edu

Computer Science

Class Instructor: SeongYong Hong

**Table of Contents**

|  |  |
| --- | --- |
| Abstract……………………………………………………………………………………... | 3 |
| Introduction………………………………………………………………………………….. | 4 |
| Planning.......................…...…………………………………………………………..……… | 5 |
| Implementing.....…….……………………………………………………………………….  Results……….....…….………………………………………………………………………  Final Thoughts...…….……………………………………………………………………….  Citations.....…….……………………………………………………………….........………. | 7  9  10  11 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Abstract**

The purpose of this project is to provide users with immediate access to news articles relating to cybersecurity. This is done by addressing the critical need for timely and relevant information in the constantly changing landscape. With the increasing number of cyber threats and the nature of technological advancements, staying informed is essential for cybersecurity professionals. By offering a streamlined and customizable tool, this project ensures that professionals are equipped with the latest news and updates; therefore, enabling them to respond proactively to emerging threats and trends.

Cybersecurity professionals operate in an environment where even a moment's delay in information can result in significant consequences. The constantly evolving nature of malware, ransomware, vulnerabilities, and data breaches requires staying on the cutting edge of industry developments. This project supports that mission by aggregating news articles through the use of predefined or user instructed keywords and RSS feeds, ensuring that only the most relevant information is presented. The ability to filter content based on specific interests or organizational needs empowers users to focus on topics that are most critical to their work.

**Keywords:** Cybersecurity, Linux, Bash, RSS

**Introduction**

The script is designed to streamline the process of finding reliable cybersecurity news by having default and customizable settings. Using predefined categories such as “cybersecurity”, “hack”, and “exploit” combined with news sites such as ESET Security Blog and CNN can result in the most up-to-date articles being recommended for the user. If the user does not like these websites or words, they are free to change them as well. This can be done in the application by selecting custom words and allowing it to be stored into the configuration preference file located in the system created by the program.

A diagram of a website

Description automatically generated

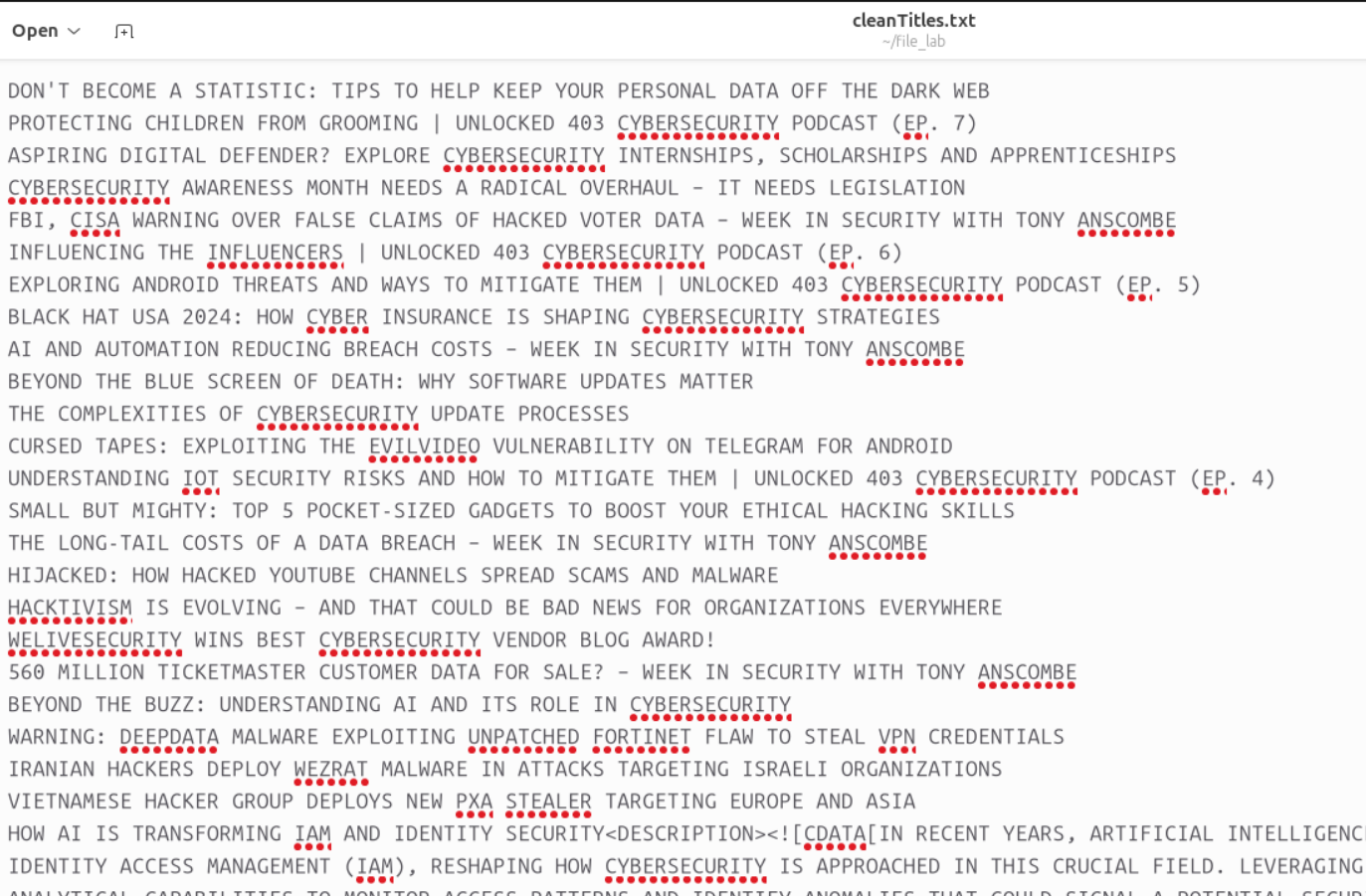
**Figure 1: Flowchart of Application**

**Planning**

As seen in figure 1, a user loads this information into the program which then begins the scraping process. The scraping process is performed once every hour to align itself with the most current topics being provided. This is performed through the use of RSS feeds. RSS feeds are a standardized way of formatting news articles, and it allows us to make them easily scrapable. It is possible to scrape these articles through the method of using the grep function inside of the program. Using these in figure 2, a user can mention they are looking for specific keywords in the titles and once found, they can be put into a text file to be referenced later for our news feed.

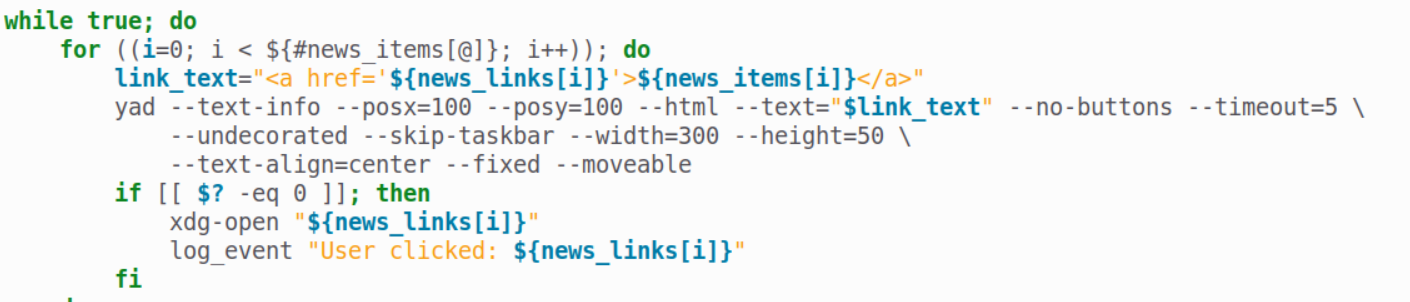


**Figure 2: Pulling Title Matches to Keywords**

**Figure 3: Pulled Titles from using GREP**

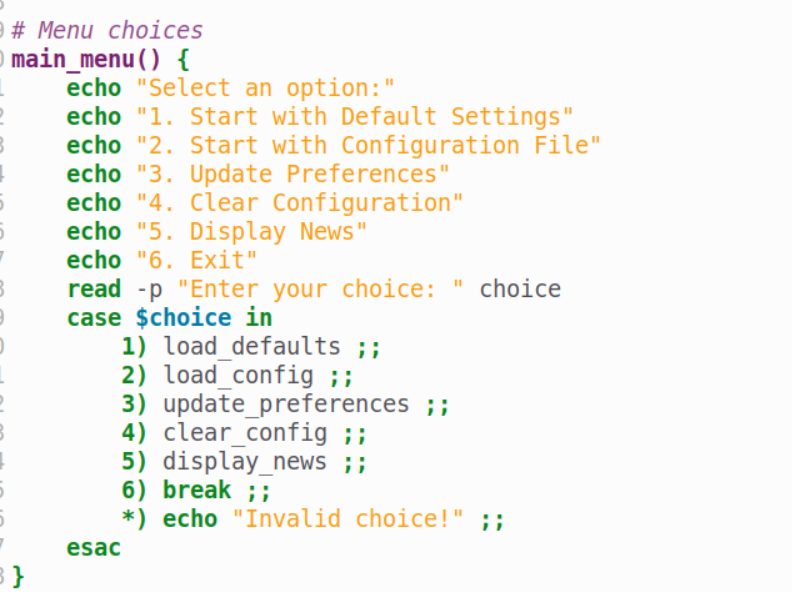
**Implementing**

Once this process is completed and we have all of the keywords inside of a text file it is time to display them to the user. This process is completed using the program YAD (Yet Another Dialog) which is a graphical user interface-based package that allows us to display a current news title and a link to access it.

**Figure 4: YAD Displaying Links**

A benefit of using YAD over other graphical interfaces is that it supports HTML tags. The reward of using such a method means that it is possible to integrate the link of the story with the actual title of the article. This is a benefit in user simplicity. A user does not have to manually copy and paste a link into a search engine. All that is needed is to click the notification and it will take you directly to the page of the article.

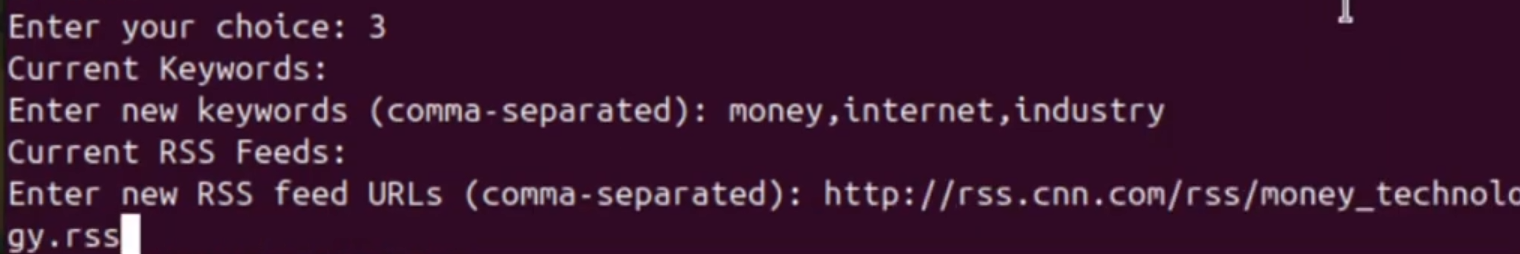
This leaves us with how we get to the point of interacting with the program. This is completed using a terminal-driven menu that allows a user to decide on how to use the program initially. A user has 5 choices, ranging from downloading news from default words, custom words, resetting the custom word configuration file, adding to the custom file, and running the news headline. This choice-based menu system allows a user to set up their headliner however they wish.

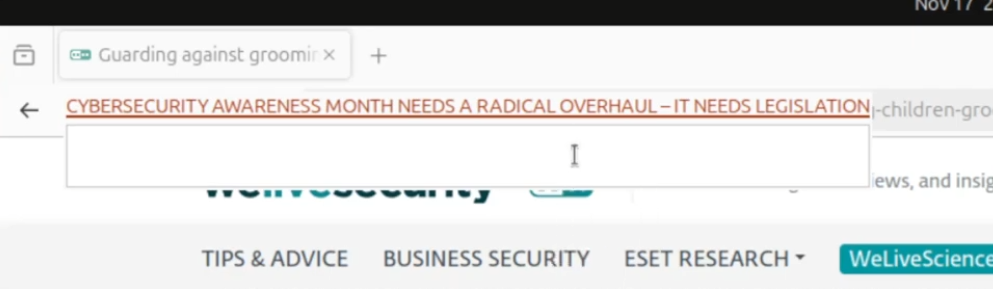


**Figure 5: Menu System**

**Results**

This system then results in a system that allows a user to configure their news however they like, and have it displayed on their system where they can click on subjects. As seen in figure 6, a user has the ability mentioned of adding customized keywords and RSS feeds that make this project robust and useful. The purpose of the project is to be accessible to cybersecurity professionals and adjust to their needs and topics, so this addition adds more ability for that user.

**Figure 6: Adding Custom Keywords and URLS to Configuration**

**Figure 7: Cybersecurity Alert End Result**

**Final Thoughts**

I believe with this project I was able to successfully make a system that was able to cover the needs of a cybersecurity specialist and keep them up to date on recent news. It is difficult to follow all news websites so being able to combine them into one streamlined place is a goal I achieved exceptionally.

There was a lot of learning about utilizing grep to pull and search through articles for keywords and how to transform that data into usable formats for the graphical display interface. A big challenge and problem faced with this project was finding modern RSS feeds as there is a mix of outdated and unsupported formats, so trying to make a scraping program that was able to identify bad quality feeds was a challenge posed during this project. Overall, however, I was able to overcome this issue and create a project that was able to give cybersecurity professionals a tool they can use to succeed.

**Citations**

* "The cURL Command Tutorial." *cURL*. <https://curl.se/docs/tutorial.html>. Accessed 17 Nov. 2024.
* "How Do RSS Feeds Work?" *RSS.com*. <https://rss.com/blog/how-do-rss-feeds-work/>. Accessed 17 Nov. 2024.
* "YAD – Yet Another Dialog – Guide." *YAD Guide*. <https://yad-guide.ingk.se/>. Accessed 17 Nov. 2024.
* "grep Command in Unix/Linux with Examples." *GeeksforGeeks*. <https://www.geeksforgeeks.org/grep-command-in-unixlinux/>. Accessed 17 Nov. 2024.