# Source Scan

CSC468-02 CloudLab Project

Sayo Ade-Oyetayo, David Desjardins, Jason Kost, Andrew Roche, Mayur Suresh

Summary

Source Scan is a program designed to streamline news-searching through reddit. It utilizes a python based Reddit API worker that scans selected subreddits based on key words submitted by the user, and stores its findings in a redis database. The javascript based WebUI quieres the redis database and displays the top reddit submissions, as well as the posts original source, to the user for their perusal. The python worker, redis database, and javascript WebUI combined create an easy to use cloud-based experience for the user to quickly receive the top reddit news as it happens.

Chapter 1: Intended Application Components
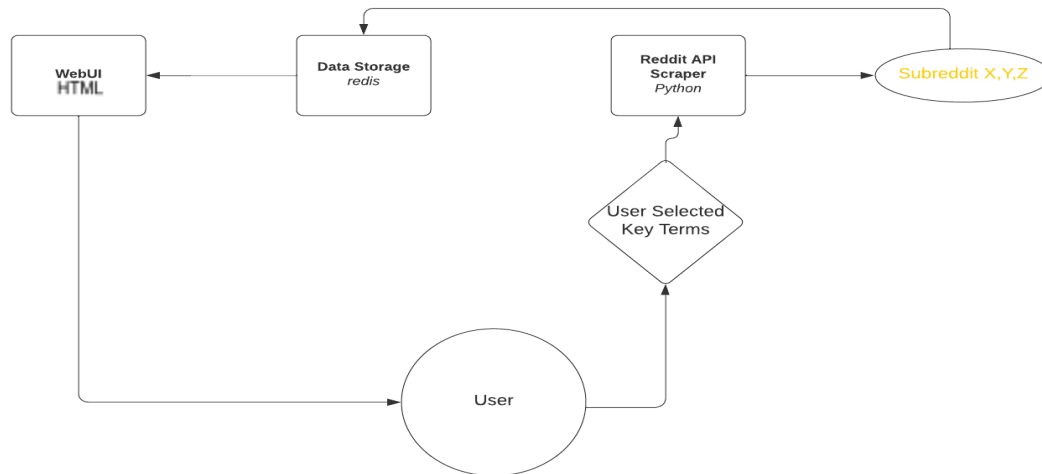
- ● User submits their selected key terms through the Reddit API
- ● Reddit API scans through selected news subreddits and pulls matches to key terms
- ● Matched terms of interest are stored in a database
- ● WebUI queries Redis and displays chosen submissions to the user

Chapter 2: Purpose and Goals

In the year 2022, the public cannot assume the news stories will originate from mainstream news sources. Information across internet sites abounds, and breaking stories often arise on online message boards directly from the source before reaching mainstream network news channels. While the public may wish to read or learn about current stories, gathering information from the internet proves a cumbersome task for an individual, given the sheer amount of news one could obtain from numerous places on the internet.

Reddit is one such board that not only receives news updates first, but often is itself the source of noteworthy trends and stories, whether they be political, social, or economic. As a place of interest, the public requires the means to readily access relevant posts, comments, and stories as they break. The proposed software solution to this problem operates in three distinct pieces – an API Reddit scraper, a Redis data storage bank, and a User Interface – with all these components we will have a Reddit bot that will take the client's request and return a set of submissions which match key terms set by the user.

The API Reddit scraper, which will be designed in the python language, takes a list of subreddits and uses keywords provided by the user to gather all pertinent information surrounding the provided topic. Given the numerous subreddits without credibility within Reddit, and given concern for avoiding fake news, the subreddits scraped by the program will be predetermined 'news source' subreddits. These selected subreddits include, but are not limited to, subreddits like r/news, r/global news, and r/politics. Instead of individually searching each subreddit and sifting through irrelevant posts which are unrelated to their desired key terms, the user is presented with only relevant submissions, from each news-based subreddit. This program will not only ease and streamline the news searching process for Reddit users, but it will also allow users to avoid potential "fake news" subreddits which may be otherwise opinionated or non-factual.

Our Redis database is designed around the concept of key terms inputted by the client and holds the gathered information after it is filtered. The database is called after the use of the scraper, and is needed for the WebUI's query, in order to return the results to the user.

Finally, the HTML user interface (WebUI) ensures the user the opportunity to filter information, while presenting them with a display that is easy to navigate and understand. The WebUI is something users can interact with, allowing them to provide key terms to the scraper. It queries the database after the scraper is deployed, returning submissions to the user that they can then view the details of.

The three-piece program aims to get Reddit news – what people are talking about relevant to what the user wishes to see. By creating a program that can identify and filter for useful information based on submitted key terms, immediately returning that resulting information into a database, and displaying that information in the database to the user, Redditors can save themselves browsing time, without compromising their news intake.

Chapter 3: Intermediate Milestones

The project has seen progress in the area of developing the python code for the worker that finds the top submissions from selected subreddits. The code can pull top posts from the selected subreddits as intended, but one challenge area surrounding the API is the filtering out of old posts. While the program looks to find popular or "hot" submissions within the subreddits, it is set up to search within a recent time frame. Active solutions to the problem are currently in the works, with one potential solution being to place limits on what exact timeframe will be selected for the actual "scan."

Another area of progress for the project has been with integrating the github into the CD/CI pipeline. The progress is mostly limited to the connection of the two, but given that the project is unfinished, the pipeline produces nothing more than an error. While some of the issue is due to the lack of a working API, other issues stem from the unfinished WebUI. The WebUI is still under development, which also impacts the Dockerfile and .Yaml file production. Dockerfiles have been uploaded to the repository, with the file for the worker being complete. The Dockerfile for the WebUI is a shell template, but uploaded as well. The .docker-compose.yaml file similarly depends on the WebUI, and is thus unfinished. The areas that depend on the WebUI are not presenting as many challenges, as they are on hold until the progress can be made. Once we do this, we will also be able to establish a connection with the Kubernetes pipeline.

Although progress was made, we have run into some shortcomings throughout the process of this project. One of the first issues that came upon our group was building the .yaml files. Confused about when to use build or image for a service, we concluded that it depends on the needs of that service and what we want it to do. In fact, we don't even perform manual builds because the CI does it for us (e.g., running docker-compose up in step 4 of the Docker Compose episode). Another reason why we are not able to completely finish off the .yaml files is because we have yet to integrate any portion of our user-interface or front-end. We are still currently facing some problems with our Reddit API. While we are able to gather top posts from multiple different subreddits about a specific topic, we have yet to find a way to filter out older posts. The Reddit API has been fairly simple to configure so far. They provide many clear-cut functionalities that can be performed via Python, making for a simple code. Because we ran into this problem so recently, we believe that if we continue our search through the numerous examples of Reddit APIs online, a solution will be found soon.

Chapter 4: Final Evaluation

        The team initially overestimated the functions of our webUI. Using nothing other than HTML to build the project interface and its templates, which can be accessed via the worker, the team concluded that containerizing proved redundant, as the webUI folder contained nothing to build. The templates within were moved into the worker container for the sake of organization.When deployed locally, the application performed acceptably. The team successfully created a list of posts from multiple subreddits based on a search from the user. The team filtered out old posts from the search query by implementing the Python module "import datetime." Given the app's intention to act as an alternative news source, the team believed the most appropriate option was to provide users with the most popular posts from that week. The team retrospectively concluded that a future implementation of new features, such as permitting user selection of timeframe, may be worth pursuing.

        One of the notable struggles facing the team arose from the team's attempt to successfully integrate a database. Originally, the design team assumed implementation would prove simple, as producing the container came with few difficulties. While the production of a container using docker compose proved simple, the team lacked any members with adequate experience in actually implementing the database with the worker. The team believes it to be beneficial for future implementations to pursue other means of database implementation, such as attempting to substitute mySQL in the place of Redis as an active database.While there is a notable absence in the number of dockerfiles, due in part to the team's late notice of the lack of necessity for a webui dockerfile - the team successfully utilized docker compose to build both of the containers. The dockerfiles built successfully and ran without error, after copious time and effort allotted for troubleshooting bugs. The major struggles came from implementing flask within the dockerfile, a problem the team eventually solved. The program, despite working locally, did not display after using docker despite a lack of errors in building and running dockerfiles.

        The issue of containerized deployment likely originated within the implementation of flask through the python code. There exist several contributing factors to the inability to solve the issue. One of the first difficulties that faced the team was its overall lack of experience using flask. The second difficulty came through a faulty translation of the formatting through css/html with flask. The greatest issue contributing to the inability to solve the problem was the team's overemphasis on a local deployment for most of the development cycle, leading to issues with using docker and a containerized deployment. The result was a "page not found" error when the containers were built and ran through docker. The team determined that the issue - still actively being studied and investigated - arose from within the team's app.py worker code, and was not related to any dockerfiles. As of writing this chapter, the team was unable to solve the issue, despite numerous attempts.

Overall, the scope of this project was not aligned to the team's initial priorities. The team placed excessive emphasis on design, and insufficient emphasis on testing deployment. The design also failed to incorporate sufficient cloud-based elements, including a limited number of dockerfiles due to design choices, and inevitably a lack of necessity of deploying through Kubernetes given the limited number of containers. The team, in retrospect, would focus on designs which included a webUI component to be containerized, and would similarly focus more effort on running attempts and troubleshooting, rather than perfecting localized deployment and design.

Adesayo Ade-Oyetayo

adesayo.oyetayo@gmail.com ·

(484)-470-3951

EDUCATION
MAY 2022

WEST CHESTER UNIVERSITY
BACHELOR OF SCIENCE IN COMPUTER SCIENCE
MINOR IN APPLIED STATISTICS
COMPUTER SECURITY CERTIFICATE

EXPERIENCE
JUNE 2014 – CURRENT
OFFICE ASSISTANT, *MARJOY VENTURES*
COLLINGDALE, PA
- Performed various office tasks, such as answering phones and faxing and printing as well as engaging with customers and their needs.

JULY 2017 – OCTOBER 2017

BRAND ASSOCIATE, *OLD NAVY*

DARBY, PA
- In charge of keeping clothing sections organized and fully stocked.
- Helped customers with their clothing questions and needs.

AUGUST 2018 – OCTOBER 2018
SALES ASSOCIATE, *WEST CHESTER UNIVERSITY BOOKSTORE*

WEST CHESTER, PA
- Worked the cash register.
- Assisted students in finding books and materials for select classes.

AUGUST 2019 – CURRENT
RESIDENT ASSISTANT, *WEST CHESTER UNIVERSITY*

WEST CHESTER, PA
- Facilitate a community atmosphere that is conducive to the academic, social, and personal development of residents.
- Enforce university policy to ensure safety and wellbeing of residents.

EXTRACURRICULAR INVOLVEMENT
JANUARY 2013- CURRENT
VOLUNTEER, *CHRIST TEMPLE WORLD OUTREACH CENTER*
DARBY, PA
- Volunteer with feeding the surrounding community.
- Assist and lead in teaching the youth and creating events centered around youth in the surrounding area.

SKILLS AND QUALIFICATIONS

- Fluent in English and Yoruba, Basic understanding in Spanish
- Trained in FISH Philosophy
- LGBTQIA+ Advocacy Training
- Mental Health First Aid Certification.
- Python
- Java
- Haskell

David Desjardins   David_A_Desjardins@mac.com
8 Hickory Drive, Glen Mills PA   610-425-8084

Education

***West Chester University**, West Chester PA*
Bachelor of Science in Computer Science
Computer Security Certificate
Expected graduation spring 2022

- 3.88 GPA
- Dean's list every semester

Skills

- Java
- C
- Python
- Linux
- Cloud computing
- Git
- Ocaml
- Assembly
- Powershell

Experience

Busser/Food Runner 06/2017 - 11/2018
AMC Dine-in Theaters
112 Wilmington Pike, West Chester, PA 19382

- Responsible for cleaning the movie theaters in between each showing as well as bringing food to people in theaters
- Promoted to busser lead after good results

Bartender/Cocktail Server
AMC Dine-in Theaters
112 Wilmington Pike, West Chester, PA 19382

- Prepared cocktails to order for guests in the theaters
- Communicated with customers at the bar and filled their orders

Cyber Security Intern 05/2021 - 08/2021
Penn Medicine Cyber Security Team
3400 Spruce Street Philadelphia, PA 19104

- Helped implement change in USB policy through Crowdstrike using java and splunk queries
- Helped find off domain windows servers using windows powershell
- Helped identify unenrolled DUO users and helped them enroll
- Shadowed many different employees to learn about their jobs
- Participated in biweekly Penn test security challenges
- Trained using SANS and tryhackme

IT temp 02/2022 - current
Bellevue Community Center
510 Duncan Rd, Wilmington, DE 19809

- Helped with various IT related projects and consultations

Jason Kost   jasonkost256@gmail.com
3 Wensel Circle, Schwenksville PA   215-859-1232

Education
West Chester University, West Chester PA
Bachelor of Science in Computer Science
Minors in Mathematics and Civic & Professional Leadership
Expected Date of Graduation May 2023
- Member of the Honors College
- Dean's List every semester

Skills
- Strong understanding of Object-Oriented Java Programming.
- Experience navigating computer systems, operating within LinuxOS, and working in C, Python, and Ocaml programming languages.

Leadership Experience
Friar's Society, West Chester University, PA
Member - *11/21-Present*
- Service organization assists local groups including but not limited to Holiday Family, Holy Trinity Church, and the Community Warehouse Project.

Work Experience
Music Service Learning, West Chester, PA
Grant Writing - *09/2021 - 12/2021*
- Collaborated with writing team to produce grant request of $5,000.
- Independently researched and produced literature review on the value of paid internships.

Waltz's Golf Farm, Limerick, PA
Manager, Batting Cages - *04/2021 – 08/2021*
- Scheduled all employees.
- Handled all daily finances. Counted and reported income.
- Operated, maintained, and repaired all machinery.
- Addressed all employee and customer concerns and disputes.
- Learned complete work independence.

Waltz's Golf Farm, Limerick, Pennsylvania
Cashier/ Food Server - *05/2018 – 11/2020*
- Managed money, required responsibility and strong attention to detail.
- Communicated with customers and addressed customer concerns.
- Maintained facility, responsible for upkeep.

Andrew Roche

306 Surrey Pl. Macungie, Pa 18062 ▪ 484-273-6830 ▪ aroche11@ptd.net

EDUCATION

(*Pursuing*) **Bachelor of Science in Computer Science**
 ➢ **Computer Security Certificate**
Major: Computer Science
Focus: Computer Security (Security Engineering)
West Chester University of Pennsylvania, *West Chester, PA*
College of Science and Mathematics
Accredited by the National Security Agency and the Accreditation Board for Engineering & Technology
College of the Sciences and Mathematics Dean's List for the 2020 Fall Semester.
Current GPA: 3.2
High School Diploma: June, 2018
Emmaus High School, *Emmaus, Pennsylvania*
Student Athletic Honors: All-Division, All-Area Team in junior and senior year of varsity football.

WORK EXPERIENCE

Facility Maintenance, May 2019- August 2019
Lower Macungie Township
Macungie, PA 18062
 ➢ Provided upkeep for all the facilities and properties owned by the township
  ○ Painted, cleaned, and mulched the playgrounds
 ➢ Infrastructure repairs
  ○ Used power tools and some machinery to fix and maintain property
  ○ Roofing/siding
  ○ Playground equipment removal
  ○ Industrial power washing
Deli Worker, March 2018- August 2018
Macungie, PA 18062
 ➢ Dealt with customer service and communicated accommodations with customers for every order.
 ➢ Helped train new employees.
 ➢ Had to work efficiently and multi-task to keep up with large customer flow.
 ➢ Prepared and cut meat/cheese slices, salads, and cooked chicken.
 ➢ Communicated and collaborated with co-workers for various job duties.
 ➢ Self-closer responsibilities
  ○ Full department sweep (cleaning everything)
  ○ Counting inventory

COMMUNITY SERVICE/ VOLUNTEER EXPERIENCE

Set-up helper/ticket seller, Citizens Fire Co. #2, Emmaus, PA (July 2016-2020)
 ➢ Helped organize budget plan
  ○ Worked on hiring best entertainers at the lowest cost.
  ○ Helped with pricing for food and games.
 ➢ Set up food stands and tents
 ➢ Assisted parking and running ticket booth

Technology Skills

Programming: Began writing with Java.

- ➢ Designed, implemented and evaluated secure network systems.
  - ○ Analyzing different kinds of attacks (Stack overflow, Format String, ROP, Return-to-libc, GOT Overwrite, etc.)
  - ○ Studied some basic defense mechanisms like ASLR and StackGuard, and how hackers can disable/avoid them.
- ➢ Object-oriented programming
- ➢ Designed and implemented algorithmic solutions to program problems and testing those solutions.
  - ○ Implemented recursion, lists, stacks, queues, linked lists, trees, hashing, searching and sorting algorithms, and the evaluation of algorithm efficiency.
- ➢ Worked with arrays and arrays of objects, sorted, and searched through arrays, recursion, and aggregate objects.
- ➢ Later introduced to C and Python.
- ➢ Familiar with assembly language for IA-32 compatible processors and how to read compiler-generated assembly language code.
- ➢ Gained an understanding of being able to manipulate the byte-representation of different data types (int, float, char…)
- ➢ Able to understand how processes are loaded into memory and memory placements of stack, heap, and code within the virtual memory, as well as the tradeoff in accessing data stored in various memory hierarchy levels.
- ➢ Able to understand how to perform various debugging and optimization techniques.
  - ○ Use of reverse engineering to study malicious software, and its legal implications.

Systems: Windows and Linux experience.

- ➢ Able to run Linux via a virtual machine, as well as an understanding of how simple network communication works within the Linux operating system.

MAYUR SURESH

SOFTWARE DEVELOPMENT

CONTACT

- 610-864-6804
- suresh.mayur.2002@gmail.com
- West Chester PA

CERTIFICATIONS

- PCEP
- COMPUTER SECURITY CERTIFICATE FROM THE DEPARTMENT OF SCIENCES AND MATHEMATICS
- COMPETITIVE PROGRAMMING COURSE
- COMPUTER SYSTEMS COURSE

PROGRAMMING LANGUAGES

- JAVA
- PYTHON
- C/C++
- POWERSHELL
- HTML/CSS/BOOTSTRAP
- JAVASCRIPT (CURRENTLY LEARNING)

SKILLS

- MULTI-TASKING
- COMMUNICATION
- TIME MANAGEMENT
- FLEXIBLE
- OPEN-MINDED
- QUICK LEARNING HABIT

EDUCATION

West Chester of Pennsylvania

2020-2023

Bachelor of Science in Computer Science

College of Sciences and Mathematics

Accelerated Master's Program

Dean's List – Fall of 2021

West Chester of Pennsylvania

EXPERIENCE

- SURI – RESEARCH PROJECT IN THE DEPARTMENT OF SCIENCES AND MATHEMATICS (PLANNING ON DOING IT THE SUMMER OF 2022)
- HTTPS://GITHUB.COM/MAYSURESH/KATTIS
- WEBSITE IN PROGRESS (BUILDING WITH HTML/CSS AND JAVASCRIPT)

2023-2024

Master of Science in Computer Science

College of Sciences and Mathematics