

Project Phase #1 Report

Table of Contents

Section	Page
Introduction	-
Introducing the team	2
Introducing the team members	-
Alan Chan	3
Muhammad Mehdi (aka. Ali)	4
Chun Cho	5
Seong Min Jeong (aka. Max)	6
Zhen Jiang	7
Share a meal	8
Personas	-
Anthropologist (Rod Harper)	9-10
Librarian (Karen Staph)	11-12
User Stories	-
Researcher User Stories	13-17
Librarian User Stories	18-19
System Design	20
Release Plan	
Phase 1	21-22
Phase 2	23-24
Phase 3	25-26
Video	-
The 2-minute video will be added to the repository separately.	

Introducing the team

Team name: Coffee Coders

Team Number: 12



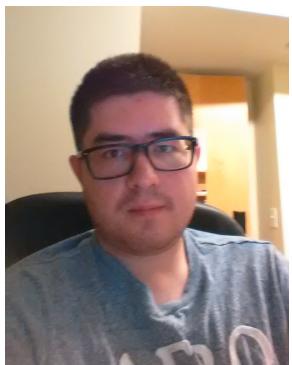
Introducing the team members

Alan Chan

My name is Alan Chan and I am a student at the Scarborough campus of University of Toronto. I have been studying computer science since grade 11 and it became my passion. Since then, I developed an interest in examining video games as an interactive art form. In my spare time, I think up ideas for various game projects, and I research intricate mechanics of game design. I have experience in Java, Python and C programming languages so I am adaptable in transferring knowledge from one skill to another. Once I start on a project, I am able to work on it for hours on end and I don't feel comfortable until the project is of optimal quality.



Muhammad Mehdi



My name is Muhammad Mehdi, and I am a third year computer science student at University of Toronto. I like to code complex mathematical algorithms in python, for example, finding the nth prime number efficiently. I also love to develop open source web technologies. I recently worked on an open source angular JavaScript plugin that localizes a web application. I also developed a web application using angular JavaScript and PhoneGap. I love working with hardware and software combined. Recently, I started working with Arduino board. Arduino is a micro controller that you can write software for, and add input sensors. I worked at Wind River in Ottawa for my first coop work-term. I was responsible for developing features for the internal testing framework. Then, I worked at BlackBerry for my second coop work-term. I was responsible for developing tools and utilities for the Localization team.

Chun Cho



My full name is Chun Hong Cho and I am a 3rd year University of Toronto student studying Computer Science. I am experienced with Java, Python, C, Unix Shell Scripting, Scheme, ML, Visual Basic, and HTML. I also enjoy creating and testing all sorts of computer programs, especially game related ones. I am very enthusiastic in programming, I could understand the concepts easily and code for hours without being tired of it. As a result, I find it very easy to learn new technologies, software and programming languages. Although many Computer Programming assignments are individual work, I really enjoy working in teams. This is because I love cooperative work. Also it provides an opportunity for me to meet new people and work together with them.



Seong Min Jeong

My name is Seong Min Jeong and I am a student at the University of Toronto Scarborough studying computer science. I choose this program because I want to develop the skills I need to be a game developer. My goal is to either work for a game company or create games with a small team of friends. In fact, I joined with a group of friends recently to work on a game project using the Unity game engine. Creating games in Unity became my new hobby as this is what I'm striving for. I have experience with a number of programming languages such as Python, Java, and C. This experience greatly improves my ability to adapt to any new programming languages,

which is essential when working with different kinds of programs. I hope that my work ethic and passion in game design will one day allow me to accomplish my goals.



Zhen Jiang

My name is Zhen Jiang. I am a third year management and technology information student at the University of Toronto Scarborough. When I learned about the features of computer science, it is advanced with technology and no end to learning. The fact that I can keep learning more is what brought me to the field of computer science and have been passionate for it since I first started. So far, I have done several projects in Python and C individually, and completed an Android application for hospital emergency system in a team of four. Outside of university, I also learn other computer science skills such as web framework on my own. I am able to pay close attention in doing one work and I always complete the project on time and efficiently. I always strive to cooperate and learn from my teammates and work together in equal terms.

Share a meal

Team Lunch



Picture of our team having lunch! We ate pizza, sausages, wraps and various other snacks.

Persona (Anthropologist)

Rod Harper



Rod Harper is a 36 old Linguistic Anthropologist who is interested in working on transnationalism and diaspora; and Israel in the Middle East as well. He is a friendly and outgoing person. Out of his work, he is delighted to share his ideas, as well as creative thoughts and arguments with his students, colleagues and friends. In addition, he is not afraid to share the adverse opinions. Harper always loves to listen to different ideas and have a deep discussion about it.

Harper received his PhD in Anthropology and Linguistics from the University of Harvard in 2008, and he is now a Senior Lecturer of Anthropology at the University of Chicago.

Harper now lives with his wife and has two children aged 2 and 6 in Chicago. He is a vegetarian. His hobby includes playing guitar and watching baseball.

The projects that Harper has done and work in progress are based on extensive ethnographic, archival and linguistic fieldwork in Israel and Palestine. The following is a list of his projects:

1. History, Settlement and Territoriality in East Jerusalem
2. Noncitizen Latinos in Israel: Domesticity and Publicity
3. Mediating Israel to a Global Public: News, Translation and Transnational Circulation

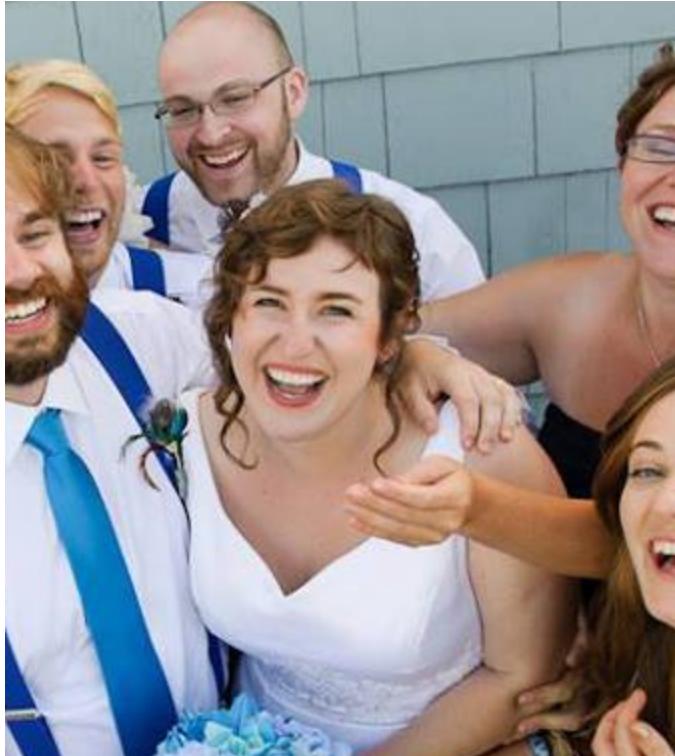
For his third project, Harper would like to use an information search system to collect all the Israeli English-language news that be used as references by one specific global

English-language news or a list of global news which he is really interested in. In order to recognize the degree of attention of collected data intuitively, Harper prefers to use a diagram to show the result.

Harper has good computer search skills as he has tried some search engines like Google and Bing but they are not efficient to make his research. He is also comfortable using applications in his smartphone.

Persona (Librarian)

Karen Staph



Karen Staph is a 35 years old librarian. She has an energetic personality who is very energetic and happy. She is service oriented, enthusiastic and idealistic about new technology and has a lot of hope for what technology is going to do for her profession and the world. Karen has three siblings, one is a 2 years younger sister and two older brothers. She is married with a son named Henry who is 16 months old and a daughter named Krissi who is 5 years old. Her hobbies are reading books and swimming.

Karen went to Dalhousie University and has a Ma, English and Literature, General and Master of Library & Information Science (MLIS). She also went to Trent University and graduated with an Honours Bachelors Degree in English/History. She comes from a humanities background as it was her hobby before becoming a librarian. Being a librarian became her new passion as she took a lot of interest in open software.

Karen is able to speak fluently in English and German. Her top skills consist of information Literacy, Library, Digital Libraries, Library Instruction, Library Science, Cataloging, Information Retrieval, Research, Electronic Resources and Digital Preservation. She also has extensive knowledge on Metadata Standards, Library Management, Social Media, Teaching and Online Research. Her wide range of skills allowed her to work in multiple libraries and have contributed to many projects including the Islandora Documentation project. The project became comprehensive for users, administrators and developers.

Her experience as a librarian and technology turned her to be more knowledgeable than an average computer user. She is able to supervise usually 5 people for her work. Karen is really into the Agile Project Management. She tries to be in the same shoes as a developer and apply techniques that are outside of software development. Her love for technology drives her to learn more like a child that always admire new technology.

User Stories

Legend:

Priority: Low, Medium, High

Estimate: 1, 2, 3 (number of weeks)

Researcher User Stories

As Alejandro (a researcher), I want the report to be accessible on the web, so that I can review the report from my home.

Priority: High

Estimate: 1

Confirmation: I can see the report on a web browser from my home.

As Alejandro (a researcher), I want to modify twitter sources one by one, so that I can monitor activities between users for my research.

Priority: High

Estimate: 2

Confirmation: Let the user type a html link and then press enter / click a button to add it to the list.

As Alejandro (a researcher), I want to be able to plot different twitter keywords, so that I can graph different keywords in a single page/graph.

Priority: High

Estimate: 2

Confirmation: I can plot the keywords onto a graph.

As Alejandro (a researcher), I want to be able to modify the news website URLs, so that I can add/delete new websites from the list one by one.

Priority: High

Estimate: 1

Confirmation: I can add new website urls, and I can also delete them.

As Alejandro (a researcher), I want to be able to view the collected data in a Venn diagram, so that I can better assess the report.

Priority: Medium

Estimate: 1

Confirmation: I can plot the results into a Venn diagram.

As Alejandro (a researcher), I want to be able to run the extractor through a webpage, so that I can run the extractor without installing any dependencies.

Priority: High

Estimate: 2

Confirmation: I can go onto the browser, and kick-off the extractor.

As Alejandro (a researcher), I want to be able to see the progress of the extractor through a webpage; so that I can track how long will the extractor take.

Priority: Medium

Estimate: 1

Confirmation: I can see the progress on my browser.

As Alejandro (a researcher), I want to get notified when the extractor is finished extracting the data from twitter and news websites, so that I can continue my other tasks while the extractor is running in the background.

Priority: Low

Estimate: 1

Confirmation: I get an email that says that the extractor is done.

As Alejandro (a researcher), I want to login as a staff, so that I can view the report on the collected data from the extractor.

Priority: High

Estimate: 2

Confirmation: I am able to login into the report using my own credentials.

As Alejandro (a researcher), I want to logout as a staff, so that my report is kept secure and I can access it again.

Priority: High

Estimate: 1

Confirmation: I am able to log off safely and securely.

As Alejandro (a researcher), I want to be able to delete older entries of the collected data, so that I can clean up the unrelated entries.

Priority: Medium

Estimate: 2

Confirmation: I can delete the older entries from the UI.

As Alejandro (a researcher), I want to be able to see the actual article itself, so that I can read it and obtain information.

Priority: Medium

Estimate: 1

Confirmation: I am able to click the link from the list, and preview the article.

As Alejandro (a researcher), I want to be able to search for a keyword within the articles to get a list of external links matching that keyword, so that I can see where the external links are heading too in the article.

Priority: Medium

Estimate: 2

Confirmation: I am able to enter keywords into a prompt and a list of articles shows up which contains those keywords

As Alejandro (a researcher), I want there to be details on the author and published date, so that I can look up the author's biography.

Priority: Low

Estimate: 1

Confirmation: This information will be displayed below the article title.

As Alejandro (a researcher), I want there to be a search function, so that I can look up a specific news article.

Priority: Medium

Estimate: 1

Confirmation: I can look up things for a specific news article through search.

As Alejandro (a researcher), I want to be able to see which category the article is in, so that I can look into the category and find similar articles.

Priority: Low

Estimate: 1

Confirmation: Category will be displayed at the top of the title.

As Alejandro (a researcher), I want to have a history of my previous searches, so that I can use them again in the future.

Priority: Low

Estimate: 1

Confirmation: Can see previous history list when searching

As Alejandro (a researcher), I want to see a paragraph to describe the final results, so that I can easily understand the variables or graphs or patterns.

Priority: Low

Estimate: 2

Confirmation: After clicking search button, the program will output a paragraph under the graph to describe the details of the collected data. The paragraph may contain all the variables that appear in the graph.

As Alejandro (a researcher), I want to see the results as soon as possible, so that I will not spend lots of time to wait for the system to display the results.

Priority: Low

Estimate: 2

Confirmation: After clicking search button, the program will display the result in 3 seconds.

As Alejandro (a researcher), I want to remove global English-language news sources in batches (multiple at once), so that I can adjust my sources efficiently for my research.

Priority: Low

Estimate: 1

Confirmation: Let the user type multiple html links (separated by a delimiter) and then press enter / click a button to add them to the list.

As Alejandro (a researcher), I want my sources to collect data on what is cited most frequently when I start a watch on my sources, so that I can make analysis on the data for my research.

Priority: Medium

Estimate: 2

Confirmation: Program should get information accordingly for every source

As Alejandro (a researcher), I want to be able to view the collected data in a line graph, so that I can make better analysis of the data and present them efficiently.

Priority: High

Estimate: 1

Confirmation: Generate a line graph to represent the collected data

As Alejandro (a researcher), I want to be able to view the collected data in a pie graph, so that I can make better analysis of the data and present them efficiently.

Priority: Medium

Estimate: 1

Confirmation: Generate a pie graph to represent the collected data

As Alejandro (a researcher), I want to be able to view the collected data in a bar graph, so that I can make better analysis of the data and present them efficiently.

Priority: Medium

Estimate: 1

Confirmation: Generate a bar graph to represent the collected data

As Alejandro (a researcher), I want to be able to sort my sources by alphabetically (descending/ascending order), so that I can collect data on matched sources for my research.

Priority: Low

Estimate: 1

Confirmation: Sort the sources alphabetically (descending/ascending order) correctly

As Alejandro (a researcher), I want to mark specific sources that from the list I have entered, so that I can monitor them more closely.

Priority: Low

Estimate: 1

Confirmation: Mark chosen sources with a flag and add it to a group so that the user can view that group again later

As Alejandro (a researcher), I receive system or email notifications from the marked sources if their activity goes above a threshold, so that I know when activities for that source goes above a certain threshold.

Priority: Low

Estimate: 2

Confirmation: Get a notification or email when activity goes above a certain threshold

As Alejandro (a researcher), I want to know how many hits something gets, so I can have data.

Priority: Low

Estimate: 1

Confirmation: Able to see the hits next to the news/twitter source

Librarian User Stories

As Kirsta (a librarian), I want the collected data to be outputted in a “best-practice” standard, so that I can maintain the data’s stability and reuse in future.

Priority: High

Estimate: 3

Confirmation: Able to collect data in the best-practice standard

As Kirsta (a librarian), I want to be able to change the format of the output data, so that I can easily export the collected data into different archives that use different format, e.g. xml, csv etc.

Priority: Medium

Estimate: 2

Confirmation: Able to format data and export in the different formats

As Kirsta (a librarian), I want to be able to login as a librarian, so that I can download the collected data in a specified format.

Priority: High

Estimate: 2

Confirmation: I can log into the system with librarian credentials

As Kirsta (a librarian), I want to be able to collect the graphs generated from the system, so that I can maintain the data stability and use it in the future.

Priority: High

Estimate: 3

Confirmation: Produce a file in best standard format which contains all the graphs

As Kirsta (a librarian), I want to have all the files associated with the project in a nice, organized format so that I can keep this project in my archives and have my programmers continue to use and adapt it.

Priority: High

Estimate: 2

Confirmation: The program will be in one folder with 3 subfolders: backend, frontend, user-interface.

As Kirsta (a librarian), I want to be able to obtain the data collected in .warc format, so that I can store it in my library system.

Priority: High

Estimate: 3

Confirmation: Generate a .warc file which contains all our data

As Kirsta (a librarian), I want to be able to look up a specific article so that I can see what it looked like in the past.

Priority: Low

Estimate: 1

Confirmation: Able to search up old articles

As Kirsta (a librarian), I want guests to have access to the data, so that they can only view the results without making any changes to the data.

Priority: Medium

Estimate: 2

Confirmation: Guest has the ability to log in with their guest credentials

System Design

Brainstorm

- Create a system that allows the user to provide a list of news/twitter sources to watch. It should be able to display graphs, extract data from the sources, and describe the relationships between the sources to the user. The user can also get notifications about sources if they are active. The system should also allow users to extract data, e.g. download the raw data from the storage.

System Metaphor

Our system is like a Private Investigator / Detective Agency. A customer of the agency works together with a detective, in the same sense that our system works together with the user to solve the problem. The main reason for a customer to approach the agency is because they want the Detective to start an investigation on the customer's sources. For example, a husband might want a Detective to follow or investigate a list of men who he thinks is possibly engaged in a relationship with his wife. He will provide a list of men as the source. This is similar to how the user will provide a list of sources for our system to watch. The Detective will find out relationships between the men and the customer's wife, also any other information to draw conclusions and report back to the customer. Our system will be providing data, showing graphs and describing relationships between the sources. The customer might also want to be notified immediately if suspicious activity is going on for one of the source. Likewise, our program will be able to notify the user when a source's activity is abnormal or high. At the end of the investigation, the customer might want a copy of all the evidence the detective has obtained. Similarly, our system will allow users to extract the data it has collected.

Naming Components/Classes:

- Source
- Graph
- Data
- Relationship
- Notification/Alert
- Report

Release Plan

Summary:

We have separating our release plan into three phases, 3 weeks per phase. Also placed the highest / most important user stories in the earlier phases, as well as top of the phase list.

Phase 1:

1. **As Alejandro (a researcher), I want to login as a staff, so that I can view the report on the collected data from the extractor.**
Priority: High
Estimate: 2
Confirmation: I am able to login into the report using my own credentials.

2. **As Kirsta (a librarian), I want to be able to login as a librarian, so that I can download the collected data in a specified format.**
Priority: High
Estimate: 2
Confirmation: I can log into the system with librarian credentials

3. **As Alejandro (a researcher), I want to logout as a staff, so that my report is kept secure and I can access it again.**
Priority: High
Estimate: 1
Confirmation: I am able to log off safely and securely.

4. **As Alejandro (a researcher), I want to be able to run the extractor through a webpage, so that I can run the extractor without installing any dependencies.**
Priority: High
Estimate: 2
Confirmation: I can go onto the browser, and kick-off the extractor.

5. **As Alejandro (a researcher), I want to be able to modify the news website URLs, so that I can add/delete new websites from the list one by one.**
Priority: High
Estimate: 1
Confirmation: I can add new website urls, and I can also delete them.

6. **As Alejandro (a researcher), I want to modify twitter sources one by one, so that I can monitor activities between users for my research.**
Priority: High
Estimate: 2
Confirmation: Let the user type a html link and then press enter / click a button to add it to the list.

7. As Alejandro (a researcher), I want the report to be accessible on the web, so that I can review the report from my home.

Priority: High

Estimate: 1

Confirmation: I can see the report on a web browser from my home.

8. As Kirsta (a librarian), I want the collected data to be outputted in a “best-practice” standard, so that I can maintain the data’s stability and reuse in future.

Priority: High

Estimate: 3

Confirmation: Able to collect data in the best-practice standard

9. As Kirsta (a librarian), I want to be able to obtain the data collected in .warc format, so that I can store it in my library system.

Priority: High

Estimate: 3

Confirmation: Generate a .warc file which contains all our data

10. As Kirsta (a librarian), I want to have all the files associated with the project in a nice, organized format so that I can keep this project in my archives and have my programmers continue to use and adapt it.

Priority: High

Estimate: 2

Confirmation: The program will be in one folder with 3 subfolders: backend, frontend, user-interface.

11. As Alejandro (a researcher), I want to be able to view the collected data in a bar graph, so that I can make better analysis of the data and present them efficiently.

Priority: Medium

Estimate: 1

Confirmation: Generate a bar graph to represent the collected data

Phase 2:

1. As Alejandro (a researcher), I want to be able to plot different twitter keywords, so that I can graph different keywords in a single page/graph.
Priority: High
Estimate: 2
Confirmation: I can plot the keywords onto a graph.

2. As Alejandro (a researcher), I want to have a history of my previous searches, so that I can use them again in the future.
Priority: Low
Estimate: 1
Confirmation: Can see previous history list when searching

3. As Kirsta (a librarian), I want to be able to collect the graphs generated from the system, so that I can maintain the data stability and use it in the future.
Priority: High
Estimate: 3
Confirmation: Produce a file in best standard format which contains all the graphs

4. As Alejandro (a researcher), I want to be able to view the collected data in a line graph, so that I can make better analysis of the data and present them efficiently.
Priority: High
Estimate: 1
Confirmation: Generate a line graph to represent the collected data

5. As Alejandro (a researcher), I want to be able to view the collected data in a pie graph, so that I can make better analysis of the data and present them efficiently.
Priority: Medium
Estimate: 1
Confirmation: Generate a pie graph to represent the collected data

6. As Alejandro (a researcher), I want my sources to collect data on what is cited most frequently when I start a watch on my sources, so that I can make analysis on the data for my research.
Priority: Medium
Estimate: 2
Confirmation: Program should get information accordingly for every source

7. As Alejandro (a researcher), I want to be able to view the collected data in a Venn diagram, so that I can better assess the report.
Priority: Medium
Estimate: 1
Confirmation: I can plot the results into a Venn diagram.

8. As Alejandro (a researcher), I want to be able to see the progress of the extractor through a webpage; so that I can track how long will the extractor take.
Priority: Medium

Estimate: 1

Confirmation: I can see the progress on my browser.

9. As Alejandro (a researcher), I want to be able to see the actual article itself, so that I can read it and obtain information.

Priority: Medium

Estimate: 1

Confirmation: I am able to click the link from the list, and preview the article.

10. As Alejandro (a researcher), I want to be able to search for a keyword within the articles to get a list of external links matching that keyword, so that I can see where the external links are heading too in the article.

Priority: Medium

Estimate: 2

Confirmation: I am able to enter keywords into a prompt and a list of articles shows up which contains those keywords

11. As Alejandro (a researcher), I want there to be a search function, so that I can look up a specific news article.

Priority: Medium

Estimate: 1

Confirmation: I can look up things for a specific news article through search.

12. As Alejandro (a researcher), I receive system or email notifications from the marked sources if their activity goes above a threshold, so that I know when activities for that source goes above a certain threshold.

Priority: Low

Estimate: 2

Confirmation: Get a notification or email when activity goes above a certain threshold

13. As Kirsta (a librarian), I want to be able to change the format of the output data, so that I can easily export the collected data into different archives that use different format, e.g. xml, csv etc.

Priority: Medium

Estimate: 2

Confirmation: Able to format data and export in the different formats

14. As Alejandro (a researcher), I want to see a paragraph to describe the final results, so that I can easily understand the variables or graphs or patterns.

Priority: Low

Estimate: 2

Confirmation: After clicking search button, the program will output a paragraph under the graph to describe the details of the collected data. The paragraph may contain all the variables that appear in the graph.

Phase 3:

1. As Alejandro (a researcher), I want to remove global English-language news sources in batches (multiple at once), so that I can adjust my sources efficiently for my research.
Priority: Low
Estimate: 1
Confirmation: Let the user type multiple html links (separated by a delimiter) and then press enter / click a button to add them to the list.

2. As Alejandro (a researcher), I want to be able to see which category the article is in, so that I can look into the category and find similar articles.
Priority: Low
Estimate: 1
Confirmation: Category will be displayed at the top of the title.

3. As Alejandro (a researcher), I want there to be details on the author and published date, so that I can look up the author's biography.
Priority: Low
Estimate: 1
Confirmation: This information will be displayed below the article title.

4. As Alejandro (a researcher), I want to get notified when the extractor is finished extracting the data from twitter and news websites, so that I can continue my other tasks while the extractor is running in the background.
Priority: Low
Estimate: 1
Confirmation: I get an email that says that the extractor is done.

5. As Alejandro (a researcher), I want to mark specific sources that from the list I have entered, so that I can monitor them more closely.
Priority: Low
Estimate: 1
Confirmation: Mark chosen sources with a flag and add it to a group so that the user can view that group again later

6. As Alejandro (a researcher), I want to be able to sort my sources by alphabetically (descending/ascending order), so that I can collect data on matched sources for my research.
Priority: Low
Estimate: 1
Confirmation: Sort the sources alphabetically (descending/ascending order) correctly

7. As Alejandro (a researcher), I want to see the results as soon as possible, so that I will not spend lots of time to wait for the system to display the results.
Priority: Low
Estimate: 2
Confirmation: After clicking search button, the program will display the result in 3 seconds.

8. As Kirsta (a librarian), I want guests to have access to the data, so that they can only view the results without making any changes to the data.

Priority: Medium

Estimate: 2

Confirmation: Guest has the ability to log in with their guest credentials

9. As Alejandro (a researcher), I want to know how many hits something gets, so I can have data.

Priority: Low

Estimate: 1

Confirmation: Able to see the hits next to the news/twitter source

10. As Kirsta (a librarian), I want to be able to look up a specific article so that I can see what it looked like in the past.

Priority: Low

Estimate: 1

Confirmation: Able to search up old articles

11. As Alejandro (a researcher), I want to be able to delete older entries of the collected data, so that I can clean up the unrelated entries.

Priority: Medium

Estimate: 2

Confirmation: I can delete the older entries from the UI.