

# Final Project Proposal: Superhero Scraper (September 2019)

Chase Brunette-Pak, Casey Lemon, Noah Branch, and Sai Manikonda

## I. OBJECTIVE

The objective of the final project, Comic Scraper, is to create an application that will search through a comic series or universe for a specific character(s) the user defines and create a biography for them based off their dialogue and body language. The application will analyze all the available dialogue that their character would say or think and will take into account that mentions the character and creates a biography for the individual. This information will be stored in the database. Our primary focus will be on gathering the information from the Detective Comics universe, due to the availability of their comics within the repositories we had found.

## II. THE MOTIVATION FOR THE PROJECT

The incentive for choosing this project was due to the general interests in comics. As readers, we all know there are times where you jump into a series and there are tie-ins from other series and universes. These tie-ins can be fellow heroes or villains that have a backstory that we missed because we have not read their story arcs. Also, it gets annoying when the characters have a conversation that references the particular tie-in's stories. To solve this, we wanted to create this application to solve that gap of knowledge of these tie-in characters. This will allow readers to view a quick bio and potentially relevant information about a character without having to look at a completely different story arc. If interests grow from these character's other story arcs readers may want to know where to start in that series. That knowledge of where to begin to find the relevant information can be a bit cumbersome. The biography though will hold the tie in character series where the reader can again begin a biography for that character within that other arc. In conclusion, the reason for the application is to help readers, like us, to be able to read our series without having to compromise more time reading other series characters that tie-in.

## III. DETAILED DISCUSSION OF DATA

There are multiple online repositories that host a variety of different comics series and universes. The data that will be used are specific series that are made from Detective Comics for this project. We will create a database of characters, based on the user's input, that holds recently created bios of said characters. These issues and characters would be listed in the user's interface to learn about.

When creating the database, we will have to find all the unique characters of each individual series and create a profile for each series. The profile will list the characters and the times they are either mentioned or spoken about. We do this because, in each series, certain characters have a new version of themselves which have their own unique traits (Ex. Batgirl being cripple verses not crippled). This allows us to ensure that we aren't bleeding in data from another issue/universe.

When the data has been scraped from all the issues, then the application will look at which specific character(s) we are to look out, and filter through the fetched data to find relevant information to build the profiles. When that is finished the profiles will be stored in the browser cache. We planned on only using a handful of cached data due to the minimum storage we have by using the user's browsers cache to store this information. This would allow the user to look at some of the previous character(s) they had searched for and not have to wait for the scraper to run again for this information.

The data would filter out eventually. If the user is constantly looking for new characters, the cache would begin to remove the older profiles to make room for the newer ones. This allows users to access only the most recent profile builds.

## IV. RESPONSIBILITIES OF EACH TEAM MEMBER

Currently, each member of the team will be taking in tasks that will be marked on our Trello board. This will help us stay organized and keep track of who is working on what at any given time. Team members will need to finish a certain number of tasks per iteration or milestone. It makes sense to have a team member working with the user interface, maybe two team members on the actual implementation of scraping information from the websites and repositories, and potentially another working with putting that information on the database. Team members will probably rotate out to work and focus on different implementations of the project. For example, the person working with the user interface may devote some time into helping with scraping. As that scraped information will be displayed on the user interface and Vice Versa. Likewise, for the person working with the database. Also, as we move along it may make more sense for one part of the project to get more attention. Again, this is why Trello will work in our favor. While each team member may have a focus, we can easily see what they are trying to achieve by using the Trello board and see where their current focus lies.

## V. A TIMELINE OF MILESTONES

- Gathering reliable repositories

- Creating a scraper to find an individual series
  - Add the feature to identify various text boxes
    - Boxes such as dialogue or inner thoughts
  - Add the feature to identify certain characters
- Create a tool to analyze characters data
  - Add the feature to understand the dialogue
    - Feelings towards other characters
    - Feelings on certain matters
  - Add the feature for picture analysis
    - Used for character physical traits
    - List known associates
    - Wardrobe style
- Create a web GUI
  - Add a feature to show the progress of the analysis
  - Connect to web's cache to store previously analyzed data

## VI. EXPECTED OUTCOME

This project is fairly ambitious to get the entire GUI up and running. The realistic expected outcome is to have the scraper be able to identify a specific character within a given series. From there, the application should be able to give known associates and personal traits from the image analysis. Varying on the amount of progress we can make from the text analysis will be able to give us things such as character personality, emotional states in the series, or feelings towards other characters.

The finished application, which may or may not be finished by the end of the term, will allow the user to search a DC character(s) within the database and pull their information. Once entered there will be a progress bar that shows the estimated time remaining for the program and scrap and create their profile(s) and store them into a cache for later usage.