

INFO 3300/5100 - Project 2 Final Report

Nikolas Lillios, Mike Phillips and Yidan Xu

J.R.R.Tolkien's works, such as The Lord of the Rings, The Hobbits and The Silmarillion, have become popular over the years and have a wide readership. Many Middle-earth enthusiasts have since created a large number of fanfictions under the influence of Tolkien's works. In this project we look at fun facts about Middle Earth themed fanfictions, such as the number of published stories and most popular characters pairings over the last decade. We also compared the trend of fanfictions based on Tolkien's work to fanfictions based on other popular novels.

Since there was no relevant dataset online, we scraped data from three major fanfiction websites: a03 (<https://archiveofourown.org>), fantasyfiction.net and LOTR wiki using Python, Selenium, and BeautifulSoup. In each website we scraped title, rating, word count, chapter count, characters, relationships from thousands of fanfiction based on Tolkien's works, then parsed it into json format for future analysis. The Python scripts used for scraping are available on Github.

We were all very curious about character pairings hence decided to display relationships in one of the graphs. We originally experimented with tidy trees and network graphs but neither capture the relationship effectively. We then were inspired by the chord diagram that was recently taught in class. Since one of the benefit of chord diagram is that "allow to visualize weighed relationships between several entities"(<https://www.data-to-viz.com/graph/chord.html>), using chord diagram for character pairings in our case allows the character relationships to be understood at a glance.

Once we had the chord diagram, we realized there were too many flows and it is not easy for users to see if two characters are connected. As a result we added interactive feature to the chord diagram: when user's mouse is on the ring of a character, we highlight this character's rings and flows, as well as characters paired with this character, and dim everything else. This effect will disappear once the user's mouse moves away. We also realized that many of the flows between the characters were so narrow and thin that it was difficult to see them even when they were highlighted. We then came up with the idea to draw a tree graph showing all the connections to a character next to the chord diagram every time user had their mouse over a ring.

Another thing we thought was interesting to discuss is the character frequency over time. We chose circular bar chart to visualize the data. We added two interactive features to the circular bar chart to help users better understand the data. The first one is a time slider. Above the bar chart there is a time slider, user can use the slider to set the year, and the circular bar chart will update its content accordingly. We originally thought about displaying 20 mini bar charts, each representing a year, or creating buttons for each year and update the graph when user clicks on the button. After we tested out it is clear that time slider is the most elegant and effective way to represent the data.

Besides the general trend provided in the bar chart we would also like the user to see the actual number of each character frequency. Therefore we decided to add a tooltip to each bar: when user's mouse rested a bar, the color of the bar changed from purple to yellow and a tooltip popped up showing the frequency of the character.

Besides internal trends within Middle-Earth themed fanfictions, we also wanted to see how popular Middle-Earth themed fanfictions are compared to fanfictions based on other themes. Therefore in the last graph we created a line chart showing the number of fanfiction from nine popular themes published over years. The graph is busier than we expected, therefore we added an interactive feature to allow user to see the name of the theme when they click on one of the lines on the graph. From the graph we can tell Middle-Earth themed fanfiction are not as popular as other themes such as Harry Potter and Twilight.

The story of our data is to portray the trend of Middle Earth themed fanfictions over years, and we found three interesting trends, one from each graph we have. The chord diagram shows us that in almost all of the character pairings, both sides of the pairing are from the same story. This may be because there is a 60 year gap between the beginning of Lord of the Rings and the end of the Hobbit, or it may be because there is more room for development between characters from the same story.

The second trend is that contrary to the conventional wisdom that middle earth is tantamount of Lord of the Rings, in fanfiction world characters from the Hobbit has become just as popular as those from Lord of the Rings in the last decade. As we can see in the bar chart of character frequency, all of the most popular characters prior to 2012 came from Lord of the Rings, while in the year 2013, four of the top five most popular characters came from the Hobbit. This could be because there was one Hobbit Trilogy movie released in 2012 and 2013 each, but even long after the movie have come and gone their popularity remains: since 2013, about half of the top ten most popular characters have come from The Hobbit.

The trend of the line chart is, as we discussed above, showing that in fanfiction world Middle-Earth theme may not be as popular as many Middle-Earth enthusiasts expected. Even at its peak in 2013, the number of Middle Earth fanfiction published is still far behind Harry Potter, Twilight and Marvel Universe and Pokemon.

Outline of Contributions

Mike :

Writing a python script to scrape data from fanfiction websites

Nikolas:

Experimenting with different visualizations/ways to display concurrent data

Assisting with final design process/script/layout

Contributing to report writing/cosmetic fixes for final design

Minor cosmetic fixes throughout the project

Yidan:

Analyzed, interpreted, and documented data into the write up

Resources:

The ideas for the individual scatter plots was based on code from:

<https://cornell.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=13123485-52ad-42b6-9baa-ad0301493c5e>

https://www.d3-graph-gallery.com/graph/circular_barplot_label.html

<https://www.twilio.com/blog/build-interactive-bar-chart-taylor-swift-lyrics>