Visualizing the Text of Philip Pullman's Trilogy "His Dark Materials"

Dr Tim Regan & Linda Becker

Article Summary

- Sreeraj Rimmalapudi

Context - The paper is trying to use the analytical and display power of the computer to enhance the reading and understanding of the text in the book. This paper aims to provide benefits to authors, readers, publishers, agents, critics, academics and academic reader who have a different understanding of the text from the author and the abstract visualization might help in visualizing these multiple theories. I believe that their contribution helps in better understanding of the book among the readers.

Contributions - I believe that the visualization that they are trying to build that helps in forming theories from the perspective of the reader and the author is interesting.

Structure - They made designs by focusing on two aspects - the difference in language of characters, its change over time and linguistic theme used in the series. They visualized the word and character distribution. Their first sketch looked at distribution of words and their connectedness. This was done to give a sense of rhythm of related characters. The second sketch was their attempt to visualize character word plot. It helped them to answer the questions - what form the data might take and what visual dimensions is required to plot differing data or to reinforce existing data. Their third sketch was to plot theme progression and lastly they used the text themselves and visualized it. They wanted to draw the structures with the words themselves instead of drawing structures based on relationship between words.

Results - It was made available to closed group of the author, publications and fans and no result were provided in the paper.

Methods - The results presented in the paper is incomplete due to the following reasons. The visualizations produced are not available to try on the web as they have not invested the development effort necessary to produce a robust and secure

implementation that would cope with multiple concurrent users. Secondly they have not explored the legal or commercial implications of placing such a visualization of an incopyright popular book series online. I believe that their work would be well received in the public and would provide a lot more result.

Technologies - The paper uses multiple technologies to build the book visualization. They use Java and Adobe Illustrator for programming. The authors uses databases to store the data. Due to multiple issues, they have not made it available in the web. In my opinion, their contribution would have scaled up well to the web, making it accessible to everyone.

Figures - Their first sketch was visualization of word distribution using character names. They took the text and highlighted the occurrence of the character's name in each chapter. Their second sketch was a character word plot with character name at the center and words around it. The third sketch was trying to describe plot progression. It tried to highlight different part of the text with different colour to depict a theme. The final sketch was a circle made of text.

The used these sketches for their final complete visualization. Their first visualization is a character flower (the center of the flower contains the character name). There are small circles, termed as life belts by the authors, around the center which tells us about the occurrence of the character throughout the series. These are connected to buds with a red line. The buds represent the words following the character name. The distance reveals the probability of the word following the character name. For ex - A word identifying the surname of the character will have small distance. The size of the buds depict the frequency of those words following the character name.

Their second visualization looks at whole text. It plots the rhythm of character occurrence by figuring out its location in the text and placing a dot next to it. These dots are connected by a line. If u back away from the text and get a bird's eye view, it will show a different detail about the character like whether they were mentioned in the book.

Confusion - The paper mentions them working on "book text" rather than books so as to notify the readers that they are not trying to build a system that tries to understand books. While it seems intuitive as to why they used this word, I feel a little disturbed when they use word distributions to visualize, since a distribution by itself might help the system understand a little about the book.

Evaluations - The paper mentions exploring the efficacy and usefulness by hosting a workshop with the author, the publishing companies and fans. The paper doesn't mention the results of those evaluations. I think a better evaluation would to deploy this online(if they worked on web version) and let the comments of the users figure out its usefulness. The paper also mentions several further works like adding machine learning to improve for better analysis of the words around the character, scaling up to multiple books and adding interactivity to their work. I believe adding other features like sentiment of the words spoken by the character and additional advancement in language processing can help with the problem that they are trying to solve. This short paper is cited by one other article.

Relevance - The visualizations presented in the paper and the contributions that they provide seems appealing enough to incorporate those ideas into a project related to books, however, our task is more generic and is not particular to a single book. It requires us to check the influence of books and its words on people and not specifically focusing on characters transitions and progression in the book. We are trying to visualize the effects the book has, while paying little detail to the text inside the book.