

Text Analytics and Natural Language Processing: Why The Lord of The Rings is so fascinating?

Daniel Torres



How tools, such as Sentiment Analysis and Natural Language Processing, can generate valuable insights that allow us to go beyond the words and focus on ideas, feelings, and emotions.

The restriction in social activities, combined with the cancellation of main sporting events, such as the Olympic Games, Football and Soccer Leagues, have made streaming services one of the main sources of entertainment worldwide.

According to a survey by Kagan, the media research unit within S&P Global Market Intelligence; the four streaming options preferred by the public are Hulu, Netflix, Disney +, and Amazon Prime Video. Among these four players, the one that stands out for its incredible growth in the last year is Amazon, reaching an astonishing 150 million subscribers, a figure close to the main player in the industry, Netflix, which has 167 million subscribers.

One of the reasons for this growth is that Amazon is betting big, investing millions to generate quality content in response to the support received by the public. One of the projects that generate the most expectation is the series based on the book written by J.R.R. Tolkien: The Lord of the Rings, which made me wonder why this story is so fascinating.

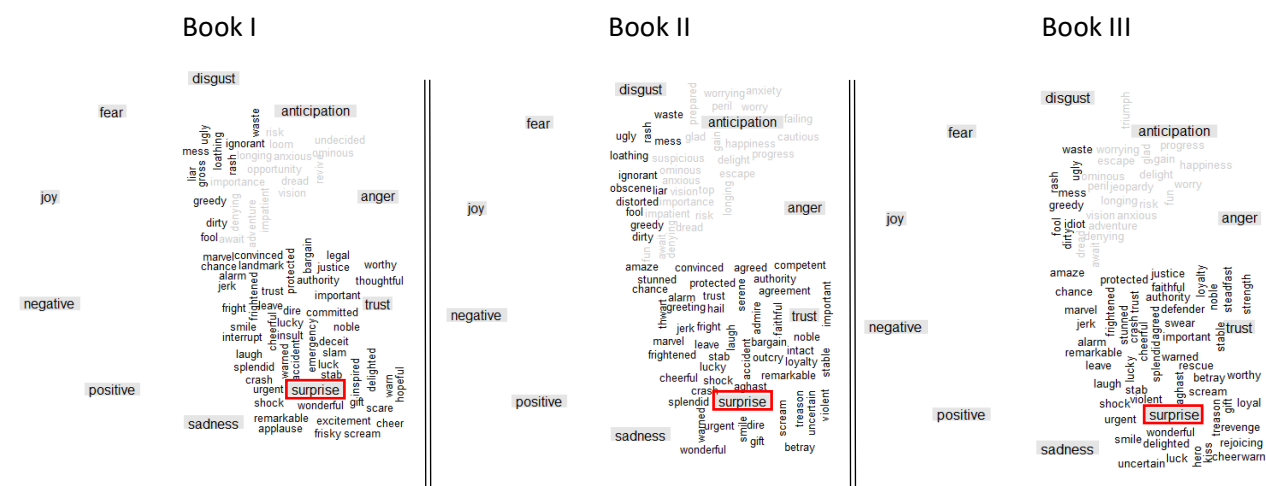
To answer this question, we can take many paths, analyze the narrative style, the characteristics of the universe created by Tolkien, the development of the characters, and more, but this analysis will always have a considerable level of subjectivity. Although it is very difficult to eliminate this subjectivity when analyzing a book or any text in general, we can reduce this aspect by taking the analysis to a more quantitative field, where the frequencies and correlations will help us to have a different approach when analyzing a text. It is in this attempt to quantify the characteristics of a book, where tools, such as Natural Language Processing (NLP) and Sentimental Analysis, come into play.

NLP is part of artificial intelligence and computer science, which helps computers extract, manipulate and understand the human language, and Sentiment Analysis is the process of determining the emotion associated with a text in order to have a better understanding of the ideas and attitudes behind the words.

In this article, I will try to answer the question posed in the title using the R programming language as well as NLP and Sentimental Analysis. Additionally, I will use three types of graphics that will allow a clearer view of the words used by the author and the feelings transmitted with these.

The first graph is made up of word clouds and aims to identify which is the most relevant feeling in each book:

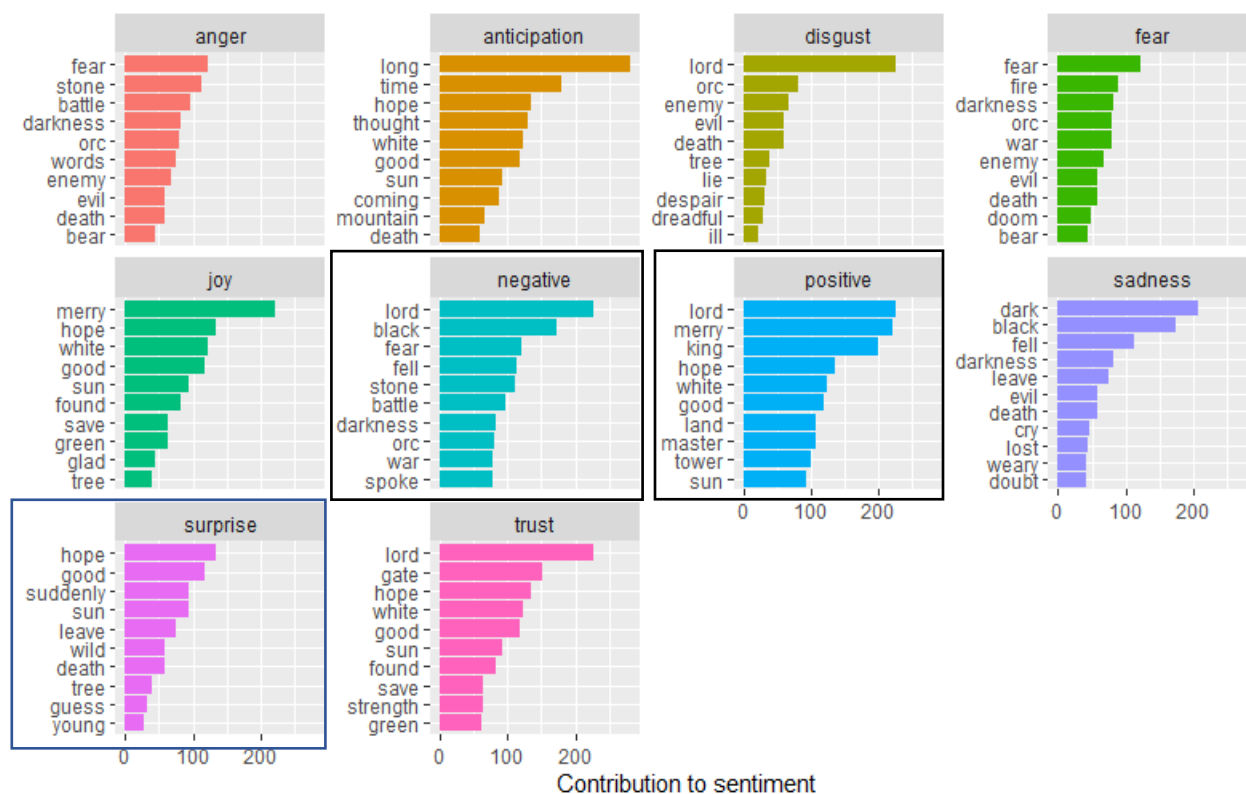
Word Clouds Plot



The three books have a similar structure based on the words associated with each feeling. Most of the relevant words in the books are related to the feelings of surprise, trust, and anticipation. Among the three feelings, the one that is surrounded by more words and, therefore, the most important, is surprise.

This sentiment is fundamental throughout the saga, but it is not the only important characteristic of the story because there are many more nuances in Tolkien's narrative; that is why the story is full of ups and downs, contrasts, mixed feelings and, of course, surprise. Hence, I decided to do a deeper analysis, quantifying the frequency of the words associated with each sentiment, for which I used bar plots and the 'NRC' database, which is available in the Tidytext package. The result is shown in the following graph:

Bar plots

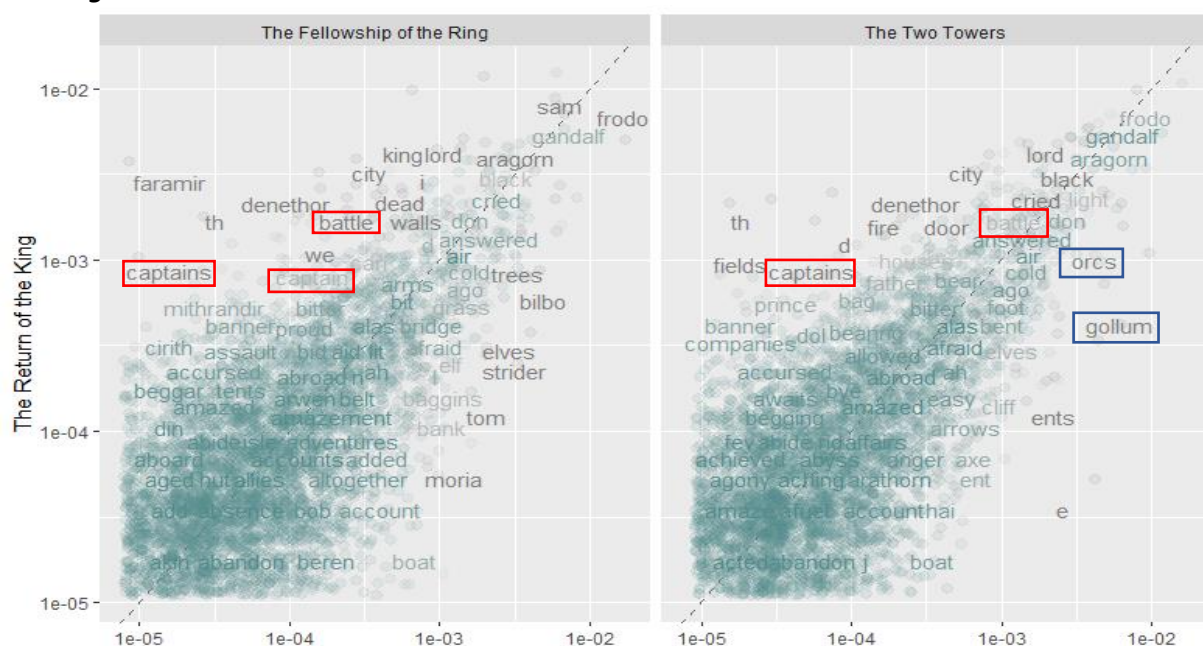


As seen in the previous graph, there is a wide variety of feelings found in the books. An interesting aspect of this graphic is that the frequency of words related to surprise is lower compared to other feelings. This leads to the question of why the feeling of surprise is the most important, according to the word clouds, if the frequency of the words associated with this feeling is not big but quite the opposite.

To understand this peculiarity, it is necessary to observe the other feelings shown in the bar plots. Positive and Negative sentiments are of great importance, according to the frequency of the words in these sentiments. A possible explanation that arises from the way words are used is that the author wanted to generate amazement by mixing positive and negative feelings; this mix is what generates the perfect contrast to intensify the emotions conveyed in the books. This is key to maintaining the readers' attention and the surprise effect, which, as we mentioned earlier, is an important element in the history created by J. R. R. Tolkien.

Finally, I wanted to compare the three books based on the most used words, which is why I included all the words of the books, not only those related to feelings. For this task, a correlogram was plotted (see the figure below). A correlogram is a tool that allows analyzing the relationship of a pair of variables, in this case, words.

Correlogram



The correlogram shows that two of the most frequent words used in the third book (The Return of the King) in comparison to the first two are "battle" and "captains," both words associated with war. This makes sense if we consider that the war where the fate of middle earth is decided occurs in the last book, while in the previous books, it is explained how each side is preparing for a battle that feels imminent. This aspect is paramount because this battle is the ideal outcome for a story developed in the middle ages, an era characterized by great battles and fighting.

Another essential point observed in the correlogram is that the words "gollum" and "orcs" are less frequent in the third book. Although the orcs are secondary characters that only become important in the war and the fact that they are mentioned less in the last book is understandable, it is not the case for Gollum, who has a fundamental role towards the end of the story. But if we consider that revelations and shocking elements are primordial in this story, we can infer that in the case of Gollum, the author wanted to take the surprise element until the end by giving vital importance to a character that seems irrelevant after the second book.

Conclusion and main takeaways

Taking into consideration that about 90% of the data generated in the world is considered unstructured data and a big portion of that is made up of text, being able to transform this data into information that can be analyzed becomes a primary task to understand people's opinion. But at the same time, it is a titanic task if we consider the large amount of data in the form of text that is generated every day; therefore, tools, such as Sentiment Analysis and NLP, make this task easier and faster, generating valuable insights that allow us to go beyond the words and to focus on ideas, feelings, and emotions.

References

J. R. R. Tolkien (2008). *The Lord of The Rings*. HarperCollinsPublishers

Julia Silge and David Robinson (2017) Text Mining with R: A Tidy Approach. (pp. 12–65). Printed in the United States of America

Peter Csathy (2020, January 31). Amazon Prime Video: The Stealthy, Ominous Streaming Force. Retrieved from Forbes:

<https://www.forbes.com/sites/petercsathy/2020/01/31/amazon-prime-video-the-quiet-ominous-streaming-force/#1d6c40621f1a>

Rick Marshall (2020, March 4) Amazon's Lord of the Rings series: Everything we know so far. Retrieved from Digital Trends:

<https://www.digitaltrends.com/movies/amazon-lord-of-the-rings-series-news-cast/>

Jon Swartz(2020, April 21) Netflix may have edge on competition as coronavirus keeps people looking for new shows. Retrieved from Market Watch:

<https://www.marketwatch.com/story/netflix-in-the-age-of-covid-19-streaming-pioneer-may-have-new-edge-on-competition-2020-04-07>

Kristian Bannister (2018, February 26) Understanding Sentiment Analysis: What It Is & Why It's Used. Retrieved from Brandwatch magazine:

<https://www.brandwatch.com/blog/understanding-sentiment-analysis/>