

UDK: 821.111.09
811.111'42
Prethodno priopćenje
5. V. 2022.

MARTIN KOSTELEJ*

MARINA BAGIĆ BABAC**

TEXT ANALYSIS OF THE HARRY POTTER BOOK SERIES

Based on worldwide popular novels about Harry Potter written by J. K. Rowling, this study provides an original insight into sentiment distribution in the series based on text analysis. It shows what sentiment appears in the book series and how it changes throughout the chapters. Our analysis confirmed that the Harry Potter book series are more negative than positive. It means that 111 chapters out of 200 have a negative sentiment. Another contribution is finding out what the story is about using text analysis. Our findings confirm that an automated approach to book analysis mostly is in accordance with the reader's impression. Consequently, natural language processing techniques can help infer what each chapter of the book is about.

Keywords: book analysis; Harry Potter; sentiment analysis; text mining; natural language processing

* Faculty of Management Science and Informatics, University of Žilina, Slovakia, martin.kostelej@fer.hr.

** Faculty of Electrical Engineering and Computing, University of Zagreb, marina.bagic@fer.hr

1. INTRODUCTION

The Harry Potter series is widely known all around the world. It all started on the first of September 1998 when was the first book released. The Series was written by writer J. K. Rowling from the United Kingdom. Harry Potter became very popular for all age categories. In 2001 started movie series were filmed with books and popularity started to grow even faster (Dempster et al., 2016). In the next years, Harry Potter became an absolute phenomenon (Crysel et al., 2015). It started just with a book, but it created much more. According to this book series, movies, games, clothes, board games, and theme park attractions, were created (Biswas, 2014).

Harry Potter books are widely popular not just for children or youngsters but even adults like them. A study from 2012 found that 55 percent of young adult novels are bought by adults (Fris, 2013). This boom was caused by Harry Potter books, which became hit and adored by both children and adults. So, this made it acceptable for adults to read books that are ostensibly for children. Harry Potter series raised the attention of children for long books. When it became so popular, fans wanted to buy new Harry Potter books no matter what. The last four books were even longer than 700 pages and children still loved them. Publisher and children's writers noticed changes in middle-grade novels. It expanded by 115.5 percent between the years 2006 and 2016. These years were the Harry Potter novels' longest (Liang, 2015). It can be concluded that this series phenomenon truly changed the world, and it keeps its popularity even after years (Bowers, 2004). In 2021, it is almost 22 years after the first book release.

In this study, for the purpose of analyzing the Harry Potter books series, text analysis is used. Text analysis is an automated process of understanding text data and making it easier to manage (Harmon-Jones, 2000). It can provide certain insights into text which are required by the user (Medhat et al., 2014). Text analysis is widely used in processing social media, comments, survey responses, books, etc. (Rao, 2016). It can be used

for example for analyzing comments on some videos to find out what the comments are about or to make certain statistics about these comments. There are a lot of possibilities because we live in an age when we have much data everywhere and it is growing so fast without signs of slowing down any time soon. Everyone is commenting on something, posting, writing blogs, and sending messages (Tripathi et al., 2016). When one wants to use some of these data, (s)he wants to have it in clear and easily readable form, and that is the reason why is text analysis used (Silge & Robinson, 2018).

There are 293.6 billion emails sent every day (Jiang et al., 2017). If one needs to read, process, tag, or sort this number of emails, this process must be made automatically. We can use technology to automate some of the tasks. Text analysis can help in business in terms of accuracy and scalability (Avdić & Bagić Babac, 2021). Businesses can extract some specific information from text, for example, keywords, names, or some information about the company. Then, given parts of the text can be categorized into some topics, or classified by sentiment (positive or negative) by using words in a text (Mohammad & Turney, 2010).

The basic text analysis methods are word frequencies or collocations (Guerra et al., 2014). Word frequency can be used for several cases, for example, to analyze expressions that customers use most frequently, e.g., if the word delivery is used most often, that means there are some issues with company delivery. Collocation can be used for connecting words together because of their meaning (Kawade & Waghmare, 2017). In the text, we can try to find some bigrams (two adjacent words) which can have different meanings when these words are together. For example, if we have the word "good", we know its sentiment is positive but if there is the word "not" before "good", the sentiment is negative.

More advanced methods are text classification and sentiment analysis (Lin et al., 2008). Sentiment analysis is used for finding out if the text is positive or negative (Bandhakavi, et al. 2017).

It can be used for finding out if our comments on some videos are more negative or more positive. Text classification is the process that assigns predefined tags to unstructured text. It is considered one of the most useful natural language processing methods (Kramer et al., 2014).

Based on the Harry Potter book series, the purpose of this study is to apply natural language processing techniques to find useful insights into the huge amount of text data, i.e., to analyze the topics and sentiment of the proposed novels. More specifically, the aim is to infer what the story of chapters is about using text mining and compare the results with the real story and the characters important to the story. In order to find out what sentiment appears in the book series and how it changes throughout chapters, three different lexicons were used for sentiment analysis of each book, and the results were compared. In addition, the most positive and most negative books are shown as the result of sentiment analysis.

2. RESEARCH METHODOLOGY

The first step of text analysis was to get book series using an R package called *harrypotter*. This package provides access to the full text of all Harry Potter books. Each text is in a character vector where each element represents one chapter. This package contains the following books:

- *Harry Potter and the Philosopher Stone*
- *Harry Potter and the Chamber of Secrets*
- *Harry Potter and the Prisoner of Azkaban*
- *Harry Potter and the Goblet of Fire*
- *Harry Potter and the Order of the Phoenix*
- *Harry Potter and the Half-Blood Prince*
- *Harry Potter and the Deathly Hallows*

As preprocessing steps, the story was tokenized (divided text into words) and converted into lowercase. Then, stopwords were removed to prepare the text for the sentiment analysis.

Sentimental analysis is a method used for understanding the social sentiment of a brand, product, or service (Banerjee & Dutta, 2015) from the

text. It is the most common text classification tool (Canales & Martinez-Barco, 2014). For example, the comments to a video or review of a product are categorized into negative and positive using sentiment analysis. Thus, interested stakeholders can get an insight into what a good part of their product is or what should be improved (Hochschild, 1983).

To find out what are emotions of text are, sentiment lexicons are used (Bagić Babac & Podobnik, 2016). Sentiment lexicons contain the words with assigned emotion(s) to them, such as happiness, sadness, frustration, etc. (Jurafsky & Martin, 2015). The lexicon-based approaches to sentiment analysis are approaches that use one or several lexical resources to detect emotions detection (Mohammad et al., 2009). Among these approaches, there are keyword-based approaches (Kampić & Bagić Babac, 2021) that are based on predetermining a set of terms to classify the text into emotion categories (Medhat et al., 2014).

In this study, text analysis is used to find the most used words in the Harry Potter book series to infer the story and main characters of books. Then, the sentiment analysis is used to detect positive sentiment. More specifically, data preprocessing and processing are implemented using the R programming language. More specifically, R libraries used for these purposes are *readr*, *plyr*, *stringr*, *stringi*, *magrittr*, *dplyr*, *tm*, and *plotly*. The package used for sentiment analysis is *tidytext* (Silge & Robinson, 2018).

3. RESULTS OF TEXT ANALYSIS

In Figure 1, the most frequently used words in the whole series are shown. It can be noticed that most of them are the names of the main characters of the story. In the first place is Harry by the far. He is followed by his best friends Ron and Hermione as expected. The main negative character called Voldemort is in 23rd place. It should not be considered as something weird, but his name was marked as banned once during the whole series and everybody was afraid to call him by his name, thus 23rd place is high for this name.

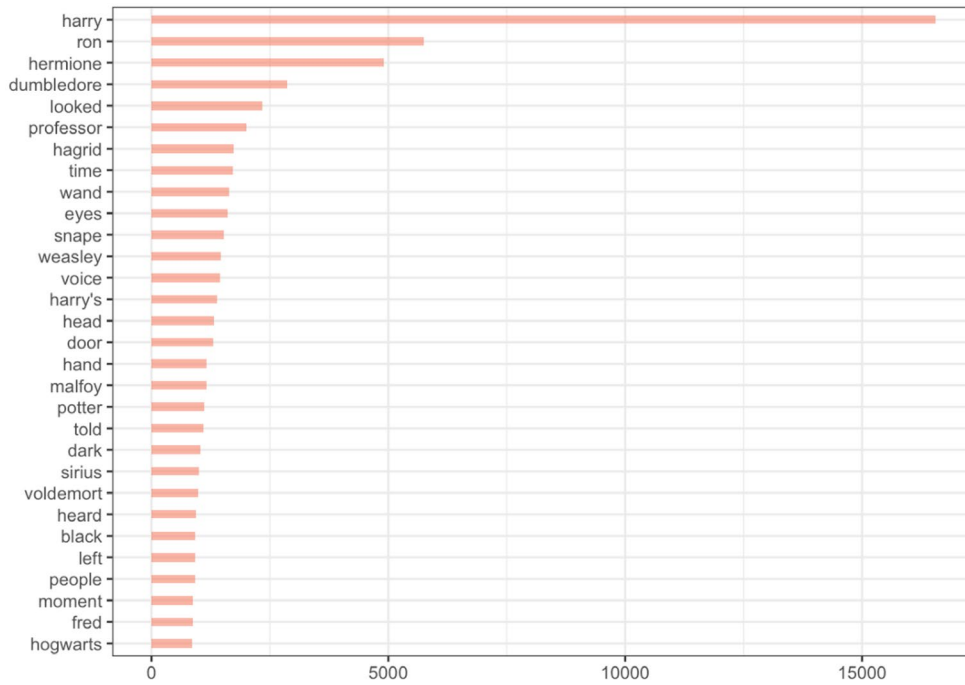


Figure 1.
The most frequently used words in the whole series

To learn more about the story from the used words, the names are filtered from the set of most common words as shown in Figure 2.

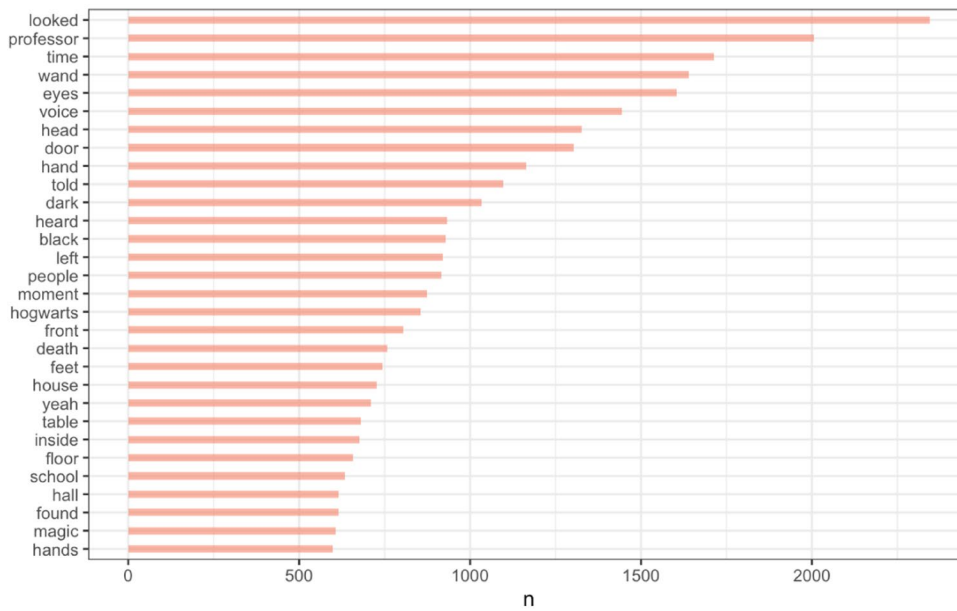


Figure 2.
The most frequently used words in the whole series filtered by names

It can be noticed that the words “professor” and “school” are frequently used. So, even without reading the books, it can be concluded that the main part of the story happened in some schools. In the fourth place is the word *wand* and even the word *magic* was used so many times. Because of this, we can know this book is about some wizard school. In the 19th place is the word *death* and in the 11th is *dark*, which implies this book series is sometimes probably negative because of the negative sentiment of these two words.

Next, the most used words for each book separately are shown in Figure 3. It is not surprising that every book puts Harry in the first place, but there are changes in every book because of the story changing. Different characters appear in every book in the top 10. For example, in 5 books out of 7 (except the first and last one) there is always a professor of defense against the dark arts:

- *Chamber of Secrets: Lockhart*
- *Prisoner of Azkaban: Lupin*
- *Goblet of Fire: Moody*
- *Order of the Phoenix: Umbridge*
- *Half-Blood Prince: Snape*

This implies they are always a big part of the story. Except for the main three characters (Harry, Hermione, and Ron), here is another character who is in the top 10 often (5 books out of 7). This character is called Dumbledore, the headmaster of the wizarding school of the story. He is mentioned the most times in the sixth book (almost 1000 times), in which his character is killed. Even after his death, he is still in fifth place in the next book. Because of that, we can conclude that he is one of the most important characters in the story.

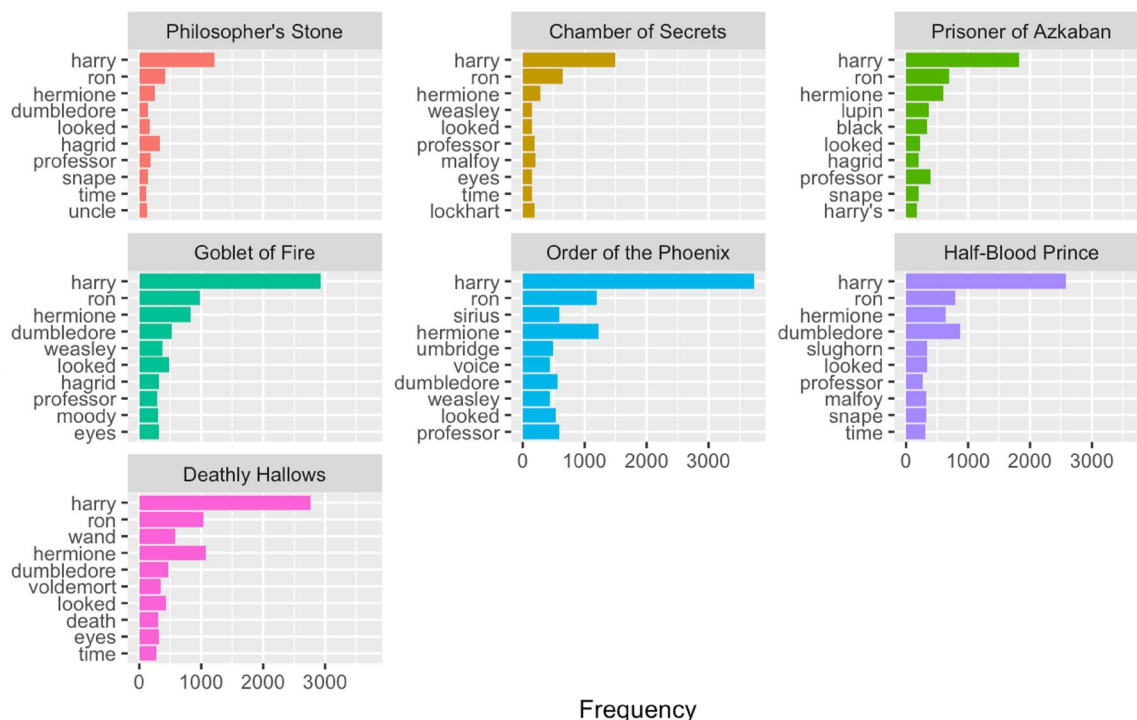


Figure 3. The most frequently used words per book

Next, the sentiment changing during every book is explored. For the purpose of sentiment analysis, a lexicon-based approach was used for analyzing each chapter of the book (counting how many negative and positive words are in each chapter). Then, a comparison of the results from different lexicons is provided. Three different sentiment lexicons were used, namely AFINN, Bing, and NRC.

The Bing lexicon has 6786 words, that is 2005 positive and 4781 negative words. Figure 4 shows

how the sentiment changes throughout the book based on the Bing lexicon.

NRC lexicon offers more sentiment types than the Bing lexicon (for example, trust, fear, sadness, anger, and joy), so to compare the results with other lexicons, we filtered negative and positive words comparing lexicons. If we are working with positive and negative it has 5636 words, that is 3324 negative and 2312 positive words. Figure 5 shows how the sentiment changes throughout the book based on the NRC lexicon.

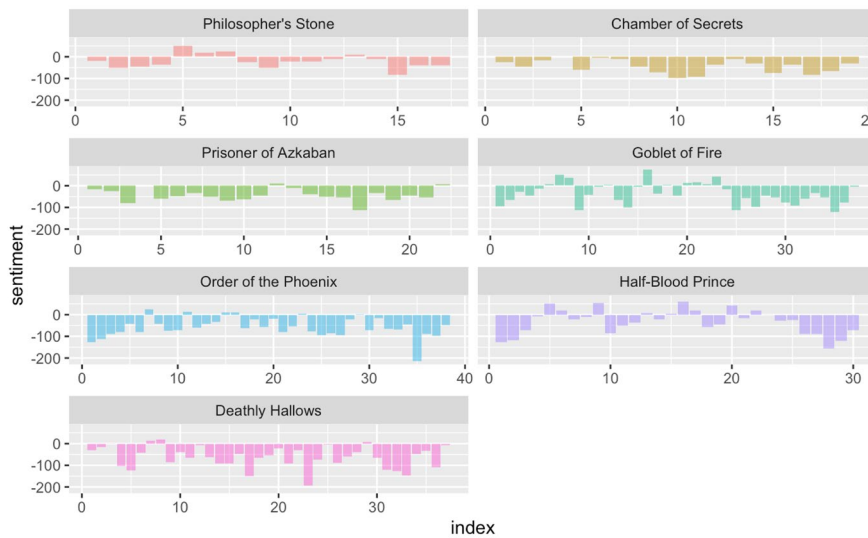


Figure 4. Sentiment changes throughout the book based on the Bing lexicon

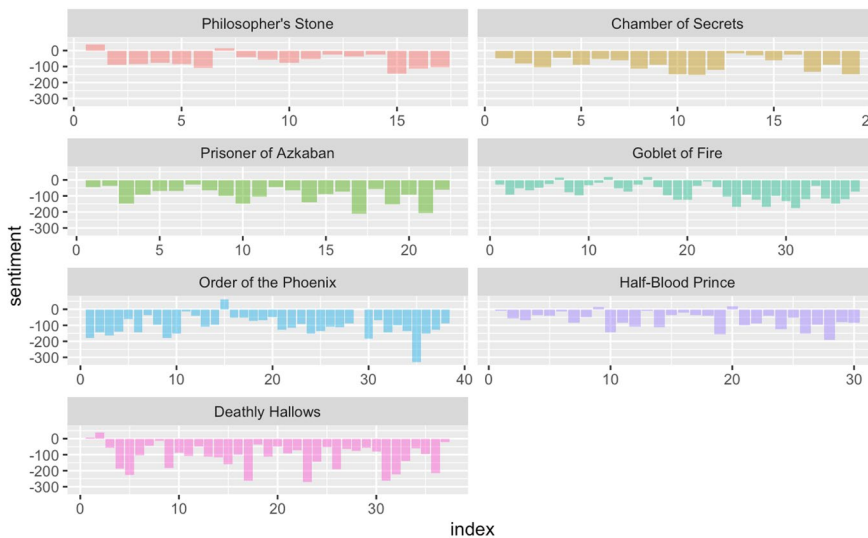


Figure 5. Sentiment changes throughout the book based on the NRC lexicon

AFINN lexicon does not use just negative or positive marks but has different levels of positivity and negativity (-5, -4, -3, ..., 3, 4, 5). So, it should be more precise in determining the level of positivity or negativity in chapters. However, it should be noted that it has only 2477 words. Figure 6 shows how the sentiment changes throughout the book based on the AFINN lexicon.

When trying to find the most positive books comparing all the lexicons and measured values, the winners are the first and sixth books. But if we want to compare this to reality, the first book is positive, because it was written in an optimistic way, but the lexicons do not seem to be very successful in finding sentiment in the sixth book. The sixth book is dark and full of sad moments

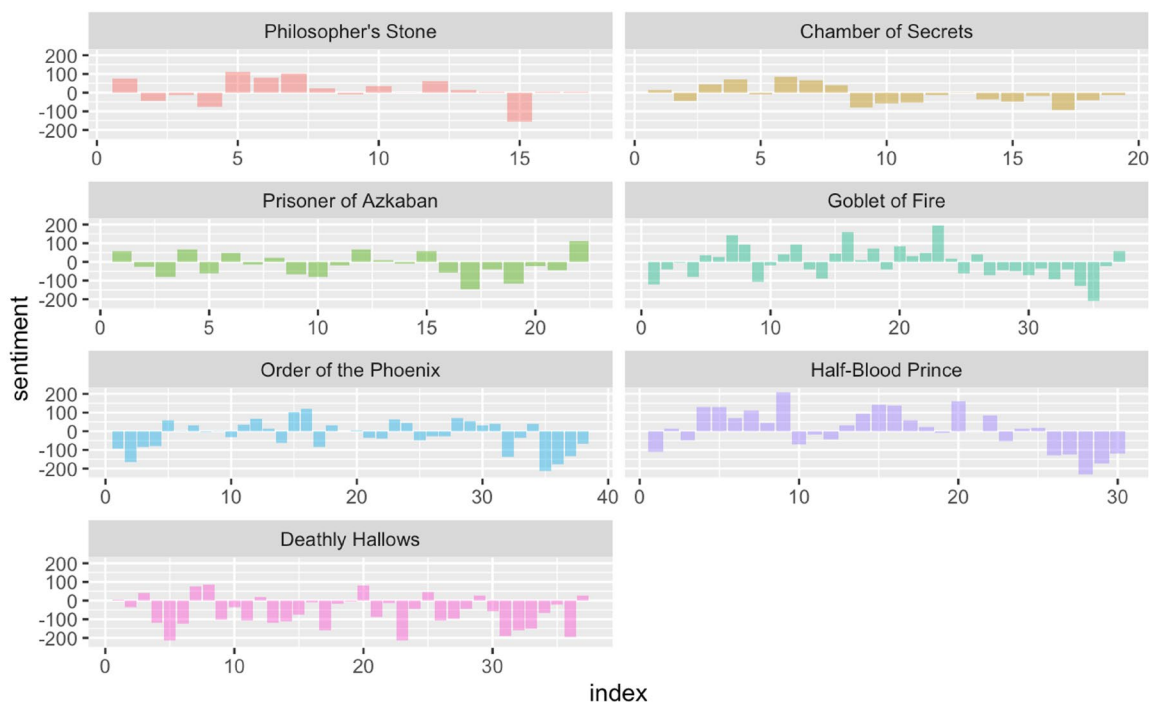


Figure 6. Sentiment changes throughout the book based on the AFINN lexicon

From these figures, it can be noticed that the whole series is almost all the time more negative than positive, especially based on Bing and NRC lexicons. The Bing and NRC lexicons most of the time show similar results, with the exception that Bing has more chapters in the positive part (for example, in fourth and sixth parts has Bing higher positivity). AFINN lexicon shows way more positive chapters than other lexicons and even higher positivity numbers in chapters. It can be explained by the different types of the lexicon. It seems more accurate because of the wider scale for positivity and negativity even when this lexicon has least words than others.

preparing everybody for the end of the series. At the end of the sixth book, all lexicons are accurate because, at the end of book, one of the main characters of the series (Dumbledore) dies.

Finding the most negative book of the series is clear because all the lexicons show that this was the last book. It is correct in the last book was the highest number of dead characters and the darkest atmosphere.

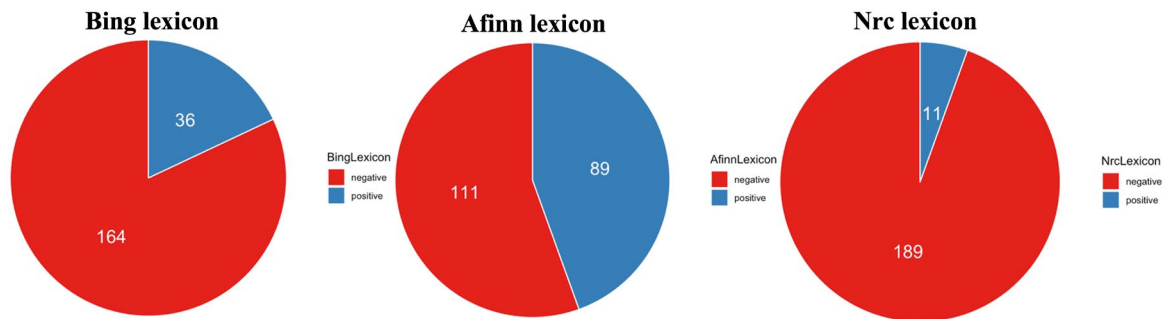


Figure 7. The number of positive and negative chapters in series

The number of positive and negative chapters in the series is shown in Figure 7. These pie charts show that all books together have 200 chapters. When comparing all three lexicons, it can be noticed that there is a significant difference when we are looking at how many chapters were more positive or negative. The major difference is between AFINN and NRC lexicon. AFINN lexicon shows that 89 chapters are positive. That means almost half of the chapters. But when we look at the NRC lexicon, it shows only 11 positive chapters, which is 5.5 % of all chapters.

When we compare these results to reality, it can be concluded that the most accurate is the AFINN

lexicon because the whole series has some good and some bad parts. It can be said that half of the story is positive, the other half is negative and the AFINN lexicon is quite close to these numbers.

Finally, the most frequently used words in each chapter of a book are used to infer what the story is about. These words including names reveal which characters are involved in the story. When further filtering these words by the character names, the story can also be revealed. Figure 8 shows the results from this analysis based on the first book of the series called Harry Potter and Philosopher's stone.

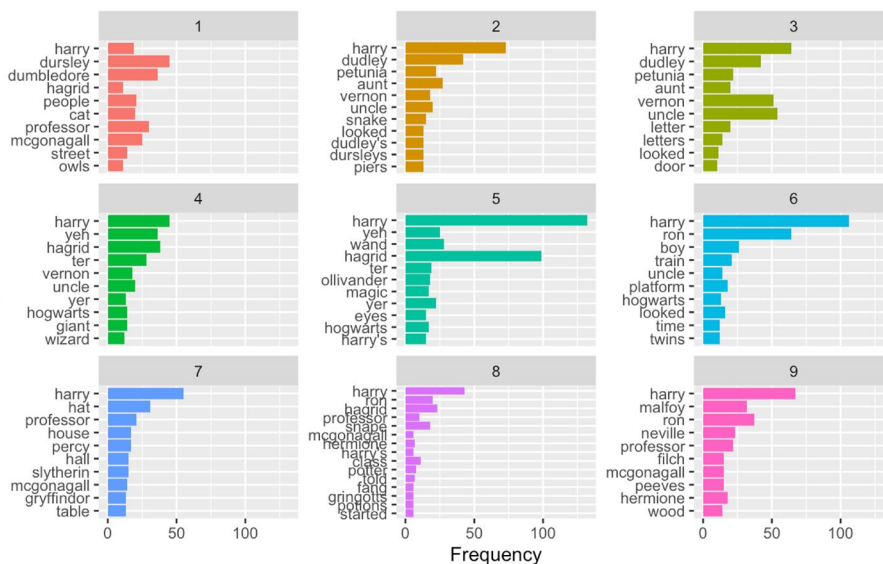


Figure 8. The most frequent words from the first book of the series called Harry Potter and Philosopher's stone

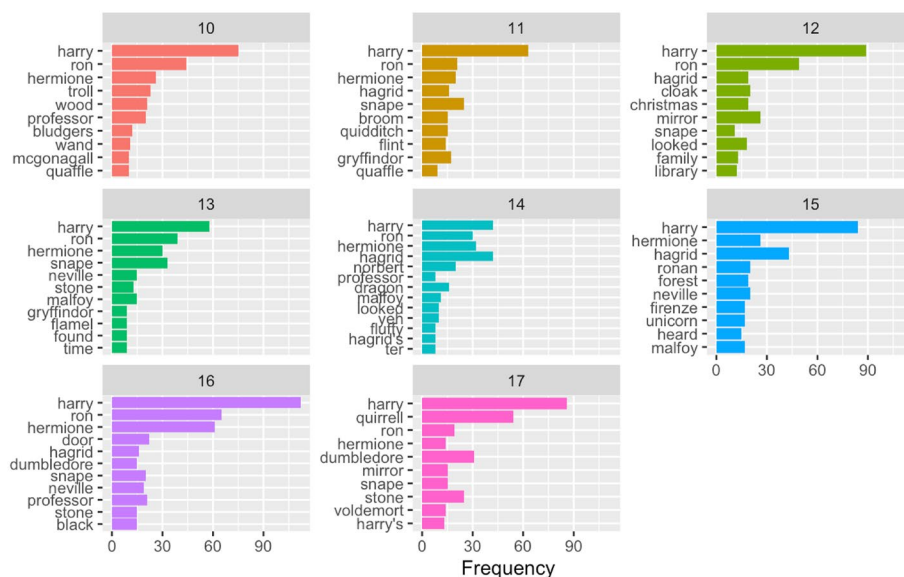


Figure 8. The most frequent words from the first book of the series called Harry Potter and Philosopher's stone

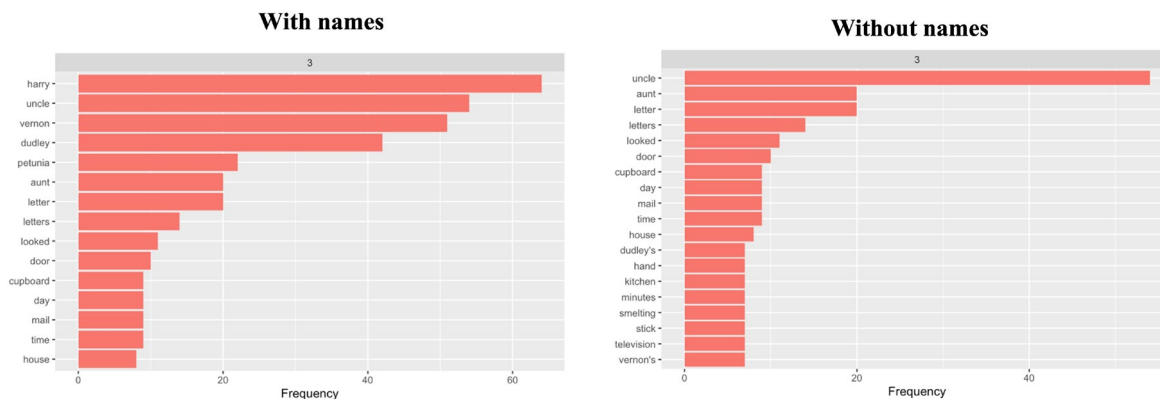


Figure 9. The most frequent words from the 3rd chapter

The main characters in the third chapter are Harry, Vernon, Petunia, and Dudley. From Figure 9, the most frequently used words except names are *uncle* and *aunt*. That makes sense because it is uncle Vernon and aunt Petunia of Harry Potter, thus, we learn that this chapter is about Harry and his aunt and uncle. In the third and fourth

place are *letters* and *letters* and in the ninth is word *mail*. Because of that, we know this chapter is about some important sent letter. And it is correct because this chapter is basically about arriving acceptance letter sent from wizarding school to Harry Potter. But his uncle and aunt keep destroying all letters coming to Harry.

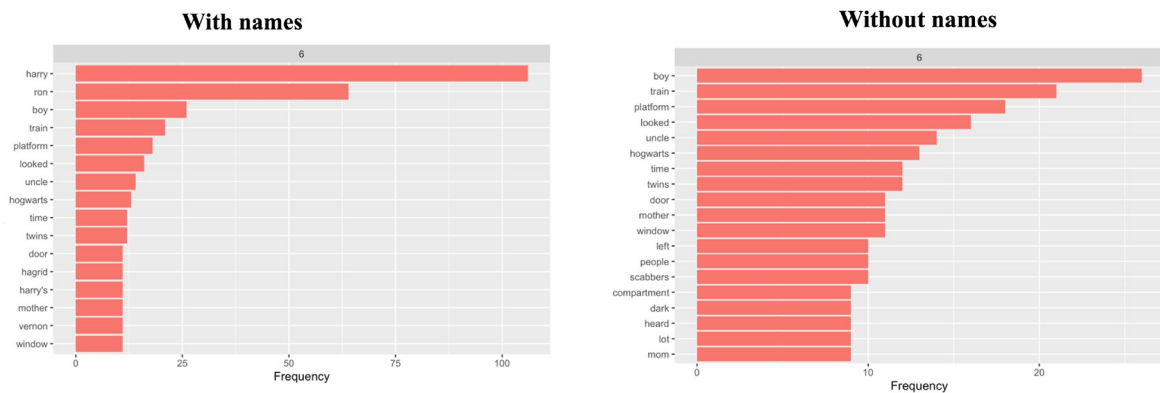


Figure 10. The most frequent words from the 6th chapter

The main characters of the sixth chapter are Harry and Ron. Other characters appearing here are Vernon and Hagrid. Name Ron appeared the first time in this chapter and it's in the second place. From the diagram without names in Figure 10, it can be noticed that the word *boy* is in the first place. So, we can connect it with a new boy named Ron. Because of this, we can deduce that this chapter is about meeting Harry with Ron for the first time. The second and third place are the words *train* and *platform*. Even the word *compartment* is mentioned often, which means they are traveling somewhere by train. When checking the real story described in the 6th chapter, it matches this. Harry is going to the train station to travel to the wizarding school called Hogwarts and he meets Ron on the train. They are together in one compartment and everybody who knows the story of the series knows that they'll become

best friends. They meet Hermione in this chapter too for the first time, but her name is not in the most frequently used words, so we could find out something about this part of the story.

The main characters are Harry, Quirrell, and Dumbledore. Mentioned a lot of times too: Ron, Hermione, and Hagrid. In the diagram without names in Figure 11, the first place took the word *stone*. The name of this book is called *Harry Potter and Philosopher's stone*, so this chapter is mainly about this *stone*. We can see that in the second place is the word *mirror* followed by the words like *pain*, *kill*, and *died*, so it can be deduced that this chapter has negative sentiment, and somebody was killed.

When compared to the real story of the last chapter, the main part of the story is about Harry

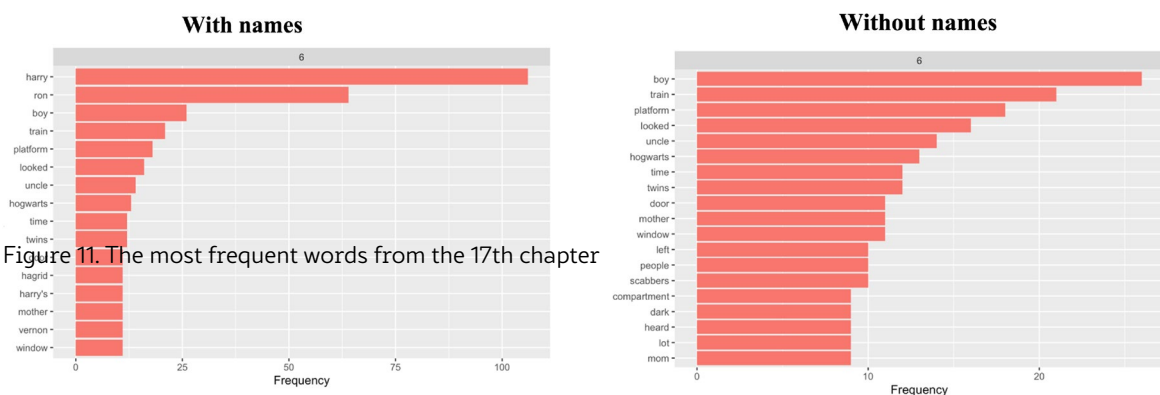


Figure 11. The most frequent words from the 17th chapter

trying to stop professor Quirrell from stealing the philosopher's stone, which could be used by the main negative character Voldemort. If somebody wanted to get this stone, he had to use a special mirror which was mentioned so many times in this chapter. Harry was successful in saving the stone from stealing it by professor Quirrell. Professor Quirrell is dying in this chapter killed by Harry in self-defense.

4. CONCLUSION

The main objective of this study was to find out what sentiment is in Harry Potter books and from the results of the analysis, it was revealed that the chapters of books are more negative than positive. The most positive book is the first one (Philosopher's Stone) because it has 11 chapters in positive part, 5 is negative, and one neutral, thus 64.7% of the book is positive. The second most positive is the sixth book (Half-Blood Prince) 56.7% positive chapters (17 chapters). The third most positive one is Goblet of Fire with 45.9% of positive chapters. The most negative book of the Harry Potter novels is the last book (Deathly Hallows) with a huge percentage of negative chapters. 33 chapters out of 37 are negative (89.2%). The second place belongs to the Prisoner of Azkaban with 63.6%. Overall, it was found that the Harry Potter book series are more negative than positive since 111 chapters out of 200 are negative.

In addition, one of the goals of this study was to find out which lexicon for sentiment analysis scored the best in finding out negativity or positivity of books. After the sentiment analysis by three lexicons (Bing, AFINN, and NRC), it was found that closest to the reality was the AFINN lexicon. An advantage of this lexicon is that more than one value for negativity and positivity is used, that is value -5 is for the most negative words and 5 for the most positive words. Thus, it was more accurate than other lexicons. Each lexicon had a different number of positive chapters in Harry Potter novels. When comparing NRC and AFINN lexicon, the difference is significant as the AFINN lexicon had 89 chapters positive and NRC 11. By using the AFINN lexicon, it was found out

that the most positive book is the first one and the most negative is the last one. The results of this lexicon are in line with the real story of the book series.

The first book was written in the most positive way. Everything was written mainly for the children's audience. Author Rowling was trying to welcome everybody in the wizarding world in an optimistic and peaceful way. In contrast with the first book is the last one with a dark atmosphere, where a war is described, and a lot of characters die.

From the most frequently used words, it was concluded that the main characters are Harry, Ron, Hermione, and Dumbledore, and the novels are about magic and wizard school. In analyzing specific chapters in the first book, some basic information about the story and characters is found. In the third chapter, it was found out that this chapter is about Harry, his aunt, and his uncle. The main part of the story consists of some important letters. In the sixth chapter, it was found out that it is about meeting Harry with Ron for the first time and about traveling somewhere by train. In the last chapter of the book, it was deduced that it is mainly about Harry and Quirrell and stone which is already mentioned in the name of the book. It was also found that somebody was killed in this chapter.

A limitation to our study is lower accuracy under specific conditions. For example, finding out what a story is about in these novels was not as accurate as expected. It is not straightforward to infer the topic from the story based only on the most frequently used words. It suffices to find out which characters are the most important in each chapter and certain basic information about the chapter. For example, it was concluded that somebody died in the chapter based on the words like *die*, *death*, and *kill*. However, this technique was not efficient if we want to know more information about chapters. Therefore, more complex techniques such as machine learning (Poch Alonso & Bagić Babac, 2022) and deep learning (Cvitanić & Bagić Babac, 2022) should be explored and implemented in future work.

The results presented in this study can be further used by marketers and decision-makers involved in any book or other kind of campaign. By showing certain statistics and sentiments related to the textual resource obtained by using text analysis, deeper insights into the quality of any kind of text including social media content can be learned and used for different purposes and domains of interest (Puh & Bagić Babac, 2022).

REFERENCES

- Avdić, D. & Bagić Babac, M. (2021). Application of Affective Lexicons in Sports Text Mining: a Case Study of FIFA World Cup 2018, *South Eastern European journal of communication*, 2; 23-33.
- Bagić Babac, M. & Podobnik, V. (2016). A sentiment analysis of who participates, how and why, at social media sport websites: How differently men and women write about football, *Online Information Review*, 40(6), 814-833, doi: <https://doi.org/10.1108/OIR-02-2016-005>.
- Bandhakavi, A., Wiratunga, N., Massie, S., & Padmanabhan, D. (2017). Lexicon Generation for Emotion Detection from Text, *IEEE Intelligent Systems*, 102-108.
- Banerjee, S., Dutta, U. (2015). Detection of Emotions in Text: A Survey, *International Journal of Advanced Engineering and Global Technology*, 3(12), 1436-1439.
- Biswas, P. (2014). Exploration of Magic Realism: Harry Potter novels in perspective. *International Journal of English Language, Literature and Humanities*, 2(4).
- Bowers, M. A. (2004). *Magic (al) Realism*. Routledge Taylor & Francis group, London and New York, 2004.
- Canales, L. & Martinez-Barco, P. (2014). Emotion Detection from the text: A Survey, *Proceedings of the Workshop on Natural Language Processing in the 5th Information Systems Research Working Days*, Faculty of Systems Engineering, National Polytechnic School of Ecuador, Quito, Ecuador, 37-43.
- Crysel, L. C., Cook, C. L., Schember, T. O., & Webster, G. D. (2015). Harry Potter and the measures of personality: Extraverted Gryffindors, agreeable Hufflepuffs, clever Ravenclaws, and manipulative Slytherins. *Personality and Individual Differences*, 83(10), 174-178.
- Cvitanović, I. & Bagić Babac, M. (2022). Deep Learning with Self-Attention Mechanism for Fake News Detection // Combating Fake News with Computational Intelligence Techniques / Lahby M. ; Pathan AS.K. ; Maleh Y. ; Yafooz W.M.S. (ur.). Switzerland: Springer, 205-229. doi:10.1007/978-3-030-90087-8_10
- Dempster, S., Oliver, A., Sunderland, J., & Thistlethwaite, J. (2016). What has Harry Potter Done for Me? Children's Reflections on their Potter Experience. *Children's Literature in Education*, 47(2), 268-269.
- Friis, C. (2013). Abandoned Children in Literature: The Orphans in J.K. Rowling's Harry Potter and the Philosopher's Stone. *Languages & Literatures*, 68(3), 1-10.
- Guerra, P.C., Meira W.J., & Cardie, C. (2014). Sentiment Analysis on Evolving Social Streams: How Self-Report Imbalances Can Help, WSDM'14, ACM, New York, USA.
- Harmon-Jones, E. (2000). A cognitive dissonance theory perspective on the role of emotion in the maintenance and change of beliefs and attitudes, In N. Frijda, A. Manstead, & S. Bem (Eds.), *Emotions and Beliefs: How Feelings Influence Thoughts* (Studies in Emotion and Social Interaction, 185-211), Cambridge: Cambridge University Press.
- Hochschild, A. (1983). *The Managed Heart: Commercialization of Human Feeling*, University of California Press, Berkeley, CA.
- Jiang, D., Luo, X., Xuan, J., & Xu, Z. (2017). Sentiment Computing for the News Event Based on the Big Social Media Data, *IEEE Access*, 5, 2373-2382.
- Jurafsky, D. & Martin, J.H. (2015). *Speech and Language Processing*. Chapter 21: Lexicons for Sentiment and Affect Extraction, Prentice Hall.
- Kampić, M. & Bagić Babac, M. (2021). Sentiment Analysis of President Trump's Tweets: From Winning the Election to the Fight against COVID-19, *Communication management review*, 6 (2), 90-111. doi:10.22522/cmr20210272
- Kawade, S. & Waghmare, K.C. (2017). A Survey on Identification of Emotion from Text Corpus, *International Journal of Innovative Research in Computer and Communication Engineering*, 5(3), 3882-3886.

- Kramer, A.D., Guillory, J.E., & Hancock, J.T. (2014). Experimental evidence of massive-scale emotional contagion through social networks, *Proceedings of the National Academy of Sciences*, 111(24), 8788–8790.
- Liang, S. (2015). *Magic realism in beloved*. *Campus English*, 12(01), 254-255.
- Lin, K. H.-Y., Yang, C., & Chen, H.-H. (2008). Emotion classification of online news articles from the reader's perspective, *International Conference on Web Intelligence and Intelligent Agent Technology*, IEEE, 1, 220–226.
- Medhat, W., Hassan, A., Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey, *Ain Shams Engineering Journal*, 5(4), 1093-1113.
- Mohammad, S., Dunne, C., & Dorr, B. (2009), Generating high-coverage semantic orientation lexicons from overly marked words and a thesaurus, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 599-608.
- Mohammad, S.M. & Turney, P. (2010). Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon, *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, LA, California.
- Poch Alonso, R. & Bagić Babac, M. (2022). Machine learning approach to predicting a basketball game outcome, *International Journal of Data Science*, 7(1), 60-77.
- Puh, K. & Bagić Babac, M. (2022). Predicting Sentiment and Rating of Tourist Reviews using Machine Learning, *Journal of Hospitality and Tourism Insights* (in press)
- Rao, Y. (2016), Contextual sentiment topic model for adaptive social emotion classification, *IEEE Intelligent Systems*, 31(1), 41–47.
- Silge, J. & Robinson, D. (2018). *Text Mining with R*, O'Reilly.
- Tripathi, V., Joshi, A., & Bhattacharyya, P. (2016). Emotion Analysis from Text: A Survey, *Center for Indian Language Technology (CFILT)*.

ANALIZA TEKSTA SERIJE KNJIGA O HARRY POTTERU

SAŽETAK

Na temelju svjetski popularnih romana o Harryju Potteru, koje je napisala J. K. Rowling, ovaj rad pruža originalan uvid u distribuciju sentimenta pomoću analize teksta. Prikazuje kakav se sentiment pojavljuje u seriji knjiga te kako se mijenja kroz poglavlja. Ova analiza potvrdila je da u seriji knjiga o Harryju Potteru prevladava negativni sentiment, odnosno 111 poglavlja od 200 ima negativan sentiment. Drugi je doprinos otkrivanje o čemu se radi u knjizi pomoću analize teksta. Naši zaključci potvrđuju da je automatizirani pristup analizi knjige uglavnom u skladu s dojmom čitatelja, stoga tehnike obrade prirodnoga jezika mogu uvelike pomoći u zaključivanju o čemu se radi u svakome poglavlju knjige.

Ključne riječi: analiza teksta, Harry Potter, sentiment analiza, rudarenje teksta, obrada prirodnoga jezika.