The Words are not Enough

An Investigation into the Viability of Textual Complexity as a Feature for Recommendation Systems

Mees van Smaalen

Supervisor: M.S. Pera

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
January 29, 2023

Name of the student: Mees van Smaalen Final project course: CSE3000 Research Project
Thesis committee: M.S. Pera, W.P. Brinkman

Abstract

Reading is an essential skill for any child to learn, and finding enjoyment in it can greatly contribute to developing proper reading comprehension. Finding the books they like could prove to be difficult. Utilizing collaborative filtering recommender systems to recommend books to children is a tricky task, the lack of user feedback makes it difficult to accurately recommend books they would enjoy. Using content based recommender systems might be preferable, but what book features could a recommender system like this base recommendations on? This research explores the idea of utilizing the textual complexity of books and their descriptions as such a possible feature. By evaluating how accurate readability formulas can predict the age a book is intended for, how the variability and length of sentences vary per age and analysing the difficulty of words used, this paper finds that the descriptions of books intended for younger audiences might not be aimed at them, but instead at their parents. These findings imply that basing recommendations of the textual complexity of book descriptions might not be the most useful feature to base recommendations of.

1 Introduction

Developing reading skills is crucial in a child's upbringing, otherwise they possibly experience difficulties in their later life with regards to academic development and reaching their full potential (Lyon, 2002) [6]. Naturally children show more enthusiasm for activities they like. Therefore it is important to find books that fit their age and interests.
The use of recommender systems has quickly become widespread and deployed across various forms of media; movies, music, books, etc. (Kalifeh & Al-Mousa, 2021) [4]. Most of these recommender systems have traditional users in mind, they will provide feedback to the system in the form of ratings and reviews. When this feedback is not present for the system, due to the fact that users may not be 'standard' users (e.g. disabled or non-adult), the act of recommending becomes more difficult. This calls for another approach, instead of using methods that focus on the user, e.g. collaborative filtering, a recommender system that operates on domains with non-standard users could lean more towards content based filtering to garner better results.

One such domain that this might be a preferable solution for is children's books. Children are non-standard users due to the fact they don't often leave ratings and reviews, if they do the usability of this feedback may be limited. A child that is growing up may develop new tastes in books rapidly within a short period of time.
The act of recommending children's books is no small task. One of the challenges of recommending children's books is what content to base a recommendation on. Developing a recommendation system for children's books is not the purpose of this research and lies beyond its scope, instead this research will investigate one possible feature that a future recommender system could base recommendations on, namely textual complexity.

The main focus of this investigation will be on the descriptions of books. A book's description is an important criterion for selecting a book to read and will always be available to the party employing a recommender system, e.g. retailers and libraries. This research will focus on the textual complexity of children's book descriptions and ask the question: *Does the language used in books and their descriptions match the age of the children it's intended for?* Answering this question will lead to a general idea whether the textual complexity of books and their descriptions are an adequate indication of the age of the reader, which would tell if the textual complexity is a viable feature for children book recommender system to incorporate in their recommendation algorithm. For investigating this question, it is important to define the factors that could make a text complex. Therefore some sub-research questions arise: To what extent do complex words impact the textual complexity? Is the length and variation of sentences used relevant? And what existing readability formulas perform the best in analysing book descriptions?

This research can be understood as an extension of the research done in the paper *"Don't Judge a Book by its Cover": Exploring Book Traits that Children Favor* (Milton et al., 2020) [8]. Milton et al. analyse the book title by comparing the terminology in the title to data which indicates the average age where children tend to learn specific words, known as Age of Acquisition (AoA) [5]. This research found that children pick books based on titles with vocabulary appropriate to their age. They believe children pass by books with titles that they cannot comprehend. The question remains: Do the results they found on analysing the title extend to the book description? This might not naturally be the case as book descriptions are often not written by the author, instead this is often the responsibility of the publisher. Therefore additionally this research will attempt to evaluate book descriptions by comparing the analysis of the descriptions to the analysis of some full texts from books and investigate if descriptions provides a good reflection of the textual complexity of the book.

Section 2 goes more in depth on the research this paper builds upon. In section 3 the method used in this research is explained in detail. The main body of this paper consists of section 4, in which the results of the research are shown and discussed. Section 5 discusses what measures where taken to ensure the research was conducted in a responsible manner. Section 6 rounds out the paper by drawing conclusions and discussing future work.

2 Related Literature

In *"Supercalifragilisticexpialidocious: Why Using the "Right" Readability Formula in Children's Web Search*

Matters" (Allen et al.,2022) [1] the Textstat Python library was utilized to analyse the effectiveness of the various readability formulas it contains on web searches by children. They concluded that the choice of readability formula impacted the accuracy and that the right readability formula needs to be chosen for a specific scenario. The Textstat library has been applied in this research as well. Various readability formulas were employed to investigate which formula would perform the best for the scenario of children's book descriptions. Whereas Allen et al. found that the more traditional readability formulas like Flesch-Kincaid perform primarily inferior to newer ones like Spache on the domain of web searches, this might not be the case for book descriptions since the language in web searches is strictly modern.

The main idea that is being improved upon is the comparison of book titles to Age of Acquisition data done by Milton et al. [8]. The improvement lies in extending this method to not just the title of a book but also incorporate the book description and investigate if the conclusions from Milton et al. are also applicable on a more extensive text. Milton et al. found that children prefer books with titles that correspond to the reading level of their age, furthermore that they would actually pass over literature where they could not comprehend the title.

'Age-of-acquisition ratings for 30,000 English words' (Kuperman et al., 2012) [5] describes the relevance of the data they collected. Kuperman et al. state "we confirmed that AoA is an important variable to control in word recognition experiments.". The original paper regards over 30,000 words, however an expanded data set exists which contains an additional 20,000 words published as a blogpost [7]. This is the data set used in this research and was also employed by Milton et al in their experiments. This additional dataset contains words that Kuperman et al. did not regard for the original paper, these include: pronouns, number words, adverbs, nouns mostly used as names and inflective forms of words used in the original dataset.

3 Method

3.1 Analysis Procedures

For most textual analysis the Textstat[10] Python library was employed. This library is able to analyse text based on a plethora of readability algorithms. Close to all of these were used on all book descriptions to attain a broad readability analysis. The purpose of deploying many textual complexity algorithms was to evaluate which algorithm would prove to be the closest fit to the intended reading age of the books from the data. The best fitting readability formula was determined and from that formula, the error terms per age bucket were investigated, to explore if children preferred books where the description or full text contained more or less complex language than the algorithm suggests.

Sentence length could be a possible factor in children's preferences in books. To research this aspect the sentences used in books and descriptions were examined by taking the average sentence length and variation of the text. These were evaluated per age bucket and t-tests were performed to uncover similarities. The results of these evaluation could display certain trends among age buckets.

For the analysis of individual terms Age of Acquisition data was gathered. Comparing each word used to Age of Acquisition data, a word complexity analysis was performed on a description and full text scale, in contrast to Milton et al. who performed this method of analysis on a smaller scale by analysing only the book title. To increase the accuracy of this analysis, stop words were removed from each text in the description and full text data sets. Per age bucket the distribution of the Age of Aqcuisition scores were inspected to discover similarities and trends.

While the main purpose of the research focused on finding meaningful results from analysing a book's description, in addition the research compared the language used in the description of various books to the actual language used in these books by analysing the full texts of books. The main idea that lies behind performing this analysis is to evaluate how well a book description actually represents a book's contents.

3.2 Data Description

Data was amassed from various sources. One data set used to gather descriptions was sourced from from Thomas Konstatin's data set on Kaggle [9]. This looked to be a promising data set containing over 3000 records, however unfortunately most of them turned out to be duplicates and only 41 unique records existed. These unique records did possess fine quality data for analysis, therefore these records have been analysed.

The biggest source was a data set from Goodreads. This set provided a wealth of records for analysing descriptions. This set had to be trimmed down due to the fact that not all of the records were usable. The Goodreads data set contained many non-English books, these books were filtered out because the Textstat library might not provide sufficient analysis with some languages. Above that, the Age of Acquisition data only contains English words and would be unusable with non-English books. While the Goodreads data set possessed a substantial quantity of book records, the quality of the records was not always a given. One of the main attributes of interest was the description of the book and for part of the records the description included was insufficient, it was either missing or too short for it to be worth investigating. The Textstat library performs more reliably when the text consists of at least 3 sentences, therefore entries with descriptions containing 3 sentences or fewer were deemed unfavourable and omitted. The largest concern turned out to be the lack of age data. A large part of the Goodreads data did not possess sufficiently structured information for determining ages, as a result a large part of the data set had to be discarded. In the end the Goodreads dataset ended up consisting of 2855 good

Age Bucket	Descriptions	Full Texts
[0-5)	725	3
[5-8)	1207	8
[8-12)	811	50
12+	153	9
Total	2896	70

Table 1: Amount of Data Entries per Age Bucket

quality book records which were analysed.

The full book texts were gathered from Wikisource. Due to copyright laws the only full texts available for analysis were stories that exist in the open domain, as a result most of these stories are from the past century in contrast to the other gathered data which includes books published more recently. The data from Wikisource was also not large in numbers, to make matters worse some of the books had missing texts, eventually there were 70 records fit for analysis.

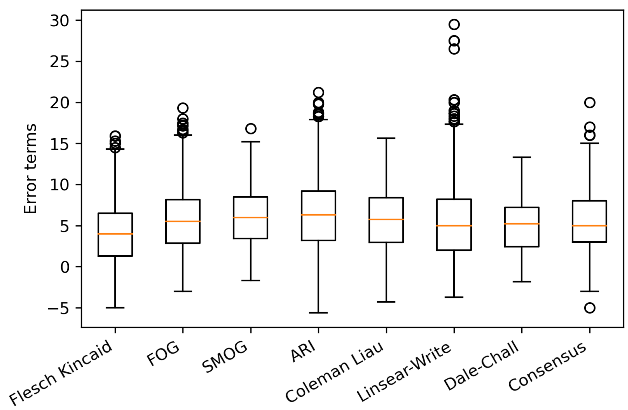
The Kaggle and Goodreads data set were combined to form the description data set. The Wikisource was the only full text data set. For the purpose of analysing across varying ages the records were each assigned an age bucket corresponding to their intended age range. Four age buckets were constructed: [0-5), [5-8), [8-12) and 12+. The amount of entries per age bucket can be found in table 1.

4 Results and Analysis

4.1 Description Analysis

Each book description from this data set was analysed with almost every readability algorithm that Textstat possesses. Allen et al. [1] stress the importance of determining the right readability formula. To investigate which algorithm has the closest fit with the interest age the **Root Mean Squared Error (RMSE)** was computed for every book and expressed in age, meaning that an RSME of 3 reflects that on average an algorithm misjudges the complexity by 3 years. It is important to note that most readability algorithms provided their score in the form of an American school grade, for computing the RMSE the interest age was mapped to the respective grade accordingly. The error terms were computed by comparing the result of an algorithm with the actual age bucket of the book, where a positive error term exemplifies that an algorithm has rated a text to be more complex than the age bucket it is associated with and vice versa for a negative error term. The error terms of all algorithms are mapped in Figure 1. The performance of the algorithms varied per book description, however in general it was found that the **Flesch-Kincaid** readability algorithm performed

Figure 1: Error Terms of all applied Readability Formulas applied on Book Descriptions



the best on the combined data set with a RSME of **5.4**. The Flesch-Kincaid readability algorithm was developed by J.P. Kincaid and his team for the US navy in 1975 and has seen widespread use. The algorithm they developed works according to the formula:

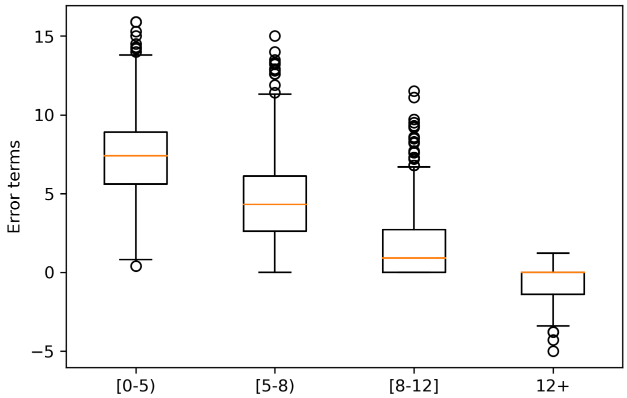
$$0.39\left(\frac{\text{total words}}{\text{total sentences}}\right) + 11.8\left(\frac{\text{total syllables}}{\text{total words}}\right) - 15.59$$

What makes this result interesting is that when researching good fitting readability formulas on some book excerpts Allen et al. observed that Flesch-Kincaid produced a higher error rate compared to other readability formulas like Dale-Chall. This further reinforces the idea that different algorithms produce better or worse results depending on the intended use. However it also provides early evidence that a book description might not be reflective of the complexity of a books content.

After having determined the most applicable readability formula for the whole data set, the performance of the Flesch-Kincaid was examined closer by investigating the execution of the algorithm per age group. The results are shown in Figure 2, something that should be noted is that the 12+ age bucket has some positive error terms, these exist because for the 12+ bucket an upper boundary of 18 was chosen, motivated by the idea that the book descriptions that were analysed were aimed at children.

What immediately stands out from Figure 2 is the high error terms for younger ages. Though, this does coincide with expectation, for it seems to be extremely unlikely that children at younger ages pick their own books, instead it will be the parents who will read the description and determine if a book is suitable for their children. This is similar to a result which Miller et al. [8] observed when analysing book titles, they attributed this to parents reading along with their children, which is about the same explanation for the results found here. While the error terms are high at younger ages there is a strong downward trend in the higher age buckets, resulting in a RSME of **1.323** for the 12+ age bucket, and

Figure 2: Error Terms of Flesch-Kincaid on Book Description per Age Bucket



Age Bucket	Sentence Length Average	Sentence Length Variation
[0-5)	15.806	63.340
[5-8)	16.544	66.237
[8-12)	17.393	81.186
12+	17.830	77.651

Table 2: Mean of the Average and Variation of Book Description Sentence Lengths across Age Buckets

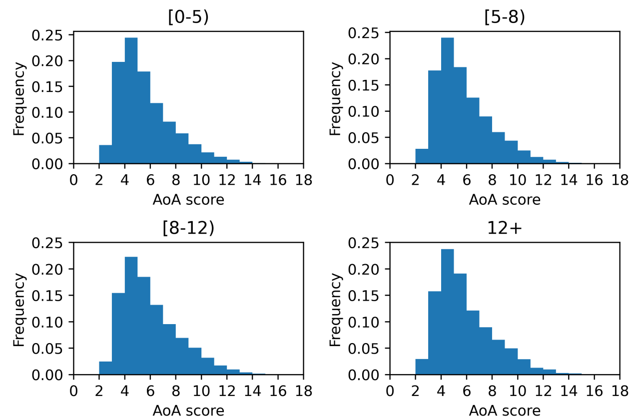
the distribution of these error terms tightly centered around 0. The more curious aspect of these results reside in the fact that the error terms for the [5-8) and [8-12) age buckets are exclusively positive and have some high outliers. There are several possible explanations for these findings, it might be that Flesch-Kincaid has a bias for this data set where it consistently rates book descriptions to be more complex than they are, it might also be due to low quality descriptions, however these explanations base themselves on the results being faulty, instead the results could be interpreted as evidence that children favour books where the description is more intricate compared to what their age suggests.

Another aspect of the descriptions that was investigated were the sentences, mainly in their length and how they vary across the age buckets. For each age bucket the mean of the average sentence length and variation were calculated, the results can be found in Table 2.

The expectation was that the descriptions of books for children in the younger age buckets would have short sentences with lower variation to reflect the simpler contents of the book. While this can be observed as there is an increase in average sentence length and in general also in variation, the evidence is not cogent. The increase in sentence length average when a jump in age buckets occurs is never more than around 5%. However when performing t-tests for each adjacent jump, the null hypothesis where the mean is the same is rejected[1] for each jump excluding the age buckets of [8-12) and 12+. This suggests that at lower ages the sentences of book descriptions are shorter.

For analysing the complexity of the words themselves, the individual terms of each text were compared to the Age of Acquisition data, per age bucket the distribution of the resulting scores are displayed in Figure 3. The scores represent the age at which a child learns the corresponding word.

Figure 3: Distribution of Age of Acquisition Scores From Descriptions per Age Bucket



At first observation the distributions all seem to be alike. However when performing a t-test, the null hypothesis is rejected for every possible comparison between distributions. Table 3 shows the 25th, 50th and 75th percentiles of the data per age bucket.

From this table there is growth visible in the jumps between age buckets. What is noticeable is how low the scores are where most of the data is located. Only in the lowest age bucket it appears that most of the terms are above the Age of Acquisition of the child it's intended for. This fits with the earlier results and strengthens the idea that it is the parents who pick out the book and at whom the description is targeted. The reason for these low results is likely due to the fact that in contrast to a title that contains more uncommon words to represent the book, a description also contains plenty of general language even after having filtered out the stop words.
Something to be observed is the lower scores in the 12+ age bucket compared to the [8-12) age bucket. This could either be the data set, however another motivation could be that for the purpose of storytelling the language used does not have to be more complex than what an 8-12 year old can understand.

[1] $\alpha = 0.05$

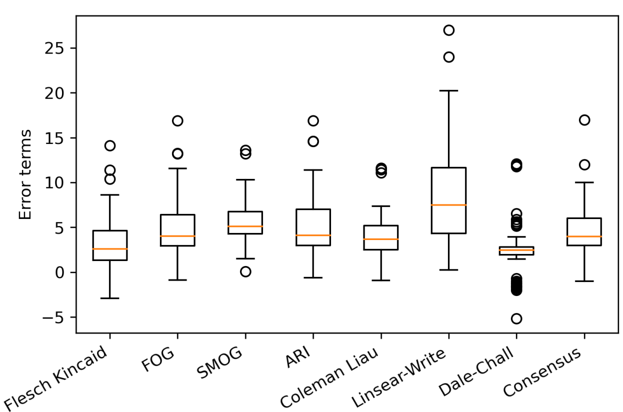
Age Bucket	25th	50th	75th
[0-5)	4.06	5.11	6.74
[5-8)	4.20	5.25	6.94
[8-12)	4.35	5.43	7.28
12+	4.3	5.32	7.11

Table 3: Percentiles of the Age of Acquisition Results from Descriptions across Age Buckets

4.2 Full Text Analysis

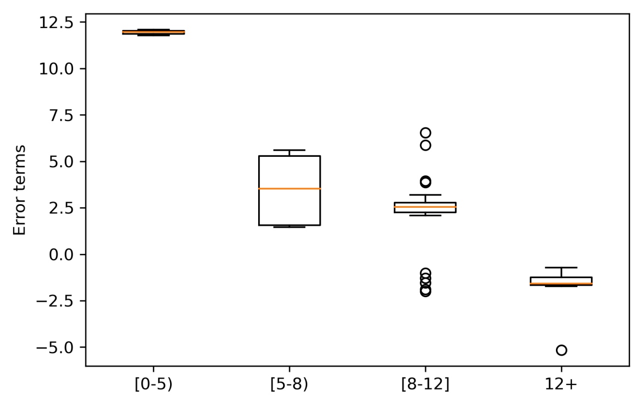
Analysing the data set gathered from Wikisource containing the full texts of books was achieved in the same manner as the book descriptions. Again employing the RSME to determine the best performing readability algorithm. The error terms each readability algorithm applied to the whole data set are displayed in Figure 4.

Figure 4: Error Terms of all applied Readability Formulas applied on Full Texts of Books



The most intriguing result from Figure 4 is the Flesch-Kincaid boxplot, as mentioned in subsection 4.2 Allen et al. found that Flesch-Kincaid was not a high performing algorithm on their data set of book excerpts, however on the Wikisource data Flesch-Kincaid once again achieved a relatively low RSME of **4.35**, and while a book description is something different than the actual book text, an excerpt is the same domain as the full text only differing in length. The most likely reason for this behaviour would be he difference in data used, it is possible that the book excerpts Allen et al. analyzed are from more recently published books. This curious observation does invite more research for future work. While the RSME of Flesch-Kincaid was relatively low, the algorithm with the lowest RSME was the **Dale-Chall** algorithm with an RSME of **3.748**. The Dale-Chall readability formula is based on Flesch-Kincaid, however it also uses a list of words that fourth graders are familiar with. The original Dale-Chall algorithm was first published in *A Formula for Predicting Readability* (Dale & Chall, 1948) [3], however the algorithm was later updated in *Readability Revisited: The New Dale-Chall Readability Formula* (Dale & Chall, 1995)[2] which is the one Textstat utilizes. Before diving deeper into the error terms of the Dale-Chall algorithm it is worth noting how all algorithms exhibit a lower RSME compared to the description analysis, suggesting that the readability formulas work better on a larger body of text or provides further evidence that the complexity a children's book description is not fully reflective of the complexity of the actual book.

Figure 5: Error Terms of Dale-Chall applied on Full Text per Age Bucket



In Figure 5 the error terms for the Dale-Chall algorithm applied on full texts are shown. With only 3 books in the corpus that fit in the lowest age bucket there can't be drawn any conclusions with confidence, however some observations of this full text analysis appear to be homogeneous with the results found by analysing the book descriptions. The same downward trend can be observed in the higher age buckets that was also present in the book description analysis. Does this imply that the complexity of a book description is reflective of the complexity of its content? That cannot be said with confidence due to the differences. For one the actual error terms of the full text analysis are significantly lower than with the description, above that negative values are also more prevalent in full text analysis than they were in the description, especially in the 12+ age bucket where the full text readability of each book was deemed more elementary than the ages of the children.

As with the book descriptions, the sentence lengths of the full texts were also analysed, these results are shown in Table 4.

The first thing that has to be pointed out is the huge variation in sentence length in the [0-5) age bucket and how the average sentence length is also longer. As mentioned before this bucket only contains 3 books, the culprit is Tommy Thumb's Songbook. An edition from 1815 containing children nursery

Age Bucket	Sentence Length Average	Sentence Length Variation
[0-5)	23.840	567.812
[5-8)	15.979	138.391
[8-12)	16.181	159.726
12+	17.566	173.192

Table 4: Mean of the Average and Variation of Full Book Text Sentence Lengths across Age Buckets

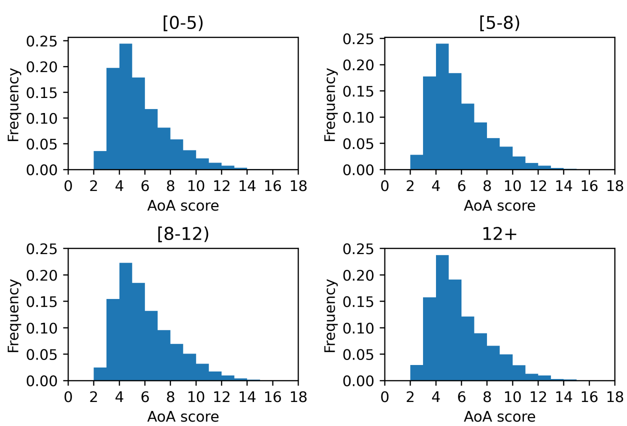
Age Bucket	25th	50th	75th
[0-5)	3.89	4.60	5.95
[5-8)	3.98	4.93	6.25
[8-12)	4.19	5.17	6.94
12+	4.11	5.16	6.94

Table 5: Percentiles of the Age of Acquisition Results from Full Texts across Age Buckets

rhymes. Due to a lack punctuation all songs became one long sentence, hence how it could expand the sentence length average and variation to a preposterous degree. With the small sample size in this age bucket no proper conclusion could be reliably drawn. Examining the other age buckets, the same trend can be discerned as seen with the book descriptions where there seems to be a small increase in both sentence length average and variation as the ages increase. It should be noted that since the full texts of books were analysed the set of all sentences was significantly larger compared to the set of sentences from the book descriptions, as a result the sentence length variation is exceedingly higher, meaning that most of the books contained both short and longer sentences. This is also observed when performing t-tests among the age buckets, the null hypothesis is never rejected. Therefore it can not be concluded that sentences become longer for books read by older children.

The results of the Age of Acquisition analysis are found in figure 6.

Figure 6: Distribution of Age of Acquisition Scores from Full Texts per Age Bucket



The results show similar behaviour as with the description analysis. The distributions appear similar however, as was the case with the descriptions, the null hypothesis of a t-test is rejected for every possible comparison.

Looking at the percentiles of the results figured in Table 5, the same trends can be remarked as with the descriptions. However what stands out is that all percentiles have lower results in regard to their corresponding score resulting from the description analysis. This reinforces the idea that since the full texts possesses a massive body of text compared to a description, the ratio of general language to the subject related words becomes larger and thus the Age of Acquisition percentile data decreases. The other motivation that the language used for children storytelling stagnates around ages [8-12) is also supported, as the same observation can be made that the scores at the percentiles for the 12+ age bucket are once again equal to or below the scores from the [8-12) bucket.

5 Responsible Research

There were a few steps taken to ensure the research was conducted with integrity. All data collected is accessible for anyone and was exclusively used with the purpose of conducting this research. For investigating the full texts of books only stories were used that exist in the public domain and are free of copyright. Due to the fact that this researched focused on the contents of books and not on actually profiling users, there was no identifying user information needed and thus privacy was ensured. To avoid the possible impression of cherry-picking data all adjustments to the data set have been clearly motivated to give insight on why some data was omitted or used. All addressed literature included in this paper have their sources in the reference section and can be accessed, however it should be noted that some of the discussed literature require a login or academic credentials. Every bit of programming was done with Python, the libraries that were used have either been mentioned or have widespread use (Numpy, Pandas, etc.), as a result this research should be reproducible by anyone.

6 Conclusions and Future Work

The research performed investigated the textual complexity of book descriptions across varying ages with the purpose of researching the viability of using this as a possible feature in children recommender systems.
Applying different readability formulas found that on the

book description data set the Flesch-Kincaid algorithm performed the best. However from taking a closer look it became clear that the readability formula is increasingly accurate for books intended for older children. This trend, while interesting, is not a positive for the viability of textual complexity as a recommendation feature. When the intended age of a book is unknown and must be predicted, difficulties arise due to the fact that it would be hard to discern if a book predicted to be for older children is actually for older children or is a book for younger children that has been misjudged. The theory for these common misjudgements is reasoned to be caused by book descriptions for a younger audience not being geared towards that younger audience but instead to their parents who would pick out books for their young children.

From analysing sentences the conclusion could be drawn that there is an increase in sentence length and variation, possibly meaning that descriptions with on average longer and more varied sentences signifies a book intended for older audiences. However this increase is not large enough to exhibit strong evidence that this is indeed the case.

Looking at individual terms with the Age of Acquisition analysis strengthens the idea that the description of books is not intended to be read by younger audiences. Additionally there are strong hints that the difficulty of words used in these descriptions for storybooks stagnates when children get older. This would make it more difficult for a possible recommender to accurately gauge the age of children when they are getting older.

Using another corpus containing the full texts of books, allowed for investigating the textual complexity of full texts. Most of the same trends could be observed as with the description analysis. The actual numbers differed from the description. In terms of error rates of the readability formulas the accuracy of the readability formulas was improved substantially. Sentences contained a great deal of increased variation, in part due to the limited amount of text a description has. The results of the individual term analysis coincided with the results from the description analysis, only with lower values indicating that the ratio of general language to the subject related language is greater compared to smaller bodies of text. Drawing definite conclusions from the full texts is especially difficult due to the limited amount of books in the lowest age bucket. However similar trends were observed and the difference in results could be rationalized, hinting that while the descriptions aren't fully reflective of the full texts, there may indeed be a relation between the textual complexity of an actual book to the book description, though for this to be said with confidence further research is required.

Circling back to the main purpose of the research: *Does the language used in books and their descriptions match the age of the children it's intended for?*. From this research no strong conclusions can be drawn that this is the case. This conclusion is mainly motivated by the idea that the book description of book intended for younger children is not aimed at these children but instead at their parents. When children become older the description becomes more reflective of the intended audience of the book. Leading back to the recommender systems, is textual complexity of book descriptions a valid feature to base recommendations on? According to the outcomes of this research, the textual complexity of book descriptions might not be the most optimal feature to utilize in recommender systems, as it will lead to errors in recommendation.

This research came with its fair share of limitations, future research could address this and arrive at a different conclusion. The big limitation regarding data could be addressed, the data set consisted of nearly 3000 records, however in the grand scheme this is still not a large quantity of records. The full text book data was even more lacking, both in size and quality. With only 70 full texts analysed no conclusions could be drawn with much confidence, these books where all stories in the public domain and therefore somewhat dated. Future research might be able to gather a private set of full texts from more recently published books, which would allow for more accurate research. Method-wise, additional readability analysis methods could be added to improve on the analysis done in this paper. With a larger corpus of books which also contains information on the age of children reading them would also allow for more specific age buckets, increasing the confidence for drawn conclusions.

References

[1] Garrett Allen, Ashlee Milton, Katherine Landau Wright, Jerry Alan Fails, Casey Kennington, and Maria Soledad Pera. Supercalifragilisticexpialidocious: Why Using the "Right" Readability Formula in Children's Web Search Matters. In Matthias Hagen, Suzan Verberne, Craig Macdonald, Christin Seifert, Krisztian Balog, Kjetil Nørvåg, and Vinay Setty, editors, *Advances in Information Retrieval*, pages 3–18, Cham, 2022. Springer International Publishing.

[2] Jeanne Sternlicht Chall and Edgar Dale. *Readability revisited: The new Dale-Chall readability formula*. Brookline Books, 1995.

[3] Edgar Dale and Jeanne S Chall. A Formula for Predicting Readability: Instructions. *Educational Research Bulletin*, 27(2):37–54, 1948.

[4] Sewar Khalifeh and Amjed Al-Mousa. A book recommender system using collaborative filtering method. In *ACM International Conference Proceeding Series*, pages 131–135. Association for Computing Machinery, 4 2021.

[5] Victor Kuperman, Hans Stadthagen-Gonzalez, and Marc Brysbaert. Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44(4):978–990, 2012.

[6] G.Reid Lyon. Reading Development, Reading Difficulties, and Reading Instruction Educational and Public Health Issues. *Journal of School Psychology*, 40(1):3–6, 2002.

[7] 'Marc', 'Victor' 'Kuperman', and 'Hans' 'Stadthagen-Gonzalez'. Age-of-acquisition (AoA) norms for over 50 thousand English words, 4 2012.

[8] Ashlee Milton, Levesson Batista, Garrett Allen, Siqi Gao, Yiu Kai D. Ng, and Maria Soledad Pera. "don't Judge a Book by its Cover": Exploring Book Traits Children Favor. In *RecSys 2020 - 14th ACM Conference on Recommender Systems*, pages 669–674. Association for Computing Machinery, Inc, 9 2020.

[9] Thomas Konstatin. Highly Rated Children Books And Stories, https://www.kaggle.com, 2020.

[10] Alex Ward. https://github.com/textstat/textstat, 2022.