Trait Analysis to Facilitate Children's Books Recommender Systems

Extracting perspective and sentiment traits from book fulltext and descriptions

Stijn Swart[1]

Supervisor: M.S. Pera[1]

[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: Stijn Swart
Student ID: 4590589
Final project course: CSE3000 Research Project
Thesis committee: M.S. Pera, WP. Brinkman

Abstract

Recommender systems are a useful tool for matching readers with books. However, the lack of user data from children, both due to privacy concerns as well as a low incentive to leave reviews, results in existing systems proving inadequate at recommending to the youth. It has been shown that children have a different relation regarding emotion at different ages. We also theorize that the perspective that a book is told from varies depending on target audience age. In this paper we examine traits that could be used for recommendation instead of review data. By examining both the description of books as well as the fulltext we obtain a set of traits relating to perspective and sentiment. Comparing these traits among books written for different age groups we observe a subset of these traits that show potential significance in age categorization. By performing a combination of empirical examination and significance testing we find that both book descriptions and book fulltext contain perspective and sentiment traits that show significance when comparing books written for different ages. Because traits obtained from book fulltext present a higher quantity of significant instances when compared those obtained from descriptions, we conclude that analysing fulltext for traits shows promise when considering a recommender system that aims to use book content.

1 Introduction

Many platforms, from the shows you watch on Netflix to the articles you find on Google Scholar use recommender systems. In general, these platforms rely on recommendation strategies based on collaborative filtering [25], where the suggestions are made based on information about the user. More specifically, user based collaborative filtering algorithms look for similarity between users and then recommend items based on items that neighbouring users liked[21]. In most cases this is preferred, since a lot of data about the user is available. However, if a system is recommending to children, this data is less plentiful [10]. Some platforms like Spotify and Netflix have a rich cache of data to use even in the case of children [14]. In the field of book recommendation, where this user-based data is mostly reviews and rating left by said users. One of the most popular website to write book reviews is Goodreads [28]. However, of the people leaving reviews on this website, only 14% is under the age of 18[24]. Indicating a relatively low review count when comparing to adults, and as such a relative lack of data. This makes collaborative filtering a less viable method within the field of book recommendation toward children when compared to adults.

Because of this lack of data attempts have been made at recommending to children using differing methods. One such algorithm, CBRec, is based on matrix factorization and collaborative filtering [18]. It still uses a significant amount of data about the user. Another attempt at fitting books to the preference of children is through a system where parents indicate both the preferences their child has as well as the preferences they as parents have when reading to their children. [9]. For example, the child might like books about animals, while the parent might prefer books about family. This method of data acquisition makes it easier for the system to incorporate the children's' preference as the data is supplied by adults. Both of these systems still focus on the user. The system uses topic modeling to find books containing topic related to the ones supplied by the parents [27]. While this is a very useful tool it does not allow for recommending to a large group as each individual needs to input a plethora of data.

Another popular approach examines content for recommendation generation. Content-based recommending tries to find similarities in items and then based on one item, similar items can be recommended [26], allowing an algorithm based on this concept to match users to items without requiring the information of other users [15], and as such seems more fitting for a context where little user information is available. In the case of both systems some user information is often still required.

When approaching the problem a from a bit further away we can see a need for user data that does not require the user to input pre-existing preferences. This means an approach to children recommendation could use user data that broadly describes the targeted user without requiring pre-existing user review data. Examples of details about users that are disjoint from review data are things like age, country of origin, gender and level of education: details that do not directly involve books. By focusing on this type of user information, we can omit the review data requirement. As such it is easier to target a group of users instead of one specific user.

Studies have suggested that books that are of interest to children change as the children age [7]. This implies that finding traits in books that differ between books depending on the target age could aid in targeting children of different ages.

Studies suggest that children at various ages process emotions differently [12]. This means that traits relating to emotion could be a subset of the traits discussed before. We also theorize that children relate to the characters in books differently at various ages. For example, a book for a very young child might contain a story about a fox doing simple things. While a story written for a teenager contains things more relatable to the life of the child.

This gives us two different traits for our empirical examination. The traits regarding emotion will be referred to as sentiment traits. While the traits regarding book perspective will be referred to as perspective traits.

When considering these aspects the main observation that can be made is that these traits are contained in the content of a book. This means that research into these topics will most likely yield more fruitful results if conducted on the full text of books. However for comparison, attempting to extract the same traits from book description will yield an interesting avenue for comparison.

In order to investigate perspective and sentiment aspects that could, in the future, assist in the potential construction of a recommender system based primarily on data found in books instead of user review data we question ***What sentiment and perspective-related traits can be extracted from the contents of children's books and the descriptions thereof that could to inform the recommendation process targeting children of different ages?***

In our quest to address this concern, we outline the following RQs:

RQ1 Which perspective-related traits can be extracted from book descriptions and full text?

RQ2 Which sentiment traits can be extracted from book descriptions and full text?

RQ3 Can perspective-related traits be used to group books by age category?

RQ4 Can sentiment traits be used to group books by age category?

RQ5 How do the traits from the description differ from the traits found in the full text?

2 Related Work

In this section, we discuss related literature offering context and informing our work.

2.1 Sentiment analysis

Researchers have allocated efforts performing sentiment analysis on book content. Of note, Jacobs [13] produced an emotional profile for characters in the Harry Potter books using the SentiArt tool [16]. The study also classified text segments into categories, with the categories being: "Joyful," "Fearful," and "Neutral" [13]. While the results of this paper did show a possibility to create a sentiment profile for characters, it focused on the contents of a single book instead of comparing books based on sentiment. It also required manual selection of the characters to profile. This means it is less suited for large scale profiling. Furthermore, the classification of the text segments into only 3 categories limits the amount of data points the segments can be compared with. And while this was not the aim of that research, for the problem at hand we might require more points of comparison. Sentiment analysis is more commonly done in the field of social media. While this might not translate one-to-one to book analysis the amount of research done in the field is much greater. As such tool originally used for social media analysis might be useful for books as well. In a study performed by Bhooshjan et al. [5]. The sentiment of around 1000 tweets containing a certain hashtag was categorized. They used the text2emotion library to look at levels of different emotions in tweets [2]. The benefit of using text2emotion over a classifier is that the text2emotion library gives a weight for 5 different emotions, instead of classifying the input text into one of the 5. This allows for more nuanced comparison.

2.2 Perspective

As perspective is related to relatability one avenue of exploring perspective traits is by looking at gender identity in books. Because people of different gender relate to different perspectives that can be found in books. The influence of gender on reading habits also has some pre-existing studies. An examination by McGeown et al. [17] tried to find a connection between gender identity and reading habits. They found that feminine traits were more closely associated with reading motivation and engagement with neutral books when compared to masculine traits. However, this study looked at the extent to which readers identified with said traits, and not at the presence of said traits in books.

Another way to approach perspective is by looking at traits emerging from the writing style used in a book. One existing study by Wyvile et al. tries to find a connection between the narrative perspective of a book and the personal relationship the reader develops with the characters in the book [30]. The study mostly contrasts the writing style of children's literature with books written for adults. This leaves some room for exploration within the bounds of children's literature alone.

3 Methodology

Here we go into detail on what experiments we use to answers the research questions posed. Furthermore, we describe the data collection and processing method required to facilitate performing said experiments. To aid reproducibility, the code used has been published at the provided url: *https://github.com/Stevelet/sentiment-and-gender-trait-extraction.git*.

3.1 Experiments

To access what traits can be extracted we propose the following experiments to assert usefulness of potentially extracted traits. We define these experiments at the hand of the research questions posed.

Which perspective-related traits can be extracted from book descriptions and full text? Since the aim of the perspective-related traits is to find the ease with which the reader can relate with the protagonist/story that is being written, we will look at words in the books that indicate perspective. To accomplish this we create a lexicon of words to compare the book content to. Furthermore we compare different methods to extract names text and see if the implied gender of these names can be determined by performing an empirical examination on the results of these methods.

Which sentiment traits can be extracted from book descriptions and full text? Based on the research done into similar instances of exploration we found two methods of acquiring sentiment traits from text that seem promising for the sentiment trait extraction. These two are the Text2Emotion library and the SentiArt project. After gathering the required corpus we apply both of these and compare the acquired traits using a correlation test. This will indicate if the sentiment extraction depends heavily on the tool used or if the significance is not affected by the tool selected. Furthermore, in the case of fulltext, we look into the progression of sentiment traits throughout a book to see if any trends emerge when comparing sentiment across book chapters.

Can perspective-related traits be used to group books by age category? We hypothesise that books are written in different perspectives depending on the age of the target audience. To test this hypothesis we allocate the books in our dataset into age groups and look for trends arising from these groupings. By performing a Tukey HSD test to compare the groups we test for differences in mean value to find perspective traits that can provide significant distinction [19].

Can sentiment traits be used to group books by age category? As stated before children have a different relation to emotion at different ages. Because of this we try to extract traits from books that are targeted at children of different ages. To do this we compute the average distribution of some sentiment values across books found in these ages. Then we look at trends that can indicate a connection between the recommended age and the sentiment traits found in books for that age. After that we perform a Tukey HSD test to assert a difference in sentiment value between the different age groups. While this on its own cannot show if the traits can be used for distinction it can be a good indication of potential differentiation.

How do the traits from the description differ from the traits found in the full text? To investigate trait difference between those found in descriptions and the ones found in fulltext we use the results of the previous experiments. We look at the salient traits found in each of them and compare between the descriptions and fulltext. We also compare the mean values and the amount of significant instances of these mean values.

3.2 Age group selection

As previously stated, the interest of children in the type of books they read changes as they age. However, as not each child develops at the same speed [23] we aim to cluster books based on the age group they target instead of targeting specific individual ages. We select 4 different age ranges as our categories based on the following criteria: reading development and public library availability. As children age their reading skills develop, using this progression in skill development a couple of groups can be defined. The CDC defines the following stages of age progression: infant, toddler, preschooler, middle childhood, young teen and teenager [33]. With the corresponding age ranges as follows: 0-1, 2-3, 3-5, 6-8, 9-11, 12-14 and 15-17. When looking at the children's literature index found on Wikisource we find another group of age categories [4]. The ranges found here are 0+, 3+, 5+, 8+ and 12+. While these categories do not explicitly state an upper range, we use the start of the next category as the upper limit of each category, as it ensures no books occur in a category twice. By combining the given categories from both of these sources with the availability of books in the Wikisource database for their stated categories we arrive at the final set of age groups. The groups are as follows: [0-5), [5-8), [8-12) and 12+. Where [0-5) combines infants, toddlers and preschoolers. The group [5-8) represents early childhood. The [8-12) group captures children in middle childhood. And finally, the 12+ group, which represents the teenage stage. The infants, toddlers and preschooler group combination is needed in this case due to the limited set of books available for the ages 0, 1 and 2.

Age Category	**Wikisource**	**Goodreads**
[0-5)	31	1006
[5-8)	18	985
[8-12)	72	1010
12+	25	114

Table 1: Assembled TraitSet ordered by age groups

3.3 Data Collection

The empirical exploration we conduct is based on public domain books. We focus on these specific items as, in general, many books are protected from unpaid distribution by publishers. These books are gathered from Wikisource, a publicly hosted and maintained database of public domain books [4]. Wikisource has an index containing the children's literature hosted on the site. We collect all books found in this index as long, as the format of the book has been validated by Wikisource moderators, in order to ensure consistency between fulltext data points. We also gathered a dataset from Goodreads containing descriptions of children's books written by reviewers. For each English book that is contained in either of these two sources we collect the following: either a description or the fulltext of a book depending on if the source is wikisource or goodreads, the book title, the author, the recommended reading age and the year the book was published. We then grouped them on recommended reading age into the groups as described in the previous paragraph. This leaves us with a dataset of books as can bee seen in figure 1. From this point onward we will refer to this set of books as TraitSet .

From TraitSet the sentiment and perspective traits are extracted by the methods as described below.

3.4 Sentiment Trait Extraction

We extract sentiment traits from books in TraitSet using Text2Emotion, a frequently used tool [11; 6; 31]. This tool is a good fit for the analysis of sentiment traits as the reproducibility of the research is important for the potential implementation of a recommender system based on the results. Furthermore, because of the frequent usage, the results of this research are more comparable to the results other researchers have achieved.

Text2Emotion takes a paragraph of text and returns a vector of emotional value in a [0, 1] range for the emotions: anger, fear, happiness, sadness, and surprise. In this case, we apply the Text2Emotion tool to each chapter of each book in TraitSet . This allows us to both get a single average sentiment value for a book as well as an indication of how each sentiment progresses throughout said book. For the fulltext instances in TraitSet we split this book into chapters. Then we run each chapter through the Text2Emotion library. This gives us a sentiment value for each chapter in the book. In the case of a book description we get one sentiment vector for the entire description.

We go through the same process using the SentiArt project. Which gives us the mean sentiment value for each chapter and description in the same emotion categories.

3.5 Perspective Trait Extraction

For the perspective aspects of books we extract 6 different dependent values by creating a breakdown of the writing perspective of the text we analyse. These values are categorized as follows: first person words, second person words, third person male words, third person female words, male names and female names.
In the case of the first, second and third person words we can use a lexicon approach as options for first, second and third person words are finite. This means we look for the following words in the text analysed and map those words to their corresponding category as seen in table 2 based on the pronouns and perspective words used by Zibri-Hertz et al [32]. We once again separate the fulltext books in TraitSet into chapters and look for the words shown in Table 2. For each description in TraitSet we also gather said words.

Lexicon words	Mapped perspective word
i, my, mine, we, our, ours	first person
you, your, yours, they, their, theirs	second person
he, him, his	male third person
she, her, hers	female third person

Table 2: Different perspective words and their mapping to first, second and third person

For the male and female a lexicon approach is not sufficient as the names used in literature differ from those used by real people. This is especially the case in children's fantasy literature [8], a genre in children's literature preferred by around 55% of children [29].

As such we require a different approach. First we can use a Named Entity Recogniser model (NER), more specifically the Stanford NER model. This type of text analysis model is adept at labeling different words in a paragraph of text[20]. After filtering the dataset on words the NER recognises as names we apply another model, namely the Naive Bayes Classifier.

The Naive Bayes Classifier is trained to differentiate name gender. This is accomplished using a dataset containing 5.5 million names from the USA gathered between 1910 and 2013 and their most common associated gender as found on the Social Security Administration site of the USA [3]. This allows the model to classify the different names in the books into Male and Female. We use this classifier instead of using a lexicon based on the same data due to books containing names not included in the training set. While a classifier like this is generally used to classify into one specific category, we use the percentile distribution among different categories instead. This compensates for names that are gender ambiguous. An example of this is the name Ali, see excerpt in figure 1, which the classifier genders as female. While this name, in this context, should be classified as male. Using the percentile distributions helps mitigate this error.

Combining the perspective words found using the lexicon approach with the name genders extracted using the NER and Naive Bayes classifier we get a breakdown of the 6 proposed perspective values per book. We group these books using the age categories as described above.

Presently The Sheik came and parted the rugs. He glared through the dim light of the interior. The Sheik jerked his thumb toward Ali ben Kadin and addressed Meriem. "I am getting old," he said, "I shall not live much longer. Therefore I have given you to Ali ben Kadin, my brother."

Figure 1: Excerpt from "The Son of Tarzan"

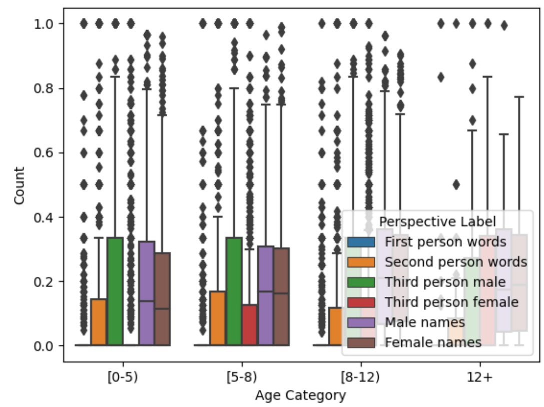

Figure 2: Book description gender/perspective grouped by age

3.6 Analysis

To answer the questions as posed in the introduction we need to ascertain the trends that emerge from the acquired traits. As state in the experiments we compare the means of the traits across the different categories and use a Tukey HSD test using the statsmodels python library [22] to find out if the difference between the means is in any way statistically significant. The tests as performed can be found in appendix A of this report.

4 Experimental Results and Discussion

Here we present the results of the analysis performed on TraitSet based on the experiments as described before.

4.1 Extracted perspective traits

By processing TraitSet using the methods and experiments described in the perspective traits part of the methodology section, we get a breakdown of the character gender distribution as well as a pronoun count for each text. This gives us 2 points of information: the gender that is likely the most prevalent in the book, and the perspective the book is probably told from. The perspectives can be visualized per age category to get an overview for the most common perspectives in each age category, see figure 3 and 2. We separate the fulltext and description perspectives into distinct graphs. Because talking about a character by name counts as a third person perspective mention, the male and female names are included in the graph for perspective.

Description perspective traits When looking at the figure 2 for the book descriptions one can see the descriptions mostly use names to talk about the characters occurring in

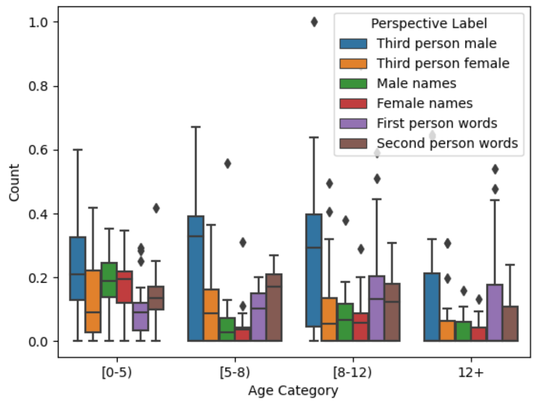

Figure 3: Book fulltext gender/perspective grouped by age

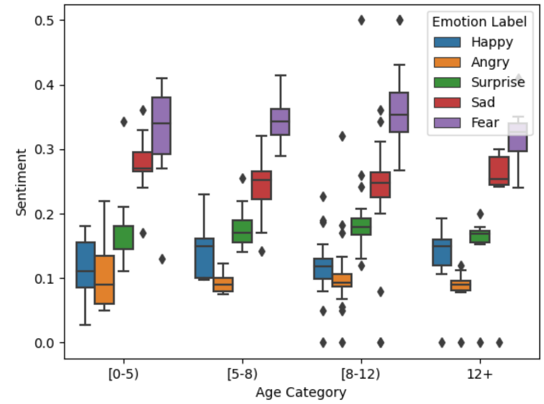

Figure 4: Sentiment values found per fulltext book, grouped by age.

books. Something of note is that the variance of usage differs greatly between the third person male and third person female pronouns. Furthermore, the usage of first person perspective is noticeably less frequent, which is most likely due to the nature of descriptions being told from the perspective of the description writer towards the characters in the book. When describing a book the author of said description might address the reader of the description, second person perspective, but it seems less common for a writer to talk about themselves.

Fulltext perspective traits By observing figure 3 a couple things stand out. It appears that the most common point of view used in books across all age categories is third person male. In each of the age categories in TraitSet male name are more common than female names. This most likely indicates that there are more books with a male protagonist than ones with a female protagonist. An interesting distinction can be seen between the [0-5) and [5-8) categories when regarding names. When moving from the former to the latter category, the relative usage of names goes down and the use of pronouns goes up. This could indicate that books for younger children are more inclined to using name instead of pronouns as the book has to reiterate names for clarity. We will look into the significance of this in the age category comparison discussion.

4.2 Extracted sentiment traits

Sentiment extraction tool comparison. After extracting traits using both Text2Emotion and SentiArt we compare the tools to see how much the extraction is dependant on the tool selected. We perform a correlation tests between the traits as found using both tools. The matrix representation for this can be seen in appendix B. For none of the traits the correlation coefficient goes above 0.25. This means the tool selected has a major impact on the results obtained. Going forward all of the results presented are the ones acquired with the Text2Emotion library, as empirical examination suggests that in the case of TraitSet the Text2Emotion method gives a more accurate representation of the emotion found in the corpus.

Fulltext sentiment traits When looking at the fulltext sentiment trait graph, figure 4, a couple thing stand out. First of all, when looking at the mean of the sentiment in each age

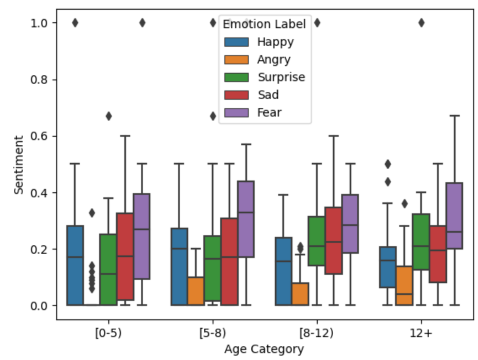

Figure 5: Sentiment values found per book description, grouped by age.

category, the fear sentiment is the most prevalent sentiment in each of them. The order of the most prevalent sentiments is also the same for each: fear followed in order by sadness, surprise, happiness and finally anger. The levels in which these sentiments appear in each category does vary per category.

Description sentiment traits Regarding figure 5 a couple observations can be made. Once again the fear sentiment is the most prevalent sentiment in each age category. The same order of prevalence also holds where fear is followed in order by sadness, surprise, happiness and finally anger. In general the sentiment values found in description are not very strong. While the extraction method gives a value between 0 and 1, all of the mean values are below the 0.4 mark. This could indicate that descriptions tend to be factually descriptive instead of emotional

Sentiment progression. In figure 6 the average progression throughout the books in the age category of [8-12). This is the only category with a large enough corpus to perform an analysis that spans multiple chapters. We hypothesized that there would be a decrease in the average sadness as the book progressed while the average happiness would increase when approaching the end. This can be observed in the figure. However, there is no indication of significance to support this trend.

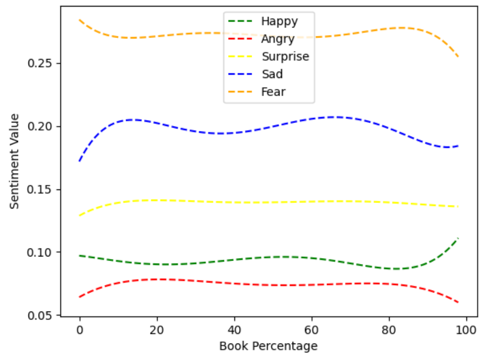

Figure 6: Sentiment progression in books for children between the ages of 8 and 12

4.3 Perspective traits for age grouping

As described in the methodology section we use a Tukey HSD test to compare the means of the perspective traits found in both the fulltext and description of the books in TraitSet . The results of these tests can be found in the appendix. The paragraphs below will include subsections of the performed tests containing significant details.

Group 1	Group 2	Perspective type
[0-5)	12+	Male names
[0-5)	[5-8)	Male names
[0-5)	[8-12)	Male names
[0-5)	12+	Female names
[0-5)	[5-8)	Female names
[0-5)	[8-12)	Female names
[0-5)	12+	Second person words
[5-8)	12+	Second person words
[8-12)	12+	Second person words
12+	[8-12)	Male third person pronouns
12+	[8-12)	Female third person pronouns

Table 3: Fulltext age group pairings with a significant difference in perspective value mean

Fulltext perspective traits When looking at the results of the mean difference test shown in figure 3 plenty of pairings with statistical significance appear. Interestingly for both the male and female names the difference between the [0-5) and all of the other pairings is shown to be significant. This likely mirrors the point discussed in the extracted perspective trait section where we observe a decrease in name usage as books are written for an older target audience. Another interesting observation can be made when looking at the second person perspective words. All of the pairings that show significance in the case of this trait are between the 12+ and another category. The mean value of the 12+ category for this trait lies at 0 while this is not the case for the other categories. This could allow this trait to be useful in age grouping based on perspective traits.

Description perspective traits When considering the significance tables for the perspective traits for book descriptions it becomes apparent that the difference between means is rarely different enough to be of note. The only cases where any significance can be observed is when looking at the female and male name usage, which shows a difference in mean when comparing the [8-12) category to the [0-5) and [5-8) categories. Interestingly the female third person pronouns show potential significance when comparing the [0-5) and 12+ categories. However, this is not the case for the male third person pronouns.

Group 1	Group 2	Perspective type
[0-5)	[8-12)	Male names
[5-8)	[8-12)	Male names
[0-5)	[8-12)	Female names
[5-8)	[8-12)	Female names
[0-5)	12+	Female third person pronouns

Table 4: Description age group pairings with a significant difference in perspective value mean

4.4 Sentiment traits for age grouping

To explore the trends that can be found in sentiment traits we compare said traits among the 4 age categories present in TraitSet . The following paragraphs explore these trends based on the previously shown figures for sentiment as well as the Tukey HSD results for tests performed on the extracted sentiment traits. The full results can be found in appendix A.

Group 1	Group 2	Emotion
[0-5)	[8-12)	Fear
[0-5)	[8-12)	Happy
[0-5)	[8-12)	Surprise
[5-8)	[8-12)	Surprise
12+	[8-12)	Anger
12+	[8-12)	Fear
12+	[8-12)	Happy
12+	[8-12)	Sad
12+	[8-12)	Surprise

Table 5: Fulltext age group pairings with a significant difference in sentiment mean

Fulltext sentiment traits. In order to investigate trends that could indicate distinctions between the categories in the fulltext we look different age group pairs that show significance difference in mean. Figure 5 shows us that the categories that have a significant difference in sentiment are mostly seen in differentiating the [8-12) category to other categories in the case of fulltext. All 5 sentiment values can potentially be used to differentiate books between the [8-12) and the 12+ category. 3 of the 5 sentiment values can potentially be used to differentiate the [0-5) and [8-12) categories. This means that there is no single emotions that can, according to this specific investigation, be used to differentiate between all 5 categories, as that would require this emotion to have a significant difference in mean in all 6 of the possible group pairings.

Description sentiment traits. When comparing sentiment traits found in book descriptions among age categories some

Group 1	Group 2	Emotion
[0-5)	[5-8)	Happy
[0-5)	[5-8)	Surprise
[0-5)	[8-12)	Happy
[0-5)	[8-12)	Surprise
[0-5)	12+	Happy
[5-8)	[8-12)	Happy
[5-8)	[8-12)	Surprise

Table 6: Description age group pairings with a significant difference in sentiment mean

trends arise. While the sentiment values follow the same order of prevalence seen in the fulltext sentiment analysis, the inter-category mean comparisons show different pairings. The cases where the means of a sentiment value was significantly different within a pairing of two age groups can be seen in figure 6. This figure shows us that the only two sentiments that show any significant difference in mean between age categories are happy and surprise. The happy sentiment shows a lot of promise in category distinction as it has a significant mean difference in 4 out of the 6 different age pairings.

4.5 Differences between traits found in fulltext and descriptions

Here we assess the differences between the traits between the two types of text sources. By comparing the prevalence of both perspective and sentiment traits among descriptions and fulltext we find some trends. Namely that for both perspective and sentiment the mean value is, on average, higher when looking at fulltext compared to descriptions. Another potential point of distinction arises from the results of the Tukey HSD analysis. The description perspective traits contained 5 significant pairings while the fulltext perspective traits contained 11 significant pairings. The same trend can be observed in the case of sentiment traits, although to a lesser extent. The description sentiment traits contain 7 significant pairings, while the fulltext sentiment traits contain 9 significant pairings.

5 Responsible Research

Public book collection The main challenge of doing large scale research on books is the collection of a corpus to do research on. In this specific analysis the required corpus consisted of the fulltext of public domain books. Of the 3 main sources of public domain books, Wikisource, Project Gutenberg [1] and Archive.org, wikisource was the only resource that had an api and allowed for automatic data gathering. Project Gutenberg does not explicitly forbid automated data gathering but does have a ban policy if frequent requests are detected, making it less excusable to scour their website for fulltext books. While gathering books from Project Gutenberg using automated tools would probably yield the largest and most usable formatted data the more ethical choice is Wikisource.

Reproducibility As mentioned before to aid in reproducible the code for the experimental setup used in this paper is available on GitHub. Furthermore, all of the datasets and libraries used in the research are publicly available.

"Once upon a time there were four little Rabbits, and their names were—Flopsy, Mopsy, Cotton-tail, and Peter. They lived with their Mother in a sand-bank, underneath the root of a very big fir-tree."

Figure 7: Excerpt from Tales of Peter Rabbit

I opened my eyes and looked around, trying to make out where I was. It was after sun-up, and I had been sound asleep. Pap was standing over me, looking sour—and sick, too. He says: "What you doin' with this gun?" I judged he didn't know nothing about what he had been doing, so I says: "Somebody tried to get in, so I was laying for him."

Figure 8: Excerpt from Huckleberry Finn

6 Discussion

In the following section, we discuss the acquired result in the context of the posed sub questions.

Which perspective-related traits can be extracted from book descriptions and full text? The perspective traits as extracted by our methodology seem consistent with empirical examinations of the books in TraitSet . The combination of NER and Naive Bayes also proves adequate at recognising and connecting gender to names. The names aswell as the pronouns in the category of first, second and third person can be extracted and do show some interesting trends.

Which sentiment traits can be extracted from book descriptions and fulltext? The method, as described in this paper, proves adequate at extracting anger, sadness, happiness, surprise and fear for each of the chapters in a book as-well as giving these values for the description of a book.

Can perspective-related traits be used to group books by age category? When looking at the perspective traits that were extracted, the place where the most notable age category distinction can be found is the male and female names. More specifically, in the case of fulltext when comparing the bucket containing the books for ages [0-5) to the one containing the ages [5-8). As discussed before the usage of names in books relative to pronouns used goes down noticeably between these categories. A potential hypothesis for this phenomenon is that books for young children are commonly fairy tales or poems about characters when situations that happen to said character are described, see figure 7. While a book for a slightly older age category might be more focused on following the point of view of a character instead of describing what happens to them, see figure 8.

Can sentiment traits be used to group books by age category? When looking at the traits extracted from the fulltext, a notable distinction can be made between the age categories of [8-12) and 12+/[0-5), which also happen to be the categories containing the most fulltext books. This leads us to believe that the method could aid in potential recommendation when applied to a larger corpus.

How do the traits from the description differ from the traits found in the full text? Notable differences and similarities can be regarded when looking at the sentiment values of the fulltext sentiment compared to the sentiment found in book descriptions. One thing of note is the wider range of values in the descriptions versus the fulltext. One reason for the difference could be the writing style of children's book authors versus a more varied range of styles employed by review writers. Another reason for this difference might lie in the fact that the books used for the fulltext analysis were, on average written, in 1891, while the average publishing date of the books that the descriptions were written for lies around 2005. Additionally, book descriptions may only highlight certain aspects of the book, leaving out important plot points to prevent the reader from learning about critical plot points found in the book. This could lead in a difference in sentiment between the book and the description thereof.

When looking at difference in perspective traits when comparing the book fulltext and descriptions it becomes apparent that they differ greatly. This is one again most likely due to a difference in writing style. Furthermore, when describing characters a writer is more likely to use the 3rd person perspective for a description.

When looking at the quantity of traits that show statistical significance the observation can be made that more of the traits extracted from the fulltext show a potential significance when compared to the descriptions. This hold for both the sentiment and the perspective traits. This would suggest that, in general, analysis on fulltext is more fruitful when the aim of the analysis is to aid in categorizing books by age.

7 Limitations

Below we briefly mention limitations that we observed in conducting the proposed research work.

Book dataset imperfections The dataset collected from Wikisource, while public, is fairly limited. The collected dataset contained 146 books, which led to a need to combine age categories. Because the books are submitted to a public site by volunteer moderators the books are not consistently formatted, making for quite a bit of malformed data. While it would definitely be possible to write a more effective sanitizer the fulltext as used in the current analysis has some imperfections, this includes inconsistent line breaks, which could have influence on the sentiment analysis performed.

Foreign names in books The Naive Bayes classifier performs very well on names commonly occurring in America it struggles to correctly classify names that are not in that subset. Some of the books in TraitSet are from countries that have names not commonly present in America and as such the gender trait data collected is definitely framed by the limitations of the dataset used to train the gender classifier. We compensated for this using the percentile distribution, however in the future a different approach could be explored.

8 Conclusions and Future Work

This paper aims find What sentiment and perspective-related traits that can be extracted from the contents of children's books and the descriptions thereof. The findings show that potential sentiment traits that can be extracted from fulltext and descriptions are Anger, Fear, Happiness, Sadness, and Surprise. While these traits do not provide any significant information when looking at individual age groups. However, some evidence of inter-group relations derived from these traits can be found. These trends seems more apparent from the fulltext traits when compared to the description traits. Sentiment progression is not found to be useful as of right now but does lend itself to future study.

Pronouns and name gender can be sufficiently extracted from both fulltext and descriptions in the form of first, second and third person perspective. Furthermore names can be recognized and categorized by gender. While these traits seem to be of potential use when considering fulltext the same cannot be said so easily when looking at those same traits obtained from book descriptions.

Two main things can be concludes from looking at the examination performed in this paper. Firstly, there are significant perspective and sentiment traits to be found in both book description as well as fulltext. Secondly this paper showcases the usefulness and importance of fulltext when compared to descriptions. All of the comparisons made in this paper suggest a real use for fulltext in the field of book recommendation. We encourage future work focusing on obtaining the proposed traits from a larger corpus.

In the wrapping paragraphs we explore a couple of options for future work to expand upon our claim made in the previous paragraph.

Different data source. While the source we use in this paper is the natural choice for performing the analysis at hand within the given time frame, it would be interesting to see the results if the same process were to be applied to a data set found on a data source with a larger corpus. The same analysis applied to this larger corpus could lead to finding more traits that show significance.

Sentiment progression. The results obtained from the sentiment progression visualization based on the books in the [8-12) category seem promising. And while, due to a lack of fulltext books, no claims can be made about the usefulness of the sentiment progression as a point of data in book recommendation this aspect seems promising and calls for further investigation. Applying the described methodology to a larger corpus, as mentioned above, could lead to interesting results.

References

[1] Project gutenberg. (n.d.). www.gutenberg.org, 2023.

[2] aman2656. Text2emotion. https://github.com/aman2656/text2emotion-library, 2022.

[3] SSA America. Popular baby names in the usa. https://www.ssa.gov/OACT/babynames/limits.html.

[4] Timothy K Armstrong. Rich texts: Wikisource as an open access repository for law and the humanities. *U of Cincinnati Public Law Research Paper*, (10-09), 2010.

[5] Sooraj Bhooshan, R Praveen Pai, and R Nandakumar. A sentiment analysis of a boycott movement on twit-

ter. In *Computer Networks and Inventive Communication Technologies*, pages 313–324. Springer, 2022.

[6] Sorelys Sandoval Díaz, Jessied María Marriaga Shaik, Juan Camilo Gómez Santofimio, et al. Intelligent execution of behaviors in a nao robot exposed to audiovisual stimulus. In *2018 IEEE 2nd Colombian Conference on Robotics and Automation (CCRA)*, pages 1–6. IEEE, 2018.

[7] Julie Dwyer and Susan B Neuman. Selecting books for children birth through four: A developmental approach. *Early Childhood Education Journal*, 35:489–494, 2008.

[8] Lincoln Fernandes. Translation of names in children's fantasy literature: Bringing the young reader into play. *New voices in translation studies*, 2(2):44–57, 2006.

[9] Anupama Garla. Story-times: Building a content-based children's book recommender geared towards parents. *Towards Data Science*, 2020.

[10] Emilia Gómez, Vicky Charisi, and Stéphane Chaudron. Evaluating recommender systems with and for children: towards a multi-perspective framework. In *Perspectives@ RecSys*, 2021.

[11] Bhoomika Gupta and Sunita Daniel. Sentiment analysis and topic modeling from tweets about the covid-19 vaccine. *Mathematical Statistician and Engineering Applications*, 71(3s):675–696, 2022.

[12] Catherine M Herba, Sabine Landau, Tamara Russell, Christine Ecker, and Mary L Phillips. The development of emotion-processing in children: Effects of age, emotion, and intensity. *Journal of Child Psychology and Psychiatry*, 47(11):1098–1106, 2006.

[13] Arthur M Jacobs. Sentiment analysis for words and fiction characters from the perspective of computational (neuro-) poetics. *Frontiers in Robotics and AI*, 6:53, 2019.

[14] Hilda K Kabali, Matilde M Irigoyen, Rosemary Nunez-Davis, Jennifer G Budacki, Sweta H Mohanty, Kristin P Leister, and Robert L Bonner Jr. Exposure and use of mobile media devices by young children. *Pediatrics*, 136(6):1044–1050, 2015.

[15] Pasquale Lops, Marco De Gemmis, and Giovanni Semeraro. Content-based recommender systems: State of the art and trends. *Recommender systems handbook*, pages 73–105, 2011.

[16] matinho13. Sentiart. https://github.com/matinho13/SentiArt, 2021.

[17] Sarah McGeown, Hannah Goodwin, Nikola Henderson, and Penelope Wright. Gender differences in reading motivation: Does sex or gender identity provide a better account? *Journal of Research in Reading*, 35(3):328–336, 2012.

[18] Yiu-Kai Ng. Cbrec: a book recommendation system for children using the matrix factorisation and content-based filtering approaches. *International Journal of Business Intelligence and Data Mining*, 16(2):129–149, 2020.

[19] Tetsuya Sakai. Statistical reform in information retrieval? In *ACM SIGIR Forum*, volume 48, pages 3–12. ACM New York, NY, USA, 2014.

[20] Branislava Šandrih, Cvetana Krstev, and Ranka Stankovic. Development and evaluation of three named entity recognition systems for Serbian - the case of personal names. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 1060–1068, Varna, Bulgaria, September 2019. INCOMA Ltd.

[21] J Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. Collaborative filtering recommender systems. In *The adaptive web*, pages 291–324. Springer, 2007.

[22] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[23] Linda Kreger Silverman. The construct of asynchronous development. In *Charting a New Course in Gifted Education*, pages 36–58. Routledge, 2017.

[24] T J Slee. Who is the average goodreads user? https://www.goodreads.com/author_blog_posts/14538341-who-is-the-average-goodreads-user-you-ll-be-surprised, Jan 2017.

[25] Xiaoyuan Su and Taghi M Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009, 2009.

[26] Robin Van Meteren and Maarten Van Someren. Using content-based filtering for recommendation. In *Proceedings of the machine learning in the new information age: MLnet/ECML2000 workshop*, volume 30, pages 47–56, 2000.

[27] Ike Vayansky and Sathish AP Kumar. A review of topic modeling methods. *Information Systems*, 94:101582, 2020.

[28] Mengting Wan, Rishabh Misra, Ndapa Nakashole, and Julian McAuley. Fine-grained spoiler detection from large-scale review corpora. *arXiv preprint arXiv:1905.13416*, 2019.

[29] Amy Watson. Preferred book genre among children by age group and type in italy 2018.

[30] Andrea Schwenke Wyile. The value of singularity in first-and restricted third-person engaging narration. *Children's Literature*, 31(1):116–141, 2003.

[31] Saijun Zhang, Meirong Liu, Yeefay Li, and Jae Eun Chung. Teens' social media engagement during the covid-19 pandemic: A time series examination of posting and emotion on reddit. *International Journal of Environmental Research and Public Health*, 18(19):10079, 2021.

[32] Anne Zribi-Hertz. Anaphor binding and narrative point of view: English reflexive pronouns in sentence and discourse. *Language*, pages 695–727, 1989.

[33] Jennifer M Zubler, Lisa D Wiggins, Michelle M Macias, Toni M Whitaker, Judith S Shaw, Jane K Squires,

Julie A Pajek, Rebecca B Wolf, Karnesha S Slaughter, Amber S Broughton, et al. Evidence-informed milestones for developmental surveillance tools. *Pediatrics*, 149(3), 2022.

A Tukey HSD test results

The tables below show the results of some Tukey HSD tests performed on the sentiment and gender/perspective trait data displayed in the figures in the result section. The "meandiff" column shows the difference in means between the groups in each row. Based on the "p-adj" column and the "lower" and "upper" bounds columns the null hypothesis of there not being a difference in mean between the groups can be rejected. The rejection can be seen in the "reject column" and indicates potential statistical significance.

A.1 Description sentiment traits

Table 7: Description happy tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	0.0837	0.0003	0.031	0.1365	True
12+	[5-8)	0.0526	0.0516	-0.0002	0.1054	False
12+	[8-12)	0.0219	0.7096	-0.0309	0.0747	False
[0-5)	[5-8)	-0.0312	0.0002	-0.0506	-0.0118	True
[0-5)	[8-12)	-0.0618	0.0	-0.0811	-0.0426	True
[5-8)	[8-12)	-0.0307	0.0003	-0.05	-0.0113	True

Table 8: Description angry tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	-0.0059	0.9747	-0.042	0.0301	False
12+	[5-8)	-0.0138	0.7579	-0.0499	0.0223	False
12+	[8-12)	-0.0169	0.6244	-0.0529	0.0192	False
[0-5)	[5-8)	-0.0079	0.4172	-0.0212	0.0053	False
[0-5)	[8-12)	-0.011	0.1407	-0.0241	0.0022	False
[5-8)	[8-12)	-0.0031	0.9341	-0.0163	0.0102	False

Table 9: Description surprise tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	-0.0336	0.3763	-0.0875	0.0203	False
12+	[5-8)	-0.0018	0.9998	-0.0557	0.0521	False
12+	[8-12)	0.0183	0.8195	-0.0356	0.0722	False
[0-5)	[5-8)	0.0318	0.0002	0.012	0.0516	True
[0-5)	[8-12)	0.0519	0.0	0.0322	0.0716	True
[5-8)	[8-12)	0.0201	0.0451	0.0003	0.0399	True

Table 10: Description sad tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	0.0423	0.1968	-0.0127	0.0973	False
12+	[5-8)	0.0365	0.3209	-0.0185	0.0915	False
12+	[8-12)	0.0287	0.5353	-0.0262	0.0837	False
[0-5)	[5-8)	-0.0058	0.8824	-0.026	0.0144	False
[0-5)	[8-12)	-0.0136	0.3047	-0.0336	0.0065	False
[5-8)	[8-12)	-0.0078	0.7549	-0.028	0.0124	False

Table 11: Description fear tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	-0.0521	0.157	-0.1162	0.012	False
12+	[5-8)	-0.0398	0.3821	-0.1039	0.0244	False
12+	[8-12)	-0.0311	0.5977	-0.0951	0.033	False
[0-5)	[5-8)	0.0123	0.5356	-0.0112	0.0359	False
[0-5)	[8-12)	0.021	0.0962	-0.0024	0.0444	False
[5-8)	[8-12)	0.0087	0.7763	-0.0148	0.0322	False

A.2 Fulltext sentiment traits

Table 12: Fulltext happy tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	0.0096	0.9327	-0.0323	0.0516	False
12+	[5-8)	0.0171	0.7942	-0.0312	0.0653	False
12+	[8-12)	0.0455	0.0073	0.0093	0.0817	True
[0-5)	[5-8)	0.0074	0.9754	-0.0388	0.0536	False
[0-5)	[8-12)	0.0359	0.0306	0.0024	0.0694	True
[5-8)	[8-12)	0.0285	0.2773	-0.0126	0.0696	False

Table 13: Fulltext angry tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	0.0172	0.6098	-0.0192	0.0535	False
12+	[5-8)	0.0092	0.9396	-0.0326	0.051	False
12+	[8-12)	0.044	0.0021	0.0126	0.0754	True
[0-5)	[5-8)	-0.0079	0.9553	-0.048	0.0321	False
[0-5)	[8-12)	0.0268	0.0823	-0.0023	0.0558	False
[5-8)	[8-12)	0.0347	0.0591	-0.0009	0.0704	False

Table 14: Fulltext surprise tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	0.0216	0.7371	-0.0334	0.0767	False
12+	[5-8)	0.0236	0.7663	-0.0397	0.0869	False
12+	[8-12)	0.0792	0.0002	0.0317	0.1268	True
[0-5)	[5-8)	0.002	0.9998	-0.0587	0.0627	False
[0-5)	[8-12)	0.0576	0.0047	0.0136	0.1016	True
[5-8)	[8-12)	0.0556	0.0407	0.0016	0.1095	True

Table 15: Fulltext sad tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	0.0374	0.643	-0.0453	0.1202	False
12+	[5-8)	0.0262	0.8908	-0.069	0.1214	False
12+	[8-12)	0.0997	0.0022	0.0283	0.1712	True
[0-5)	[5-8)	-0.0112	0.9886	-0.1025	0.08	False
[0-5)	[8-12)	0.0623	0.0728	-0.0038	0.1284	False
[5-8)	[8-12)	0.0735	0.0905	-0.0076	0.1547	False

Table 16: Fulltext fear tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	0.0008	1.0	-0.1213	0.1229	False
12+	[5-8)	-0.0014	1.0	-0.1418	0.139	False
12+	[8-12)	0.116	0.0248	0.0106	0.2215	True
[0-5)	[5-8)	-0.0022	1.0	-0.1369	0.1324	False
[0-5)	[8-12)	0.1152	0.0135	0.0176	0.2128	True
[5-8)	[8-12)	0.1175	0.0566	-0.0022	0.2372	False

A.3 Description gender/perspective traits

Table 17: Description first person words tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	-0.0205	0.5645	-0.0611	0.0201	False
12+	[5-8)	-0.0171	0.7001	-0.0578	0.0235	False
12+	[8-12)	-0.0213	0.533	-0.0619	0.0193	False
[0-5)	[5-8)	0.0034	0.9374	-0.0115	0.0183	False
[0-5)	[8-12)	-0.0008	0.9991	-0.0156	0.0141	False
[5-8)	[8-12)	-0.0042	0.8904	-0.0191	0.0108	False

Table 18: Description second person words tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	0.0622	0.165	-0.0152	0.1396	False
12+	[5-8)	0.0678	0.1099	-0.0096	0.1453	False
12+	[8-12)	0.0435	0.4722	-0.0339	0.1208	False
[0-5)	[5-8)	0.0057	0.9561	-0.0228	0.0341	False
[0-5)	[8-12)	-0.0187	0.3227	-0.047	0.0095	False
[5-8)	[8-12)	-0.0244	0.1218	-0.0528	0.004	False

Table 19: Description third person male tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	0.0035	0.9996	-0.0865	0.0935	False
12+	[5-8)	0.0191	0.9482	-0.071	0.1091	False
12+	[8-12)	0.0048	0.9991	-0.0852	0.0948	False
[0-5)	[5-8)	0.0156	0.6211	-0.0175	0.0486	False
[0-5)	[8-12)	0.0013	0.9996	-0.0316	0.0341	False
[5-8)	[8-12)	-0.0143	0.683	-0.0473	0.0188	False

Table 20: Description third person female tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	-0.079	0.0313	-0.153	-0.0049	True
12+	[5-8)	-0.0634	0.1238	-0.1375	0.0107	False
12+	[8-12)	-0.0616	0.141	-0.1357	0.0124	False
[0-5)	[5-8)	0.0156	0.455	-0.0116	0.0428	False
[0-5)	[8-12)	0.0173	0.3517	-0.0097	0.0444	False
[5-8)	[8-12)	0.0018	0.9984	-0.0254	0.029	False

Table 21: Description male names tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	-0.0448	0.28	-0.1094	0.0197	False
12+	[5-8)	-0.0413	0.3544	-0.1059	0.0233	False
12+	[8-12)	-0.0046	0.9978	-0.0691	0.0599	False
[0-5)	[5-8)	0.0036	0.9803	-0.0201	0.0273	False
[0-5)	[8-12)	0.0403	0.0001	0.0167	0.0638	True
[5-8)	[8-12)	0.0367	0.0004	0.013	0.0604	True

Table 22: Description female names tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	-0.0437	0.2572	-0.105	0.0175	False
12+	[5-8)	-0.0222	0.7876	-0.0835	0.0391	False
12+	[8-12)	0.0126	0.9518	-0.0486	0.0739	False
[0-5)	[5-8)	0.0215	0.0675	-0.001	0.044	False
[0-5)	[8-12)	0.0564	0.0	0.034	0.0787	True
[5-8)	[8-12)	0.0349	0.0004	0.0124	0.0573	True

A.4 Fulltext gender/perspective traits

Table 23: Fulltext third person male tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	0.1135	0.1721	-0.0299	0.257	False
12+	[5-8)	0.1215	0.2015	-0.0383	0.2813	False
12+	[8-12)	0.1605	0.0038	0.0403	0.2807	True
[0-5)	[5-8)	0.0079	0.9992	-0.1493	0.1652	False
[0-5)	[8-12)	0.047	0.7226	-0.0699	0.1639	False
[5-8)	[8-12)	0.0391	0.8787	-0.0973	0.1755	False

Table 24: Fulltext third person female tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	0.0892	0.0217	0.0095	0.169	True
12+	[5-8)	0.057	0.3431	-0.0318	0.1459	False
12+	[8-12)	0.0399	0.4089	-0.0269	0.1067	False
[0-5)	[5-8)	-0.0322	0.7736	-0.1196	0.0552	False
[0-5)	[8-12)	-0.0493	0.2023	-0.1143	0.0156	False
[5-8)	[8-12)	-0.0172	0.9355	-0.093	0.0587	False

Table 25: Fulltext male names tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	0.1603	0.0	0.1034	0.2171	True
12+	[5-8)	0.0394	0.3719	-0.0239	0.1027	False
12+	[8-12)	0.0432	0.0904	-0.0045	0.0908	False
[0-5)	[5-8)	-0.1209	0.0	-0.1832	-0.0585	True
[0-5)	[8-12)	-0.1171	0.0	-0.1634	-0.0708	True
[5-8)	[8-12)	0.0038	0.9979	-0.0503	0.0578	False

Table 26: Fulltext female names tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	0.1603	0.0	0.093	0.2276	True
12+	[5-8)	0.0243	0.8336	-0.0506	0.0992	False
12+	[8-12)	0.0467	0.1408	-0.0096	0.1031	False
[0-5)	[5-8)	-0.136	0.0	-0.2097	-0.0622	True
[0-5)	[8-12)	-0.1135	0.0	-0.1683	-0.0587	True
[5-8)	[8-12)	0.0224	0.7982	-0.0415	0.0864	False

Table 27: Fulltext first person words tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	-0.0008	1.0	-0.0933	0.0917	False
12+	[5-8)	-0.0122	0.9898	-0.1152	0.0908	False
12+	[8-12)	0.0473	0.3888	-0.0302	0.1249	False
[0-5)	[5-8)	-0.0114	0.9913	-0.1128	0.0901	False
[0-5)	[8-12)	0.0482	0.3478	-0.0272	0.1235	False
[5-8)	[8-12)	0.0595	0.2969	-0.0284	0.1475	False

Table 28: Fulltext second person words tukey test

group1	group2	meandiff	p-adj	lower	upper	reject
12+	[0-5)	0.0805	0.0062	0.0175	0.1434	True
12+	[5-8)	0.0766	0.0263	0.0065	0.1468	True
12+	[8-12)	0.0629	0.0125	0.0101	0.1156	True
[0-5)	[5-8)	-0.0038	0.9989	-0.0729	0.0652	False
[0-5)	[8-12)	-0.0176	0.8092	-0.0689	0.0337	False
[5-8)	[8-12)	-0.0138	0.9324	-0.0737	0.0461	False

B Tool correlation matrix

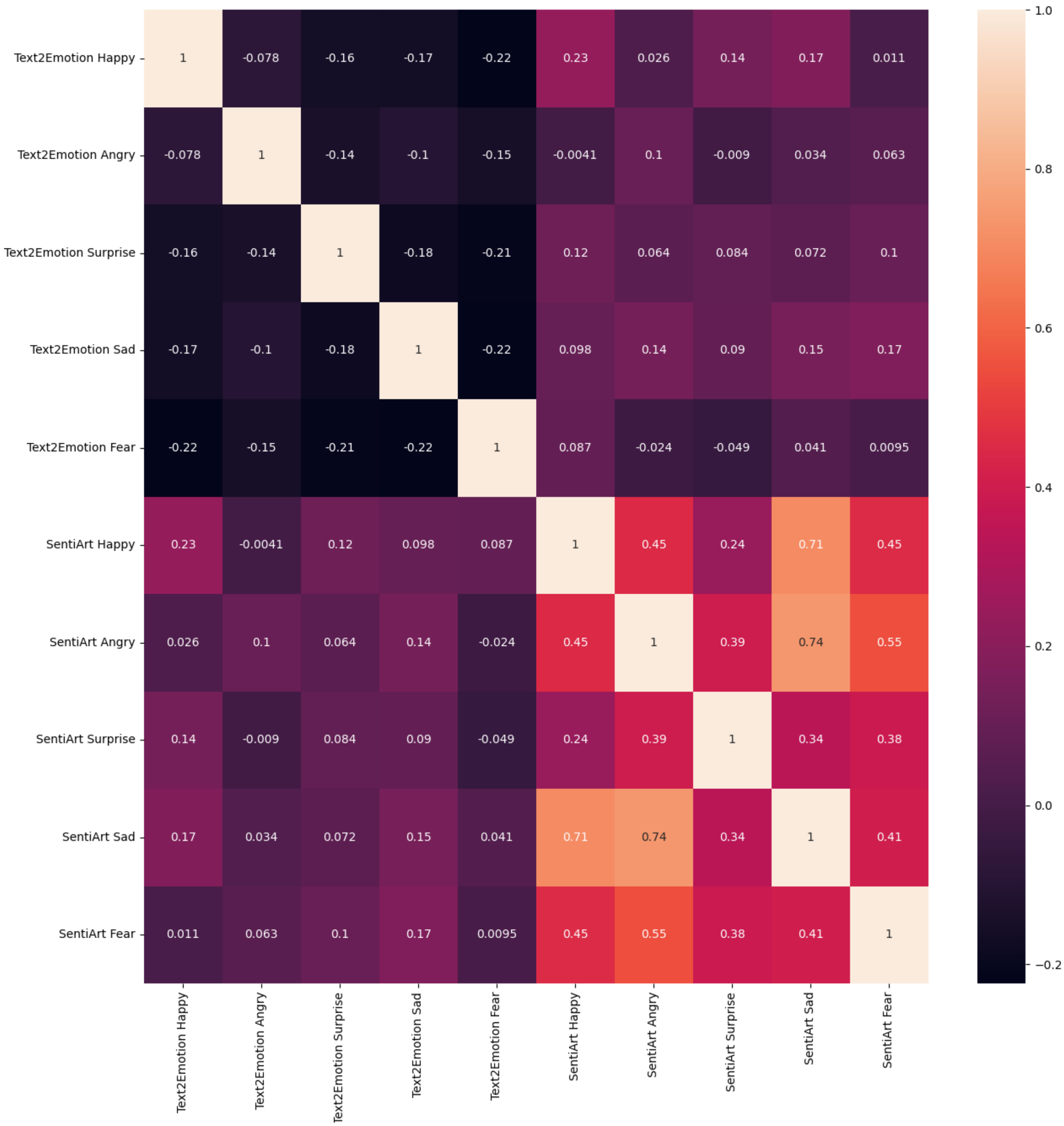

Figure 9: Correlation matrix for SentiArt and Text2Emotion