KU LEUVEN
FACULTEIT INGENIEURSWETENSCHAPPEN

Designing transparent visualization techniques for mobile book recommender systems

Felix Cammaerts

Thesis submitted for the degree of Master of Science in Engineering: Computer Science, option Artificial Intelligence

Thesis supervisor:
Prof. dr. K. Verbert

Assessors:
Dr. D. Sileo
Dr. A. Rafique

Mentor:
Dr. Francisco Gutiérrez

Academic year 2020 – 2021

Preface

This thesis is my last deadline in my Computer Science master's degree. It was unfortunately a tough year where the consequence of the COVID-19 pandemic were omnipresent. Classes being more often online than on-campus, not seeing my friends, not being able to go eat at Alma, etc.

These consequences could also be felt in the research done here. The initial research goal was to use the Hololens of the research group and do further research on the PHARA grocery store assistant. I spent a few months trying my best to get this research going. Experimenting with the Hololens, contacting a lot of physical stores as possible research places, literature study on guidance systems, AR development in Unity etc. However in the end it was not considered feasible anymore to continue in this direction as the COVID-19 situation was not coming to a an end in the foreseeable future.

This meant I had to look for a different research in which I would be able to do the evaluation in a completely remote manner. The research I went for allowed to do the evaluation completely remotely by developing a mobile application. Further possibility to expand this to an AR environment was still possible. In fact the library of CBA was willing to let me do AR research at their library. Henceforth the reason that in this research a mobile book recommender application is developed. Nonetheless SMEC approval for this would have been too challenging and henceforth I continued with just developing a mobile phone application.

Looking back at how the development of the mobile phone application went I still had a lot of fun even though this was not the initially intended research. I found it very interesting to learn how to use Meteor and to actually deploy my own mobile phone application. I am almost certain that I will be creating more applications in my spare time.

For me it was also very nice to be introduced to the Augment research group. My mentor Dr. Francisco Gutiérrez helped me out a lot during each and every phase of the research. Starting from finding a good research question up to the point of finalizing the text. I would not have been able to have done it without him.

I also enjoyed the user-centered design in which I was able to conduct interviews to evaluate my application. This was a nice way of still seeing people during the pandemic and not feeling completely isolated.

The courses Fundamenten van de Mens-Machine Interactie and Informatie Visualisatie also helped me a lot with the work done in this research. For which I would also like to thank Prof. dr. K. Verbert for providing us with these very interesting and fun courses.

Lastly I would of course like to thank all the people who participated in my user studies and provided my with very important critical feedback and suggestions.

Contents

Preface **1**

Contents **3**

Samenvatting **5**

Abstract **7**

List of Figures **11**

List of Tables **14**

1 Introduction **17**

2 Related work **21**
- 2.1 Recommender systems 21
- 2.2 Visualization techniques 22
- 2.3 Book recommender applications 31

3 Design process **37**
- 3.1 User-centered design 37
- 3.2 Visualizations 37

4 Pilot study **43**
- 4.1 Evaluation of visualizations 43
- 4.2 Prototype of application 46
- 4.3 Summary 50

5 Implementation **53**
- 5.1 Code 53
- 5.2 Dataset 53
- 5.3 Database 54
- 5.4 Recommender system 54
- 5.5 Cold-start problem 55
- 5.6 Visualizations 58
- 5.7 Deployment 64

6 First user study **65**
- 6.1 Tasks 65
- 6.2 Questionnaire 67
- 6.3 Results 67

7 Final user study **77**
7.1 Interview setup 77
7.2 Results 82
8 Conclusion **103**
9 Further research **107**
9.1 Recommender system 107
9.2 Interactive visualizations 107
9.3 New/existing visualizations 107
Bibliography **109**

Samenvatting

Aanbevelingssystemen worden vaak als zwarte dozen beschouwd. Ze bieden de gebruikers aanbevelingen aan op basis van hun vorig voorkeuren maar bieden geen uitleg aan waarom dat deze aanbevelingen precies gemaakt zijn. Dit verlaagt de gebruikerstevredenheid met betrekking tot aanbevelingssystemen.

Een mogelijk oplossing om deze gebruikerstevredenheid toch te verhogen is door het aanbieden van een visuele uitleg aan de gebruiker om uit te leggen waarom aanbevelingen gemaakt werden. Deze visuele uitleggen worden visualisaties genoemd. Er bestaan al veel visualisaties voor het uitleggen van aanbevelingen naar de gebruikers toe. Echter zijn de meeste van deze visualisatie hoofdzakelijk op tekst gebaseerd en gemaakt voor computer applicaties.

Dit onderzoek stelt vijf visualisaties voor gebaseerd op de literatuur en evalueert deze in termen van hun informatie adequaatheid, transparantie en overtuigingskracht. Deze visualisaties worden geëvalueerd in de context van een volledig werkzaam boek aanbevelingsapplicatie. Deze visualisaties maken gebruik van de *features* van het aanbevelingssysteem, in dit geval de genres van de boeken. Om het effect van het gebruik van deze *features* te kunnen begrijpen werd er ook een *baseline* visualisatie voorgesteld. Deze bevat geen *features* in zijn uitleg.

Sommige van deze visualisaties geven een uitleg op een numerische manier. Dit betekent dat ze getallen bevatten zoals bijvoorbeeld een percentage match. Andere visualisaties geven een uitleg op een categorische manier. Bijvoorbeeld een visualisatie dat enkel genres toont, zonder hier een getal mee te associëren, wordt aanzien als een categorische visualisatie.

De onderzoeksvragen luiden als volgt:

RQ1. Zorgt het gebruik van een visualisatie als uitleg voor een aanbeveling voor een verhoogde gebruikerstevredenheid van het aanbevelingssysteem?

RQ2. Zorgen de vijf voorgestelde visualisaties voor een beter informatie adequaatheid, transparantie en overtuigingskracht dan de *baseline* visualisatie?

RQ3. Is er voor de eindgebruiker een voorkeur naar een numerische of een categorische visualisatie?

In dit onderzoek werd er gebruik gemaakt van een *user-centered design* waarin de feedback van de eindgebruiker van belang is in elke iteratie van het ontwikkelingsproces.

De visualisaties zijn losjes gebaseerd op visualisaties gevonden in de literatuur, terwijl de applicatie zelf losjes gebaseerd is op andere bestaande boek aanbevelingssystemen zoals Amazon en goodreads books. Het aanbevelingssysteem zelf is gebaseerd op *content-based filtering*.

De voorgestelde visualisaties voor dit onderzoek zijn:

- *Bar charts*: staafdiagram dat voor elk genre toont hoeveel de gebruiker zijn voorkeuren overeenkomen met dat genre.
- Venn diagram: Venn diagram dat toont welke genres het aangeboden boek bevat en welke genres tot de gekende voorkeuren van de gebruiker behoren. In de overlap worden de overeenkomstige genres getoond.
- *Other books*: gerelateerde boeken worden getoont die de gebruiker in het verleden een goede rating heeft gegeven.
- *Double bar chart*: twee soorten staafdiagrammen worden getoond voor elk genre. Eén toont de de match van de gebruiker met dat genre. De andere toont de match van het aanbevolen boek met dat genre.
- *Link strength*: de zelfde informatie als bij de *double bar chart*, maar deze wordt hier op een andere manier weergegeven. Deze manier is meer compact, maar moeilijker te lezen.

Een *low-fidelity* prototype werd gemaakt voor de visualisaties en de applicatie. Dit *low-fidelity* prototype werd geëvalueerd in een initiële studie om feedback erover te krijgen. Zowel feedback over de visualisaties als de applicatie werd verzameld. Om de visualisaties te evalueren werd een vragenlijst opgesteld die door 52 mensen werd ingevuld. Voor de applicatie werd een think-aloud study uitgevoerd met vijf deelnemers.

Met de verkregen feedback werd het *low-fidelity* prototype van de applicatie en de visualisatie aangepast. Met deze aangepaste versie werd het *high-fidelity* prototype gemaakt. Dit *high-fidelity* prototype werd geëvalueerd met een think-aloud study. Hier werden 16 deelnemers voor gerekruteerd. Deze studie hielp bij het vinden van overblijvende gebruiksproblemen in de applicatie, de visualisaties en hun achterliggende algoritmes.

De feedback verkregen uit deze gebruikersstudie werd dan op zijn beurt gebruikt om het *high-fidelity* prototype aan te passen naar de finale versie. Met deze finale versie werden de visualisaties zelf geëvalueerd. Dit werd gedaan door 51 deelnemers te interviewen en hun mening te vragen over de visualisaties. Na het interview werden ze ook gevraagd een vragenlijst in te vullen.

De visualisaties doen het allemaal beter dan de *baseline* op het vlak van interactie adequaatheid en transparantie. Echter enkel de Venn diagram en *link strength* visualisaties doen het beter op het vlak van overtuigingskracht dan de *baseline* visualisatie. Uit de interview data dat verzameld werd was het mogelijk om een *thematic analysis* uit te voeren. Uit deze *thematic analysis* was het mogelijk om richtlijnen voor het ontwikkelen van nieuwe visualisaties te vinden. Geen duidelijke voorkeur naar categorische of numerische visualisaties werd gevonden. Voor elke visualisatie werden zwaktes en sterktes gevonden.

Verder onderzoek kan gebruik maken van deze richtlijnen bij het ontwerpen van (nieuwe) visualisaties voor mobiele applicaties. Het is ook mogelijk om verder onderzoek te doen op de voorgestelde visualisaties met andere aanbevelingsalgoritmes en het interactief maken van deze visualisaties.

Abstract

Recommender systems are often considered black-box systems. They provide users with recommendations based on their previous preferences but provide no explanation as to why these recommendations have been made. This reduces the user satisfaction with regards to recommender systems.

One possible way to increase this user satisfaction is by providing the user a visual explanation as to why the recommendations have been made. These visual explanations are called visualizations. There already exist a lot of visualizations that explain recommendations to users. However, most of these visualizations are text-based and developed for computer applications.

This research proposes five visualizations based on the literature and evaluates these in terms of information adequacy, transparency and persuasiveness. These visualizations are evaluated in the context of a fully functional book recommender system. The visualizations make use of the features of the recommender system, in this case the genres of the books. To understand the impact of using these features a baseline visualization was also developed which contains no features in its explanation.

Some of these visualizations use a numerical approach to explain the recommendation. This means they contain numbers such as a percentage match. Other visualizations use a categorical approach to explain the recommendation. For instance, a visualization that only mentions the genres, without any numbers associated to it, is considered a categorical visualization.

The research questions are as follows:

RQ1. Does providing a visualization explaining a recommendation to the end user, increase the user satisfaction of the application?

RQ2. Do the five proposed visualizations perform better than the baseline when considering the interaction adequacy, transparency and persuasiveness of recommender systems?

RQ3. Does the end user prefer a numerical or categorical approach towards visualizations?

In this research a user-centered design approach is used in which the feedback of the end user is important at each stage of the development process.

The visualizations are loosely based on visualizations found in the literature, while the application is loosely based on other existing book recommender applications such

as Amazon and goodreads books. The recommender system is based on content-based filtering and has been custom-made for this research.

The proposed visualizations in this research are:

- Bar charts: bar charts indicating for each genre how much the user preferences match with this genre.

- Venn diagram: Venn diagram showing which genres the recommended book contains and what genres are considered the user's preferences and the overlap between these two.

- Other books: related books that the user has liked in the past.

- Double bar chart: two types of bar charts are displayed for each genre. One displays the match of the user with that genre and one displays the match with the recommended book with this genre.

- Link strength: the same information as double bar charts is used but displayed in a different manner. This manner is more compact but harder to read.

A low-fidelity prototype of the visualizations and the application was made. This low-fidelity prototype was evaluated in a pilot study to get some initial feedback about the low-fidelity prototype of the application and the low-fidelity prototype of the visualizations. For the visualizations a questionnaire was made which was filled in by 52 people. For the application a think-aloud study was conducted with five participants.

With this feedback the low-fidelity prototype of the application and the visualizations were revised. These revised versions were used to create a high-fidelity prototype. This high-fidelity prototype was also evaluated with a think-aloud study. Here 16 participants were recruited. This study helped find any remaining usability issues of the application and the visualizations as well as the algorithms used for the visualizations.

The feedback gathered in this user study is used to revise the high-fidelity prototype of the application and redeploy it to its final version. In this final version the visualizations itself are evaluated. This was done by interviewing 51 participants and asking their opinion about the visualizations. After the interview they were also asked to fill in a questionnaire.

The visualizations all perform better than the baseline when considering the interaction adequacy and transparency. However only the Venn diagram and link strength visualizations perform better than the baseline when looking at the persuasiveness. From the interview data gathered from the interview of the final user study it was also possible to do a thematic analysis. Out of this thematic analysis guidelines for the development of new visualizations in mobile phone application were found. No clear preference between categorical or numerical approach to the visualizations has been found. For each visualization weaknesses and strengths were also discovered.

Further research can make use of the guidelines found in this research when developing (new) visualizations for mobile phone applications. It is also possible to further research the proposed visualizations with different recommender system algorithms and to make these visualizations interactive.

List of Figures

2.1 Inel's wordcloud, overlap fraction and overlap wordcloud. [19] 22
2.2 Millecamp et al's user interface for a music recommender app. [26] . . . 23
2.3 Bostandjiev et al.'s user interface for the TasteWeights system [6] 24
2.4 The three interfaces designed by Millecamp et al. [27]. 25
2.5 Tsai and Brusilovsky's visualizations. [43] 26
2.6 TalkExplorer interface as seen in Verbert et al. [46] 27
2.7 Hierarchy Visualization for Group Recommender Systems by Wang et al. [47] 27
2.8 The three types of explanations designed by Biglic and Mooney. [4] . . 28
2.9 Screenshots of the Shopr application from Lamche et al. [23] 29
2.10 More detailed explanation about the color feature in the Shopr application which was developed for the research conducted by Lamche et al. [23] 30
2.11 Screenshot of the LibraryThing desktop version when viewing the details for The Hunger Games book. 31
2.12 Screenshot of the LibraryThing desktop version when viewing the recommendations provided for my user account. 32
2.13 Screenshot of the LibraryThing desktop version when viewing the details for the Harry Potter and the Sorcerer's Stone book. 32
2.14 Screenshot of the Goodread books web application when viewing the recommended book for my profile and focusing on the details of The Lightning Thief book when hovering over the cover. 33
2.15 Screenshot of the Goodread books web application when viewing the full details page of The Lightning Thief book. 33
2.16 Screenshot of the Amazon web application when viewing the recommendations made for my user profile. 34
2.17 Screenshot of the Amazon web application when viewing the book details page for the book Das Geheimnis von Green Lake. 34
2.18 Screenshot of the Amazon web application when viewing the similar books listed on the book details page for the book Das Geheimnis von Green Lake. 35

3.1 Design process for the application and visualizations of this research. . . 38
3.2 The five visualizations designed for this research. 39
3.3 The baseline visualization designed for this research. 40

4.1 Revised visualizations after the results of the questionnaire had been collected. 45
4.2 All screens of the initial low-fidelity prototype. 47
4.3 Screens of the revised low-fidelity prototype 48
4.4 Revised low-fidelity prototype after user feedback. 51

5.1 Screenshots of the initial visualization in the initial version of the application. 58

6.1 Screens of the initial version of the application. 66
6.2 Results of the Resque questionnaire for the high fidelity prototype provided as a boxplot. 71
6.3 System usability scale score as depicted by [1] 72
6.4 System usability scale scores boxplot for the first user study based on the SUS questionnaire answered by the participants. 72
6.5 Results of the NASA TLX questionnaire for the high fidelity prototype provided as a boxplot. 73

7.1 Screens of the revised version of the application as used for the final user study. 78
7.2 Screenshots of the revised visualizations in the application as used for the final user study. 78
7.3 Responses to the Resque questions. A score of 5.0 indicates fully agreeing and a score of 1.0 indicates fully disagreeing 83
7.4 Responses to the first visualization-specific resque question for the six visualizations. A score of 5.0 indicating fully agreeing while 1.0 indicating fully disagreeing. 87
7.5 Responses to the second visualization-specific resque question for the six visualizations. A score of 5.0 indicating fully agreeing while 1.0 indicating fully disagreeing. 88
7.6 Responses to the third visualization-specific resque question for the six visualizations. A score of 5.0 indicating fully agreeing while 1.0 indicating fully disagreeing. 88
7.7 System usability scale score boxplot for the final user study. 89
7.8 System usability scale score boxplot for the first user study (on the left) compared to the final user study (on the right) based on the SUS questionnaires answered by the participants. 90
7.9 Rankings for the visualizations as provided by the participants for the final user study. 91

List of Tables

4.1 Questions asked during the questionnaire for each of the different visualizations . . . 43
4.2 Results of the questionnaire, all numbers are presented as a percentage . 44
4.3 Feedback for the visualizations during the pilot study and implemented solutions. . . . 46
4.4 List of tasks given to the participants of the think-aloud study 49
4.5 Demographics of the participants who took part in the think-aloud study of the low-fidelity prototype. . . . 49
4.6 Issues with the application when using the low-fidelity prototype. These issues arose during the think-aloud study . . . 50
4.7 Possible solutions to the problems found via the think-aloud study. These are also the solutions that were implemented. . . . 51

5.1 Example of how genres are placed in the csv file after preprocessing. . . 54
5.2 The selected representative books, their IDs, the genres they consist of and the amount of ratings received as calculated with algorithm 2. . . . 57

6.1 The tasks given to the participants to evaluate the high-fidelity prototype. 66
6.2 Latin square used for the evaluation of the high-fidelity prototype for tasks 2, 3 & 4. . . . 67
6.3 Latin square used for the evaluation of the high-fidelity prototype for tasks 7 & 8. . . . 67
6.4 Questionnaire provided to the participants at the end of the think-aloud study for the high-fidelity prototype. . . . 68
6.5 Issues with the application when using the high-fidelity prototype. These issues arose during the think-aloud study of the high-fidelity prototype . 69
6.6 Solutions to the issues with the application as found during the high-fidelity think-aloud study. . . . 70
6.7 Feedback about the different initial visualizations in the initial version of the application. . . . 74
6.8 Implemented solutions to the feedback about the different initial visualizations in the initial version of the application. . . . 75

7.1 The tasks given to the participants to evaluate the visualizations. 79

7.2 Latin square used for the evaluation of the visualizations for tasks 4, 5, 6, 7, 8 & 9. 79
7.3 Questionnaire provided to the participant at the end of the think-aloud study for the final user study. 80
7.4 User profiles of the first 20 participants of the final user study. 84
7.5 User profiles of participants 20 through 40 of the final user study. 85
7.6 User profiles of the last 11 participants of the final user study. 86
7.7 Rankings for the visualizations as provided by the participants for the final user study. 91
7.8 Relative rankings for the visualizations as provided by the participants for the final user study. 91
7.9 Themes that were identified in the first step of the thematic analysis. . 92
7.10 Themes for the thematic analysis after reducing the overlapping themes. 93
7.11 Possible improvements for the different visualizations as suggested by the participants of the final user study. 99

Chapter 1

Introduction

In a society where more and more alternatives for the same product are offered to customers, it becomes increasingly difficult to make a satisfactory decisions. In fact people tend to spend longer on decisions if they feel like they already spent a long time making them and hence it starts feeling like an important decision. This phenomenon is called decision quicksands. Not only is the importance of a decision strongly influenced by the person's perceived difficulty of the task. An at first glance seemingly easy task that turns out to be more difficult than expected will get a person more sucked in to the decision making up to the point where they even voluntarily seek more information. [34]

These decision quicksands can be found in many daily aspects of life such as going to the store as well as deciding which book to borrow next from the library. Recommender systems are designed to help us with these decisions. However most recommender systems provide users with an overload of information, while research has showed us that when we go grocery shopping we usually only consider a few key factors when making our decisions. [20]

During the time of writing we are still in the midst of the COVID-19 pandemic with ubiquitous consequences being felt around the world. One of those consequences is that online shopping has become the norm as a lot of leisure stores, such as libraries and gift shops, have been closed a lot lately. Moreover recent research has shown that there is a positive effect on the hedonic motivation of buying books via on online bookstore due to the pandemic. [29] For grocery stores research has also shown that the amount of willingness of people to go to a physical store when the number of cases are rising in the region decreases, whilst it increases when the numbers decrease over the last 2-week period. [13]

The importance and usage of e-commerce has obviously increased due to this. 52% of consumers are reported to avoid brick and mortar stores while 36% even plan to avoid brick and mortar stores until they have received a vaccine. Walmart's e-commerce has seen an increase of 74%. [3] However this also enlarges the decision quicksand problem as store chains will show all their products on their website, whilst in the physical store they usually only store products that the locals have shown an interest in. This means that when browsing an online store, even more products are being offered and the importance of an underlying, well-working recommender system increases.

Unfortunately recommender systems are often a black-box system. This means that the recommender systems provide the user with recommendations without giving any explanation as to why. In a world where misinformation is spread at a steadily increasing pace people want to know why they are seeing these recommendations. [15] A possible solution to this is making the recommendations more transparent by, for instance, providing interfaces for the user so that they can discover how their preferences influence the recommendations.

These interfaces are also called explanations. There are seven aims usually defined for explanations. These seven aims are transparency, scrutability, trust, effectiveness, persuasiveness, efficiency and satisfaction. [39]

A lot of interfaces have already been developed and evaluated in the past. However many interfaces are text-based and made for a computer screen. [4] [17] This research focuses on visual (not purely text-based) explanations, also called visualizations, embedded into a mobile phone system.

This research will try to create mobile interactive user interfaces to explain the recommendations of a book recommender system. [35] By increasing the transparancy, it becomes possible to evaluate the effect on the 6 other aims. [39] Mainly whether the user satisfaction increases is researched here.

There exist a lot of visualization techniques. Some of them consist of a numerical approach in which numbers such as "% match" are used. Others use a categorical approach, herein no numbers are used, but rather the features of the recommender system or previously liked items.

In this research five different visualization techniques based on the literature are designed and implemented as well as a baseline visualization. These visualizations are implemented in the context of a mobile book recommender app. The recommender system for this research is also custom made and is based on content-based filtering. Content-based filtering uses features to make recommendations, in this case the recommendations are the genres of the books.

The explored visualizations can be briefly described as follows:

- Bar charts: bar charts indicating for each genre how much the user preferences match with this genre.
- Venn diagram: Venn diagram showing which genres the recommended book contains and what genres are considered the user's preferences and the overlap between these two.
- Other books: related books that the user has liked in the past.
- Double bar chart: two types of bar charts are displayed for each genre. One displays the match of the user with that genre and one displays the match with the recommended book with this genre.
- Link strength: the same information as double bar charts is used but displayed in a different manner.

Not all visualizations show the same background information. Some use a more numerical approach, while others use a more categorical approach for the explanation.

The research questions for this research are as follows:

RQ1. Does providing a visualization explaining a recommendation to the end user, increase the user satisfaction of the application?

RQ2. Do the five proposed visualizations perform better than the baseline when considering the interaction adequacy, transparency and persuasiveness of recommender systems?

RQ3. Does the end user prefer a numerical or categorical approach towards visualizations?

The research is designed in a user-centered approach in which the feedback of the end user is paramount.

A pilot study is conducted to get some initial feedback about the prototypes of the proposed visualizations. A first user study is conducted to get feedback of the initial version of the application and get some feedback about the initial implementation of the actual visualizations. With this feedback an updated version of the application is used for the final user study. In this final user study the visualizations itself are evaluated and conclusions for how to design such visualizations in a phone application are drawn.

It is concluded that the bar charts, double bar charts and link strength visualizations are best used when a more detailed explanation is necessary than a categorical explanation. The Venn diagrams allow for exploration and can increase trust to the user by confirmation of his known preferences. The other books visualization uses familiarity by the usage of book covers. The double bar chart contains redundant information while the link strength contains the same redundant information but is more compact. The double bar charts should be used when users have a lower cognition towards new visualizations, while in the other case the link strength visualization should be used.

This research concludes that providing the user with a visual explanation as to why a recommendation has been made increases the user satisfaction. The five proposed visualizations are found to perform better than the baseline when looking at interaction adequacy and transparency. However only the Venn diagram and link strength visualizations perform better than the baseline visualization in terms of persuasiveness. No clear preference between a numerical or categorical approach to the visualizations has been found.

Also a few guidelines for the design of new visualizations are provided. Most of these guidelines can also be used for computer-based applications (both desktop and browser applications), but a few are only applicable to mobile applications.

Chapter 2

Related work

2.1 Recommender systems

Recommender systems are defined as systems that provide suggestions for new items to users. [33] This is used to help users make decisions in which they do not have a sufficient amount of personal experience [32] as well as when there is an abundance of choice for a user to make [24]. Recommendations are also personalized to the user's known preferences by the system, as is the case for Netflix [12].

Recommender systems leverage the knowledge of the users and the items in the system to recommend new items to users. This knowledge often includes previously liked items and demographics of the users and product descriptions and reviews for the items. [25]

2.1.1 Types of recommender systems

There are many different types of recommender systems, which all have their advantages and disadvantages. [5]

Collaborative filtering. In collaborative filtering it is assumed that users can give ratings to the items of a system. The recommendation algorithms then tries to match users with similar interest (i.e. similar previously liked items) in order to recommend new items to those users. This means that the items liked by a user will be recommended to other users with similar interests. These preferences can also be implicit such as for instance how often a certain song has been listened to. [5]

There are three main disadvantages to collaborative filtering.

Firstly, there is the problem of data sparsity, in which the system contains so many items, that even the must active users of the system have only rated a subset of all available items. [36]

Secondly, collaborative filtering is not very scalable as it would require a lot of computational power to calculate recommendations in systems with millions of users and products. [36]

Thirdly, the cold start problem makes it difficult to make recommendations for novel users and items as too little is known about them at that point. [22]

Content-based filtering A second type of recommender system algorithm is that of content-based filtering. In content-based filtering recommendations are based on knowledge about the data (often called features) as well as knowledge about the user's profile (i.e. features from previously liked items from that user). [45]

The major disadvantage of content-based filtering is recommending items from a domain, also called cross domain recommendations. For instance a content-based filtering system for books will have difficulties recommending users touristic destinations as the features vary so greatly. [37]

Knowledge-based recommender systems Knowledge-based recommender systems make use of implicit information about users or items. This means that it apriori knows which items to recommend when certain other items have been liked by the user. The great advantage of this is that it completely avoids the cold start problem. However finding the relevant data to be able to get the required knowledge is a bottleneck. [8]

Hybrid recommender systems Hybrid recommender systems consist of combining multiple recommender system methods together. So for instance collaborative and content-based filtering can be unified into one recommender system. [48] Hybrid systems can be used to circumvent problems such as the cold start problem and the sparsity problem. [18]. An example usage of a hybrid recommender system can be found in Netflix. [12]

2.2 Visualization techniques

2.2.1 Visualizations

Video summarization

In the field of video summarization, videos are usually summarized in an automatic, non-transparent fashion. To provide the end user with transparency, Inel [19] offers visual explanations to explain the video summaries. They provide 4 different explanations. The main evaluation metrics are concept prominence, semantic coverage, distance (distance between the summary and the original video) and quantity of coverage (amount of concepts covered/not covered in a summary). 3 of the 4 different explanations are shown in figure 2.1 with the fourth one being a combination of a wordcloud and fraction.

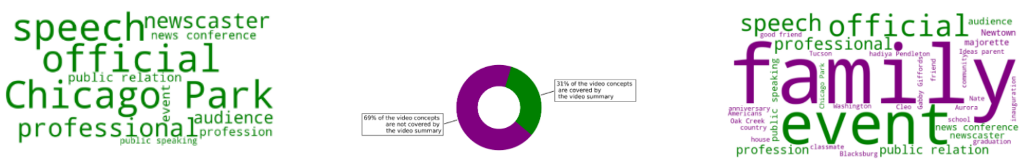


Figure 2.1: Inel's wordcloud, overlap fraction and overlap wordcloud. [19]

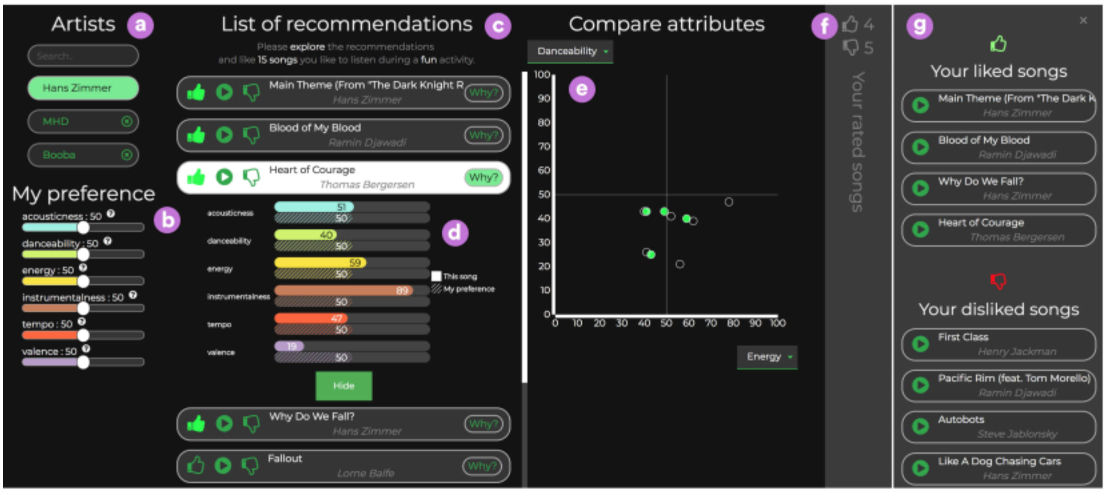


Figure 2.2: Millecamp et al's user interface for a music recommender app. [26]

Artistic Images

Another field in which research has been done on visualisation techniques is within the recommendations of artistic images. Dominguez [10] used three interfaces to research the effect of explanations to this recommender system. The first interface did not provide any explanations and was used as a baseline, the second interface provided textual explanations alongside the top-3 similar images and the third interface provided the top-1 similar image next to features' bar charts.

The researchers used two different recommender systems to provide the recommendations. The first one being DNN visual feature algorithm, which has a high accuracy but a low transparency and the second one being Attractiveness Visual Features (AVF) algorithm, which has a high transparency but low accuracy. The researchers conclude algorithms should not be studied in isolation but in conjunction with interfaces. They also use the framework provided by Knijenberg [21] to synthesize the effect of different variables in the user experience of their system.

Music recommendations

In Millecamp et al.'s research [26] a music recommendation system app is introduced to research the effects of personal characteristics on the need and effectiveness of visualizations of explanations for recommender systems.

The user interface designed for Millecamp et al.'s research is provided in figure 2.2.

Millecamp et al. concluded that personal characteristics, such as a high user cognition of the subject, influence the perceived usefulness of visualizations 2.2.

Figure 2.3: Bostandjiev et al.'s user interface for the TasteWeights system [6]

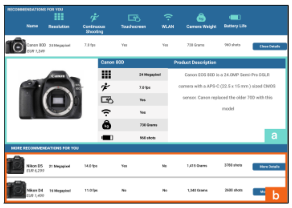
TasteWeights

Bostandjiev et al. [6] created TasteWeights. TasteWeights is an application that explains predictions based on information from the user's social media.

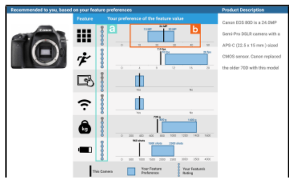
This is done via an interactive interface. This interface consists of three columns. In the context of, for instance, musical taste, the leftmost column consists of the songs the user has liked in the past. The middle column shows the context the recommender system used to generate recommendations. The final and leftmost column shows the recommendations that are made based on this context. The user can change the weights of the songs and context to understand the influence on the recommendations. The user can also hover over a song, context or recommendation to understand to which other things it is related in the visualization. [6] An example of this is provided in figure 2.3.

Camera and music recommendations

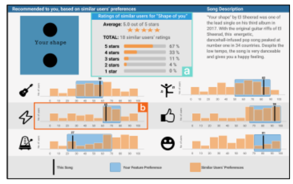
In another research conducted by Millecamp et al. [27], it was researched how the perception of users on explanations is influenced by the product domain or by users' personal characteristics. For each product domain four interfaces were designed, starting with a baseline interface. The other three interfaces were designed depending on the underlying recommendation algorithm, i.e. a content-based filtering, collaborative filtering and hybrid filtering interface were created. For this research the users were exposed to both the baseline interface with no explanations and an interface that does provide explanations. The baseline interface for camera recommendations is provided in figure 2.4a, whilst the interface with explanations for camera and music recommendations are respectively given in figure 2.4b and figure 2.4c.



(a) The baseline interface for camera recommendations designed by Millecamp et al. [27]



(b) The interface with explanations for camera recommendations designed by Millecamp et al. [27]



(c) The interface with explanations for music recommendations using the collaborative algorithm designed by Millecamp et al. [27]

Figure 2.4: The three interfaces designed by Millecamp et al. [27].

Conference support system

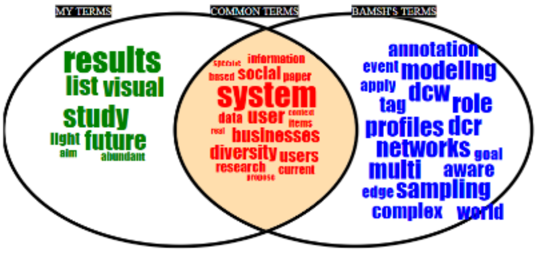
Tsai and Brusilovsky proposed a couple of visualization techniques for their conference support system trying to enhance the Conference Navigator 3 (CN3) [42]. For each of the underlying recommender systems they created five interfaces. There are five recommender systems and hence 25 interfaces were introduced. The recommender systems were each individually based on one of the following: publication similarity, topic similarity. co-authorship similarity, interest similarity and geographic distance. [43]

Venn word clouds can be used to explain the publication similarity as seen in figure 2.5a. Topical bars in which the top three topics of scholars are shown and the topical information is provided in the bar chart can be used to explain topic similarity. An example of this topical bar is found in figure 2.5b. ForceAtlas2 was one of the visualizations used for co-authorship similarity as can be seen in figure 2.5c. For the interest similarity a similar keywords interface was used as provided in figure 2.5d with dotted lines representing a weak link and full lines representing a strong link. Lastly for the geographic distance an earth-style visualization was provided (figure 2.5e). [43]

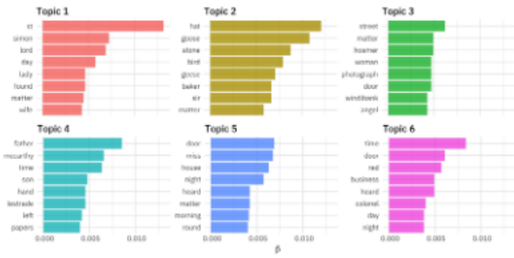
Verbert et al. [46] also conducted some research on a conference recommendation tool named TalkExplorer, which is an interactive visualization built on top of CN3. An image of the TalkExplorer interface can be found in figure 2.6. On the left side users can select tags, users and recommender agents to be added. The labeled circles represent entities such as users and tags, while the yellow circles represent individual talks. The larger bubbles that the yellow circles are part of are clusters of talks.

Linguistic analysis

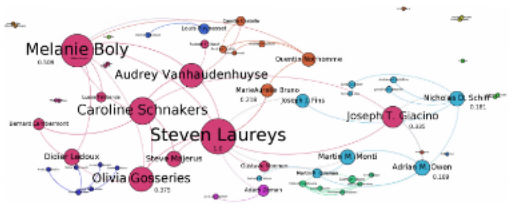
Tsai and Brusilovksy [44] also did research on linguistic analysis of the user feedback of their visualization made for the CN3 system [43]. This was done to understand the rationale behind the choices of the participants. Tsai and Brusilovsky conclude that controllable user interfaces help users make decisions more quickly while explainable interfaces make users understand the rationale behind their decisions better.



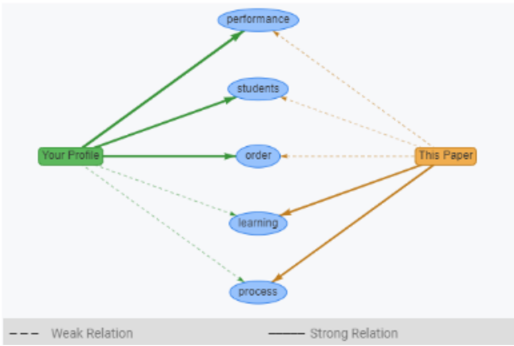
(a) Tsai and Brusilovsky's Venn wordcloud to explain the publication similarity. [43]



(b) Tsai and Brusilovsky's topical bar to explain the topic similarity. [43]



(c) Tsai and Brusilovsky's ForceAtlas2 interface to explain the co-authorship similarity. [43]



(d) Tsai and Brusilovsky's similar keywords interface to explain the interest similarity. [43]

(e) Tsai and Brusilovsky's earth-style interface to explain the geographic distance. [43]

Figure 2.5: Tsai and Brusilovsky's visualizations. [43]

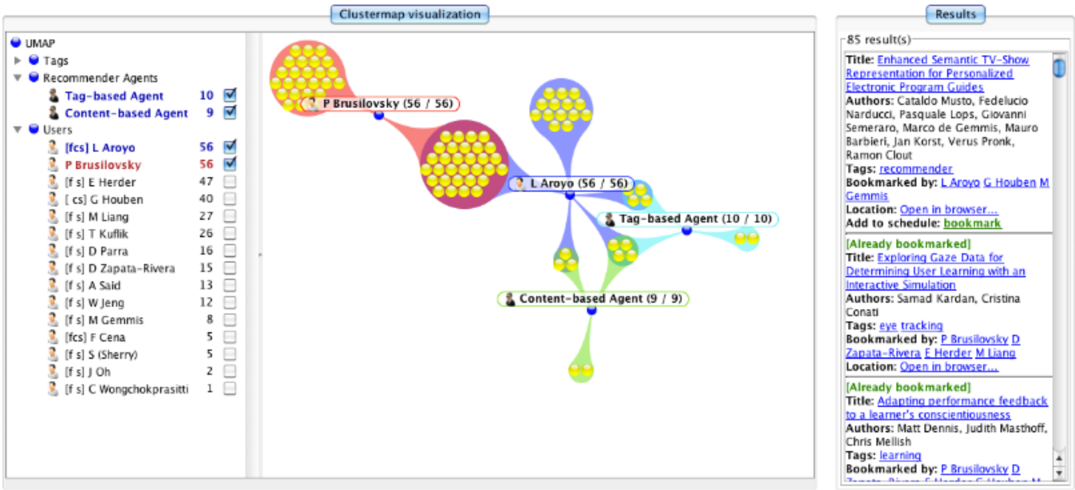


Figure 2.6: TalkExplorer interface as seen in Verbert et al. [46]

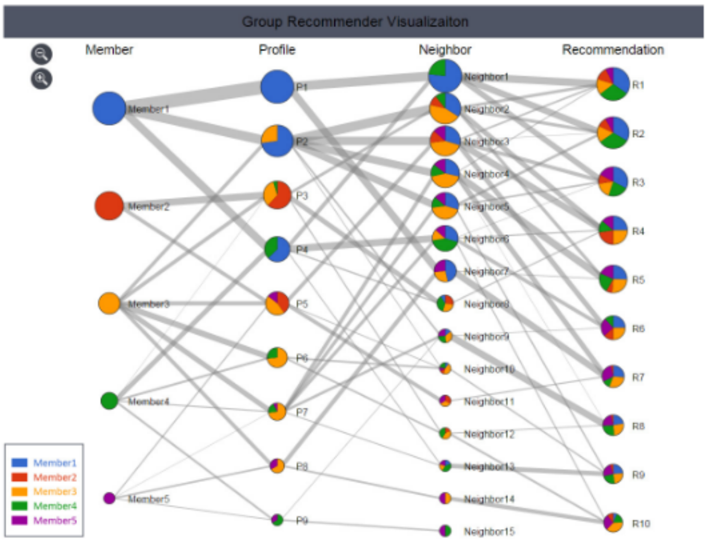
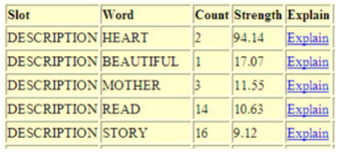


Figure 2.7: Hierarchy Visualization for Group Recommender Systems by Wang et al. [47]

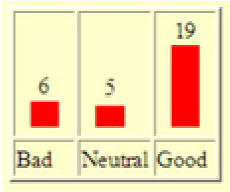
Group recommendations

Wang et al. [47] provide a hierarchy visualization for group recommender systems. This visualization is provided in Figure 2.7. Each of the different levels is the representation of the different methods the recommender used to obtain the recommendations. These are pseudouser modeling, neighbor identification and recommendation prediction.

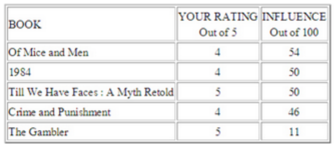


Slot	Word	Count	Strength	Explain
DESCRIPTION	HEART	2	94.14	Explain
DESCRIPTION	BEAUTIFUL	1	17.07	Explain
DESCRIPTION	MOTHER	3	11.55	Explain
DESCRIPTION	READ	14	10.63	Explain
DESCRIPTION	STORY	16	9.12	Explain

(a) A keyword style explanation as designed by Biglic and Mooney. [4]



(b) A neighbor style explanation as designed by Biglic and Mooney. [4]



BOOK	YOUR RATING Out of 5	INFLUENCE Out of 100
Of Mice and Men	4	54
1984	4	50
Till We Have Faces : A Myth Retold	5	50
Crime and Punishment	4	46
The Gambler	5	11

(c) An influence style explanation as designed by Biglic and Mooney. [4]

Figure 2.8: The three types of explanations designed by Biglic and Mooney. [4]

Book recommendation visualizations

Also for book recommendation a few visualizations have been designed in previous work. The LIBRA book recommender system from Biglic and Mooney has been developed for this [4]. In this system the users are provided visualizations which can be subdivided into three different categories. These categories are keyword style explanation, neighbor style explanation and influence style explanation. These are respectively shown in figure 2.8a, 2.8b and 2.8c. LIBRA uses independent content-based and collaborative filtering algorithms. The keyword style explanation shows the similar content between recommendations and previously liked books based on the content-based filtering algorithm. The neighbor style explanation shows how other users with similar interests perceived the recommended book. Finally the influence style explanation shows which books had the most impact on the current recommendation. [4]

The researchers found the keyword and influence-style explanations the most effective at enabling users to make accurate assessments. [4]

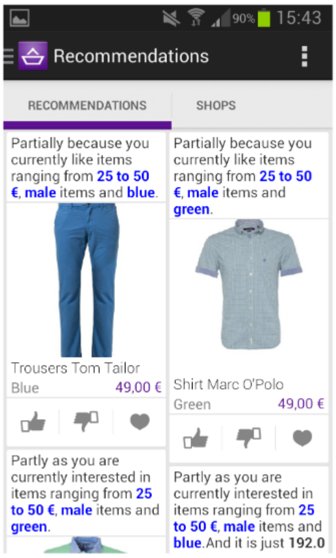
Movie recommendation visualizations

Also on movie recommender systems some visualizations have already been designed in previous work. Herlocker et al. [17] made 21 different visualizations for their collaborative filtering recommender system. A histogram showing how similar users had rated the same movie turned out to be the most effective. [17]

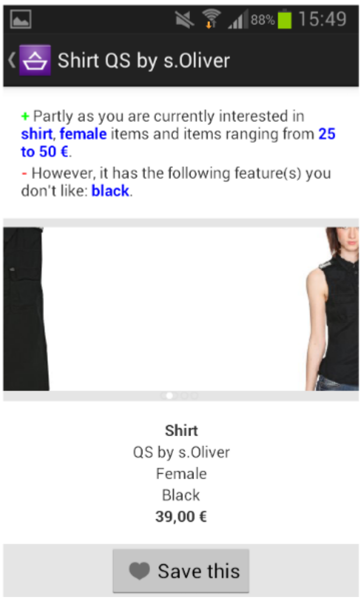
Research on visualizations

Al-Taie et al. [2] did some research on the general perception on visualizations for recommender techniques. The participants were asked to evaluate three book recommender systems and three movie recommender systems. Some important conclusions include:

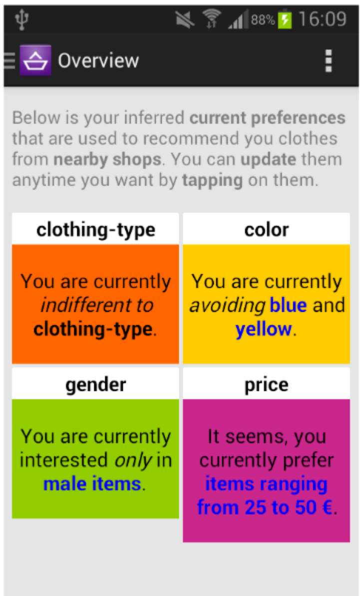
- Users are mostly concerned with the navigation and layout of the system as this greatly impacts the perceived usefulness of the system



(a) Recommendations screen of the Shopr application which was developed for the research conducted by Lamche et al. [23]



(b) Explanation screen for a recommended item in the Shopr application which was developed for the research conducted by Lamche et al. [23]



(c) User profile screen in the Shopr application which was developed for the research conducted by Lamche et al. [23]

Figure 2.9: Screenshots of the Shopr application from Lamche et al. [23]

- There is no clear preference between textual or graphical explanation style [38], however more concise textual explanations are preferred over detailed ones. [30]
- Trust of the system is the factor that mostly impacts the user satisfaction [40]
- Adding a photograph of the author increases the system's credibility [11]

He et al. [16] conducted a survey of different existing visualization techniques and the metrics they focused on. They also propose new research possibilities focusing on metrics such as privacy.

2.2.2 Mobile phone visualizations

Shopr

Lamche et al. [23] have created a mobile recommender system for clothing items. This recommender system provides the user with an explanation based on automatically generated text explanations next to an image of the recommended item as can be seen in figure 2.9a.

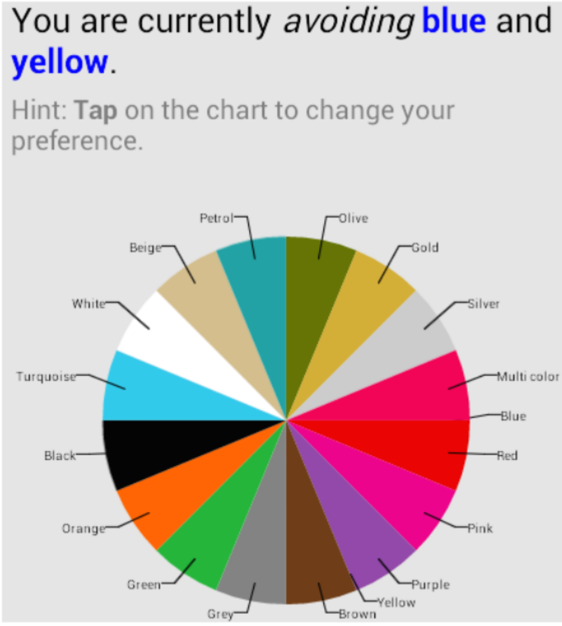


Figure 2.10: More detailed explanation about the color feature in the Shopr application which was developed for the research conducted by Lamche et al. [23]

In Shopr the user can also click on one of the recommended items in order to see a more detailed explanation. The textual explanations are clearly indicated with a + or - sign depending on whether it is a positive or negative argument for the recommendation respectively. This screen is shown in figure 2.9b. [23]

Lastly the Shopr app allows user to view their user profile w.r.t. recommender systems. This is done via an overview screen in which each different feature that is used by the recommender system is shown. For each of these features the user preferences that have been recorded by the system are displayed. This can be seen in figure 2.9c. It is also possible for the user to click on one of the features to get a more detailed explanation about that particular feature. This can be seen in figure 2.10, where a more detailed explanation for the color feature is displayed. [23]

2.2.3 Evaluation

For the evaluation of explanations, seven goals are usually considered as proposed by Tintarev et al. [41] These goals are as follows:

1. Transparency: this allows users to understand how the recommendations were generated.

2. Scrutability: this allows user to correct any mistakes made by the system's assumptions.

3. Trust: this increases the user's confidence in a system.

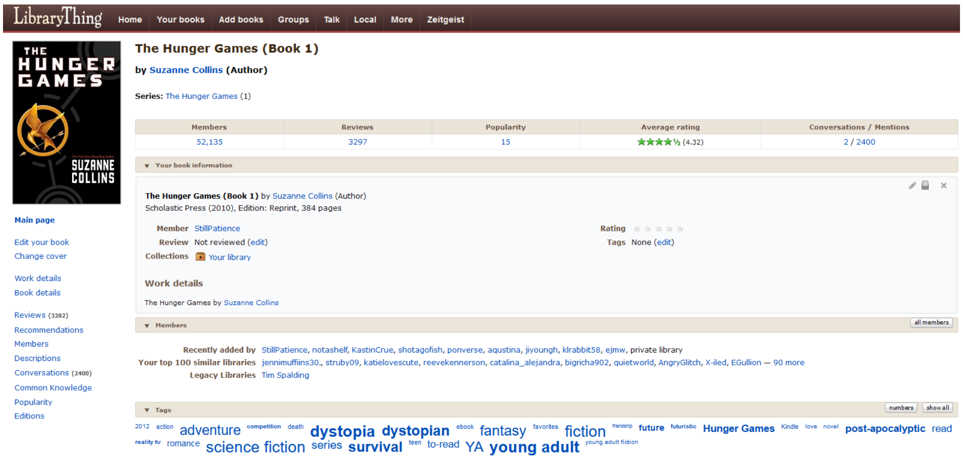


Figure 2.11: Screenshot of the LibraryThing desktop version when viewing the details for The Hunger Games book.

4. Persuasiveness: this convinces users to buy certain items.

5. Effectiveness: it is important that the explanation effectively helps the user make decisions.

6. Efficiency: explanations allow users to make faster decisions.

7. Satisfaction: explanations should allow the user to feel more satisfied of the overall system.

2.3 Book recommender applications

There already exist a lot of book recommender applications. These book recommender applications served as an example for the interface used in the application for this research.

2.3.1 LibraryThing

A first book recommender app to look at is the LibraryThing app. When viewing the details of a book in the LibraryThing system the user is provided with the tags that the book has been annotated with. These are written in different font sizes depending on the importance of each tag. This can be seen in figure 2.11. LibraryThing makes recommendations based on user's previous ratings, however they do not provide any explanation as to why recommendations have been made, as can be seen in figure 2.12. Lastly LibraryThing also provides similar books on the details page of each individual book, not taking into consideration the user's preferences, as can be seen in figure 2.13.

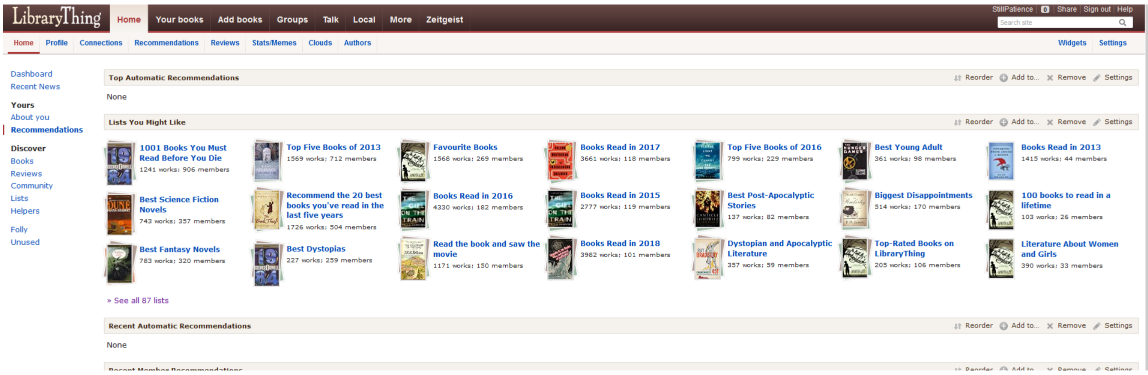


Figure 2.12: Screenshot of the LibraryThing desktop version when viewing the recommendations provided for my user account.

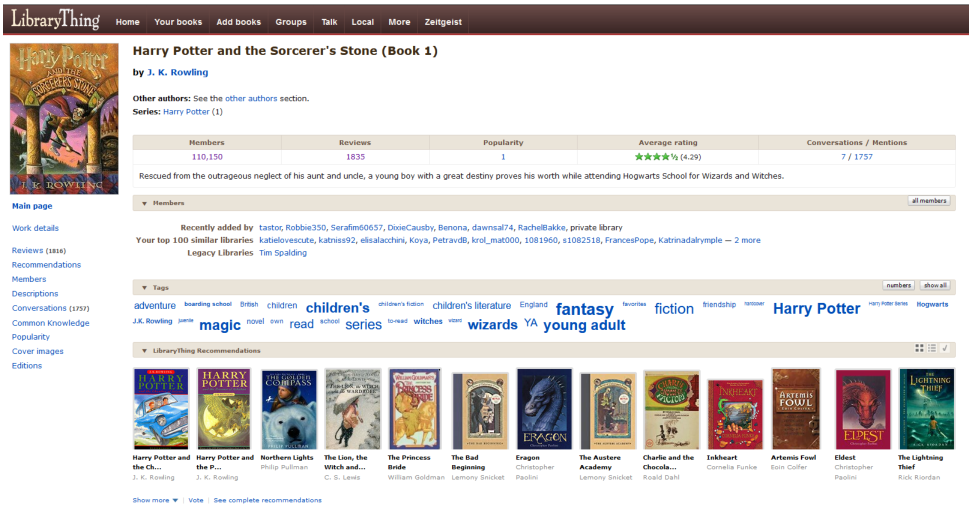


Figure 2.13: Screenshot of the LibraryThing desktop version when viewing the details for the Harry Potter and the Sorcerer's Stone book.

2.3.2 Goodread

Goodreads book is an other book recommender system. In this application recommendations are only made when a user has rated at least 20 books, this counters the cold-start problem. Goodreads books does provide some explanation as to why a book has been recommended to a user. This is done by displaying books that are related to the current book that the user had previously added to his list as can be seen in figure 2.14. Unfortunately there is no explanation provided when viewing the full details page of a book as can be seen in figure 2.15.

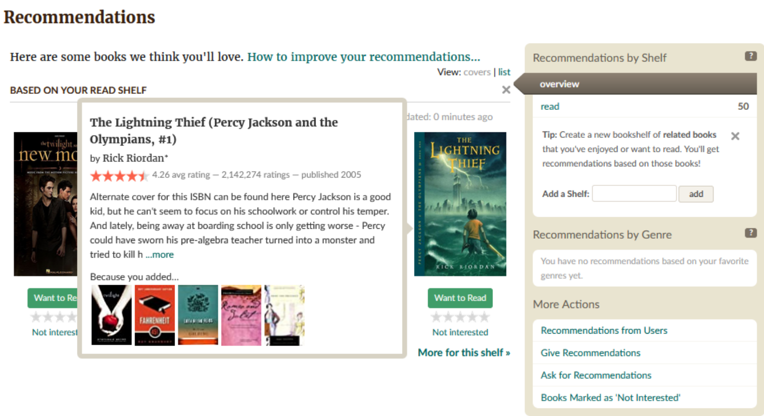


Figure 2.14: Screenshot of the Goodread books web application when viewing the recommended book for my profile and focusing on the details of The Lightning Thief book when hovering over the cover.

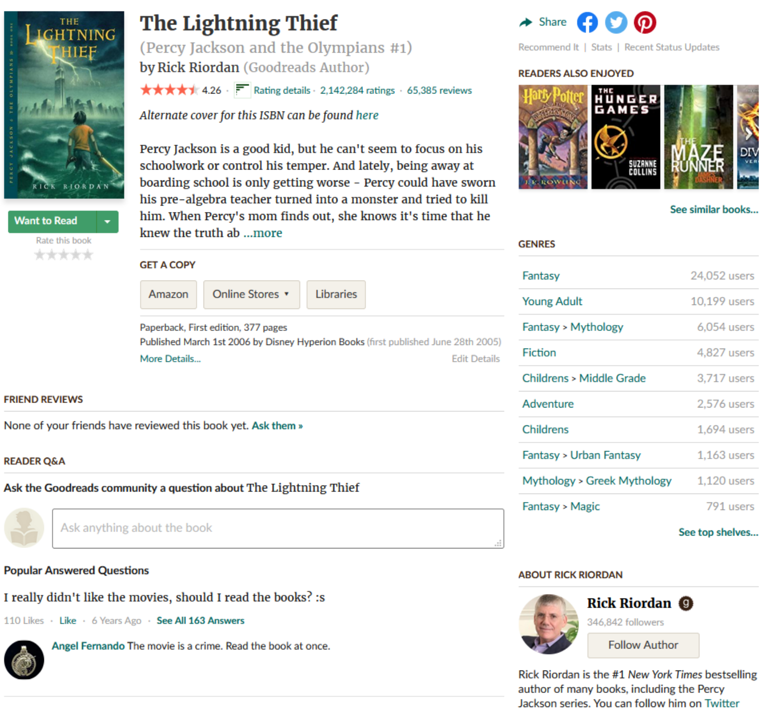


Figure 2.15: Screenshot of the Goodread books web application when viewing the full details page of The Lightning Thief book.

Figure 2.16: Screenshot of the Amazon web application when viewing the recommendations made for my user profile.

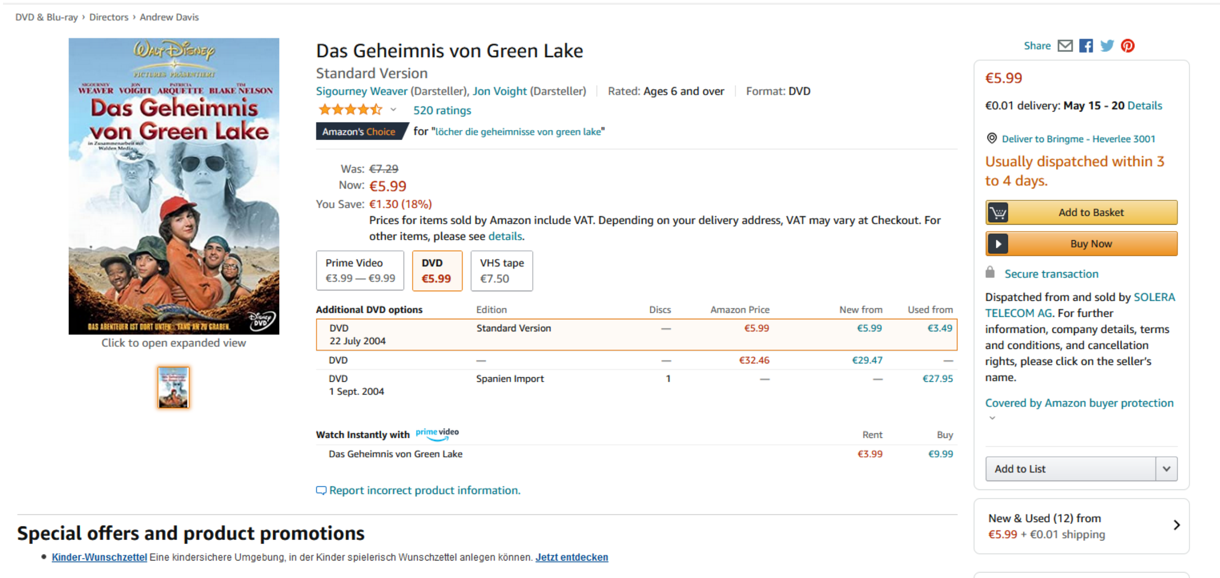


Figure 2.17: Screenshot of the Amazon web application when viewing the book details page for the book Das Geheimnis von Green Lake.

2.3.3 Amazon

Amazon is an online webshop that recommends new items to users depending on their previously viewed and bought items. Amazon also contains books and can henceforth also be considered as a book recommender system. After having added the book Löcher to my list I could see a few recommendations pop up on my home screen as can be seen in figure 2.16. When viewing one of the recommended book details pages no explanation is found as to why the book has been recommended, this can be seen in figure 2.17. However when scrolling down a bit further down the book details pages, similar books to the book being viewed are listed as can be seen in figure 2.18. This consists of books that are frequently bought together by other Amazon users as well as books that have also been viewed by other users after having viewed this book.

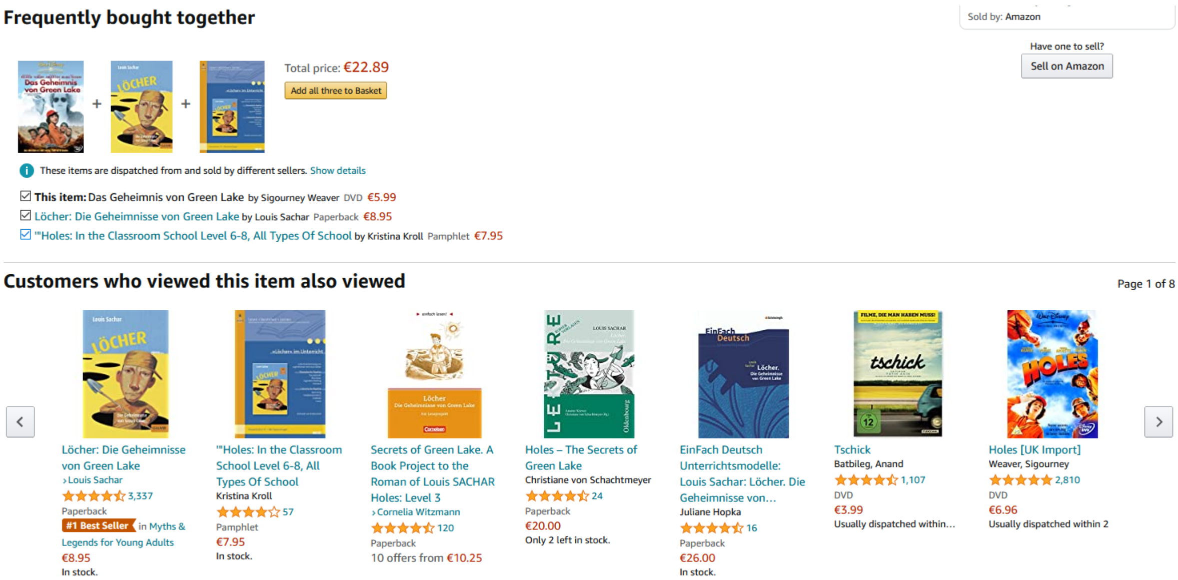


Figure 2.18: Screenshot of the Amazon web application when viewing the similar books listed on the book details page for the book Das Geheimnis von Green Lake.

Chapter 3

Design process

3.1 User-centered design

In this research a user-centered design process is used. Here in the feedback of the end user is asked at various stages of the development process and taken into account in the next step.

Low-fidelity prototype Based on the literature five visualizations are proposed in a low-fidelity prototype.

Pilot study In a pilot study some initial feedback of the low-fidelity prototype for the application and the visualizations is gathered. With this feedback the low-fidelity prototypes is revised. Using this revised low-fidelity prototype, a high-fidelity prototype is made.

High-fidelity prototype The actual application and visualizations are made with the feedback from this pilot study. In the first user study the actual application was evaluated to find any remaining usability issues as well as some feedback about the actual visualizations and the algorithms used to make them. With this feedback the application and visualizations have been updated and redeployed. The redeployed and revised application is the version which is used for the final study.

Final user study Using the revised high-fidelity prototype the final user study is conducted. In the final user study the visualizations itself are evaluated. Feedback about the interaction adequacy, transparency and persuasiveness are gathered. Further improvements to the visualizations are also found.

A schematic overview of the design process is provided in figure 3.1.

3.2 Visualizations

I introduce five visualizations that are made to help users understand the recommendations that are being made to them via the recommender system.

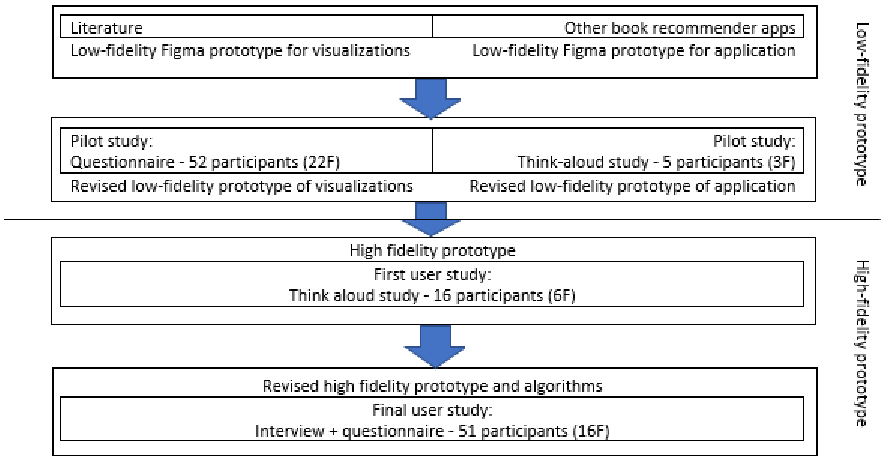


Figure 3.1: Design process for the application and visualizations of this research.

All of these are loosely based on Tsai and Brusilovsky's previous work. [43]

The first and fourth visualization are based on the topical bars as seen in figure 2.5b. The second visualization is based on the Venn wordcloud as seen in figure 2.5a. The fifth visualization is based on the keywords interface as seen in figure 2.5d. [43]

I created a low-fidelity prototype for these visualizations. A low-fidelity prototype allows me to quickly make changes in the early stages of development where major design issues are mostly prevalent. This low-fidelity prototype is made with the online modelling tool Figma[1]. The visualizations proposed for this research are provided in Figure 3.2[2]. Starting from the left the users are provided with a visualization that:

1. Shows how much the recommended book matches the genres that the person has shown a liking to, presented with bar charts. (bar chart)
2. Shows the users which genres the recommended book has in common with the user's preferences using a Venn diagram. The matching genres are placed in the eye of the Venn diagram while the non-corresponding genres are also displayed next to the eye. (Venn diagram)
3. Shows the user other books that the user has liked in the past that contain the same genres as the recommended book (other books)
4. Shows the user how much the recommended book contains of each genre and how much the user likes this genre by looking at the genres that constitute the books that the user has liked before. This is done using bar charts. For each genre a bar chart is used to represent how much of the genre the recommended

[1]More information about this tool can be found at `https://www.figma.com/`

[2]The prototype implementations made in Figma can be found at `https://www.figma.com/file/iWkZwlp8yqGYCwHTxCpaCJ/Low-fidelity-with-components-before-TA`

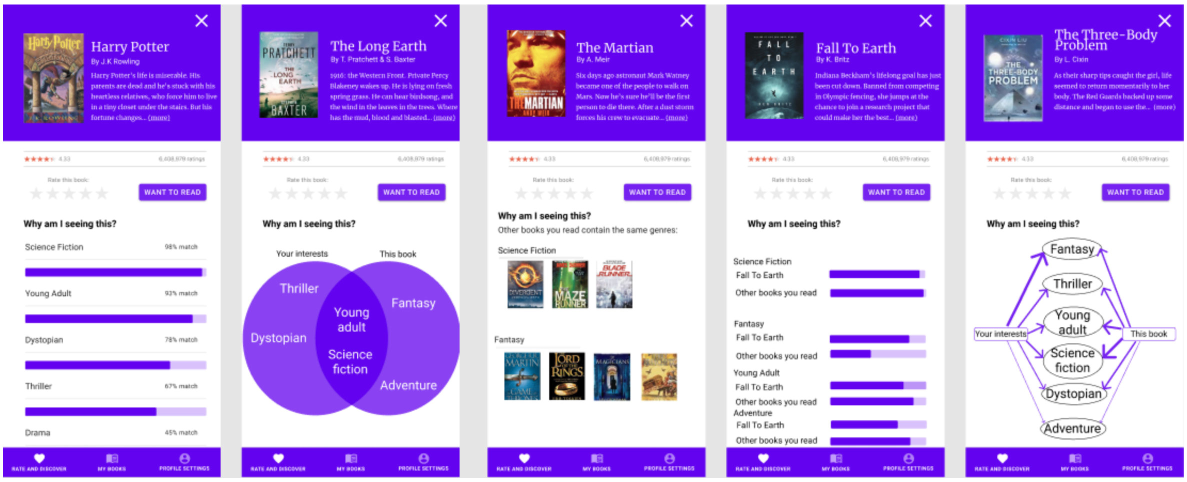


Figure 3.2: The five visualizations designed for this research.

book contains while another bar chart is used to show how much the genre has been present in the books the user has previously liked. (double bar chart)

5. Shows the user a diagram in which the genres of the recommended book and the user's previously liked books are plotted. The arrows indicate how much the recommended book and previously liked books have of the genre. A thicker line indicates that a book contains more of the genre. (link strength)

Finally I also designed a baseline visualization, which provides no explanation about the genres of the book. This baseline will allow to understand whether the other visualizations that do contain information about the features provide a better explanation. This visualization functions as a baseline for the other visualizations as it is not based on the same features (i.e. genres of books) that the other visualizations are based on. This will help ascertain the usefulness of displaying these features. This baseline can be seen in figure 3.3[3].

3.2.1 Differences in visualizations

The six different visualizations as shown in figures 3.2 and 3.3 each show the reasoning behind the recommendations in a different way. Here I analyze the differences between each of the visualizations, including the baseline.

In the bar chart visualization each genre that the recommended book contains, is represented by one line in this bar chart. This is accompanied by a % match metric, which shows the user how much % match his taste has with these genres.

In the Venn diagram visualization each genre of the recommended book is placed in the right circle. Each genre of the user's previously shown preferences is shown in the left circle. The overlap contains the genres that are both in the recommended

[3]The prototype implementation for the baseline made in Figma can also be found at `https://www.figma.com/file/iWkZwlp8yqGYCwHTxCpaCJ/Low-fidelity-with-components-before-TA`

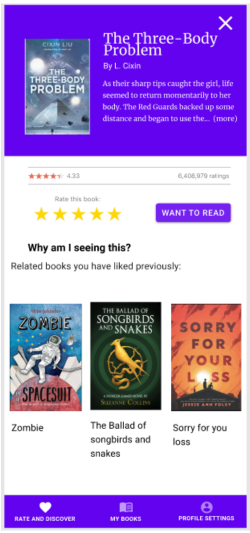


Figure 3.3: The baseline visualization designed for this research.

book and in the user's previously shown preferences. The major difference here is that no numerical value for the importance of each of the different genres is shown, as was the case for the bar chart visualization. However more contextual information is provided by the fact that the bar chart visualization only gave a % match to the different genres in the overlap, while here the genres in the non-overlapping sections are also mentioned. Therefor the left circle, showing your interests can help the user understand which preferences are already known by the system. This acts as a sort of confirmation. While the right circle, showing the interests of the recommended books allows for a more explorative approach, in which the user can see which other genres the recommended book contains that might not yet be part of his/her known preferences.

The other books visualization shows books that contain the same genre as the recommended book that the user has liked in the past. For each genre in the recommended book the visualization shows books the user has liked in the past by displaying its cover. This allows the user to understand which similar books to the recommended book he has liked in the past. Here again no quantitative data is being displayed as was the case for the first visualization. The major difference with this visualization compared to the Venn diagram visualization is that the genres here are the genres that are in the overlap of the Venn diagram, i.e. genres that are both in the user's known preferences and in the recommended book.

The double bar charts visualization uses bar charts like the first visualization. Here again the genres that are both in the recommended book and the user's known preferences are shown. For each genre, the % match to this book is displayed as well as the % match with the user's preferences. Here again a numerical approach is used such as in the first visualization. But no exploration beyond the similarities between the recommended books and the user's known preferences is possible, as was the case for the Venn diagram visualization.

The link strength visualization also uses the genres that are both in the recommended book and the user's known preferences. The arrows indicate the % match toward the user's preferences and the recommended book. This visualization displays the same amount of information as the double bar charts visualization, however it does so in a more compact, however less familiar way. It is less familiar as end user's are more likely to be familiar with bar charts than with Tsai and Brusilovsky's similar keywords interface on which this visualization is based. [43]

Finally the baseline visualization displays the books that the user has liked in the past that were related to this book. Here no numerical values are used such as in bar chart, double bar charts and link strength visualizations. The books are displayed in a similar fashion as in the other books visualization, but here there is no distinction between the different genres.

The bar chart, double bar charts and the link strength visualization are examples of categorical approaches to visualizations. While the Venn diagram and other books visualizations are thematic approaches to visualizations as they contain no numerical values.

Chapter 4

Pilot study

In this chapter the design and results of a pilot study are discussed. This pilot study is done to get feedback about the initial version of the application and visualizations. To get feedback for the initial version of the application a think-aloud study is conducted. For feedback on the initial version of the visualizations a questionnaire is used. With the feedback for both the application and the visualizations it is possible to still change them before the initial version of the application is deployed. All iterations of this research have been approved by the Social and Societal Ethics Commitee (SMEC) of the KU Leuven university[1].

4.1 Evaluation of visualizations

Questionnaire

To get feedback for the visualizations I created a questionnaire. In this questionnaire the five visualizations were shown to the participants, for each visualization the users were asked to answer the questions provided in table 4.1. These questions were answered on a five point Likert scale ranging from strongly agree to strongly disagree. For each visualization the participants were also given the option to provide extra feedback if they wanted, however this was not mandatory.

Number	Question
1	This visualization feels useful.
2	This visualization makes me understand why I see this recommendation.
3	This visualization was easy to understand.

Table 4.1: Questions asked during the questionnaire for each of the different visualizations

[1]The approval for this research can be found with the G-2021-3430-R2(MAR) code.

Question 1

Visualization	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
1	13.5	61.5	9.6	13.5	1.9
2	15.4	42.3	23.1	15.4	3.8
3	19.2	40.4	26.9	9.6	3.8
4	1.9	30.8	23.1	36.5	7.7
5	1.9	25	21.2	38.5	13.5

Question 2

Visualization	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
1	25	57.7	5.8	9.6	1.9
2	30.8	44.2	23.1	1.9	0
3	17.3	50	21.2	11.5	0
4	3.8	40.4	17.3	34.6	3.8
5	5.8	40.4	25	19.2	9.6

Question 3

Visualization	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
1	34.6	55.8	5.8	3.8	0
2	21.2	44.2	19.2	13.5	1.9
3	13.5	50	21.2	15.4	0
4	0	21.2	11.5	53.8	13.5
5	1.9	15.4	21.2	36.5	25

Table 4.2: Results of the questionnaire, all numbers are presented as a percentage

Recruitment

For the recruitment of participants to fill in my questionnaire I asked my acquaintances to fill it in. I also informed all of them about the importance of being neutral when doing so. A total of 52 people have filled in the questionnaire. 22 of these participants were female.

Results

The results of the questionnaire are given in table 4.2. The bar charts visualization has scored the best with 75%, 82.7% and 90.4% of the participants at least agreeing[2] on the three respective questions. Scoring the least is the link strength visualization with only 26.9%, 46.2% and 17.3% at least agreeing to the respective questions. The Venn diagram visualization got the most mixed reviews with reactions ranging from "the information is clear at a glance" to "the information is way too complex".

4.1.1 Revised visualizations

Thanks to the feedback from the questionnaire it was possible to improve some of the visualizations. For instance the bar charts for the books were found to take a

[2] At least agreeing means either "Agree" or "Strongly agree" has been answered by a participants

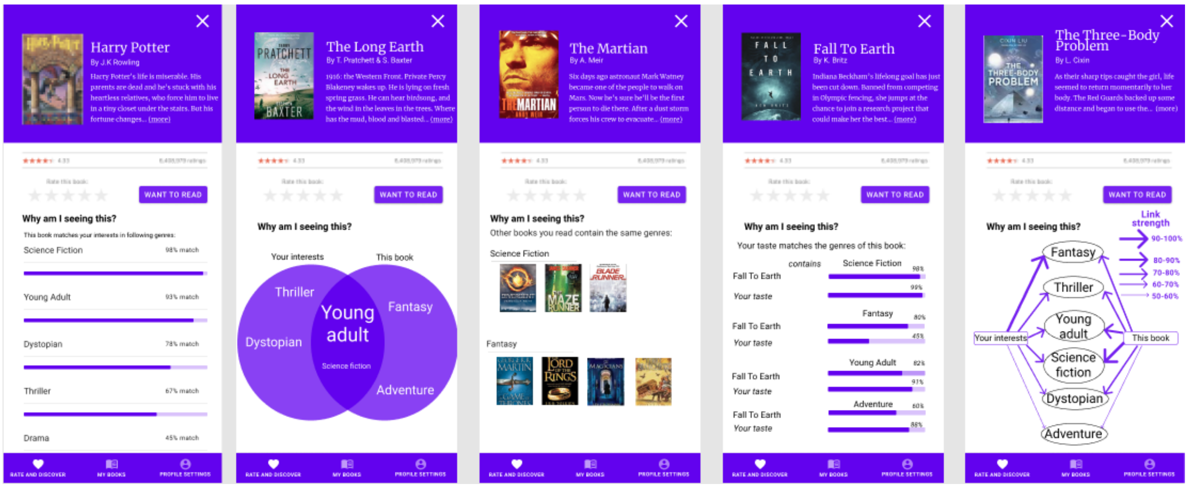


Figure 4.1: Revised visualizations after the results of the questionnaire had been collected.

lot of space on the screen and hence have been made smaller. The visualization using Venn diagrams has been enhanced by giving the genres a weight, representing the importance of those genres in the recommendation decision. The weights are visually represented by writing the genres in a bigger size when they have a higher weight. For the link strength visualization, which was the hardest to understand for the participants, a legend explaining the weights of the arrows has been added. The revised visualizations can be seen in figure 4.1[3].

All the feedback provided about the visualizations is given in table 4.3. The implemented solutions are given in the same table.

Problem two highlights that it is important to have clear parts in a visualization. In this particular case just writing "% match" without as clear meaning confuses the end users. Problem four highlights a similar issue in which again bar charts are present, but again the "% match" is not clear, in this case how a single book matches with a genre.

Problem three shows that the Venn diagram is missing numerical information and that by some users this is desired. This numerical information can be implicitly placed using different font sizes for different genres.

Some users also commented that the other books visualization helps them as it gives them a visual cue to their past experiences. Others also mentioned that the visual nature of the Venn diagrams and its familiarity as a visualization technique, helps them retrieve the information at a glance.

[3]The prototype implemented in Figma can be found at `https://www.figma.com/file/NpCN0astPv3QJXAsgt5w1F/Low-fidelity-with-components-with-TA-feedback?node-id=0%3A1`

Problem	Visualization	Feedback	Solution
1	Bar charts	The bar charts are too wide	The bar charts made less wide.
2	Bar charts	It is not clear what % match means	The text "This book matches your interests in following genres" has been added.
3	Venn diagram	It is not clear which genres are more important here	More important genres are written in a bigger font size and vice versa.
4	Double bar charts	It is not clear that a book contains the genres that are listed	The text "contains" is added
5	Link strength	It is not clear what the different widths of the arrows mean	A legend with the different percentages is added

Table 4.3: Feedback for the visualizations during the pilot study and implemented solutions.

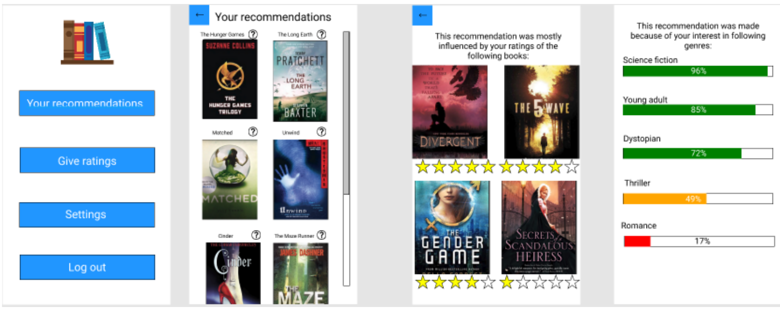
4.2 Prototype of application

4.2.1 Initial version

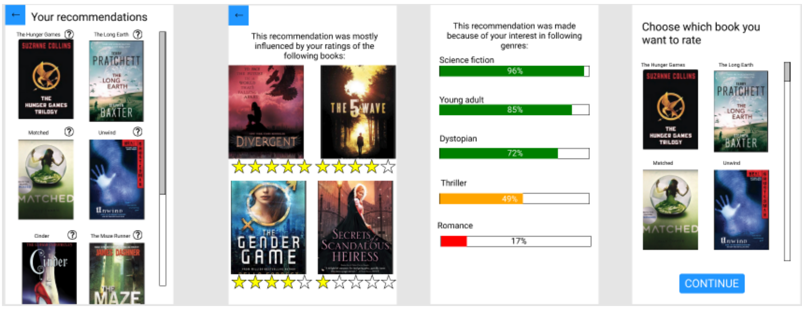
In an initial version of the prototype for the application I designed a few screens for the application which allowed users to create an account or login with an existing account. When someone is creating an account they will initially be prompted to fill in their email address and set a user name and password. After this they are given a screen containing a few books of which they can select books they like to set up their profile.

Once their account has been created or they have logged in to their existing account, the user can choose to go to view their recommended books, give ratings to books or change their profile settings. Clicking on one of the recommended books in the recommendation screen gives a pop-up showing a visual explanation as to why the recommendation was made. The screens of the prototype can be found in figure 4.2[4].

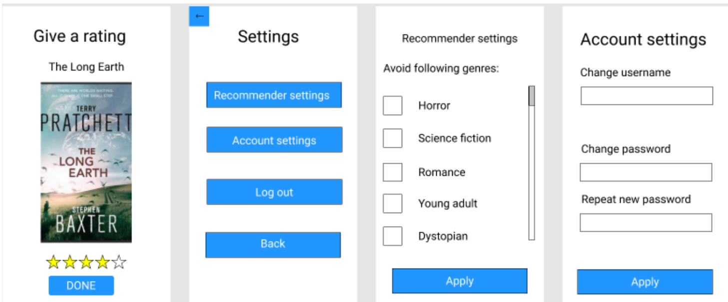
[4]The prototype implementations made in Figma can be found at `https://www.figma.com/file/4YQAL0i8YJeOU72SzQ9Fnc/Book-recommendation-app?node-id=0%3A1`



(a) Screens 1 through 4 of the initial low-fidelity prototype.



(b) Screens 5 through 8 of the initial low-fidelity prototype.



(c) Screens 9 through 12 of the initial low-fidelity prototype.

Figure 4.2: All screens of the initial low-fidelity prototype.

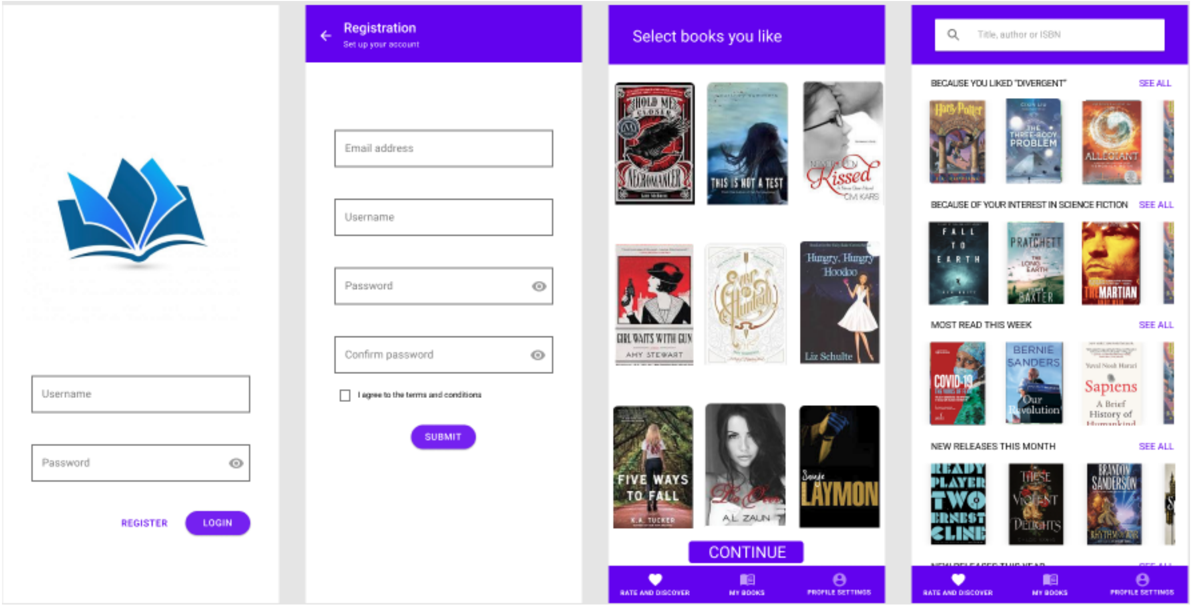


Figure 4.3: Screens of the revised low-fidelity prototype

4.2.2 Revised version

After some initial internal feedback, mainly from my thesis mentor, I created a new version working with predefined Figma components. This gives the application a more professional look. The screens of the revised version are provided in figure 4.3[5]. The screens providing visual explanations are omitted in this figure as they were already shown in figure 3.2.

4.2.3 Evaluation low-fidelity prototype application

Think aloud study

To evaluate the low-fidelity prototype of the application I conducted a think-aloud study. For this think-aloud study I provided my participants with tasks to do within the application. I also asked my participants to say everything that comes to their mind out loud when doing their given tasks. This way I was able to ascertain whether there are certain issues with the navigation of the application. In this think-aloud study I am only assessing the intuitiveness of the navigation of the application and not of the visualizations, that is what the previous questionnaire was for.

The tasks that I asked my participants to do are given in table 4.4. A total of five participants were recruited for the think aloud study, I considered this sufficient as during the fifth participant's think-aloud study no new major issues were being brought up. The demographics of these participants is given in table 4.5.

[5]The prototype implementations made in Figma can be found at `https://www.figma.com/file/iWkZwlp8yqGYCwHTxCpaCJ/Low-fidelity-with-components-before-TA`

Task	Description
1	Create an account
2	Log in with an existing account
3	Check the recommendation for the book ”Harry Potter”
4	Check the recommendation for the book ”The Long Earth”
5	Rate the book ”The Three Body Problem” five stars
6	Go to the profile settings

Table 4.4: List of tasks given to the participants of the think-aloud study

Age	Gender	Job	Amount of books read a year
21	F	Student	10
20	F	Student	1
45	M	Engineer	5
46	F	Financial Assistant	48
50	M	Engineer	10

Table 4.5: Demographics of the participants who took part in the think-aloud study of the low-fidelity prototype.

The issues with the application that were found via the think-aloud study are given in table 4.6. Most notably four out of five participants had problems finding the books of which they were told to check the recommendations, this was solved by putting the titles of the books in a textual form underneath the cover images of the books. Another major issue that became clear during the think aloud study was that participants were uncertain about when exactly they finished the registration process. This resulted in them often asking me whether they were finished before they actually were. This was solved by providing the user with a popup screen informing them that the registration process is finished. Also the end of the registration process was redefined, initially it was finished after the user has selected a few books that he/she likes from the setup screen, this has been changed so that the registration is finished as soon as the user has provided an email address, user name and password.

All the solutions that were adopted to the found problems can be found in table 4.7. No solution has been proposed for problem 11 as the initial book selection screen to create an account is supposed to contain representative books that allow the recommender system to avoid the cold start problem. This will be a very limited selection of books and a search bar should not be necessary.

4.2.4 Revised version after feedback

All the proposed solutions in table 4.7 are also the solutions that were finally implemented in the revised version of the low-fidelity prototype. The screens of the

Number	Description	Amount of users
1	The ”rate and discover” does not redirect to the home screen	1
2	It is unclear when the profile set up is finished. It looks done when the preferences are being asked	3
3	”Your recommendations” screen after the profile setup seems like a continuation of the preferences being asked	1
4	It is hard to find books solely based on the given images	4
5	The ”rate and discover” navigation bar at the bottom is confusing during the ”select your preferences” screen	2
6	There is no option to register with an email instead of a username	2
7	There is no way to actually read the terms and conditions	1
8	It is not clearly when a user is logged in	1
9	It is not clear why it is being asked to ”select books you like”	1
10	The ”SEE ALL” button does nothing	3
11	There is no search button at the ”select books you like” screen	1
12	There is no keyword search in the search bar of the recommendation screen	1

Table 4.6: Issues with the application when using the low-fidelity prototype. These issues arose during the think-aloud study

final low-fidelity prototype are given in figure 4.4[6]. Here again the screens with visual explanations have been omitted as the revised version of those was already given in figure 4.1.

4.3 Summary

In this pilot study feedback about the initial version of the application and the visualizations has been gathered. Some usability issues were found in both the application and the visualizations.

For the application users mainly wanted more information about where they were in the application and what they were supposed to do. It was for instance not clear when the registration process was finished or why the user had to select books he/she likes.

[6]The prototype implemented in Figma can be found at `https://www.figma.com/file/NpCN0astPv3QJXAsgt5w1F/Low-fidelity-with-components-with-TA-feedback?node-id=0%3A1`

Problem number	Possible solution
1	Add the possibility to use the heart to navigate to the home screen
2	Change button text from "Continue" to "Create account" for "Select books you like" screen and created a popup saying "Account setup successfully" and redefine end of account creation
3	Add "Account setup successfully" at the top of the "Your recommendations" screen
4	Add written titles below the book images
5	The "rate and discover" button navigation can be removed from the "select your preferences" screen
6	Change "username" field to "username or email" field
7	Create a link to the terms and conditions
8	Add a "log out" button at the top right corner of every screen where the user is logged in
9	Change "Select books you like" to "to understand you preferences please select a few books"
10	Let SEE ALL redirect to a complete list of recommendations
11	-
12	Add a keyword search in the search bar

Table 4.7: Possible solutions to the problems found via the think-aloud study. These are also the solutions that were implemented.

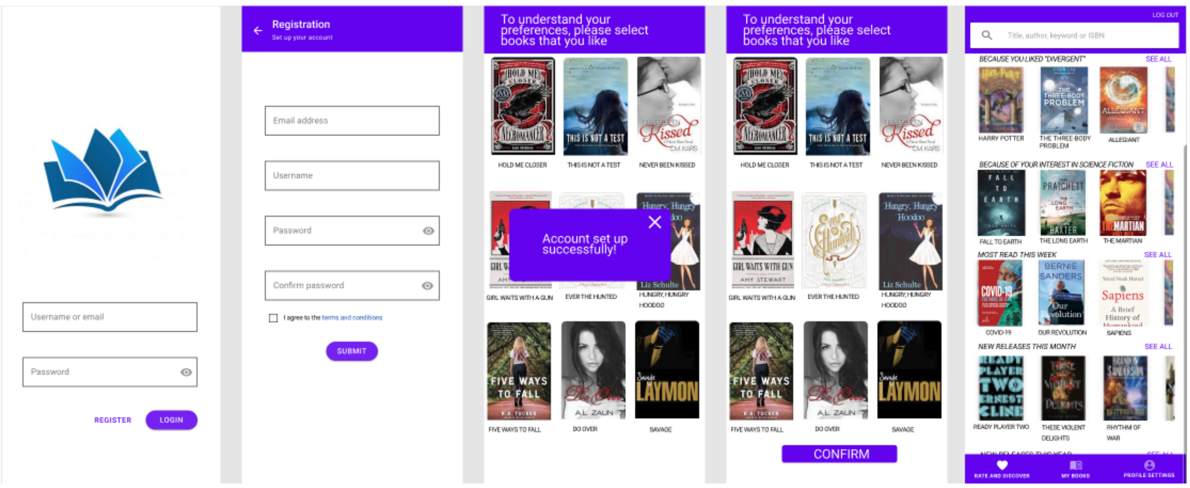


Figure 4.4: Revised low-fidelity prototype after user feedback.

For the visualizations it also became apparent that a good explanatory text as part of to the visualizations is important. Familiarity in these visualizations is also found to be important.

Chapter 5

Implementation

In this chapter the actual implementation of the application is discussed. The algorithms and frameworks for the code base are all discussed in detail here. The implementation of the visualizations and the application takes into account the feedback received from the pilot study.

5.1 Code

The code for the application has been written in the Meteor framework[1]. Meteor is a Javascript framework which allows for apps to be deployed on iOS and Android with one code base. Which for this research is a huge advantage as this allows me to send the application to anybody with a mobile phone, independent of what operating system their phone is using[2].

5.2 Dataset

As dataset I used the goodread-books-10k dataset[3] which consists of 10.000 unique books. These books are annotated with the genres they contain. There are 34.252 different book tags that can be assigned to the books. These tags are the genres the books consist of. Besides that there are also 981.756 user ratings given to the books provided by a total of 53.424 users.

The dataset contains all fields that I needed for the books (i.e. title, book id, ISBN, authors, average rating, ratings count). As the books were not directly annotated with the names of their genres, I did some preprocessing in Python. Here I matched all the books with their tags via goodreads_book_id and tag_id in the book_tags.csv file. This allowed me to create a new csv file which contains the book ids of all books and all the genres they contain as separate entries. So if for example

[1]More info about this framework can be found at `https://www.meteor.com/`

[2]At the time of writing Android and iOS have over 99% of the global mobile OS market share worldwide `https://gs.statcounter.com/os-market-share/mobile/worldwide`

[3]This dataset can be found at `https://github.com/zygmuntz/goodbooks-10k`

goodreads_book_id	genre
1	thriller
1	dystopian

Table 5.1: Example of how genres are placed in the csv file after preprocessing.

the book with id 1 contains both the thriller and dystopian genre, it was kept as two separate lines in the csv file as can be seen in table 5.1.

5.2.1 Book descriptions

Unfortunately the goodreads dataset does not contain any book descriptions linked to the books. However I did consider it necessary to have those in order to have a good book recommender application. Henceforth I wrote a Python script that uses to Google books API[4] to fetch explanations, based on ISBN. After this I added this new information to my already preprocessed dataset.

5.3 Database

As Meteor provides seamless integration with MongoDB, I opted to use MongoDB as the database for the backend of my application. In this database I stored the information of all the books (i.e. title, book id, ISBN, authors, average rating, ratings count and description), the genres each book contains and the top five similar books for each unique book (as introduced in section 5.4).

5.4 Recommender system

The recommender system I designed is based on content-based filtering. In this recommender system the genres from the books are taken as features. These genres are the top-level genres as defined by goodread books[5]. This means that a total of 94 genres are used as the features for the different books in the recommender system. The five most similar books for each unique book are calculated using the Hamming distance between their features. These five most similar books are then used to make new recommendations in which the given star rating to the books decide the weight given to the similar books. This means that if a book occurs two or more times in different similarity lists, its weight will be increased according to the different star ratings, giving it a higher rank. This rank decides in which order the books are shown (i.e. the books with the highest recommendation scores/ranks will be shown before the books with lower recommendation scores/ranks). This recommendation

[4]More information about the Google books API can be found at `https://developers.google.com/books`

[5]These top-level genres can be found here `https://www.goodreads.com/genres/list?utf8=%E2%9C%93&filter=top-level`

system is based on the recommender system used in [9] where the researchers were investigating possible solutions for the cold start problem in the context of a recipe recommender system application. The pseudo-code for this algorithm can be found in algorithm 1.

Algorithm 1: Algorithm in pseudo-code to update the recommendation scores for a user.

```
parameter : i=user ID
parameter : ratedBook=ID of the newly rated book by user i
Result: Updates the book recommendations for user i
/* Get five most similar books to ratedBook from database    */
similarBooks = getSimilarBooks(ratedBook);
for each similarBook in similarBooks do
    /* Retrieve the rating of the rated book given by the user
       */
    rating = getRating(i, ratedBook);
    if hasScore(i, similarBook) then
        /* If the user already has a recommendation score for
           this book, add the new rating to the score.        */
        addToScore(i, similarBook, rating);
    else
        /* If the user does not have a recommendation score for
           this book, create one.                             */
        addNewScore(i, similarBook, rating);
    end
end
```

5.5 Cold-start problem

To circumvent the cold-start problem the user is asked to select at least five books out of a list of 20 representative books. These selected books are used to create the initial recommendations for the new user. Even though technically speaking there is no cold-start problem when using content-based filtering I still did this to help the user find interesting books. Showing all 10.000 books would make it difficult for the users to find books that interest them. By asking a few initial questions I can guide the user to find interesting books more quickly.

5.5.1 Selecting representative books

To select the representative books I wrote an algorithm which finds the 20 most prevalent genre tuples of all books (for instance [sci-fi, dystopian, love] for The Hunger Games). This ensures that for each important genre tuple a book will be present in the representative books. With these 20 most prevalent book genre tuples the representative books are chosen by taking the books for each of these genre tuples with the highest amount of given ratings. This ensures that well-known books are selected, which increases the chances of the user knowing the books. The pseudo-code for this algorithm can be found in algorithm 2.

Algorithm 2: Algorithm in pseudo-code to update the recommendation scores for a user.

```
Result: Finds the representative books out of a list of books with genres
/* Retrieve all books from dataset                                  */
books = getAllBooks();
/* Retrieve all genres for all books from dataset                   */
bookGenres = getAllGenres();
for each book in books do
    /* Retrieve the genres that this book contains                  */
    genresTuple = bookGenres[books];
    genresTupleCounts = 0 ;
    if hasCount(i, similarBook) then
        /* If the genre tuple already has a count add 1 to the
           count.                                                   */
        genresTupleCounts[genresTuple]++;
    else
        /* If the genre tuple does not have a count yet create
           one.                                                     */
        genresTupleCounts[genresTuple]=1;
    end
end
/* Calculate the 20 most prevalant genre tuples                     */
representativeGenreTuples = get20HighestCounts(genresTupleCounts);
representativeBooks = [] ;
/* For each genre tuple find the most popular book that has
   exactly this genre tuple as genres                               */
for each genreTuple in representativeGenreTuples do
    mostPopularBook = getMostRatedBook(genreTuple);
    /* Add the most popular book to the representative books        */
    representativeBooks.append(mostPopularBook);
end
return representativeBooks
```

The books that got selected via this algorithm are provided in table 5.2. As can be clearly seen in the last column of the table each of the selected representative books is popular as they all have a lot of received ratings from users. The genre tuples that these selected representative books represent are also provided in table 5.2.

goodreads_book_id	Title	Genres	Ratings received
9	Angels & Demons	adult, fiction	2.001.311
41	The Lightning Thief	fiction	1.366.265
197	Dead Until Dark	fiction, adult, science-fiction-fantasy	420.764
203	Beautiful Disaster	adult, fiction, love	418.309
49	New Moon	love, fiction	1.149.630
42	Little Women	adult, fiction, unfinished	1.257.121
59	Charlotte's Web	animals, fiction	1.064.521
61	The Girl on the Train	fiction, dark, adult	1.008.778
136	Divine Secrets of the Ya-Ya Sisterhood	fiction, relationships, adult	465.676
65	Slaughterhouse-Five	adult, fiction, war	846.488
2	Harry Potter and the Sorcerer's Stone	science-fiction-fantasy, fiction	4.602.479
7	The Hobbit	fiction, adult, science-fiction-fantasy, unfinished	2.071.616
535	Daughter of Smoke & Bone	war, fiction	198.283
76	Sense and Sensibility	love, romantic, adult, fiction	738.894
139	Miss Peregrine's Home for Peculiar Children	unfinished, fiction	613.674
433	Dark Lover	dark, fiction, love, adult	227021
110	A Clash of Kings	war, science-fiction-fantasy, fiction, adult	523.303
251	The Cat in the Hat	fiction, poetry, animals	314.016
111	The Memory Keeper's Daughter	unfinished, fiction, relationships, adult	501.430
913	The Joy of Cooking	adult, non-fiction, fiction	102.348

Table 5.2: The selected representative books, their IDs, the genres they consist of and the amount of ratings received as calculated with algorithm 2.

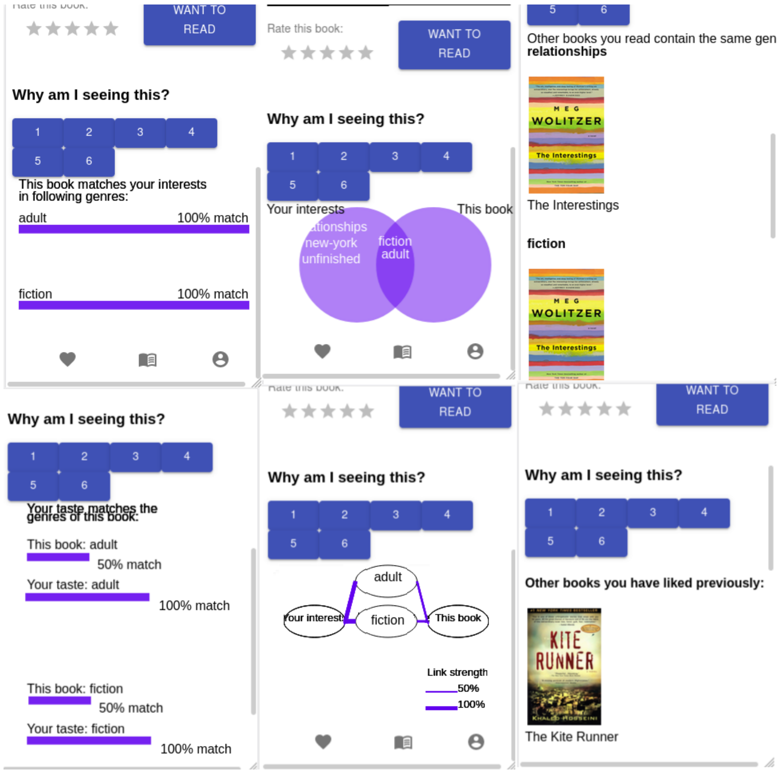


Figure 5.1: Screenshots of the initial visualization in the initial version of the application.

5.6 Visualizations

The visualizations itself have been implemented in the d3.js[6] framework. This framework can easily be integrated with Meteor via npm. The initial version of the visualizations in the actual application can be seen in figure 5.1. In the remainder of this section I explain the algorithms used to generate the visualizations.

5.6.1 Bar charts visualizations

For the calculation of the % matches in the bar charts of the bar charts visualization the common genres of the recommended books and the previously liked books of the user are taken into account. Only the previously liked books with a rating of greater than two stars provided by that user are used for this. For each genre the amount of books that the user has previously given a rating greater than two stars are counted and this number is divided by the total amount of previously liked books of that user. Only the genres with at least 50% match are kept. To avoid percentages converting to zero, as when a user rates more books, the proportion of genres might decrease, only the 100 most recently rated books are used. This also allows the visualization

[6]More info about the d3.js framwork can be found at `https://d3js.org/`.

to stay up to date with users whose preferences change over time. A more detailed description in pseudo-code of the algorithm to calculate the percentages is provided in Algorithm 3.

5.6.2 Venn diagram visualization

For the Venn diagram visualization the common genres between the books the user has previously liked and the recommended book are placed in the overlap of the Venn diagram. The genres of the books that the user has previously liked, but are not in the recommended book are placed on the left side of the Venn diagram. The genres of the recommended book that are not in the genres of the book that the user has liked in the past are placed on the right side of the Venn diagram.

To ensure that the visualization is up to date with possibly changing preferences of the user, only the last 100 ratings are considered. To also ensure that the listed genres actually interest the user only the books that received a rating greater than two from that user are considered. Finally as to not clutter the Venn diagram in each part the amount of genres is limited to the six most occurring genres for that part of the Venn diagram. As no book has more than four genres, it is not necessary to limit the amount of genres on the right side of the Venn diagram.

A more detailed description for the algorithm used for Venn diagram visualization is found in Algorithm 4.

5.6.3 Other books visualization

For the other books visualization the common genres of the recommended book and the genres of the previously liked books by the user are used. Once these genres have been calculated the previously liked books that have at least one of those genres are selected. After this they are added to the list.

To be sure the visualization stays up to date with possibly changing preferences of the user, only the last 100 ratings are used. As to not clutter the visualization with too much information, the total amount of genres is limited to the six most important genres. These are the genres that are most frequently found in the previously liked books. Again only books with a rating greater than two stars are considered.

A more detailed description for the algorithm used for the other books visualization is found in Algorithm 5.

5.6.4 Double bar chart visualization

For the double bar chart visualization the percentages for the user's taste are calculated the same way as in Algorithm 3, that was used for the bar chart visualization. These are the percentages used for the "Your taste" part of the visualization.

The percentages for the genres of the recommended book are calculated by counting the total genres in this book and taken the inverse of that. So if a book contains four genres, each genre has 25% match. These are the percentages used for the "This book" part of the visualization.

Algorithm 3: Algorithm for the calculation of the percentages of the bar chart visualization.

```
Result: Percentages for the bar chart visualization
parameter: Parameter for user, u
parameter: Parameter for the recommended book that is being viewed,
           recommendedBook
/* Get the genres of the recommended book                          */
recommendedGenres = recommendedBook.getGenres();
/* Get the 100 most recently rated books from the user             */
userBooks = ratedBooks(u, 100);
/* Variable to count the amount of books the user has given at
   least two stars.                                                */
likedBooksCount = 0 ;
for book in userBooks do
    /* Map to keep track of how often each genre occurs in the
       previously liked books                                      */
    genresCounts = {} ;
    /* Get rating given by user u to this book                     */
    rating = getRating(book, u);
    /* If the rating is greater than two stars, add the genres
       of the book to the count                                    */
    if rating > 2 then
        bookGenres = book.getGenres() ;
        for genre in bookGenres do
            likedBooksCount++;
            if genresCounts[genre] not null then
                genresCounts[genre]++ ;
            else
                genresCounts[genre] = 1 ;
            end
        end
    end
end
/* Map variable to add the percentages for each genre              */
percentages = {} ;
/* For each genre in genresCounts, add the percentages to the
   map                                                             */
for genre in genresCounts do
    /* The percentage for the genre is the count divided by the
       total amount of books the user has given a rating > 2.
       */
    percentage = genresCounts[genre] / likedBooksCount ;
    /* The percentage is added to the result if it is at least
       0.5                                                         */
    if percentage > 0.5 then
        percentages[genre] = percentage ;
    end
end
return percentages
```

Algorithm 4: Algorithm for the calculation of the genres for the different parts of the Venn diagram.

```
Result: Different parts of the Venn diagram for Venn diagram visualization
parameter: Parameter for user, u
parameter: Parameter for the recommended book that is being viewed,
           recommendedBook
/* Get the genres of the recommended book                        */
recommendedGenres = recommendedBook.getGenres();
/* Get the 100 most recently rated books from the user           */
userBooks = ratedBooks(u, 100);
for book in userBooks do
    /* Map to keep track of how often each genre occurs in the
       previously liked books                                    */
    genresCounts = {} ;
    /* Get rating given by user u to this book                   */
    rating = getRating(book, u);
    /* If the rating is greater than 2 stars, add the genres of
       the book to the count                                     */
    if rating > 2 then
        bookGenres = book.getGenres() ;
        for genre in bookGenres do
            if genresCounts[genre] not null then
                genresCounts[genre]++ ;
            else
                genresCounts[genre] = 1 ;
            end
        end
    end
    genresCounts.sort() ;
    overlapGenres = [] ;
    for genre in genresCounts do
        if genre in bookGenres and overlapGenres.length < 7 then
            overlapGenres.add(genre);
        end
    end
    /* The genres for the left part of the Venn diagram are the
       top six genres of the previously liked books              */
    leftGenres = genresCounts[0:6] ;
    /* The genres for right part of the Venn diagram are the
       genres of the recommended book                            */
    rightGenres = recommendedGenres ;
end
return leftGenres, overlapGenres, rightGenres
```

Algorithm 5: Algorithm for the calculation of the genres and books for the other books visualization .

```
Result: Books and genres for the other books visualization
parameter: Parameter for user, u
parameter: Parameter for the recommended book that is being viewed,
           recommendedBook
/* Get the genres of the recommended book */
recommendedGenres = recommendedBook.getGenres();
/* Get the 100 most recently rated books from the user */
userBooks = ratedBooks(u, 100);
likedBooks = [] ;
for book in userBooks do
    /* Map to keep track of how often each genre occurs in the
       previously liked books */
    genresCounts = {} ;
    /* Get rating given by user u to this book */
    rating = getRating(book, u);
    /* If the rating is greater than two stars, add the genres
       of the book to the count */
    if rating > 2 then
        bookGenres = book.getGenres() ;
        /* Add the book to the list of the liked books */
        likedBooks.add(book) ;
        for genre in bookGenres do
            if genresCounts[genre] not null then
                genresCounts[genre]++ ;
            else
                genresCounts[genre] = 1 ;
            end
        end
    end
    genresCounts.sort() ;
    overlapGenres = [] ;
    for genre in genresCounts do
        if genre in bookGenres and overlapGenres.length < 7 then
            overlapGenres.add(genre);
        end
    end
    /* The genres for the left part of the Venn diagram are the
       top six genres of the previously liked books */
    booksList = ;
    for genre in overlapGenres do
        for book in likedBooks do
            if genre in book.getGenres() then
                booksList[genre].add(book);
            end
        end
    end
end
return booksList
```

A more detailed description for the algorithm used for the "This book" part of double bar chart visualization is found in Algorithm 6.

Algorithm 6: Algorithm for the calculation of the percentages for the "this book" part of the double bar chart visualization.

```
Result: Percentages for "this book" for the double bar chart visualization
parameter: Parameter for the recommended book that is being viewed,
           recommendedBook
/* Get the genres of the recommended book                          */
recommendedGenres = recommendedBook.getGenres();
/* Get the 100 most recently rated books from the user            */
percentages = {};
likedBooks = [] ;
for genre in recommendedGenres do
    /* Map to keep track of how often each genre occurs in the
       previously liked books                                      */
    percentages[genre] = 1 / recommendedGenres.length() ;
end
return percentages
```

5.6.5 Link strength visualization

The link strength visualization uses the same algorithm to calculate the percentages for the "Your interests" part of the visualization as in the bar chart visualization. The code for this is provided in Algorithm 3. To avoid information overload, the amount of displayed genres is limited to the top six most important genres. These are the genres with the highest associated percentages.

For the "This book" part of the visualization, the same algorithm as for the "This book" part of the double bar chart visualization is used. This can be found in Algorithm 6.

Again only books with a rating of more than two stars are considered and only the last 100 rated books are taken into account.

Henceforth the link strength visualization uses the same algorithms as the double bar chart visualization but displays this information in a different manner.

5.6.6 Baseline visualization

Finally the baseline visualization shows the books that the user has liked in the past, that have the recommended book in its related books. These are the books that had a addToScore or addNewScore function call for the current book in algorithm 1. Only the books that were given a rating of at least two are given. A limit of the 100 last ratings is also used to ensure the visualization remains up to date with possibly changing interests of the user.

5.7 Deployment

The deployment of the application has happened on the Picasso server of the research group[7]. An apk file was generated using Meteor which could then be send to the participants of the user studies to evaluate the application. The surveys used for the first and second user study were made in LimeSurvey and are also hosted on the Picasso server.

[7]The application can be found at `picasso.experiments.cs.kuleuven.be:3325/`

Chapter 6

First user study

During a first user study I aimed to ascertain whether any major issues remained within the high fidelity prototype of the application. I also gathered some feedback about the initial visualizations in the initial version of the application. The implementation of this high fidelity prototype was described in the previous chapter.

The feedback about the visualizations during the first user study is mainly to find any remaining development issues. Possible issues with the algorithms could also be found during this first user study. Lastly the recommender system was also evaluated in this step. I did this by conducting a think-aloud study via Zoom in which the participants are called and are given a total of nine tasks to do in the application. While a participant did the tasks I took notes of where the participants were struggling with finding things in the app.

After completing the tasks the participants were also asked their general opinion about the app and whether they had any further remarks.

At the end of the Zoom call the participant is invited to a survey which contains SUS [7], NASA TLX [14] and Resque [31] questions. The SUS and NASA TLX questions help to get feedback about the application itself. The Resque questions get feedback about the recommender system and the integration of visualizations.

A few screenshots of the version of the application used for the first user study are provided in figure 6.1.

6.1 Tasks

The nine tasks provided to the users are provided in table 6.1.

To counteract the fact that some of the tasks in table 6.1 can become trivial after having successfully finished another task, I created two Latin squares. Namely for tasks 2, 3 & 4 and tasks 7 & 8. The Latin squares for these tasks are provided in tables 6.2 and 6.3.

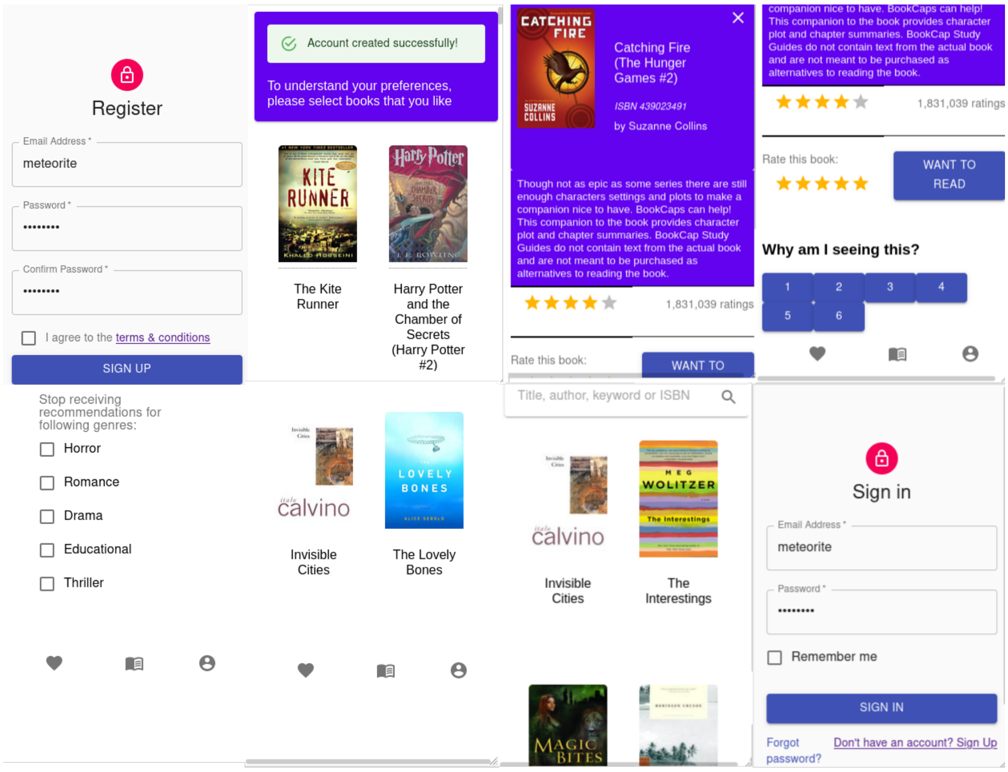


Figure 6.1: Screens of the initial version of the application.

Task	Solution
1	Create an account, in doing so select the books Secret Life of Bees and The joy luck club as books that you like.
2	Rate the book Keeping you a secret 5 stars
3	Rate the book A Reliable Wife 4 stars
4	Rate the book Breaking Dawn 4 stars
5	Mark the book Catching Fire as want to read
6	Mark the book The Brief Wondrous Life of Oscar Wao as want to read
7	Go to your profile and select that you do not want to receive recommendations from the Thriller genre anymore
8	Go to the list of books that you selected as want to read
9	Scroll through the different explanations for Catching Fire

Table 6.1: The tasks given to the participants to evaluate the high-fidelity prototype.

Order	Task	Task	Task
1	2	3	4
2	3	4	2
3	4	2	3

Table 6.2: Latin square used for the evaluation of the high-fidelity prototype for tasks 2, 3 & 4.

Order	Task	Task
1	7	8
2	8	7

Table 6.3: Latin square used for the evaluation of the high-fidelity prototype for tasks 7 & 8.

6.2 Questionnaire

The questions used during the questionnaire after the Zoom call are provided in table 6.4. These questions contain subset of the NASA-TLX [14] and Resque questionnaire [31] and the full SUS questionnaire [7]. These questions are accompanied by two small questions to understand the user profile of the participants. The actual questionnaire is hosted on the Picasso server of the research group[1].

6.3 Results

6.3.1 Demographics

A total of ten male and six female participants participated in this first user study. Most of them were students and the median age of participants was 24 years.

6.3.2 High-fidelity think aloud study

Problems

A total of 16 participants were recruited during the first user study. 16 was sufficient as no new major issues were being found within the application with the last few participants. The problems found with the application are listed in table 6.5.

Solutions

The implemented solutions to the problems found in the think-aloud study are provided in Table 6.6.

[1]The questionnaire can be found at `http://picasso.experiments.cs.kuleuven.be:3325/index.php/265169?newtest=Y&lang=en`.

User profile	
1	Please select your gender: M/F/prefer not to say
2	What is your age?
SUS questions	Please rate the following questions on a scale from strongly disagree to strongly agree.
1	I think that I would like to use this system frequently.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well integrated
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome to use.
9	I felt very confident using the system..
10	I needed to learn a lot of things before I could get going with this system.
NASA-TLX questions	Please select the answers on a scale from 1 (very low) to 10 (very high)
1	How mentally demanding were the tasks?
2	How successful were you in accomplishing what you were asked to do?
3	How insecure, discouraged, irritated, stressed, and annoyed were you?
Resque questions	Please rate the following questions on a scale from strongly disagree to strongly agree.
1	How would you rate yourself as a computer user? (tech-savviness)
2	Do you tend to trust a person/thing, even though you have little knowledge of it? (trust)
3	The items recommended to me match my interests. (accuracy)
4	The recommender explains why the products are recommended to me. (transparency)
5	I understood why the items were recommended to me. (effectiveness)
6	The information provided for the recommended items is sufficient for me to make a purchase/download decision. (persuasiveness)
7	The items recommended to me are diverse. (diversity)

Table 6.4: Questionnaire provided to the participants at the end of the think-aloud study for the high-fidelity prototype.

Number	Description	Amount of users
1	It is difficult to find a certain book in the initialization books	2
2	It is not clear that the terms and conditions had to be accepted	1
3	It is unclear that a book has received a rating after pressing the stars	6
4	It is difficult to find the "my books" tabs	1
5	The explanation at the initialization screen was not read	1
6	It is unclear that the user preferences are updated after pressing a checkbox	4
7	It is unclear that the heart icon at the bottom navigation guides the user to the home page	4
8	It is unclear when the user has completed the full setup of the account	5
9	It is unclear that a book has been added to the want to read books	2
10	The confirm selection button is not always visible in the initialization screen	7
11	It is not clear that the stars next to a book are the average rating	3
12	It is difficult to find the bottom navigation	1
13	Terms and conditions was a link to an empty page	1
14	It is unclear whether a book on the initialization screen has been selected	1
15	The checkboxes for the profile settings should be inverted	1
16	When a user does not accept the terms and conditions the page is reloaded but the credentials are lost	3
17	It is unclear that the cross icon on a book detail page closes the book detail page	2
18	The terms and conditions are opened on the same page as the app and not in a new tab	1

Table 6.5: Issues with the application when using the high-fidelity prototype. These issues arose during the think-aloud study of the high-fidelity prototype

Problem	Description
1	Sort these books alphabetically
2	The terms and condition are now indicated in red when not accepted and when the user tries to create an account without accepting them a pop up opens up
3	A popup message is now displayed for this
4	The bottom navigation is pinned and now has labels under the icons
5	Create a different page for this on which the user has to press continue before they can go further
6	Now a pop up shows that the preferences have been updated
7	This is now replaced with a house icon and a label is added under the icon
8	A popup message is now displayed for this
9	A popup message is now displayed for this
10	This button is now pinned to the bottom of the screen
11	"Average rating" is now written above the stars
12	The bottom navigation is pinned and now has labels under the icons
13	This page has now been written
14	The opacity of the book image is increased and a checkbox is displayed to indicate that a book has been successfully selected
15	These have been inverted
16	This is fixed by the fact that the terms and conditions are now opened in a new tab
17	This has been replaced with a back icon (arrow pointing left)
18	The terms and conditions are now opened in a new tab

Table 6.6: Solutions to the issues with the application as found during the high-fidelity think-aloud study.

6.3.3 Questionnaires

Resque

To analyze the responses on the resque questionnaire I plotted the responses to each of the questions on a boxplot. This boxplot is provided in Figure 6.2.

These questions were answered on a five point Likert scale with possible answers: Strongly Disagree, Strongly Agree, Neither Agree nor Disagree, Agree, Strongly Agree. Strongly Agree is encoded as a score of 5.0 while Strongly Disagree is encoded as a score of 1.0.

The first question which ascertains the tech-savviness of the users, shows that almost all participants felt confident using computers. However the second question which ascertains how easily the participants trust something new shows that we had

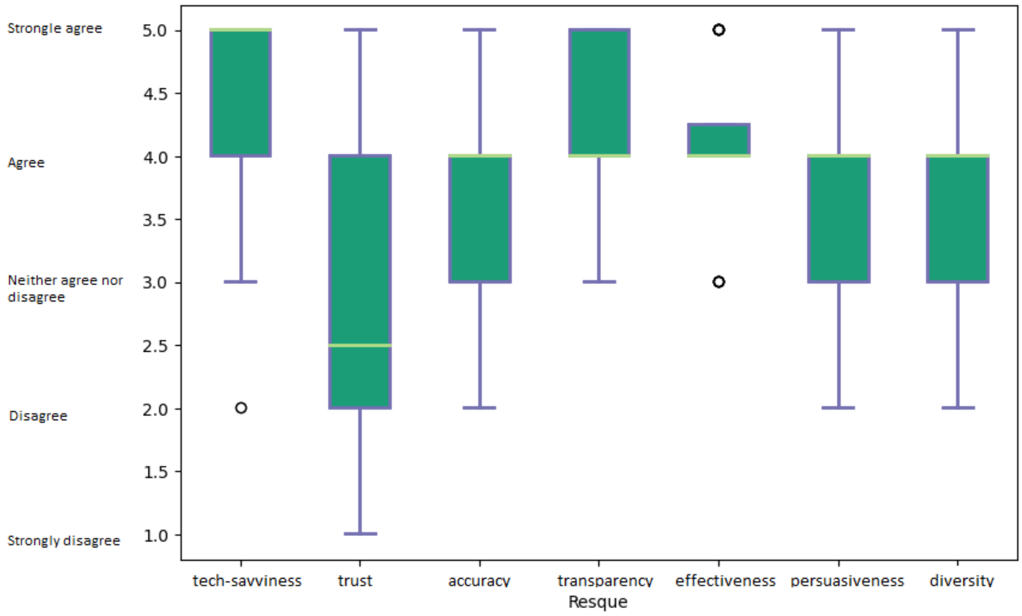


Figure 6.2: Results of the Resque questionnaire for the high fidelity prototype provided as a boxplot.

critical thinking participants with a median score of 2.5 (between Neither Agree nor Disagree and Disagree).

The third question ascertains whether the recommended items match the user's interest. The score for this question has a median of 4.0 (Agree), meaning that most participants agreed to the statement.

The fourth question is an important one as it directly concerns the provided visualizations by asking whether the recommender system explains why a recommendation has been made to the user. a median score of 4.0 (Agree) is found here, while the worst score is 3.0 (Neither Agree nor Disagree). This shows that the visualizations actually provide the user with the sense of having an explanation.

The fifth question is closely related to this by asking whether the user understood why the recommendations were made to him/her. This evaluates the effectiveness of the visualizations. Merely providing an explanation, does not mean the user will automatically understand this explanation. The scores here lay very close for all participants with almost all participants giving a score of 4.0 (Agree). This is a good indication that the visualizations are indeed effective.

The sixth question checks the persuasiveness of the system by asking the users whether they are convinced to make a download/purchase decision. Here a median score of 4.0 (Agree) is obtained, which is a good score.

The final question ascertains the diversity of the recommended items and is an evaluation metric for the recommender system algorithm itself, here again a median score of 4.0 (Agree) is obtained, which is a good score.

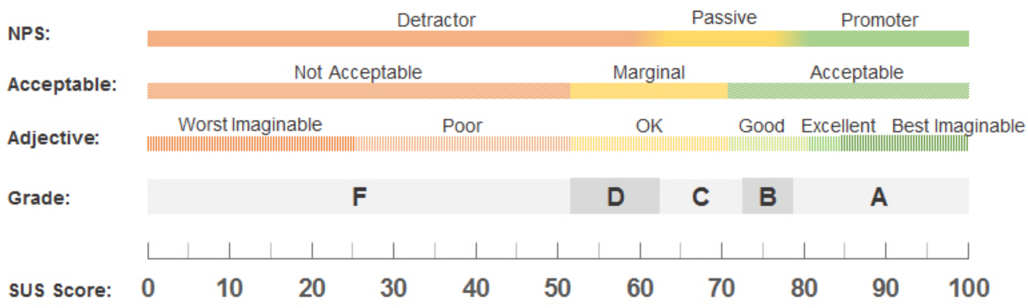


Figure 6.3: System usability scale score as depicted by [1]

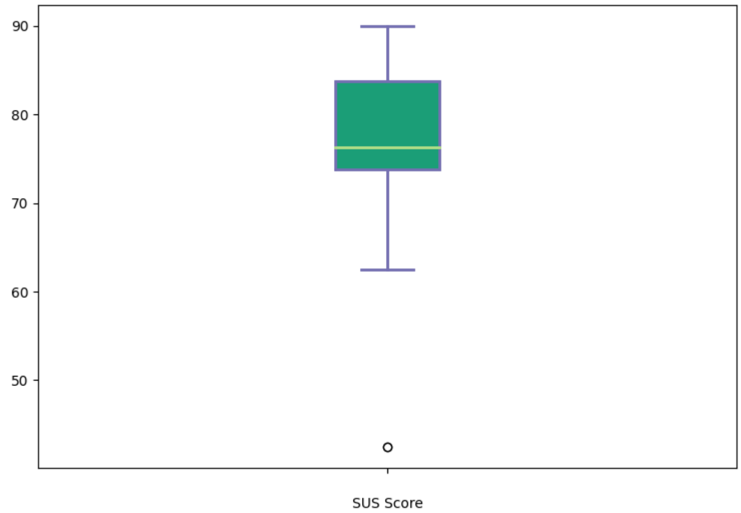


Figure 6.4: System usability scale scores boxplot for the first user study based on the SUS questionnaire answered by the participants.

System Usability Scale The System Usability Scale (SUS) gave a median score of 76 [7]. As can be seen Figure 6.3 this places the application in the B range, clearly indicating that it already has a good usability but there is still room for improvement. A boxplot of the SUS score is also provided in Figure 6.4. In this boxplot it is clear that there is one outlier, namely one participant gave the application a SUS score of 42.5.

NASA TLX For the analysis of the NASA TLX questions I also used boxplots, which are provided in 6.5. A score of one indicates the "very low" response, while a score of ten indicates the "very high" response. All participants gave a score of five or less to the first question, indicating that the tasks were not very mentally demanding. The second question has been fully answered with scores of at least six, showing that all participants considered themselves successful in the accomplishment

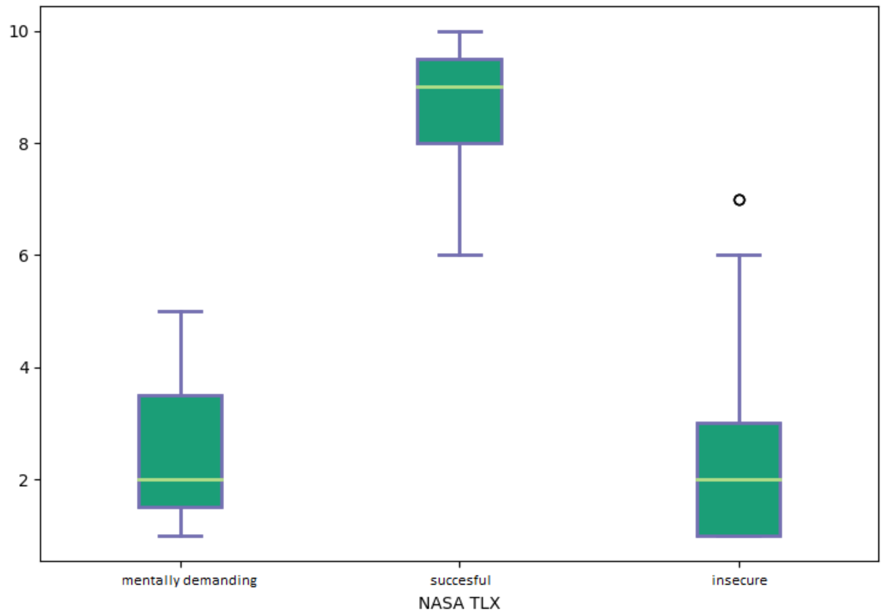


Figure 6.5: Results of the NASA TLX questionnaire for the high fidelity prototype provided as a boxplot.

of the provided tasks. Lastly the third question assessed how insecure or discouraged the participants felt doing these tasks, here an average score of two was replied, with one outlier with a score of seven. This means that some participants felt a bit insecure doing the provided tasks, but in general participants felt confident doing so.

6.3.4 Initial feedback about visualizations in app

Also some feedback about the initial visualizations in the initial version of the app was gathered.

The feedback given by the participants is provided in table 6.7. The solutions to these issues were mostly visual solutions such as changing the position of certain elements. Sometimes a solution also require a change in one of the algorithms for the visualizations. The solutions that were implemented in the final version of the application are provided in table 6.8.

The updated version of the visualizations can be seen in figure 7.2.

6.3.5 Discussion

From the results of the SUS and NASA TLX questions I can conclude that the application is working decently well but that there is still some room from improvement. These improvements are found in the feedback gathered from the think-aloud study. All the issues that arose there have been fixed.

From the results of the Resque questions it also becomes clear that the recommender system is providing accurate and diverse recommendations.

Problem	Visualization	Description	Amount of users
1	bar charts	The distance between the bar charts of different genres is large	2
2	Venn diagram	The genres that are in a smaller text size are too small to read	6
3	Venn diagram	The genres that do not fit in the circle are not readable due to a white background	7
4	double bar chart	A book should have a 100% match with a genre, as a book either contains a genre or not	5
5	double bar chart	The distance between the bar charts of different genres is large	2
6	double bar chart	The % matches are not always in the same place relative to the line in the bar chart corresponding to that genre, it is sometimes more to the right or more to the left depending on the exact percentage	1
7	link strength	Using different width of the lines for the different percentages is very difficult to read	8
8	link strength	A book should have a 100% match with a genre, as a book either contains a genre or not	5

Table 6.7: Feedback about the different initial visualizations in the initial version of the application.

Also for the visualizations itself, the Resque questions show that the visualizations are transparent, effective and persuasive. During the pilot study, these three measures were not researched in more detail yet, they merely evaluate whether giving the end user a visual explanation actually improves the three measures. Some issues and concerns with the visualizations and the algorithms used for them were also raised during the think-aloud study, these issues have been resolved.

Problem	Visual-ization	Solution
1	bar charts	The distance between the bar charts of different genres has been decreased
2	Venn diagram	The genres are now all the same text size
3	Venn diagram	The genres are now written in black
4	double bar charts	A book now has a 100% match with each genre it contains. This was implemented by changing algorithm 6 to return a 100% for each genre of the recommended book
5	double bar charts	The distance between the bar charts of different genres has been decreased
6	double bar charts	The % matches are now always in the same place relative to the line in the bar chart corresponding to that genre
7	link strength	Different colors are now used to indicate different percentages
8	link strength	A book now has a 100% match with each genre it contains. This was implemented by changing algorithm 6 to return a 100% for each genre of the recommended book

Table 6.8: Implemented solutions to the feedback about the different initial visualizations in the initial version of the application.

Chapter 7

Final user study

In the final user study mainly the visualizations itself are evaluated. This is done by conducting a final think-aloud study in which the participants can give their opinion about the different visualizations. The version of the application used in this study is the revised high fidelity prototype with the feedback of the first user study taken into account. After the think-aloud study the participant is asked to fill in a questionnaire. A few general resque questions are asked to evaluate how well the recommender system itself works, while for each visualization three specific resque questions are asked, those help to understand the usefulness of the different visualizations. In this questionnaire there is also a full SUS questionnaire, this allows me to understand whether the usability of the application actually did increase with the feedback from the first user study. A few questions to understand the user profile of the participants with regards to recommender systems are asked, in order to understand different expectations from the visualizations based on their user profile.

The participants are also asked to provide a ranking for the six visualizations.

The verbal feedback of the user is the most important followed by the questionnaire and then the ranking.

So the main goal of this final user study is to evaluate the visualizations, but with SUS questions and general resque questions I also take the opportunity to respectively evaluate the usability of the app and the recommender system.

A screenshot of the screens of the revised application for this final user study can be seen in figure 7.1. The revised visualizations can be seen in figure 7.2. Both of them have been updated with the feedback of the previous user study.

7.1 Interview setup

7.1.1 Tasks

The tasks given to the participants during the final think-aloud study are provided in table 7.1. The name of the visualizations is not said to the participant, as not to influence his opinion. This is only placed in the table for clarity to the reader.

As to not create a bias for the different visualizations, as would be the case by showing each participant all visualizations in the same order, I created a Latin square

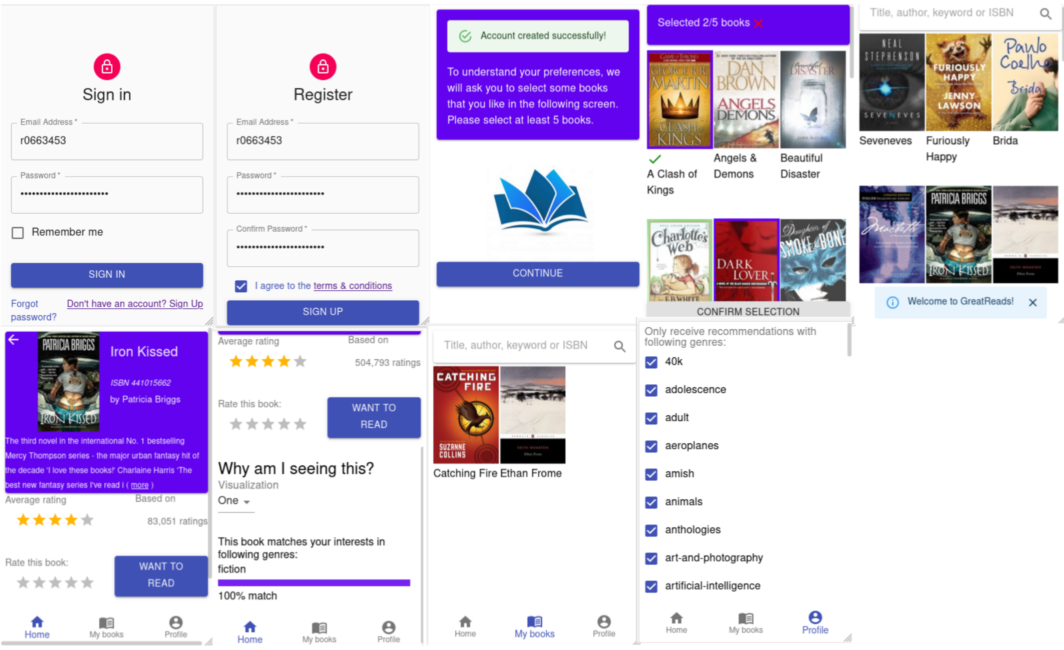


Figure 7.1: Screens of the revised version of the application as used for the final user study.

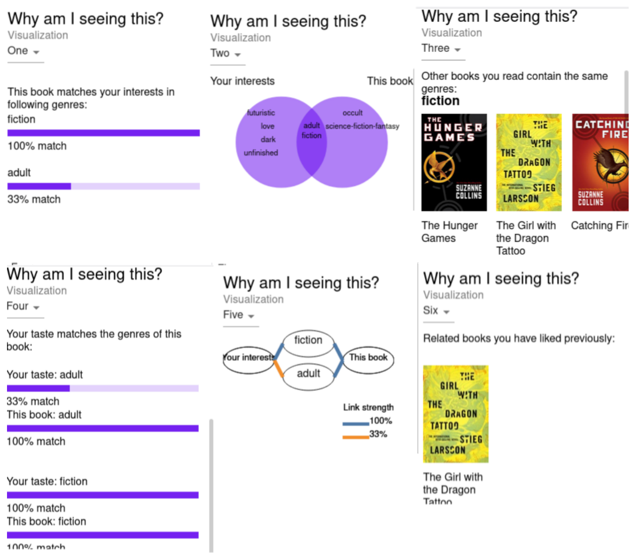


Figure 7.2: Screenshots of the revised visualizations in the application as used for the final user study.

Number	Task
1	Create an account
2	Give 3 books of your choice a rating
3	Open the details of one book
4	Look at the first visualization for this book (bar chart)
5	Look at the second visualization for this book (Venn diagram)
6	Look at the third visualization for this book (other books)
7	Look at the fourth visualization for this book (double bar chart)
8	Look at the fifth visualization for this book (link strength)
9	Look at the sixth visualization for this book (baseline)

Table 7.1: The tasks given to the participants to evaluate the visualizations.

Order	Task	Task	Task	Task	Task	Task
1	4	5	9	6	8	7
2	5	6	4	7	9	8
3	6	7	5	8	4	9
4	7	8	6	9	5	4
5	8	9	7	4	6	5
6	9	4	8	5	7	6

Table 7.2: Latin square used for the evaluation of the visualizations for tasks 4, 5, 6, 7, 8 & 9.

to counteract this effect. This bias is from the fact that the user is more likely to understand visualizations that are shown to him further in the research as he will be more acquainted with visualizations at that point. The Latin square for tasks 5 through 9 is provided in table 1.

During this think-aloud study the participant is asked to give his opinion about each of the different visualizations.

7.1.2 Questionnaire

The participant is redirected to a questionnaire consisting of user profile questions to assess the user profile of the participant. SUS questions to evaluate usability of the application itself. Some general Resque questions to evaluate the recommender system and some specific Resque questions to evaluate each evaluation separately. The questions used in the questionnaire are provided in table 5.1. The actual questionnaire has been hosted on the Picasso server of the research group[1].

The visualization-specific Resque questions were repeated for each different visualization.

[1]The questionnaire can be found at `http://picasso.experiments.cs.kuleuven.be:3325/index.php/11388?newtest=Y&lang=en`

User profile	
1	Please select your gender: M/F/prefer not to say
2	What is your age?
3	How many books do you read a year?
4	Have you ever used a book recommender system such as goodread-books or Amazon to discover new books?
5	Do you use other apps that recommend items based on your previous preferences? If so, which?
6	How often do you end up watching a recommended item/buying a recommended item?
SUS questions	Please rate the following questions on a scale from strongly disagree to strongly agree.
1	I think that I would like to use this system frequently.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well integrated
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.
General Resque questions	Please select the answers on a scale from strongly disagree to strongly agree
1	You consider yourself good with computers.
2	You tend to trust a person/thing, even though you have little knowledge of it.
3	The books recommended to me are diverse.
4	The books recommended to me match my interests.
Visualization-specific Resque questions	Please rate the following questions on a scale from strongly disagree to strongly agree.
1	The visualization explains why the book was recommended to me.
2	I understood why the book was recommended to me.
3	The information provided for the recommended book is sufficient for me to make a download decision.

Table 7.3: Questionnaire provided to the participant at the end of the think-aloud study for the final user study.

7.1.3 Interaction logs

Another possible way to understand the user understanding with regards to the visualizations could be interaction logs. Interaction logs would be logging how long a user is looking at each of the different visualizations. From this it would also be possible to draw conclusions as to how long it takes a user to understand a visualization.

However there are two main disadvantages to this approach considering the interview format. Firstly the user is asked his/her opinion about each different visualizations, which he/she communicates verbally. This implies that the time spent on one visualization is not correlated with the time needed to understand a visualization for the participant, but rather with the time needed for the participant to communicate his/her opinion. Secondly, when assuming the first issue would not be a problem, the learning curve of a participant would still make it difficult to draw conclusions from the interaction log. Namely, the participant is more likely to spend less time on the latter visualizations as he/she already has a better understanding of why the book he/she is viewing has been recommended. This is because of the fact that some of the different visualizations, show the same information as another. For example the double bar charts and link strength visualizations display the same background information (i.e. percentages) but just in a different way. This means the participant will most likely spend less long on the link strength visualization than on the double bar charts visualization if he/she has seen the double bar charts visualization prior to the link strength visualization.

7.1.4 Ranking

After the participant has seen all visualizations in the order assigned to him/her, he/she is asked to rank the different visualization from the one he/she likes the most until and including the one he/she likes the least. If a participant is unable to rank certain visualizations because in his/her opinion they rank equally, the participant is also allowed to rank two or three different visualizations at an equal position.

7.1.5 Evaluation protocol

The evaluation protocol used for this interview is provided below:

```
Tasks assigned to users.
 1. Create an account
 2. When at the home screen, select a book that is recommended and
 open the details page of that book.
 3.
    a. Select the number of the visualization given to you
    b. Repeat step 2b. until each visualization has been seen

User interview protocol (script)
Approximate duration of the interview: 20-30 minutes.
```

```
If the estimated time for the tasks is exceeded by the participant,
the task is concluded and the next task is continued.

Step 1. Call the participant through the Zoom platform (30 seconds)
Step 2. Greet and welcome the participant (10 seconds)
Step 3. Explain the purpose of this interview (1 minute)
Step 4. Let the user download the apk file on his phone (1 minute)
Step 5. Ask the user the questions of his user profile w.r.t.
recommender systems and reading books (1-2 minutes)
1. How many books do you read a year?
2. Have you ever used a book recommender system such as
goodread-books or Amazon to discover new books?
3. Do you use other apps that recommend items based on your
previous preferences? If so, which?
4. How often do you end up watching a recommended
movie/buying a recommended item?
Step 6. Let the user do the first 3 tasks in the
application: (2-3 minutes)
1. Create an account
2. Give 3 books of your choice a rating
3. Open the details of one book
Step 7. Let the user select the visualizations in the order
given to him/her.
(10-15 minutes)
Explain to the user that he/she is now free to give any feedback that
comes to mind when seeing this visualizations.
Step 8. Once each visualization has been seen and all
feedback has been given the user is asked to rank the
visualizations from the one he/she likes the most up
to the one he/she likes the least. (1-2 minutes)
Step 9. Latest comments, general opinion on the application.
What do you think? Why? (2-5 minutes).
They are thanked for their participation in the study,
they are invited to answer the post-interview questionnaire
and then say goodbye to the participant.
```

7.2 Results

7.2.1 User profile

A total of 51 participants were recruited, of which 35 were male and 16 female. The median age of the participants was 23 years.

The responses to the user profile questions are provided in tables 7.4, 7.5 and 7.6. Not many participants had prior experience with book recommender systems, only 12 of them have in fact already used such book recommender system in the past.

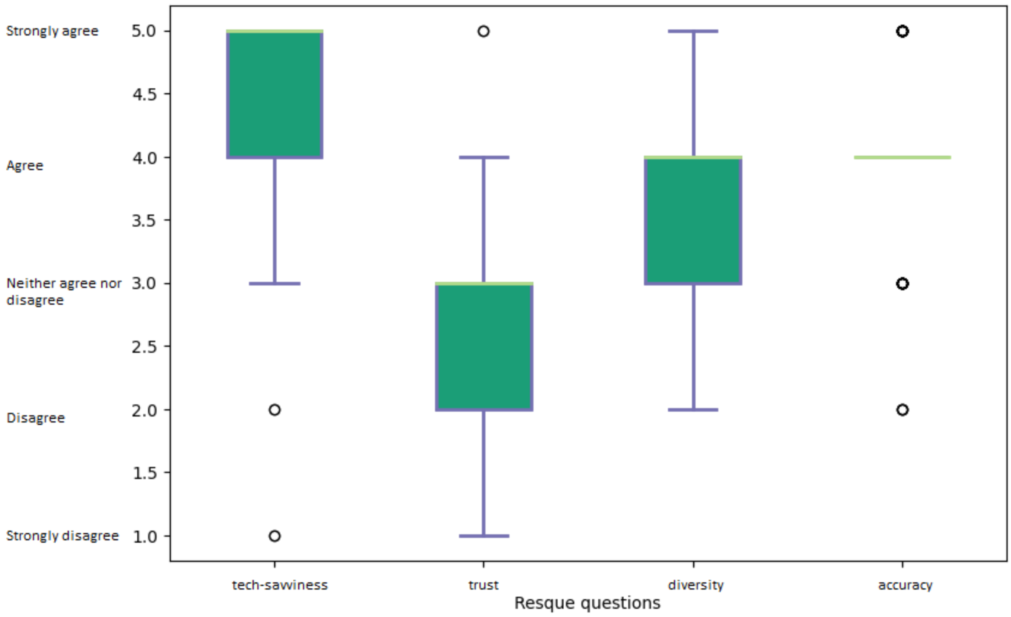


Figure 7.3: Responses to the Resque questions. A score of 5.0 indicates fully agreeing and a score of 1.0 indicates fully disagreeing

However each participant has at least experience with one or more recommender systems. 38 participants also read at least two books on a yearly basis.

7.2.2 General Resque questions

The first question ascertaining the tech-savviness of the participants has a median score of 5.0, meaning most participants consider themselves very good with computers. The second question ascertaining the trust of the participants has a median score of 3.0, meaning about half the participants tend to trust new things quickly, while the other half does not.

Finally both the third and final question have a median score of 4.0, meaning that most participants agree that the recommended books are diverse and match their interest.

These two last questions show that the recommender system is making adequate recommendations to the users. The results of these questions can be seen in the boxplots on figure 7.3.

7.2.3 Visualization-specific Resque questions

Next I interpret the results for each visualization-specific Resque question. For each of these questions a higher score is better.

Participant	Gender	Age	Books /year	Book recom -mender systems	Recommender systems	Frequent watching/buying
1	Male	21	0-1	No	YouTube, Netflix	Always, Often
2	Male	21	2-5	No	YouTube, Netflix, bol.com	Often, Often, Never
3	Male	21	5-10	No	Amazon, YouTube, YouTube Music	Often, Never, Sometimes
4	Male	24	0-1	No	Steam, YouTube, Spotify	Sometimes, Always, Always
5	Male	22	0-1	No	YouTube, Netflix, Amazon, Blinkist	Often, Often, Sometimes, Sometimes
6	Female	21	5-10	No	Wattpad, YouTube, Spotify	Often, Often, Often
7	Male	23	1-2	No	Instagram, LinkedIn, YouTube	Sometimes, Sometimes, Often
8	Male	22	5-10	goodreads	Netflix, Steam, Letterboxd, YouTube	Often, Sometimes, Often, Often
9	Male	20	5-10	goodreads	YouTube, Reddit, Steam, Spotify, Netflix, Disney+	Sometimes, Sometimes, Rarely, Often, Sometimes, Rarely
10	Female	22	5-10	No	YouTube, Netflix, Disney+	Rarely, Rarely, Rarely
11	Male	22	10-20	bol.com, Amazon	YouTube, Steam, Amazon, bol.com	Sometimes, Rarely, Rarely, Rarely
12	Male	19	20-30	No	Netflix, YouTube, bol.com	Sometimes, Rarely, Rarely
13	Female	23	0-1	No	Netflix, coolblue, bol.com	Often, Sometimes, Rarely
14	Male	22	0-1	No	YouTube, Steam, Amazon, Neftlix	Always, Sometimes, Rarely, Always
15	Male	21	2-5	goodreads	YouTube, Amazon, Spotify, goodreads	Always, Rarely, Rarely, Rarely
16	Male	23	0-1	No	YouTube, Amazon, Spotify, bol.com	Often, Never, Often, Never
17	Male	23	2-5	No	YouTube, Amazon, Spotify, bol.com, Netflix	Always, Rarely, Rarely, Rarely, Sometimes
18	Female	23	2-5	No	Spotify, Netflix, YouTube	Often, Often, Never
19	Female	23	2-5	No	YouTube, Spotify	Often, Often
20	Female	22	10-20	bol.com	YouTube, bol.com, Netflix	Often, Rarely, Rarely

Table 7.4: User profiles of the first 20 participants of the final user study.

Participant	Gender	Age	Books /year	Book recommender systems	Recommender systems	Frequent watching/buying
21	Male	18	10-20	No	Netflix, YouTube, Spotify	Sometimes, Rarely, Rarely
22	Male	21	2-5	Book depository	Book depository	Sometimes
23	Male	21	2-5	No	bol.com, YouTube, Spotify, Netflix	Rarely, Sometimes, Sometimes, Rarely
24	Female	22	2-5	No	Netflix, Spotify, YouTube, bol.com, AliExpress, Zalando	Often, Often, Often, Rarely, Rarely, Often
25	Male	22	2-5	No	Spotify, Netflix, YouTube	Never, Rarely, Always
26	Female	22	5-10	No	Instagram	Never
27	Female	18	5-10	No	Spotify, Netflix, YouTube	Always, Never, Often
28	Female	20	5-10	No	TripAdvisor, YouTube	Often, Often
29	Male	20	5-10	No	bol.com, Instagram, Pinterest	Never, Often, Often
30	Female	19	20-30	No	Spotify, YouTube, Wattpad	Always, Sometimes, Rarely
31	Male	21	2-5	goodreads	Spotify, YouTube, Amazon	Often, Often, Rarely
32	Male	24	0-1	No	YouTube, Spotify, Amazon	Often, Rarely, Never
33	Female	23	2-5	No	Netflix, Spotify	Rarely, Often
34	Male	21	10-20	goodreads	goodreads, Google News, YouTube	Rarely, Always, Always
35	Female	22	2-5	No	YouTube, Spotify, Zalando	Often, Often, Rarely
36	Male	22	20-30	No	YouTube, Spotify	Often, Often
37	Male	22	0-1	No	YouTube, Amazon, bol.com	Often, Never, Never
38	Female	47	50-70	Amazon	Amazon	Rarely
39	Male	46	2-5	No	YouTube, bol.com	Sometimes, Never
40	Male	23	5-10	goodreads	YouTube, Netflix, bol.com	Often, Sometimes, Never

Table 7.5: User profiles of participants 20 through 40 of the final user study.

Participant	Gender	Age	Books /year	Book recommender systems	Recommender systems	Frequent watching/buying
41	Male	23	0-1	No	coolblue, bol.com, YouTube, Netflix	Never, Never, Always, Sometimes
42	Male	22	0-1	No	YouTube, Spotify	Always, Always
43	Male	23	0-4	goodreads	Netflix, YouTube, Spotify	Sometimes, Always, Sometimes
44	Female	21	10-20	No	Netflix, YouTube, Deezer	Rarely, Rarely, Sometimes
45	Male	20	10-20	No	Netflix, IMDb, Amazon	Always, Often, Sometimes
46	Female	22	2-5	No	Reddit	Sometimes
47	Male	22	0-1	No	Netflix, YouTube, Spotify, bol.com, Amazon	Rarely, Sometimes, Rarely, Never, Never
48	Male	22	0-1	No	Netflix, YouTube, Spotify	Often, Sometimes, Never
49	Male	21	0-1	No	Spotify, Neftlix, HBO max, YouTube	Often, Sometimes, Rarely, Always
50	Male	22	2-5	No	Spotify, Neftlix, YouTube	Sometimes, Never, Always
51	Male	21	5-10	No	YouTube, Reddit, Netflix	Often, Sometimes, Never

Table 7.6: User profiles of the last 11 participants of the final user study.

Question 1: interaction adequacy

The first visualization-specific question asks whether the visualization explains the recommendation to the user. The Venn diagram visualization scores the best here with a median score of 5.0, while all other visualizations have a score of 4.0. The baseline visualization is performing the worst as a score of 3.0 is still within the lower quartile and a score of 1.0 is still within the lower whisker. This shows that the user agrees that the visualizations explain why the book has been recommended to him/her. All visualizations also do a better job at this than the baseline visualization. The boxplots for the responses to this question can be seen in figure 7.4.

Question 2: transparency

The second visualization-specific question ascertains whether the user understand why the book has been recommended to him/her. Here all visualizations have a median score of 4.0, meaning that most participants agree with the statement. Here

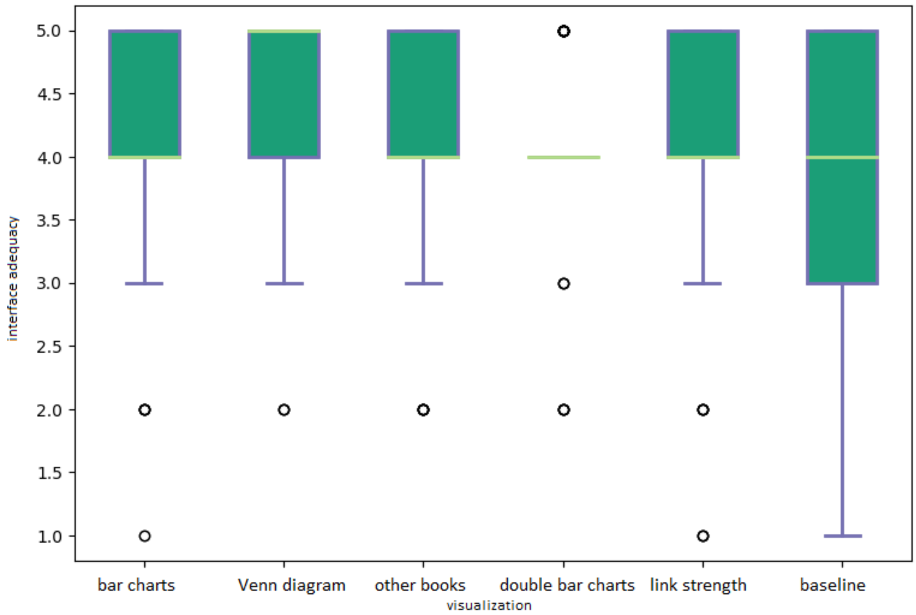


Figure 7.4: Responses to the first visualization-specific resque question for the six visualizations. A score of 5.0 indicating fully agreeing while 1.0 indicating fully disagreeing.

again the baseline visualization is performing the worst as a score of 3.0 is still in the lower quartile and a score of 2.0 is still within the lower whisker. This shows that each visualization helps the user better understand why a book has been recommended to him/her and that they do this better than the baseline visualisation. The boxplots for the responses to this question can be seen in figure 7.5.

Question 3: persuasiveness

The third visualization-specific question ascertains whether the user has been provided enough information to make a download decision. Here the bar charts and double bar charts visualizations have the lowest median score of 3.0 and all other visualizations have a median score of 4.0. The visualization scoring the best is the Venn diagram visualization as the upper quartile reaches a score of 5.0. The worst performing visualization for this question is the double bar charts visualization with a lower quartile reaching a score of 2.0. This shows that the Venn diagram and link strength visualizations are more or equally convincing than the baseline visualization and that the bar charts, other books and double bar charts visualizations are less convincing than the baseline. The boxplots for the responses to this question can be seen in figure 7.6.

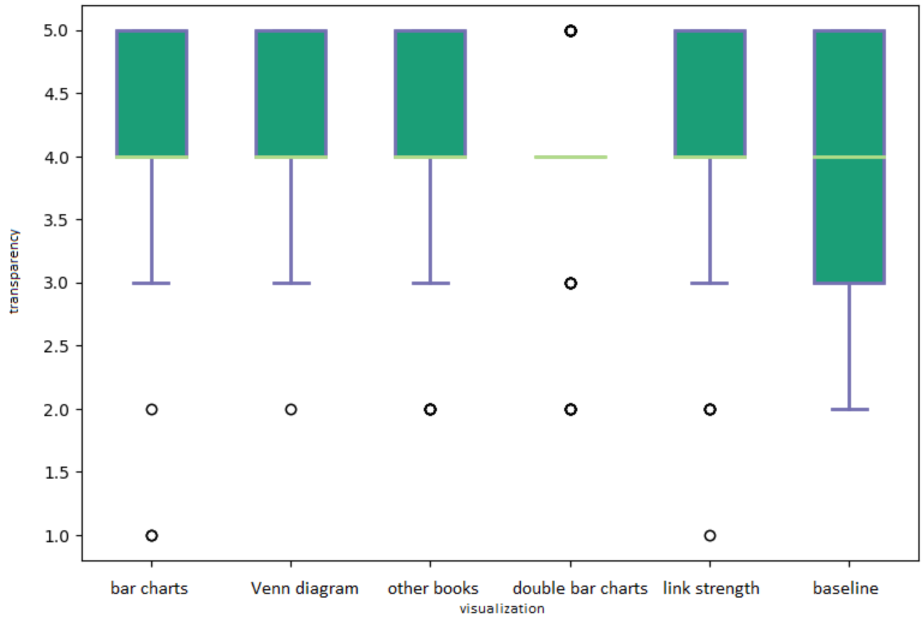


Figure 7.5: Responses to the second visualization-specific resque question for the six visualizations. A score of 5.0 indicating fully agreeing while 1.0 indicating fully disagreeing.

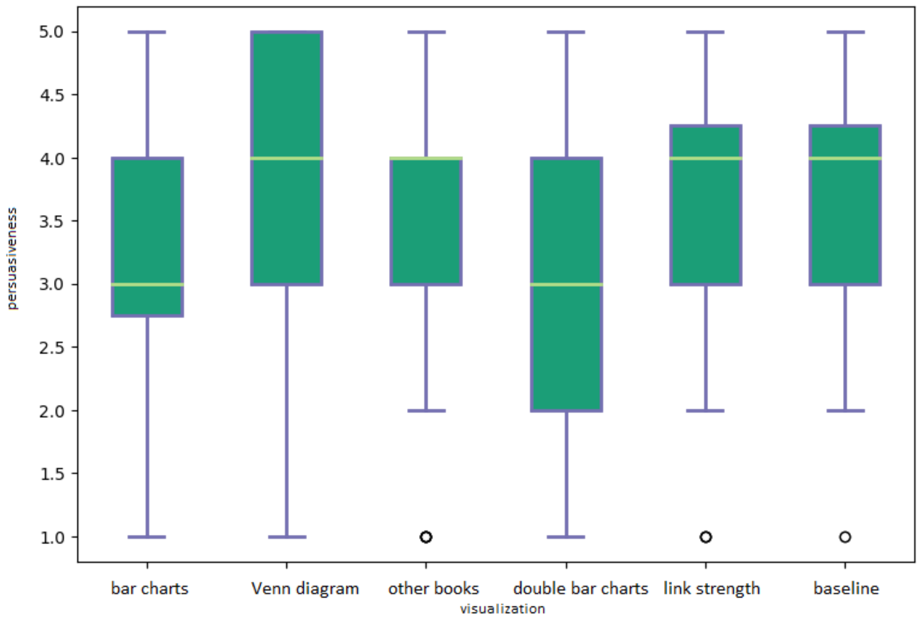


Figure 7.6: Responses to the third visualization-specific resque question for the six visualizations. A score of 5.0 indicating fully agreeing while 1.0 indicating fully disagreeing.

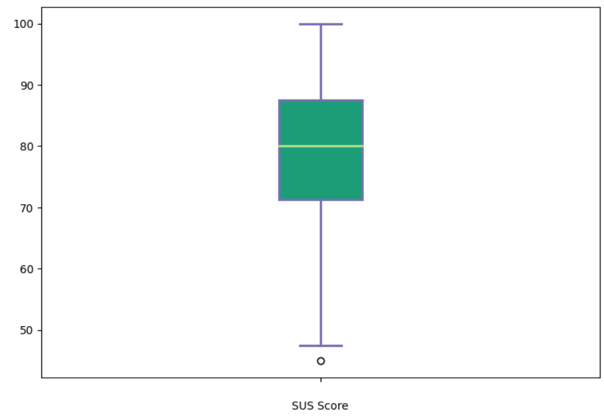


Figure 7.7: System usability scale score boxplot for the final user study.

SUS score

The SUS score of the updated application was also calculated. The median score here is 80.0. This score of 80.0 is just below the 80.8 threshold to be considered excellent as depicted in figure 6.3. It is however still an A grade, which is also considered acceptable.

A comparison between the two box plots is provided in figure 7.8. The mean score is clearly higher for the revised application. However both the first quartile and first whisker are lower for the revised application. This is probably mainly do to the higher number of participants.

To more formally verify whether the median has improved I use the Welch's t test. The standard t test is used to understand whether the median of two samples is significantly different. However the t test has a high chance of suffering of an unequal sample size and hence implying an influence on a type 1 error. The p-value for the Welch's t test is 0.66, which is a lot greater than the 0.05 threshold under which the null hypothesis is rejected. When using a t test the null hypothesis is: "the samples have a different mean". The fact that we do not reject the null hypothesis means that the data we have does not show a statistical significance to reject the null hypothesis.

Nonetheless that does not imply that no improvement has been made with the usability of the application. This merely means that it is not a significant improvement, which entails the first version of the application was already very decent.

Ranking

The ranking that the participants assigned to each of the different visualizations also shows some interesting insights. The rankings are summarized in the boxplots in figure 7.9.

Firstly looking at the median rankings, we see that the double bar charts visualization scored the worst with a median ranking of 5.0. Next both the bar charts

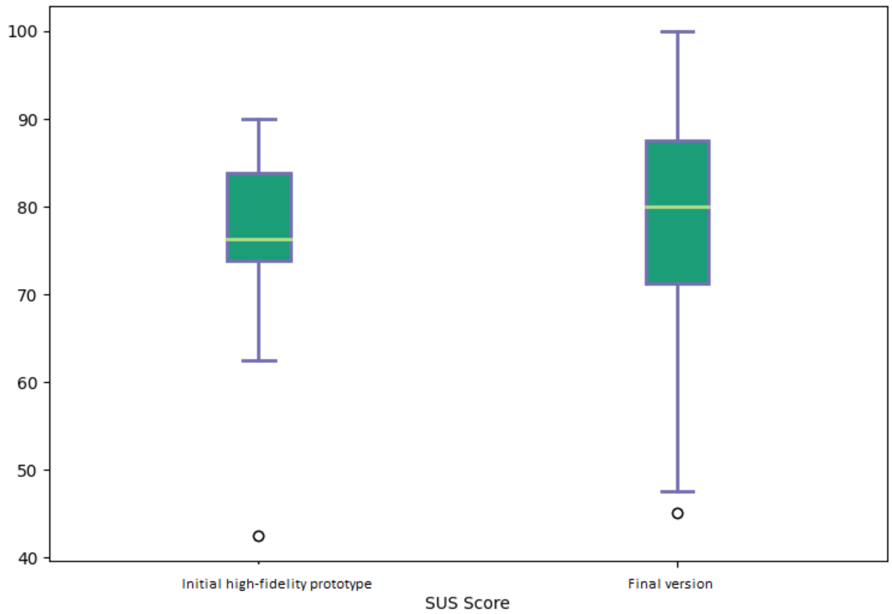


Figure 7.8: System usability scale score boxplot for the first user study (on the left) compared to the final user study (on the right) based on the SUS questionnaires answered by the participants.

and link strength visualizations have a median ranking of 4.0. Both the other books and baseline visualizations have a median ranking of 3.0. Lastly the Venn diagram visualization has the highest median ranking of 2.0.

This yields the following results for the ranking of the visualizations as can be seen in table 7.7. To be able to break the ties between the other books & baseline visualizations and the bar chart & link strength visualizations, it is possible to look at the quartiles.

Namely the lower quartile of other books visualizations is higher than the lower quartile of baseline visualization, while the upper quartile and both whiskers are the same for both. This means the other books visualization is scoring better than the baseline visualization.

Lastly to break the tie between bar charts & link strength visualizations, it is clear that the bar charts visualization has a higher lower quartile, while the upper quartile and the whiskers are the same for both visualizations. This means that the bar charts visualization scores higher than link strength visualization.

This allows to rank the visualizations relative to each other as done in table 7.8.

7.2.4 Thematic analysis

Finally I conduct a thematic analysis on the interviews. A thematic interview helps to identify patterns of themes in the interview data. [28] In a first step I created

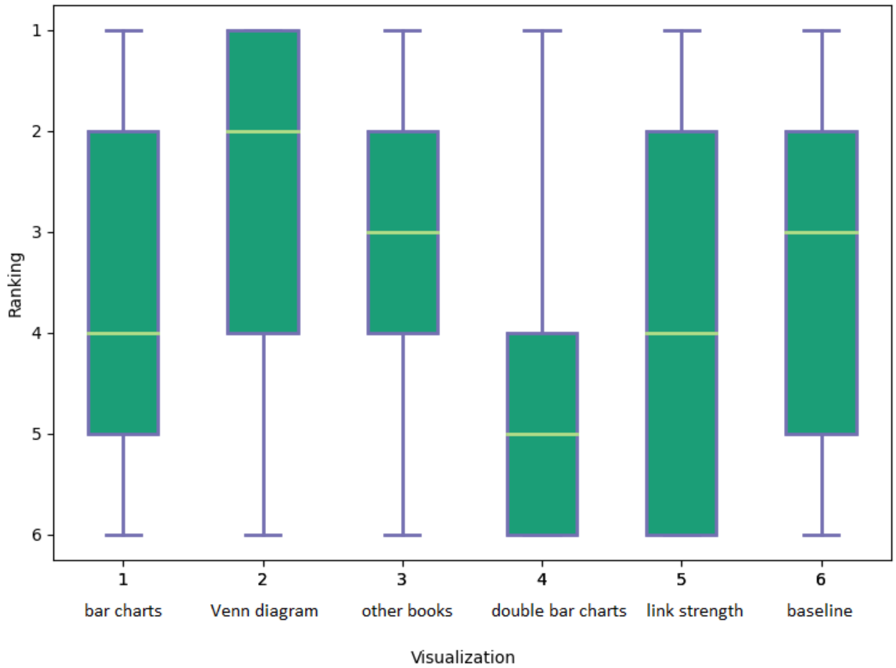


Figure 7.9: Rankings for the visualizations as provided by the participants for the final user study.

Visualizations	Average ranking
Venn diagram	2
other books & baseline	3
bar charts & link strength	4
double bar charts	5

Table 7.7: Rankings for the visualizations as provided by the participants for the final user study.

Relative ranking	Visualization
1	Venn diagram
2	other books
3	baseline
4	bar charts
5	link strength
6	double bar charts

Table 7.8: Relative rankings for the visualizations as provided by the participants for the final user study.

Number	Theme
1	Redundant information
2	Wording of explanations
3	Features of the recommender system
4	Explorative data
5	Confirmation of known preferences
6	Familiarity w.r.t. other recommended systems
7	Association/familiarity w.r.t. visualizations
8	Limitations due to phone screen
9	Repetition
10	Numerical vs categorical data
11	End user understanding w.r.t. a.o. Venn diagrams
12	Possible improvements to visualizations
13	Scalability of visualizations
14	Compactness of visualizations
15	Amount of active thinking
16	Information overload
17	User effort
18	Genres as sub-genres
19	Convincability of a visualization
20	Usability of a legend in a visualization

Table 7.9: Themes that were identified in the first step of the thematic analysis.

distinctive, non-overlapping themes that cover all the interview data. After this I analyse what has been said by the participants about all of these themes.

Identifying themes

After analysing the notes of the interviews I was able to identify recurring themes. These themes are provided in table 7.9.

Some of these themes contain some overlap and can henceforth be merged into one theme covering both of them. Namely the themes scalability of visualizations, compactness of visualizations, information overload and user effort can be merged into one theme. This theme can be called information overload & compactness of visualizations. The themes active thinking and numerical vs categorical data can also be merged into the theme numerical vs categorical data. Lastly the themes features of the recommender system and genres as sub-genres can be merged into the theme features of the recommender system. The updated, non-overlapping themes, can be found in table 7.10.

In the remainder of this section I analyze each of the themes and what the participants said about those themes.

Number	Theme
1	Redundant information
2	Wording of explanations
3	Features of the recommender system
4	Explorative data
5	Confirmation of known preferences
6	Familiarity w.r.t. other recommended systems
7	Association/familiarity w.r.t. visualizations
8	Limitations due to phone screen
9	Repetition
10	Numerical vs categorical data
11	End user understanding w.r.t. a.o. Venn diagrams
12	Possible improvements to visualizations
13	Information overload & compactness of visualizations
14	Convincability of a visualization
15	Usability of a legend in a visualization

Table 7.10: Themes for the thematic analysis after reducing the overlapping themes.

Redundant information Following quote comes from participant 1:

> "Visualization 4 (double bar charts) contains redundant information as each % match for the recommended book is always a 100%."

This quote shows that users do not want to see redundant information in the visualizations. This means that if information can be explicitly obtained from the other components, it should not be displayed explicitly. In this case 17 other participants made a similar statement to participant 1 w.r.t visualization 4 (double bar charts).

However visualization 5 (link strength) contains the same information as visualization 4 (double bar charts), it is just displayed in a different way. This means the same amount of redundancy is present as in visualization 4 (double bar charts).

Following quote comes from participant 15:

> "Visualization 5 (link strength) also contains redundant information, but here it is less prevalent due to the more compact placing of the elements. It does not bother me as much as in visualization 4 (double bar charts)."

This quote proves that the negative effect of redundancy is reduced by a more compact representation of the data. Five other participants stated a similar quote to participant 15 w.r.t. visualization 5 (link strength).

Finally when looking at a quote from participant 5:

> "Visualization 5 (link strength) is pleasing to the eye as it contains symmetry."

It becomes clear that the symmetry also counters the redundancy problem, as participant 5 did not mention anything about redundant information in visualization 5 (link strength). While participant 5 did mention the redundancy in visualization 4 (double bar charts). One other participant stated a similar quote to participant 5 w.r.t. visualization 5 (link strength).

The link strength visualization is thus best used when the end-users have prefer a compact visualization. The double bar charts are a better option when the end-users are not familiar with visualizations and would rather be confused with the link strength visualization.

Wording of explanations Participant 2 stated the following:

> “It is not clear to me what % match means in visualization 1 (bar charts).”

This shows that the usage of the words ”% match” is confusing. 12 other participants had a similar quote about visualization 1 (bar charts). After having explained to the participant the meaning of the ”% match”, the participant said:

> “In that case I would replace ”% match” with ”% of books you have liked previously contained this genre”.”

The participants also stated the same for visualization 4 (double bar charts). This quote displays the importance of using the correct words to explain a percentage in a visualization.

Features of the recommender system Participant 8 stated the following:

> “The genres displayed here are not accurate enough for me, they could be more detailed.”

This quote from participant 8 indicates that the correct selection of features is also important. However this research opted to use genres as features. Because this is not the main focus of the research, but rather a design decision, it is not further discussed.

Four other participants stated a similar quote to participant 8 w.r.t. the features of the recommender system.

Also one participant mentioned the following:

> “Science fiction is for me a sub-genre of science-fiction-fantasy and I would not expect to see both of them.”

This illustrates again that the features displayed in the visualization need to be chosen carefully.

Explorative data Participant 16 stated the following:

> "The right side of the Venn diagram shows me genres that I might not yet be interested in but that the recommended book contains."

This illustrates the explorative nature of visualization 2 (Venn diagram). The other visualizations namely only focus on the genres that are in the overlap, while visualization 2 (Venn diagram) allows the user to also see genres that are not yet in his interests.

Seven other participants made a similar quote w.r.t. visualization 2 (Venn diagram).

Henceforth the Venn diagram visualization should be used when explorative data is desired in a visualization.

Confirmation of known preferences Again in visualization 2 (Venn diagram), participant 18 mentioned:

> "The left side of the Venn diagram shows me genres that are calculated from my known preferences. This is a nice confirmation to have."

Participant 18 henceforth also highlights the importance of the left side of the Venn diagram as a means to communicate the known preferences of the user. This creates a sense of trust and familiarity.

Four other participants made a similar quote w.r.t. visualization 2 (Venn diagram).

However there is also one participant (participant 5) who made the opposite claim with his/her quote:

> "The left side of the Venn diagram feels useless to me. I would expect a bigger overlap. I would expect all genres to be at least in the overlap, if this book is recommended to me."

Participant 5 is probably only convinced about a recommendation, if the entire left side of the Venn diagram would be empty. However this is not a feasibly way of doing so, as it would take away all familiarity that users have w.r.t. Venn diagrams.

Venn diagrams henceforth can increase the trust of the end-users by the confirmation of the known preferences that it displays.

Familiarity w.r.t. other recommended systems Participant 7 noticed the following:

> "Visualization 6 (baseline) reminds me of how Amazon shows me which related articles are also available."

This raises a rather concerning issue. Namely users are already familiar with certain visual displays used in previous recommender systems. Even though for instance Amazon is actually displaying something else than this visualization. This visualization shows which books the user has rated in the past that are related to the recommended book. While Amazon shows articles that are recommended based on the interest in the current recommended article.

This implies that confusion is created due to the fact that users are already habituated to this visualization via different recommended systems, even though they might have different meanings.

One other participant made a similar remark w.r.t. visualization 6 (baseline).

Association/familiarity w.r.t. visualizations Participant 18 mentioned:

> "The image of a book in visualization 6 (baseline) gives me a feeling of familiarity. This is a book I have rated in the past. It is nice to see it show up here."

This quote illustrates that familiarity can create a sense of trust within the system. This is created by the fact that the user is shown which specific books he/she has liked in the past and that the recommended book is effectively related to those books.

Four other participants made a similar quote w.r.t. visualization 6 (baseline).

Participant 19 mentioned this, among two others, about visualization 3 (other books) as well.

Limitations due to phone screen Participant 22 mentioned:

> "The Venn diagram in visualization 2 (Venn diagram) seems rather cluttered. I imagine this is partly due to the horizontal nature of a Venn diagram, while a phone screen is usually vertical."

This highlights that a Venn diagram might not be an appropriate visualization for a phone due to the limited space and the vertical position of the phone screen.

Participant 22 was the only participant to mention this.

Participant 47 said:

> "I find it very unpractical that I have to scroll in visualization 4 (double bar charts) in order to see the entire visualization."

Here again the limitations of a phone screen are mentioned. Visualization 4 (double bar charts) is a visualization that is not very compact and henceforth requires some scrolling in most cases.

Two other participants mentioned the same issue as participant 47.

Repetition Participant 9 stated:

> "In visualization 3 (other books) I unfortunately see a lot of repetition. Books that are in multiple similar genres to the recommended book are displayed multiple times"

13 other participants stated something similar to participant 9 w.r.t. the repetition of visualization 3 (other books). On the contrary participant 17 mentioned:

> "The repetition in visualization 3 (other books) helps me see which books were more important to the decision than others."

Participant 37 similarly mentioned:

> "The repeated books in visualization 3 (other books) help me see which books contain more genres."

17 other participants stated something similar to participant 17 or 37 w.r.t. the repetition of visualization 3 (other books).

This shows that for a large part of the users too much repetition becomes annoying.

Fortunately as stated by five participants, when the system has been used for a more prolonged amount of time, this repetition is due to fall off. This is because the main reason for the repetition is that when the participants are using the system, only a limited amount of books have been rated by them. Meaning there is a high chance of one of those limited books showing up in two different genres. Because no books with a higher importance in that genre have been rated yet at that point.

Numerical vs categorical data Visualizations 2 (Venn diagram), 3 (other books) & 6 (baseline) display categorical data whereas visualizations 1 (bar charts), 4 (double bar charts) & 5 (link strength) display numerical data.

Participant 34 mentioned:

> "I rather do not see any numbers, it requires me to think too much about the meaning of those numbers."

14 other participants made a similar quote.

On the other hand participant 9 mentioned

> "The numbers help me understand the recommendations better. It provides a more detailed explanation, unlike the Venn diagram."

12 other participants made a similar quote.

This proves that no clear preference between numerical or categorical data can be found within the participants. Some users prefer seeing the numerical values, while others prefer seeing the categorical values.

One interesting remark made by participant 51:

> "In visualization 5 (link strength) I can choose whether I just look at the genres and feel satisfied with that information. If not, I can also choose to look at the legend to also see the percentages for each genre."

This is mainly possible due to the compactness of visualization 5 (link strength). Namely all genres are displayed very close to each. Also the fact that the lines are just indicated in different colors and not in different shapes or sizes, helps the user ignore those if desired.

The amount of active thinking required from the end user is also related to this theme. As already mentioned earlier with the numerical vs categorical data certain participants want to avoid the amount of active thinking. This is for most participants achieved by having a visualization which utilizes categorical instead of numerical data.

The information overload is also related to the compactness of a visualization. Participant 42 mentioned about visualization 4 (double bar charts):

> "This is too much information for me to see at one glance. The bars are too long and there are too many of them. I also have to scroll to see them all."

This again shows the importance of compactness of a visualization.

End user understanding w.r.t. a.o. Venn diagrams Some participants also raised concerns w.r.t. the end user understanding of visualizations.

For instance participant 44 said the following about visualization 2 (Venn diagram):

> "I understand this visualization, but I could imagine that not everyone is as familiar with Venn diagrams as I am."

This concern was also mentioned by one other participant.

Also for the fifth visualization (link strength) a similar concern was raised. Namely participant 46 mentioned:

> "I find this a difficult visualization to understand at first glance, because I have never seen it before. After having looked at it for a while, I do understand it."

This shows the importance of familiarity w.r.t. visualization techniques. Venn diagrams are a known visualization, however this does not mean that every end user is acquainted with them. The link strength visualization uses a rather unknown visualization technique (based on Tsai and Brusilovsky's similar keywords interface [43]). Meaning that practically all end users will not have any experience with this visualization and will not be able to interpret it at first glance.

Possible improvements to visualizations The participants mentioned a lot of possible improvements for the different visualizations. These suggestions are summarized in table 7.11. These suggestions can be implemented to improve the visualizations.

Visualization	Suggestion	Amount of participants
1	Replace ”% match” with ”% of previously liked books contain the same genre”	5
1	Place a question mark next to the ”% match” for additional information about the calculation of the percentage	2
1	When hovering over a bar chart show which previously liked books had the corresponding genre.	1
2	Use colors to indicate the difference in importance of genres	2
2	When hovering over the different genres show which previously liked books had that genre.	1
3	Invert the rows and columns, i.e. for each related book list the similar genres instead of listing the books for each similar genre.	3
3	When the user clicks a cover, redirect him/her to that book detail page	2
4	Replace ”% match” for you taste with ”% of previously liked books contain the same genre”	5
4	Replace ”% match” for this book with ”this book contains”	3
4	When hovering over a bar chart show which previously liked books had that corresponding genre.	1
5	Only keep the ellipses of the genres and indicate on the ellipses the importance of each genre. This can be done with either line width or colors.	1
5	Also show genres that are only in ”your taste” and not linking them to the recommended book and vice versa.	1
5	When hovering over a genre show which previously liked books had that genre.	1
5	Use black lines instead of colored lines for ”this book” side as it is always 100%.	2
6	When the user clicks a cover, redirect him/her to that book detail page	2

Table 7.11: Possible improvements for the different visualizations as suggested by the participants of the final user study.

Information overload & compactness of visualizations The scalability of visualizations has also been mentioned by several participants.

Participant 6 said:

> "I believe that when I start using this system more certain visualizations will get more cluttered and less readable. Probably visualization 5 (link strength) will suffer the least from this."

Two other participants raised similar concerns about the scalability of the visualizations.

Participant 6 is pointing out that not all visualizations seem very scalable. This is part of the reason why for instance for the Venn diagram the amount of genres has been limited to six. The concise format of the link strength visualization does indeed allow for a decently scalable visualization. However on the flip side, the double bar chart visualization is not as scalable. This is due to the large amount of space the double bar chart is taking.

Closely related to the scalibility of visualizations, is the compactness of visualizations. As already mentioned earlier participant 47 said:

> "I find it very unpractical that I have to scroll in visualization 4 (double bar charts) in order to see the entire visualization."

This is due to the fact that the double bar chart visualization takes a lot of space on the screen with the double bar charts it utilizes. This issue is least prevalent in the link strength visualization which is the most compact visualization.

Lastly the user effort is related to the information overload of a visualization. Participant 42 mentioned about the double bar charts visualization:

> "This is too much information for me to see at one glance. The bars are too long and there are too many of them. I also have to scroll to see them all."

This again shows that a user does not want to spend too much effort having to look at a visualization, it is in fact trying to help the user understand the recommendation quickly.

Convincability of a visualization Participant 23 mentioned:

> "Visualization 3 (other books) convinced me the most due to its familiarity in the book covers I recognized."

This emphasizes that a visualization has a higher degree of convincability when familiarity is being used.

Three other participants had a similar quote about this.

Usability of a legend in a visualization Participant 49 said:

> "In visualization 5 (link strength) I found it cumbersome to have to look for the legend to find the percentages. I would rather have them next to the lines."

This takes into question the usage of a legend. A legend gives a quick way to the user to find the corresponding value, however it is of course not directly on the visualization itself. Nonetheless when a same value is found multiple times in a visualization a legend is handy to avoid the repetition of always placing the same value in the visualization.

7.2.5 Interpretation

From all the above results the following guidelines can be made for designing visualization for mobile phone systems:

- *Avoid redundant information*
- *Choose the wording of explanations carefully*
- *Ideally use familiarity with for instance a book cover*
- *Watch out for other familiarities your users may have*
- *There is no clear preference between numerical or categorical data*
- *Ideally use visualizations that display both numerical or categorical data in a way that the categorical data can be viewed while ignoring numerical data*
- Take into account the space limitations of a phone screen when developing new visualizations
- Take into account the vertical nature of a phone screen when developing new visualizations.
- Avoid repetition, there is only so much space on a phone screen
- Using different font sizes for texts is not feasible on a phone screen

Some of these guidelines also apply to non-mobile phone systems. These are written in *italic*. The other guidelines are specific to the mobile phone environment.

7.2.6 Discussion

For this final user study I mainly evaluated the visualizations itself. I did this by interviewing 51 participants. They were allowed to use the application and freely give their opinion about the visualizations. After using the application they were asked to fill in a questionnaire consisting of SUS and Resque questions.

The SUS showed that the application was indeed slightly improved from the first user study. The general Resque questions show that the recommender system is making adequate recommendations.

The visualization specific resque questions show that all visualizations score better than the baseline visualization when looking at the intercation adequacy and transparency. However only the Venn diagram and link strength visualizations are more or equally convincing than the baseline visualization when looking at the persuasiveness.

On the information gathered during the interviews themselves I did a thematic analysis. Out of this possible analysis possible improvements to the visualizations have been found such as providing the user with a more adequate textual explanation. These improvements are summarized in 7.11.

From this feedback several guidelines for the development of new visualizations for mobile phone applications were suggested. This serves as a contribution to possible further research when developing new visualizations for mobile phone applications.

Another contribution is the evaluation of the five different visualizations. It also became apparent when which visualization is ideally used and what possible pitfalls are.

The bar chart visualization shows percentages, thus showing numerical instead of categorical data. No clear preference between numerical or categorical data was found.

The Venn diagram visualization is best used when the users are familiar with Venn diagrams. The Venn diagram visualization can increase the trust in the system as it shows what the system has already learned about the preferences of the end user. The Venn diagram visualization also show explorative data which allows the user to understand which genres he might also find interesting.

The other books visualization gives the user a sense of familiarity as it shows covers that the user has liked in the past.

The double bar chart visualization contains redundant information, this is usually considered bothersome to the end users. The link strength visualization contains the same redundant information but is more compact. It is therefor considered less bothersome. The link strength visualization however is a visualization that most end users are not acquainted to. Thus the double bar chart visualization should be used when the end users are not familiar with the link strength visualization or when they do not have a high cognition towards understanding new visualizations.

The Venn diagram and the other books visualization show categorical data, while the bar charts, double bar charts and link strength visualization show numerical data. For the end user no clear preference between these two types of data was found. However in certain situation numerical data could be preferred for a more detailed explanation for the recommendation compared to categorical data which allows for faster interpretation.

Chapter 8

Conclusion

Visualizations explaining the decision making behind a recommendation are one of the possible solutions to solving the decreased user satisfaction felt from recommender systems. This research evaluated five proposed visualizations in the context of a mobile phone book recommender system app. These visualizations are based on the literature. The main difference between the literature and this research is that the visualizations are being used on a mobile phone screen.

The five visualizations make use of the features of the recommender system to explain the recommendations. These five visualizations are compared to a baseline to analyze the impact of using the features of the recommender system. A custom-made recommender system based on content-based filtering was also made for this application.

A user-centered design approach was taken for this research. This means the feedback of the end user is taken into account at every stage of the development. This was mainly ensured by a pilot study and a first user study. To evaluate the visualizations itself a final user study was conducted.

A pilot study on the initial design of the visualization was conducted to get some feedback from the end user. Here a few selected questions from the SUS questionnaire were asked for each different visualization.

A first user study was conducted when the initial version of the application was developed. In this version the feedback from the pilot study for the visualizations was already taken into account. This first user study allowed to get some feedback about the application and get some initial feedback about the actual implementation of the visualizations and the algorithms used for them.

A questionnaire consisting of Resque, SUS and NASA TLX questions was used for this. A median SUS score of 76.0 was given to the application. The Resque results also showed that the underlying recommender system was working accordingly. Finally the NASA TLX questions showed that the user found the application relatively easy to use.

All the feedback from the pilot study and the first user study were implemented in the final version of the application.

A final user study was conducted to evaluate the visualizations itself and get feedback about them in an actual recommender system. Here again SUS questions and Resque questions were used. The SUS questions were asked to understand whether an improvement of the application was actually obtained with the feedback from the first user study. The Resque questions were split into a few general Resque questions for the recommender system itself. Also for each visualization a couple of Resque questions were asked. Namely to ascertain the interaction adequacy, transparency and persuasiveness of each visualization.

A SUS score of 80.0 was obtained with the updated version of the application, clearly showing that the usability of the application was slightly increased with the feedback from the previous user study.

All visualizations perform better than the baseline visualization when looking at the interaction adequacy and transparency of visualizations. However when looking at the persuasiveness only the Venn diagram and the link strength visualization perform better than the baseline.

A thematic analysis was also conducted on the user interviews of the final user study. From this certain guidelines could be concluded when designing visualizations for mobile phone systems. Namely:

- *Avoid redundant information*
- *Choose the wording of explanations carefully*
- *Ideally use familiarity with for instance a book cover*
- *Watch out for other familiarities your users may have*
- *There is no clear preference between numerical or categorical data*
- *Ideally use visualizations that display both numerical or categorical data in a way that the categorical data can be viewed while ignoring numerical data*
- Take into account the space limitations of a phone screen when developing new visualizations
- Take into account the vertical nature of a phone screen when developing new visualizations.
- Avoid repetition, there is only so much space on a phone screen
- Using different font sizes for texts is not feasible on a phone screen

Some of these guidelines also apply to non-mobile phone systems. These are written in *italic*. The other guidelines are specific to the mobile phone environment.

For the visualizations itself following conclusions can be drawn:

- The bar chart visualization shows numerical data and should this be used when a more detailed explanation than a categorical visualization is necessary
- The Venn diagram allows for explorative data, this can help users find genres that they might also find interesting.

- The Venn diagram can increase the trust in a system as it shows the known preferences to the end user and can henceforth be considered as a confirmation by the end user
- The Venn diagram requires the end user to be familiar with Venn diagrams and is henceforth not suitable for every target audience.
- The other books visualization gives the user a sense of familiarity as it contains book covers of books he previously liked in the system
- The other books visualization can be understood by all target audiences.
- The double bar chart visualization contains redundant information and should henceforth only be used when the users have a low level of cognition towards visualizations.
- The link strength visualization contains the same redundant information as the double bar chart visualization. However due to its more compact nature, it is less bothersome. It does require a higher need for cognition to understand new visualizations from the end users as this visualization is unfamiliar to most end users. The link strength visualization should thus be used when the users have a high amount of cognition towards new visualizations.
- The baseline also contains familiarity towards previous covers as was the case for the other books visualization.

Finally to answer the research questions:

RQ1. Does providing a visualization explaining a recommendation to the end user, increase the user satisfaction of the application? Providing the user with a visual explanation as to why a recommendation has been made increases the user satisfaction.

RQ2. Do the five proposed visualizations perform better than the baseline when considering the interaction adequacy, transparency and persuasiveness of recommender systems? The five proposed visualizations perform better than the baseline when looking at interaction adequacy and transparency. However only the Venn diagram and link strength visualizations perform better than the baseline visualization in terms of persuasiveness.

RQ3. Does the end user prefer a numerical or categorical approach towards visualizations? No clear preference between a numerical or categorical approach to the visualizations has been found.

Chapter 9

Further research

In this chapter possible further research possibilities are discussed.

9.1 Recommender system

In this research a content-based filtering algorithm was used for the recommender system. This allowed for the easy usage of features in the visualizations. Other recommender systems such as knowledge-based or collaborative filtering can also be used.

Collaborative filtering does however not make direct use of the features of the different books. This means that the visualizations need to be adjusted accordingly. However the idea of a visualization can remain the same. For instance it is still possible to use bar chart to show match percentages. Instead of using the previously liked books to calculate the percentages, the percentages of similar users can be used instead. This means the underlying algorithm is adjusted, but the visualization itself remains the same besides the adjusted percentages.

9.2 Interactive visualizations

It is also possible to make the visualizations interactive. Some of these interaction possibilities were already recommended by the participants of the final user study as shown in table 7.11. These interactions allow the user to get more information about the reasoning of the recommendation on demand. This is also very interesting in the context of mobile phone system, in which the space is limited.

9.3 New/existing visualizations

Another possible further research path could be evaluating new or existing visualizations in a mobile phone screen. With the guidelines concluded from this research it is possible to create new visualizations which take these guidelines into account as much as possible. Existing visualizations can also be easily evaluated by checking whether they follow all the guidelines listed in this research.

Bibliography

[1] 5 ways to interpret a sus score. `https://measuringu.com/interpret-sus-score/`.

[2] M. Z. Al-Taie and S. Kadry. Visualization of explanations in recommender systems. *Journal of Advanced Management Science Vol*, 2(2):140–144, 2014.

[3] A. Bhatti, H. Akram, H. M. Basit, A. U. Khan, S. M. Raza, and M. B. Naqvi. E-commerce trends during covid-19 pandemic. *International Journal of Future Generation Communication and Networking*, 13(2):1449–1452, 2020.

[4] M. Bilgic and R. J. Mooney. Explaining recommendations: Satisfaction vs. promotion. In *Beyond Personalization Workshop, IUI*, volume 5, page 153, 2005.

[5] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. Recommender systems survey. *Knowledge-based systems*, 46:109–132, 2013.

[6] S. Bostandjiev, J. O'Donovan, and T. Höllerer. Tasteweights: a visual interactive hybrid recommender system. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 35–42, 2012.

[7] J. Brooke. Sus: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, 11 1995.

[8] R. Burke. Knowledge-based recommender systems. *Encyclopedia of library and information systems*, 69(Supplement 32):175–186, 2000.

[9] F. Cammaerts, N. Dedoncker, J. Hermans, T. Martens, and T. Martens. Investigating possible solutions for the cold start problem in a food recommender application, 2020. `https://stillpatience.github.io/felix-cammaerts/files/Foodversity.pdf`.

[10] V. Dominguez, P. Messina, I. Donoso-Guzmán, and D. Parra. The effect of explanations and algorithmic accuracy on visual recommender systems of artistic images. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pages 408–416, 2019.

[11] B. Fogg, J. Marshall, T. Kameda, J. Solomon, A. Rangnekar, J. Boyd, and B. Brown. Web credibility research: a method for online experiments and early study results. In *CHI'01 extended abstracts on Human factors in computing systems*, pages 295–296, 2001.

[12] C. A. Gomez-Uribe and N. Hunt. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4):1–19, 2015.

[13] J. Grashuis, T. Skevas, and M. S. Segovia. Grocery shopping preferences during the covid-19 pandemic. *Sustainability*, 12(13):5369, 2020.

[14] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[15] T. Hassan. Trust and trustworthiness in social recommender systems. In *Companion Proceedings of The 2019 World Wide Web Conference*, pages 529–532, 2019.

[16] C. He, D. Parra, and K. Verbert. Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications*, 56:9–27, 2016.

[17] J. L. Herlocker, J. A. Konstan, and J. Riedl. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 241–250, 2000.

[18] R. Hoekstra. The knowledge reengineering bottleneck. *Semantic Web*, 1(1, 2):111–115, 2010.

[19] O. Inel, N. Tintarev, and L. Aroyo. Eliciting user preferences for personalized explanations for video summaries. *arXiv preprint arXiv:2005.00465*, 2020.

[20] V. Kalnikaitė, J. Bird, and Y. Rogers. Decision-making in the aisles: informing, overwhelming or nudging supermarket shoppers? *Personal and Ubiquitous Computing*, 17(6):1247–1259, 2013.

[21] B. P. Knijnenburg, M. C. Willemsen, and A. Kobsa. A pragmatic procedure to support the user-centric evaluation of recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 321–324, 2011.

[22] X. N. Lam, T. Vu, T. D. Le, and A. D. Duong. Addressing cold-start problem in recommendation systems. In *Proceedings of the 2nd international conference on Ubiquitous information management and communication*, pages 208–211, 2008.

[23] B. Lamche, U. Adıgüzel, and W. Wörndl. Interactive explanations in mobile shopping recommender systems. In *Joint Workshop on Interfaces and Human Decision Making in Recommender Systems*, volume 14, 2014.

[24] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou. Recommender systems. *Physics reports*, 519(1):1–49, 2012.

[25] P. Melville and V. Sindhwani. Recommender systems. *Encyclopedia of machine learning*, 1:829–838, 2010.

[26] M. Millecamp, N. N. Htun, C. Conati, and K. Verbert. To explain or not to explain: the effects of personal characteristics when explaining music recommendations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pages 397–407, 2019.

[27] M. Millecamp, S. Naveed, K. Verbert, and J. Ziegler. To explain or not to explain: the effects of personal characteristics when explaining feature-based recommendations in different domains. In *IntRS@ RecSys*, pages 10–18, 2019.

[28] D. H. Mortensen. How to do a thematic analysis of user interviews, Jun 2020.

[29] H. V. Nguyen, H. X. Tran, L. Van Huy, X. N. Nguyen, M. T. Do, and N. Nguyen. Online book shopping in vietnam: The impact of the covid-19 pandemic situation. *Publishing Research Quarterly*, 36:437–445, 2020.

[30] P. Pu and L. Chen. Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems*, 20(6):542–556, 2007.

[31] P. Pu, L. Chen, and R. Hu. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 157–164, 2011.

[32] P. Resnick and H. R. Varian. Recommender systems. *Communications of the ACM*, 40(3):56–58, 1997.

[33] F. Ricci, L. Rokach, and B. Shapira. Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer, 2011.

[34] A. Sela and J. Berger. Decision quicksand: How trivial choices suck us in. *Journal of Consumer Research*, 39(2):360–370, 2012.

[35] R. Sinha and K. Swearingen. The role of transparency in recommender systems. In *CHI'02 extended abstracts on Human factors in computing systems*, pages 830–831, 2002.

[36] X. Su and T. M. Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009, 2009.

[37] J. Tang, S. Wu, J. Sun, and H. Su. Cross-domain collaboration recommendation. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1285–1293, 2012.

[38] N. Tintarev. Explaining recommendations. In *International Conference on User Modeling*, pages 470–474. Springer, 2007.

[39] N. Tintarev and J. Masthoff. A survey of explanations in recommender systems. In *2007 IEEE 23rd international conference on data engineering workshop*, pages 801–810. IEEE, 2007.

[40] N. Tintarev and J. Masthoff. Designing and evaluating explanations for recommender systems. In *Recommender systems handbook*, pages 479–510. Springer, 2011.

[41] N. Tintarev and J. Masthoff. Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5):399–439, 2012.

[42] C.-H. Tsai and P. Brusilovsky. Providing control and transparency in a social recommender system for academic conferences. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, pages 313–317, 2017.

[43] C.-H. Tsai and P. Brusilovsky. Designing explanation interfaces for transparency and beyond. In *IUI Workshops*, 2019.

[44] C. H. Tsai and P. Brusilovsky. User feedback in controllable and explainable social recommender systems: a linguistic analysis. In *CEUR Workshop Proceedings*, volume 2682, pages 1–13. CEUR-WS, 2020.

[45] R. Van Meteren and M. 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.

[46] K. Verbert, D. Parra, P. Brusilovsky, and E. Duval. Visualizing recommendations to support exploration, transparency and controllability. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, pages 351–362, 2013.

[47] W. Wang, G. Zhang, and J. Lu. Hierarchy visualization for group recommender systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(6):1152–1163, 2017.

[48] L. Yao, Q. Z. Sheng, A. H. Ngu, J. Yu, and A. Segev. Unified collaborative and content-based web service recommendation. *IEEE Transactions on Services Computing*, 8(3):453–466, 2014.