Read it or Watch it?

Project Progress Report

Team 121 - Hollywood Critics

Lauren Hirs* Samantha Jaeger* Lauren Liou* Dylan Peters* Claudia Scott* Kelly Tallau*
*All team members are providing equal contribution to this project.

I. Introduction – Motivation

Given the number of movies that are sourced from pre-existing books, it is common for readers/movie viewers to ask the question: "Should I read the book, or should I watch the movie?" Today, researching this question must be done manually, with many separate steps. A prospective reader or movie viewer would visit individual movie and/or book review websites, Amazon reviews, and/or social media posts to get opinions on their path forward.

Additionally, box office revenue, though strong in 2018 (potentially due to the existence of MoviePass), had been on a downturn for three years before 2018 and 2019 has been disappointing thus far [1]. Movie producers often opt to adapt a screenplay from a book for lower time-to-market rates and higher likelihood of success [2]. Although adaption of a popular book can be an easier path to success, it is also easy to disappoint legions of devoted readers if the adaptation is not done well.

a. Problem Definition

The project objective is to aid both potential audiences (from book/movie) and movie producers who are trying to make the most successful book adaptation possible. Given a need for both groups to make better decisions more quickly, we combine necessary data into one location and visualize it to help draw conclusions at a glance.

II. Literature Survey

In investigating this topic, we found some basic analyses have been previously conducted to show relative book and movie scores [3, 4]. Two separate articles were published to help assess whether the book or movie tends to be better, both of which used a scatterplot to show the typical correlation [5, 6]. Their result is outdated, and their efforts did not incorporate text reviews, impact of genre, movie casting or production budget.

Luckily, there are several academic research areas with insights related to this topic. Our research focuses on the following ideas: 1) book-specific analytics, 2) movie-specific analytics, 3) consumer review analytics (as a general topic), and 4) effective visualization.

When researching for book reviews/ratings, we found that informal book reviews are a valuable source of information. Goodreads is less biased than Amazon, as Amazon reviews affect sales. Other research validated our hypothesis that review text and book features could correlate with a book's popularity. We found studies using text reviews to draw conclusions about reviewers themselves, they provided much insight on data preparation and text analysis.[3, 4, 7-9]

Meanwhile, sentiment analysis is a popular tool for movie analytics. Many researchers studied different ways of analyzing review text and reported various methodologies to build sentiment analysis from scratch (such as Naïve Bayes or AdaBoost).[10-12] Others reported on which already-available lexicons work best for movie-related terminology.[13]

Significant work was available on data preprocessing, demonstrating how to remove redundancy and identify most frequent words and sentiments.[14-17]

We also sought out literature on text topic analysis. While we investigated the use of CATHY for topic modeling, we ultimately decided to focus on Latent Dirichlet Allocation (LDA). The approach uses probabilities to discern topics from text. LDA can compress large datasets of text reviews into a much smaller set of dimensions. It is easier for future prediction purposes. LDA is a particularly effective for work with very large and unstructured user-generated content (e.g. hotel reviews).[14, 16]

Finally, we investigated interactive visualizations approaches and graphical designs for charts. Promising approaches included scatterplots with different point colors and sizes, network graphs with different line thicknesses, word clouds to highlight most important terms, animations, mental models, and botanical structures.[18-24]

Detailed literature survey per reference can be found in **Section VI**.

III. Proposed Method

Intuition / Innovations

To our knowledge there is no thorough and centralized hub for the public to compare and visualize book/movie pair statistics, ratings and public reviews, especially in a side-by-side and unbiased manner. Our primary innovations for this project are:

- Applying Latent Dirichlet Allocation (LDA) to analyze the discussion around book-movie topics, split out by positively and negatively rated reviews
  - Visualize and summarize LDA result into frequently discussed topics
  - allows users to quickly discern topics discussed by reviewers with different review polarity
- Synthesizing significant information on book/movie pairs obtained from multiple sites in one interactive visualization tool for ease of use
  - integrate and aggregate information from book/movie sources

- provide an institutive, informative and centralized platform to apprehend book/movie pair comparison and analytics
- The "Read It or Watch It" gauge which allows users to quickly identify suggestions from us based upon the normalized book and movie rating

Project Approach Description

Our proposed method has three phases: 1) data collection and cleaning, 2) text analysis, and 3) visualization.

1) Data collection and cleaning

Figure 1 displays the datasets, origin, size and the tools used to build this project. From these data, ratings were standardized to ensure data sources compatibility. Foreign languages, symbols and any punctuations were removed from reviews. Homonyms in reviews (e.g., movie vs. film) were consolidated manually, words used in book/movie titles themselves were removed, and custom stopword dictionaries were built to further remove "noise" from all reviews.

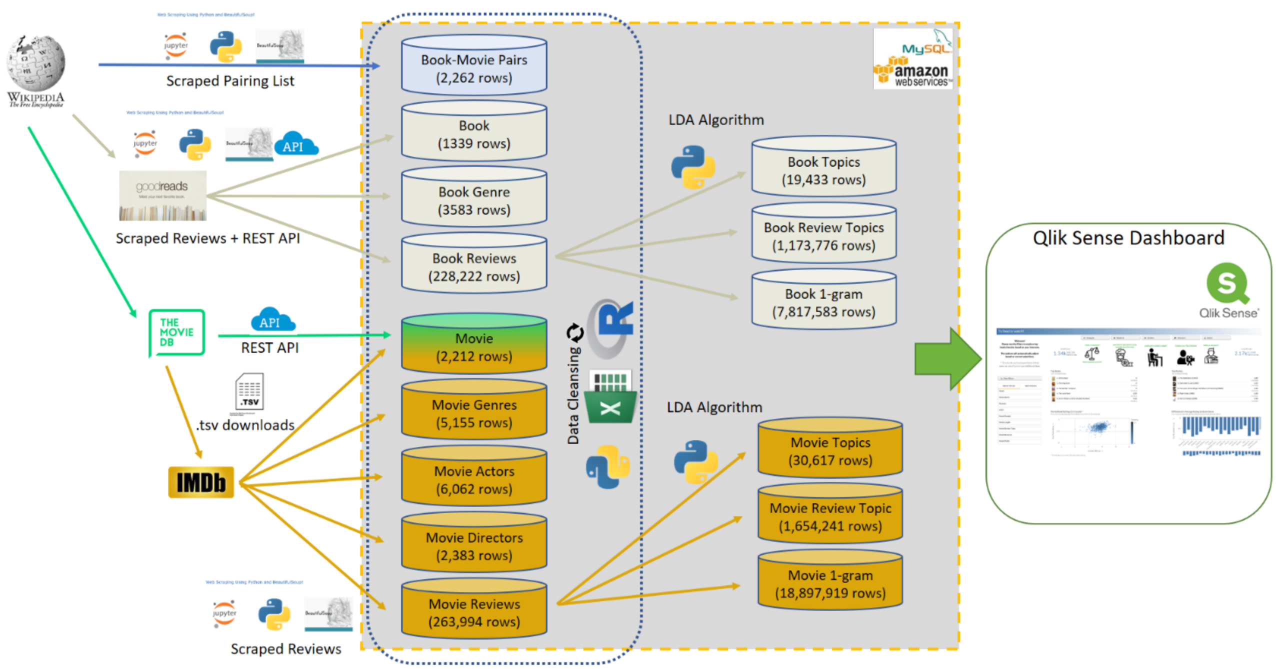

Figure 1 dataset origin, size, purpose and tools used for the Read it or Watch it project.

2) Text Analysis

For the text analysis, we have exploited 2 main paths. Literature survey suggested that there are readily available and useful lexicons to help with sentiment analysis [7, 11, 12, 15, 24, 25]. We applied text analysis packages in R along with predictive models (e.g., XGBoost) to investigate whether certain words used in reviews correlate to a review's polarity toward a book and/or movie.

We also pursued topic analysis by LDA. By consolidating the movie/book review dataset into discrete topics, we aim to 1) identify topics related specifically for comparing books to movies, and 2) summarize review topics and/or features frequently mentioned in reviews for the particular book/movie so that users can efficiently understand many reviews context without having to read them individually.

3) Visualization

For visualization, we use Qlik due to its ease of use, ability to incorporate D3 extensions, and its powerful Associative Engine that allows for speedy and thorough user-directed exploration.

The visualization landing page shows an interactive scatterplot of normalized book and movie review scores with 8 movie related filters. Upon choosing a book/movie pair the users see the text analytics and our “Read it or Watch It” (RI/WI) gauge. This gauge recommends users to 1) read the book or 2) watch the movie or 3) read/watch either. Recommendations were implemented using result from Equation 1. Upon scaling the average book rating ($m_{AR}/2$) to the same range as average movie rating (b_{AR}), the difference between $m_{AR} / 2$ and b_{AR} was calculated is divided by the minimum of the two average ratings. That fraction is transformed into a percentage. Table 1 shows the recommendation possibilities. An example of the RI/WI gauge is presented in Figure 7.

Equation 1:

$$\frac{\left(\frac{m_{AR}}{2} - b_{AR}\right)}{\min\left\{\frac{m_{AR}}{2}\ ,\ b_{AR}\right\}} \times 100\% \ ; \quad m_{AR} \text{ : average movie rating; } b_{AR}\text{: average book rating}$$

Table 1 – RI/WI gauge recommendation range

x% from reported from Equation 1	RI/WI gauge Recommendation
> 10%	Watch the movie
-10% < X < 10%	Read / watch either
> -10%	Read the book

Moreover, the networks between authors, directors, actors and authors are presented in visualization. A user who enjoys the work of an individual may benefit from seeing the related work; a movie producer can see which groupings of professionals are linked to successful adaptations.

IV. Experiments / Evaluation

Questions listed in Table 2 are those that we designed our experiments to answer or evaluate to achieve our goals for this project.

Table 2 – list of questions our project experiments are designed to answer

Questions our experiments designed to answer	Topic	Success metric
Optimal techniques to evaluate and summary the content and/or sentiment from book/movie reviews	Text analysis	N/A
Book-movie topic similarity and differences from reviews	Text analysis	Identifying book review topics about watching the movie vs. movie review topics about reading the book
Book/movie information of users' interest	Visualization	User testing
Present book-movie review comparison which help user decide to read the book or watch a movie	Visualization	User testing
Display authors, actors, directors' network for users to explore	Visualization	User testing
Intuitiveness of our interactive visualization	User Testing	User testing feedback

1) Text Analysis

Text analysis is the majority of our analytical work. Initial efforts were made to model specific uni/bigrams and movie ratings. Ideally, certain words would correlate with review polarity. The most popular uni/bigrams in the movie reviews were filtered and extracted for common words with a positive or negative connotation (e.g. "good", "best movie"). When data is fitted to linear and logistic regression, the resulted model showed poor R^2 (i.e., ~ 0.12). These results suggest that n-gram alone is not sufficient to predict review connotation.

Efforts was made to correlate AFINN (lexicon) scores with individual review rating. Individual review words (upon data cleaning) were map to its associated AFINN sentiment score using R (Figure 2). While the data exhibits obvious trend, with correlation 0.42, the R^2 is only 0.01. Result suggests that rating and average sentiment score is linearly and positively correlated but this model has high variance and cannot predict average sentiment score from the reviews. We also observe that topic words such as 'war', 'shot', or 'murder' skewed the results away from the true review polarity (positive/negative).

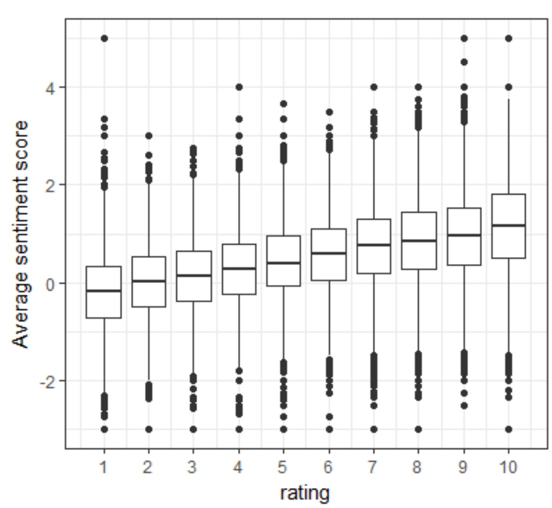

Figure 2 example boxplot correlating average sentiment score for movies vs. review rating

Lastly, we applied LDA to identify topics within our book/movie review dataset. We used coherence to determine the optimal number of topics (30) within the movie reviews (result in Figure 3). In the LDA model for movie and book reviews, we found that one topic that emerged was clearly related to book-movie comparison from reviews (Figure 4). This topic is about books and reading (Figure 4B), so we planned to use it to examine how it is related to movie reviews and ratings.

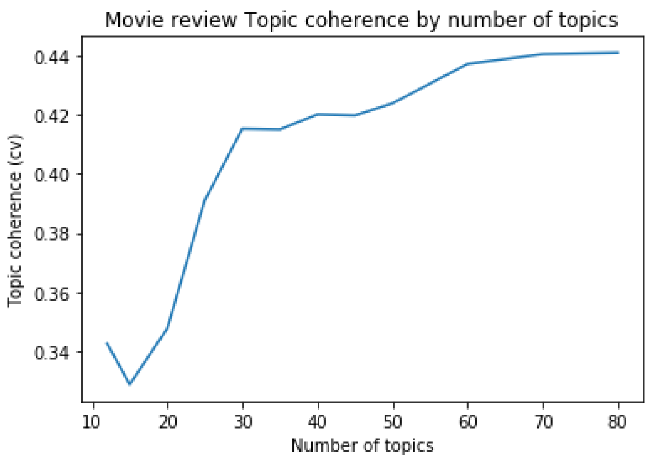

Figure 3 Topic coherence relationship with number of topics for movie reviews identified by LDA

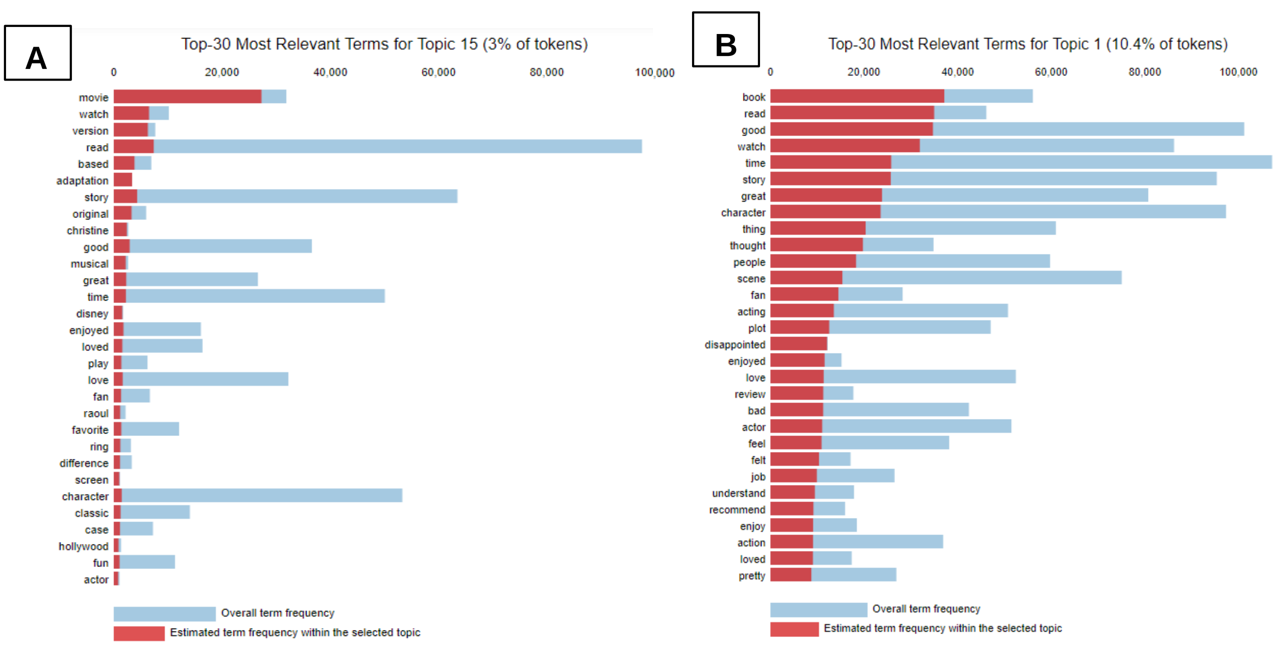

Figure 4 selected LDA result from book reviews with relevant terms (movie, watch) (in Panel A) and from movie reviews with relevant terms (book, read) (Panel B).

With LDA, top 6 topics were kept per review, based on percentage association with the review. The percentage association marks the percentage of the review that relates/discusses a topic, using the analysis of words present in the topics that also appear in the review.

We compared the LDA topics to positive and negative reviews and created a score for the book-movie comparison topic for each book and movie. With this, we discovered how often the positive and negative reviews for each item is explicitly discussed a book/movie comparison and how much of the review text is dedicated to the topic.

We compared all 30 topics vs the ratings, by filtering down to relevant reviews per topic and mapping the average of those percentages per the rating.

Starting with Topic 0, previously discovered as Topic 29 in the graph above, shows movies reviews that mentioned the word "book". There is a complex and non-linear relationship between the topic 0 and assigned average movie ratings (Avg Rating) (Figure 5). As reviewers mentioned the 'book' in a movie review, the initial sentiment of the reviewer decides the type of relationship the topic has with the assigned average movie rating. The relationship between topic percent for topic 0 vs Avg Rating seemed to be split into 3 segments – 1) Avg Rating higher than 8; 2) Avg Rating between 4-7 and 3) Avg Rating lower than 4. Segment 1 and Segment 3 can be considered as a "positive" and "negative" sentiments, "book" topic frequency in movie reviews increase with Avg Rating. Segment 2 displays the opposite behavior. We highlight this topic within the topic funnels in "compare" page (Figure 8).

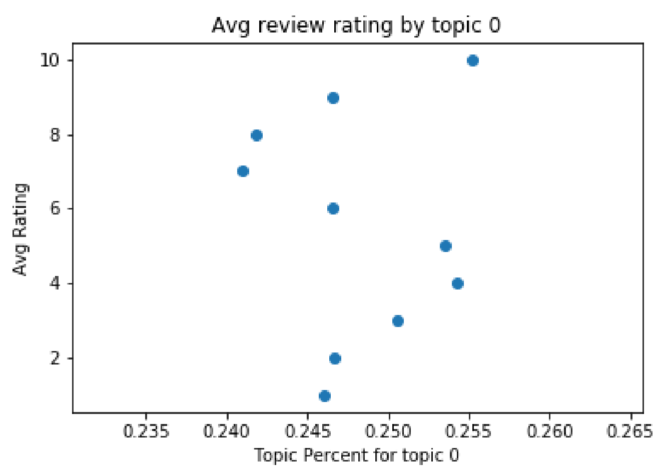

Figure 5 LDA derived Topic 0 from movie reviews that mentioned "book" vs. the reviewers' average rating.

An interesting observation from LDA analysis for Harry Potter/Lord of the Rings fans, one topic contained "harry", "lord", "ring", "potter", "Jackson" and "battle". A positive correlation between the strength of topic 7 and Avg rating is observed (Figure 6). This indicates that reviewers who discuss the book more tend to have a higher opinion of the movie.

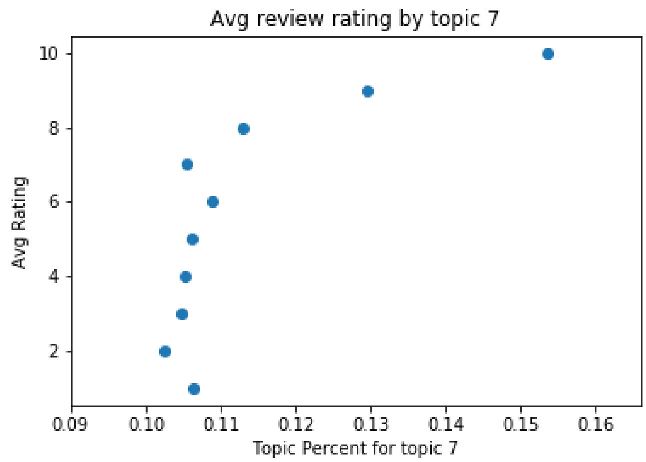

Figure 6 LDA derived result for Topic 7 from movie review. Topic 7 contains words harry", "lord", "ring", "potter", "Jackson" and "battle".

2) Visualization

Figure 7A shows the landing page for this interactive visualization. Large images, analogies and labels are used to give user a clear sense of navigation in the landing page [23, 24]. Table 3 presents the observations from perliminary user testing and visualization design implemented to reflect them. Figure 7 shows the finalized landing page design.

Table 3 – visualization design iteration, preliminary user testing observations and design purpose for visualization landing page

Observation	Design	Purpose
Users need clear sense of navigation	Reseve left of page as nagtivation + global tool features throughout the tool (Figure 7A)	Give users a sense of navigation + quick access from page to page
Desire for dyanmic book/movie information (including moive/book pair selected)	Create global tool features for customed movie, book, actor, author, director filters to both book-movie pair (Figure 7A)	Proivde ineractivity between tool and users
Different user groups might seek different information from the tool	Designed scatterplots and bar chart with interchangable x and y-axsis for average movie rating, profit and revenue (Figure 7B)	Bring custom information / statitics to diferent user group
Short explanation / guide needed for the pages avaialble in this tool	Large icon with brief description for each page listed at the top of landing page (Figure 7A)	Use large images and analogies to connect users with purpose for each page
Users wants quick search for particular item globally (e.g., book, movie, actor, director, author, etc.)	Each global tool features on the left panel has a function "search in listbox"	Permit user to search for any information within the tool
Users wants to undo their selections	"Clear filter" button added (Figure 7A) and when user double-click on selection	Users can easily and quickly undo any selections in the tool

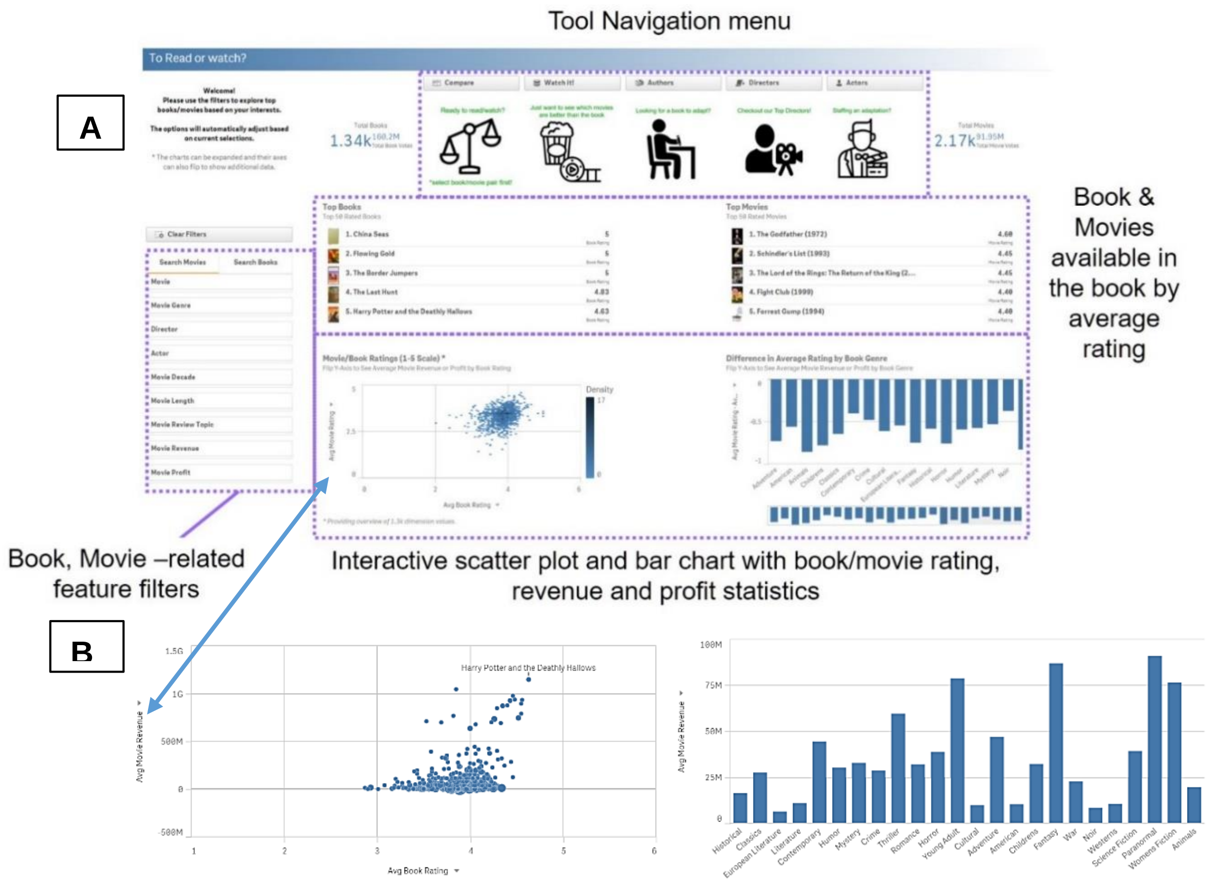

Figure 7 interactive visualization for the “Read it or Watch it” project landing page. Panel A: default options for scatter plot and bar chart. Panel B: options to change x-y axis for the scatter plot and such change is reflected in the bar chart (Note: purple dash lines are added only in this report to illustrate design concept)

Upon selecting a book-movie pair and clicking on the “compare” button, user sees our RI/WI recommendation, text analytics and topics for selected book/movie pair. Table 4 presents the observations from preliminary user testing and the design ideas for the “compare” page. Figure 8 is our finalized design for the page. An example for a book (“Nothing Last Forever”) and movie (“Die Hard”) pair is shown. The one-sentence summaries for movie influence in book reviews (and vice versa for movies reviews) and the Topic Funnels are results from LDA. Book / Movie word clouds (inspired by paper [20, 21]) are most frequent words mentioned in book/movie reviews.

In this example, “movie” and “Die Hard” were frequently mentioned in the book reviews while movie reviewers barely mention the book. This is a powerful observation that could be easily missed when users read only the book/movie reviews rather than side-by-side. The “Die Hard” discussion coverage in the book reviews aligns with the higher average movie rating and our RI/WI gauge. These features not only allow users to quickly make a decision based on multiple aspect of book vs. movie but also provide topics and/or features in the book and/or movie.

Table 4 – visualization design iteration, preliminary user testing observations and design purpose for visualization "Compare" page

Observation	Design	Purpose
Users cannot return to landing page	Created an icon and bolded label ("Home") based on recommendation from reference [23]	Users can intiutively see the "Home" button and access the landing page when desired
Users confused when there is one book and multiple movie (vice versa) but the pair was not displayed	"Total Books" and "Total Movies" labels added on the left of page was implemented (Figure 8)	To inform user to # of book and/or movies avilable in the selected pair
Users were unsure about the meaning of the text analytics results	One sentnece explanation is added to each text analytics	Provide clear description for each portion of the text analytics
Users prefer icons to signify which section is book vs. movie	Large icons added to represent book vs. movie	Give clear sense of space allocated for book vs. movie
Users want to see clear recommendation from the RI/WI guage	RI/WI guage placed in the page center with our recommendation listed above it in bolded green font	Draw user's attention to our recommendation
Users asked about how reviewers' polarity could change Movie/book influenece in the reviews + top topics displayed	Table listing "sentiment" as reviewer's polarity and the review quantity analyzed for the particular book/movie pair + dynamic filtering capability	Present data transparency and customized top topics per users' interest

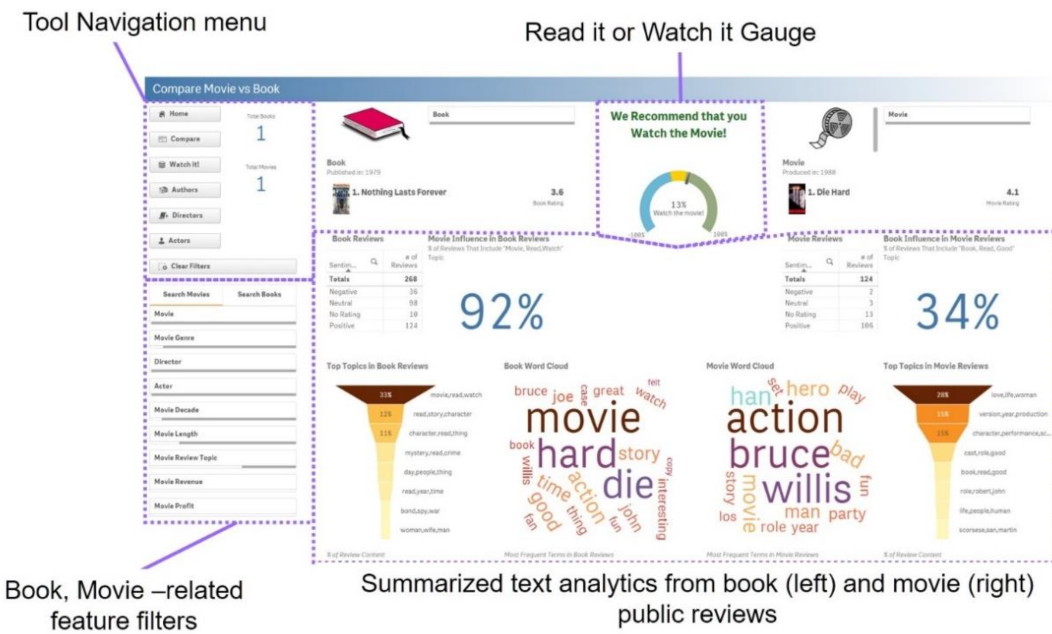

Figure 8 "Compare" page for book/movie pair (example book-movie pair used is "Nothing Lasts Forever" as the book and adapted as "Die Hard" the movie. (Note: purple dash lines are added only in this report to illustrate design concept)

We also developed interactive pages for actors(Figure 9), directors(Figure 10), and authors(Figure 11). All three pages present top actor/director/author statistics and ranking on the left and can be filtered to people of interest. The top right panel presents a rating bar chart (Figure 9 or Figure 10) or genre tiles (Figure 11) which change dynamically based upon user selection. The bottom right panel uses tree diagrams to display actor's or director's involvement in book-adapted movies (Figure 9 or Figure 10), or authors track records of books made into movies (Figure 11). The node size is proportional to average movie rating while the node color presents the average movie profit. These diagrams were inspired by examples by articles [18, 22].

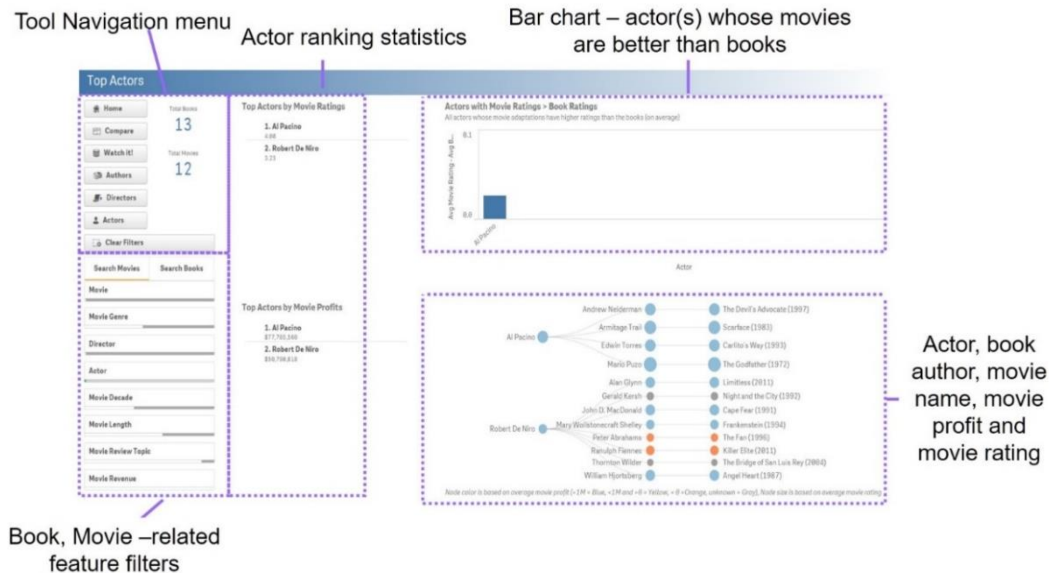

Figure 9 visualization page for top actors. (Note: purple dash lines are added only in this report to illustrate design concept)

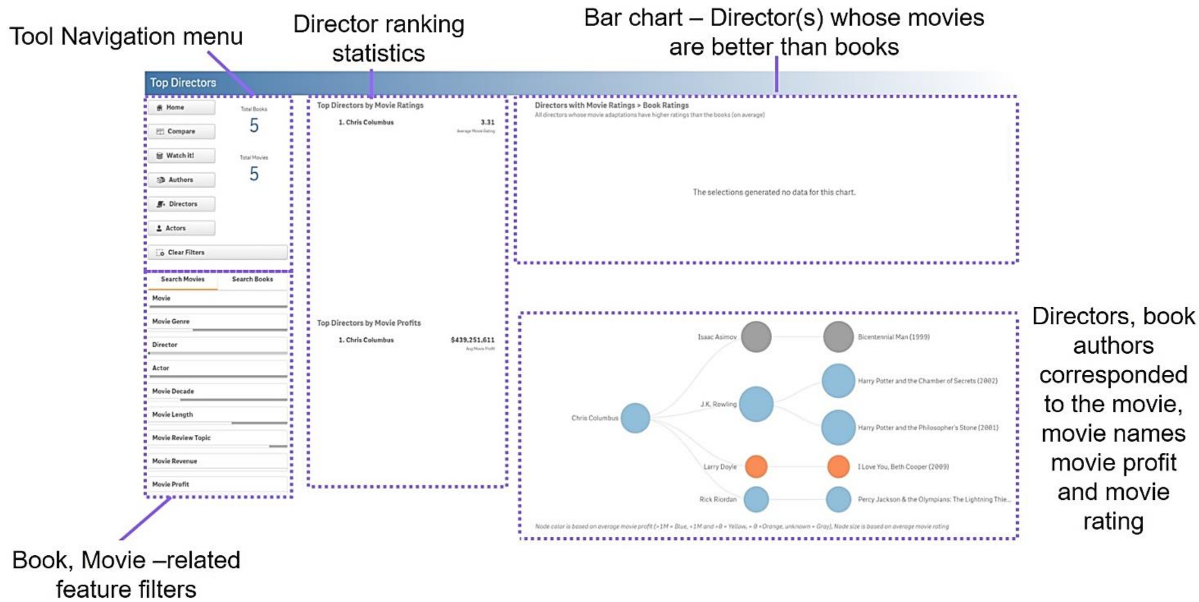

Figure 10 visualization page for top directors. (Note: purple dash lines are added only in this report to illustrate design concept)

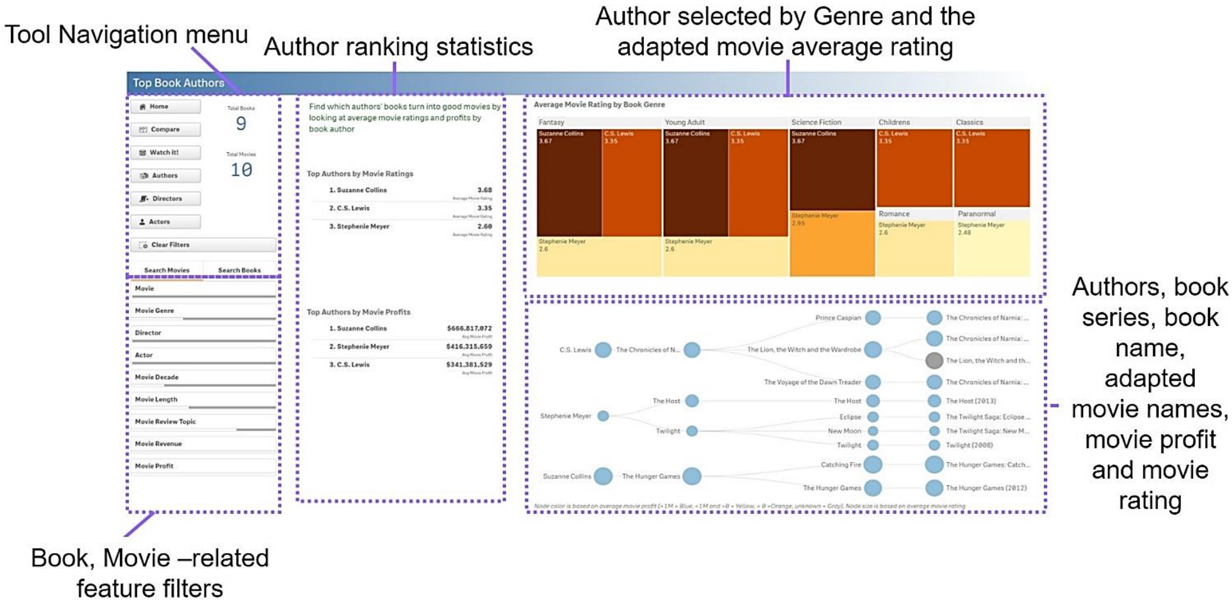

Figure 11 visualization page for top book authors. (Note: purple dash lines are added only in this report to illustrate design concept)

The preliminary user feedback suggested a page to visualize movies that our algorithm recommends over the book. Our analysis indicates that books are typically rated higher than their adapted movies. The "Watch it" page (Figure 12) allows users to quickly locate the movies with the aforementioned criteria. The top left side showed movies ranking in % movie rating higher than the book (% rating > book). This value is derived from Equation 1 and is the same value displayed in the RI/WI gauge in "Compare" page. The bar chart and tile chart are interactive – they change when movies and/or filters are selected from left panel. The bar chart offers both % rating > book and the movie length so to inform users about the movie. The tile chart displays the directors for the listed movies and their associated movie released decade.

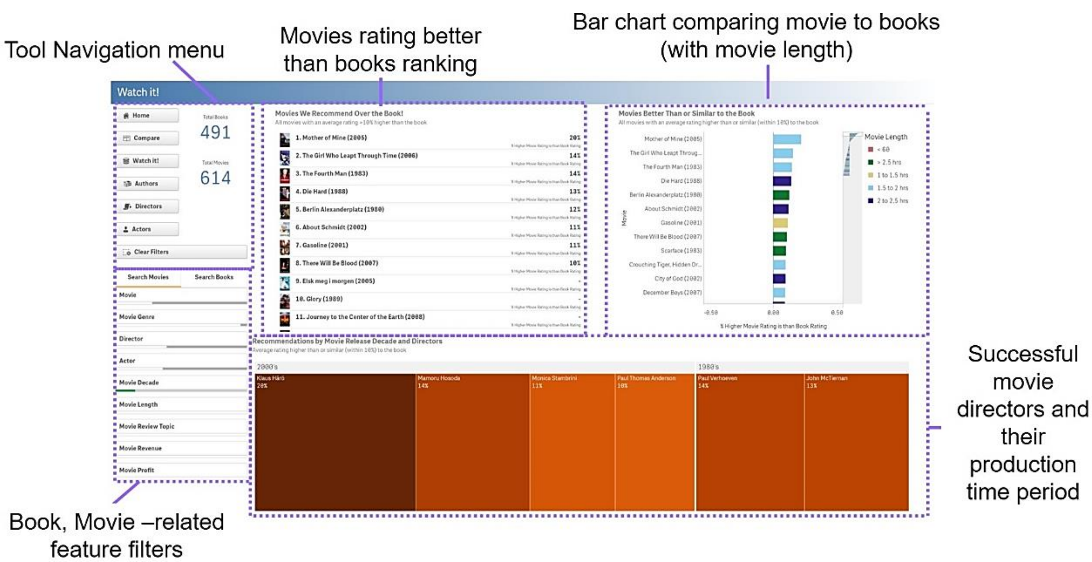

Figure 12 Watch it! Visualization page present the movies that are rated higher than their corresponded books.

3) User Testing and Survey

After finalizing LDA results for both books and movie, completing the second version of visualization design, we conduct the official user testing. Table 5 shows tasks users perform during the survey, their meaning and our project success metric defined by their response.

Table 5 – user testing survey detail and project success metric

Task	Purpose	Success Metric
Answer a few questions about their background and interest	Assess their need and typical ways to search for book-movie pairs to read/watch	N/A (for information only)
Time required to navigate to a specified book-movie pair	Overall page usability, intuitive design	< 3 minutes needed
Prompt user if they would use the page themselves to decide on reading a book and/or watching the associated movie	Learn if user interested in interacting further with the page	User wants to use the page again
Search and compare a specified book-movie pair, then answer 3 questions about what they noticed from the page	Verify that if information is clearly conveyed, does user pick up intended message	> 70% information identified

V. Conclusions and Discussion

After 6 visualization design iterations, we developed an application which provides users with self-guided experience for book-movie comparison, movie rating/profit/revenue information, exploratory pages for movies, actors, directors and authors.

20 users participated in the usability survey. 90% of users located the specified movie-book pair within 3 minutes starting from the landing page (Figure 13). 60% of the user found this process intuitive and 85% users think the information in the “Compare” page is helpful for deciding of whether to read the book or watch the movie (Figure 14). Within 5 minutes, more than 60% of the users were able to identify 4 out of 5 features within the visualization tool (Figure 15). Furthermore, we also received verbal feedback about exciting features and information about a book-movie pair. These results represent a significant improvement over the current process to find relevant information from many sources.

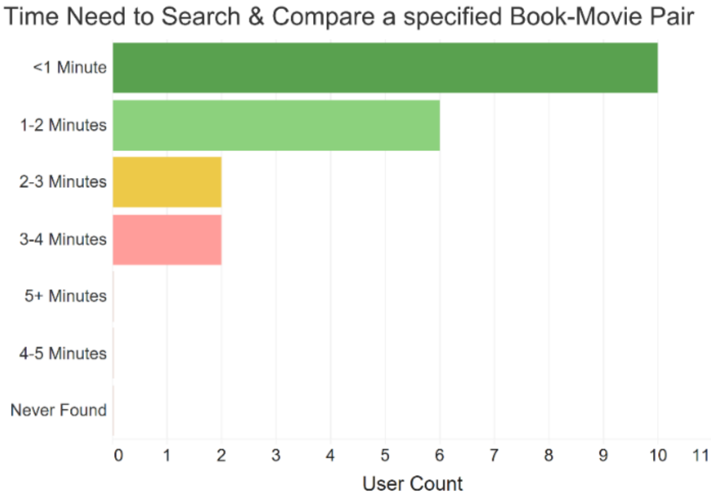

Figure 13 User testing survey for time required to navigate to a specified book-movie pair “Gone of the Wind” in visualization page

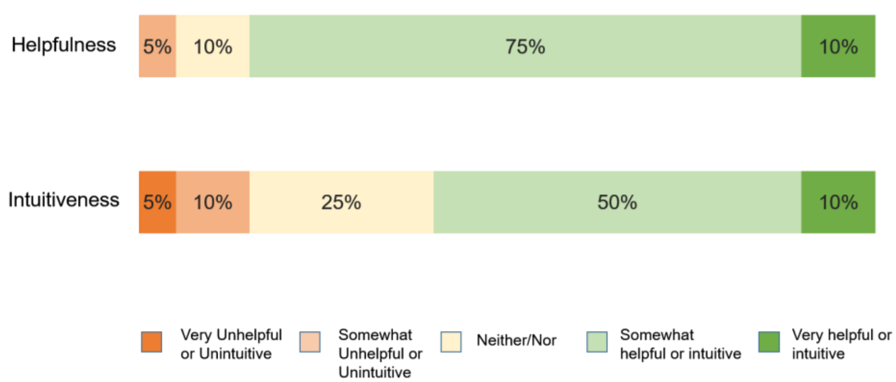

Figure 14 user feedback of helpfulness and intuitiveness for the “Read it or Watch it” application.

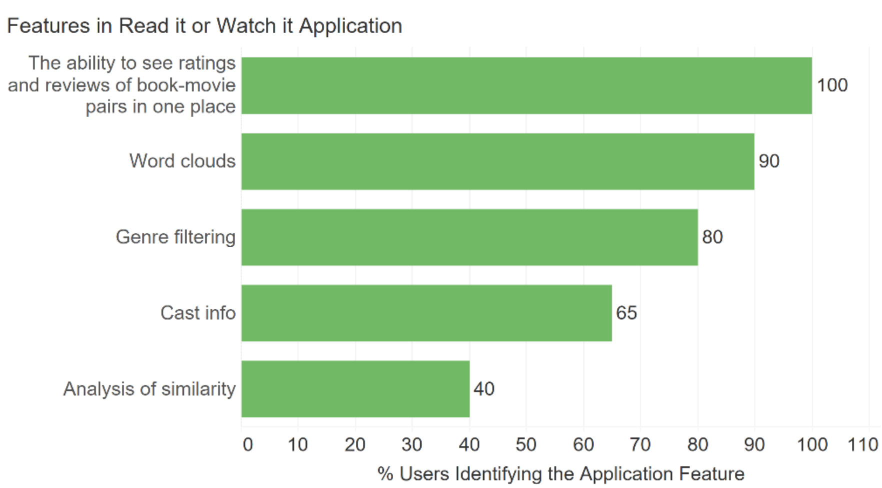

Figure 15 features identified by users within 5 minutes of exploring the application.

In conclusion, we exploited text topic analysis and interactive visualization to provide more accurate book-movie comparisons to our users. We leveraged Qlik for visualization and python scripts for LDA. The user testing survey indicates that users can find 80% of the information in the applications within 5 minutes, 90% of user need at maximum 3 minutes to locate and compare a book-movie pair. 85% users found the book-movie comparison information helpful. Overall, the user survey results align with our project success metric (Table 5) and we have accomplished the objectives we set for this project in the given timeframe. Moreover, we are happy and proud to have developed an appealing application that users enjoy using to quickly learn about book-movie pairs, including features that cannot be found anywhere else. Table 6 shows the future work recommended by us and the user.

Table 6 – future work for the “Read It or Watch It” project

Improvement area	Justification	Possible Action
User testing from personale in movie business	The application was meant for both the public and those in the movie business	Reach out to producers to conduct user testing survey
Update the book-movie pairs	Book-movie pairs can be further updated to match the recent trend	Web search and scraping from multiple sites
Portability	Qlik can only run on windows, web service cost money to users, some design limitations	Create a free web application
User intuitiveness	Following user feedback to make the navigation more intuitive	Seek expert advice for user experience (UX) to improve design
Topic Analysis Result (analysis of similarity)	Some users expressed confusion about the similarity results displayed in the “Compare page”	The free web app and UX advice can help educate user to understand similarity result
Topic Analysis by LDA	Can further optimize LDA evaluation for book/movie reviews	Combined book + movie reviews, perform LDA altogether to identify common topics

VI. Appendix: Tables from Literature Survey

Table 1 – project related literature survey for book reviews

Finding highlight(s)	Connections/Improvements	Reference
GoodReads is a more neutral book review platform compared to Amazon; pitfalls in its genre labeling	Determining review source, standardize genres so that the results are more intuitive/useful to user	[3]
Informal book reviews create societal impact for research, teaching and culture	Will use public reviews in GoodReads and IMDB rather than critics reviews	[4]
Book reviews sentiment and some book features are correlated with the book's popularity	Use sentiment analysis for books and explore correlation with movie adaptation popularity	[8]
Cross-referencing "feature tags" in book reviews to link similarities in multiple books	Features and/or tags mentioned in book/movie reviews will be explored, compare and contrast to find similarities and differences between book(s) and associated movie(s)	[7]
Procedures for preprocessing, word/opinion relevancy evaluation and machine learning models/parameters for scholarly book reviews	Similar methods will be applied to evaluate additional features for books (+ movies) type(s)	[9]

Table 2 – project related literature survey for movie reviews

Finding highlight(s)	Connections & Improvements	Reference
Pitfalls for overall sentiment analysis for movie reviews; applied three machine learning models to classify overall review polarity for movies and compare with human coded result	Evaluate effectiveness for both book-movie reviews; extract sentiment beyond overall polarity; do human check on results to assess accuracy given pitfalls	[10]
Negative reviews have stronger effect on product than positive; applied Naïve Bayes, Maximum Entropy and SVM models to classify review polarity and valency (for Korean movies)	Explore models for English reviews and extract sentiment beyond overall polarity; compare analysis results from negative reviews to overall rating;	[12]
Applied Naïve Bayes, AdaBoost algorithm and fuzzy lattice reasoning models and resulted in 73 – 89 % accuracy in classifying movie reviews overall polarity	Extract sentiment beyond overall polarity for other emotions and on specific features; spot check their results for accuracy	[11]
Identified variables to build a multivariate linear regression model to accurately predict for movie box office success using multivariate linear regression (R^2: 0.77)	Result is outdated; assess movies specifically from the book-movie pairs and determine if the similar trend follows	[15]

Table 3 – additional sentiment analysis literature survey

Finding highlight(s)	Connections & Improvements	Reference
Methods to remove redundancy while preserving infrequent and important keywords / phrases from customer review; methods on opinion word extraction, feature based analysis and features ranking	Follow similar methods to prepare and/or clean book-movie reviews, extract opinion words and feature	[17]
Extracted words/features frequently mentioned in reviews; detecting text which agrees with each other from different reviewers	Use frequency of terms, adjectives in reviews to assess book-movie reviewer polarity and follow the steps to conduct sentiment analysis	[15]
Revealed certain word sets from reviews show connections with reviewers' emotions	Apply word sets listed for book-movie reviews and compare with human evaluated results (spot test)	[26]
Naïve Bayes, SVM and Random Forest for categorizing product reviews polarity without using the entire review context; provides common sentiment phrases	Explore models and phrases for book-movie reviews by using particular words/phrases; compare results and accuracy when using full sentences (could be more effective)	[13]
Tested 26 sentiment dictionaries; aside from MPQA, most showed similar strength in results; Naïve Bayes works better	Identify or integrate sentiment dictionary from the article and compare performance with Naïve Bayes training model	[25]
Using hierarchical topic modeling process (CATHY) to identify frequent words used across different domains and the words that are domain specific	Explore the use of CATHY to find frequently used words in book-movie pair	[21]

Table 4 – text visualization literature survey

Finding highlight(s)	Connections & Improvements	Reference
Sankey plot for time dependent topic progress; circle opacity for time domain in scatter plot and keywords as text-cloud; stream chart for topic trend similarities	possible to apply text-cloud and scatterplot so show text and/or sentiment analysis results in addition to numerical ratings	[21]
Network graph line thickness for correlation strength; table with keywords, small trending chart in a column, slope/statistics in a tabular format	can use network graph to show similarity between book-movie pair, features and emotion results	[20]

Table 5 – interactive visualization literature survey

Finding highlight(s)	Connections & Improvements	Reference
Time dependent evolution of key players visualization; domain analysis via network chart with shaded color per cluster & highlighting high impact individuals	Can apply to book-movie series and/or popular cast	[18]
Multidimensional – map distances between points; SVD – map decomposed relation vectors to show variance; colors and animations enhance interpretation	May explore similar methods to reduce complex relationship among movie cast	[19]
List of basic visualization resources; “NameClarifier”: simplify and understand unclear names; a botanical structure to visualize author publications timeline	Refer to resources for visualization, may use NameClarifier for cast and author names; evaluate botanical structure for visualizing book author publications	[22]
Utilize user intuition with relatable analogies mental models and metaphor to design features/interactions - minimize errors and avoid using detailed instructions; labels give users a clear sense for navigation	Will follow these principles when designing interactive visualization for users	[23]
Visual techniques (e.g. shaded trails, arrows) for aiding transitions; bring attention to largest image – used to direct viewer on where to focus next; techniques for single-frame interactivity (mouse-over, progress bar) allowing users to control/explore information	Will follow these principles and apply elements into visualization design	[24]

References

[1] M. Surrey. (2019). *The Box Office Still Hasn't Woken Up in 2019* [Online]. Available: https://www.theringer.com/movies/2019/2/12/18221688/box-office-2019-slump-concerns.

[2] A. Liptak. (2017). *Why Hollywood is turning to books for its biggest productions* [Online]. Available: https://www.theverge.com/2017/1/26/14326356/hollywood-movie-book-adaptations-2017-expanse-game-of-thrones.

[3] S. Dimitrov, F. Zamal, A. Piper, and D. Ruths, "Goodreads Versus Amazon: The Effect of Decoupling Book Reviewing And Book Selling," in *Ninth International AAAI Conference on Web and Social Media*, 2015.

[4] K. Kousha, M. Thelwall, and M. Abdoli, "Goodreads Reviews to Assess the Wider Impacts of Books," *Journal of the Association for Information Science and Technology,* pp. 1-21, 2016.

[5] B. P. Byrne, K. Kelly, T. Reznik, and M. Gilat. (2015, 02/23/2019). *Rejoice, Snobs: The Book IS Better Than The Movie* [Online]. Available: https://www.vocativ.com/news/245040/the-book-is-better-than-the-movie/index.html.

[6] W. Hickey. (2015, 02/23/2019). *The 20 Most Extreme Cases Of 'The Book Was Better Than The Movie'* [Online]. Available: https://fivethirtyeight.com/features/the-20-most-extreme-cases-of-the-book-was-better-than-the-movie/.

[7] E. Lin, E. Shiaofen Fang, and E. Jie Wang, "Mining Online Book Reviews for Sentimental Clustering," ed, 2013, pp. 179-184.

[8] S. S. Sohail, J. Siddiqui, and R. Ali, "Feature extraction and analysis of online reviews for the recommendation of books using opinion mining technique," *Perspectives in Science,* vol. 8, pp. 754-756, 2016.

[9] A. Zuccala, M. van Someren, and M. van Bellen, "A machine-learning approach to coding book reviews as quality indicators: Toward a theory of megacitation," *Journal of the Association for Information Science and Technology,* vol. 65, no. 11, pp. 2248-2260, 2014.

[10] B. L. Pang, Lillian; Vaithyanathan, Shivakumar "Thumbs up? Sentiment Classification using Machine Learning Techniques," *Proceedings of the Conference on Empirical Methods in Natural Language Processing(EMNLP),* pp. 1-8, 2002.

[11] R. Satheesh Kumar and S. Vijayan, "Mining movie reviews - An evaluation," *Journal of Theoretical and Applied Information Technology,* vol. 56, no. 2, pp. 290-295, 2013.

[12] Y. Kim, D. Y. Kwon, and S. R. Jeong, "Comparing machine learning classifiers for movie WOM opinion mining," (in English), *KSII Transactions on Internet and Information Systems,* vol. 9, pp. 3169-3181, 2015.

[13] X. Fang and J. Zhan, "Sentiment Analysis Using Product Review Data," *Justin Journal of Big Data,* vol. Vol.2, pp. 1-14, 2015.

[14] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research,* vol. 3, pp. 993-1022, 2003.

[15] E. Cambria, B. Schuller, C. Yunqing Xia, and C. Havasi, "New Avenues in Opinion Mining and Sentiment Analysis," *Intelligent Systems, IEEE,* vol. 28, no. 2, pp. 15-21, 2013.

[16] Y. Guo, S. J. Barnes, and Q. Jia, "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation," *Tourism Management,* vol. 59, pp. 467-483, 2017/04/01/ 2017.

[17] M. Hu and B. Liu, "Mining Opinion Features in Customer Reviews," (in English), *American Association for Artificial Intelligence,* pp. 1-6, 2004.

[18] C. Chen, "Science Mapping: A Systematic Review of the Literature," (in English), *Journal of Data and Information Science,* vol. 2, no. 2, pp. 1-40, 2017.

[19] L. C. Freeman, "Visualizing Social Groups," *American Statistical Association,* pp. 47-54, 1999.

[20] P. Isenberg, T. Isenberg, M. Sedlmair, J. Chen, and T. Möller, "Visualization as Seen through its Research Paper Keywords," *IEEE Transactions on Visualization and Computer Graphics,* vol. 23, no. 1, pp. 771-780, 2017.

[21] X. Jiang and J. Zhang, "A text visualization method for cross-domain research topic mining," *Journal of Visualization,* journal article vol. 19, no. 3, pp. 561-576, August 01 2016.

[22] J. Liu, T. Tang, W. Wang, B. Xu, X. Kong, and F. Xia, "A Survey of Scholarly Data Visualization," *IEEE Access,* vol. 6, pp. 19205-19221, 2018.
[23] I. S. MacKenzie, "Human-Computer Interaction: An Empirical Research Perspective," in *Human-Computer Interaction: An Empirical Research Perspective*2 ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2013, pp. 71-111.
[24] E. Segel and J. Heer, "Narrative Visualization: Telling Stories with Data," *IEEE Transactions on Visualization and Computer Graphics,* vol. 16, no. 6, pp. 1139-1148, 2010.
[25] A. J. Reagan, C. M. Danforth, B. Tivnan, J. R. Williams, and P. S. Dodds, "Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs," *EPJ Data Science,* journal article vol. 6, no. 1, pp. 6-28, October 30 2017.
[26] M. Almashraee, D. Monett, and A. Paschke, "Emotion Level Sentiment Analysis: The Affective Opinion Evaluation," in *Workshop on Emotions, Modality, Sentiment Analysis and the Semantic Web*, Heraklion, Greece, 2016, pp. 1-12.