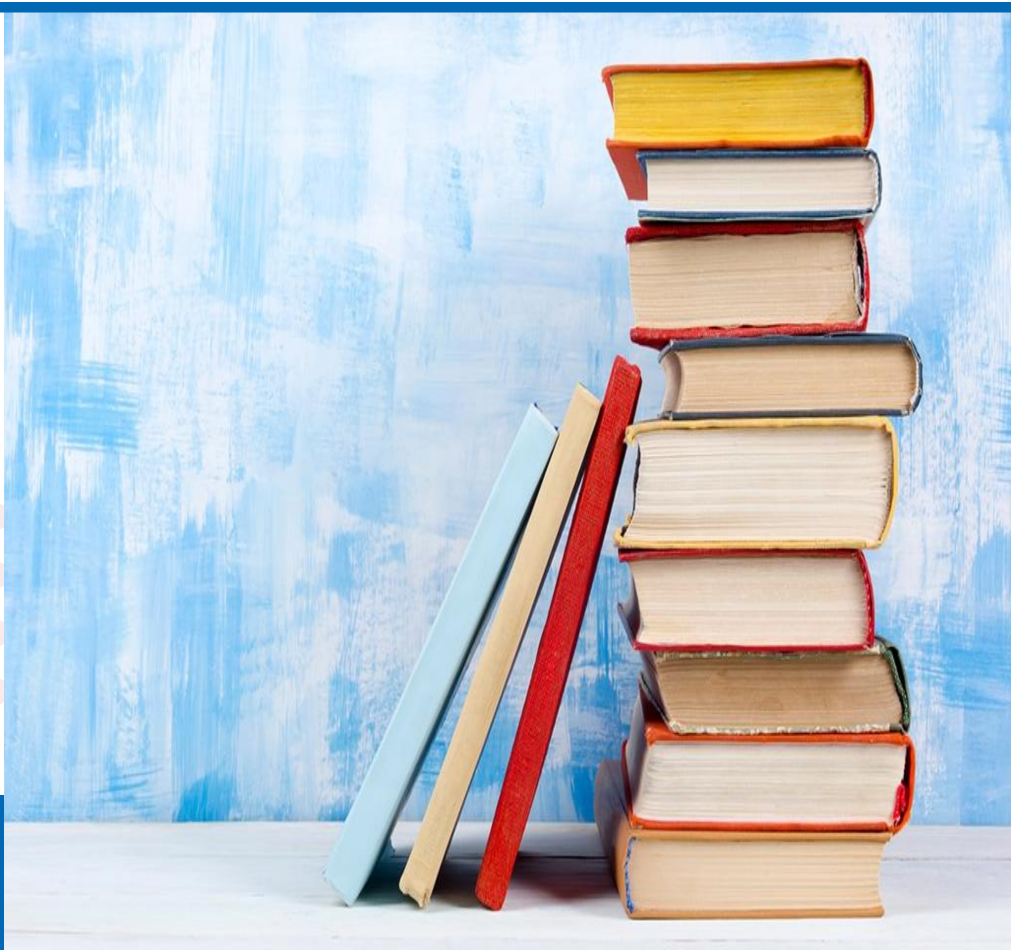
Adaptive and Explainable Search for Comics (AESC)

Supervised By (D.K.E Department)
Sayantan Polley
Prof. Dr.-Ing Andreas Nürnberger



Agenda

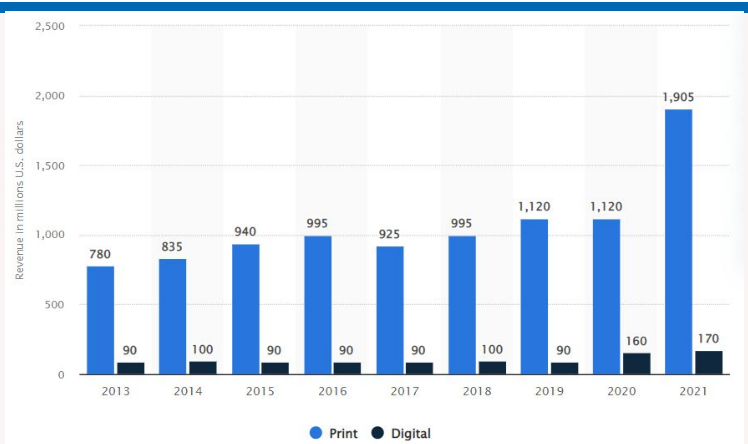
Introduction

01 Comics are a complex multi-modal art form consisting of sequential graphics with short textual dialogue

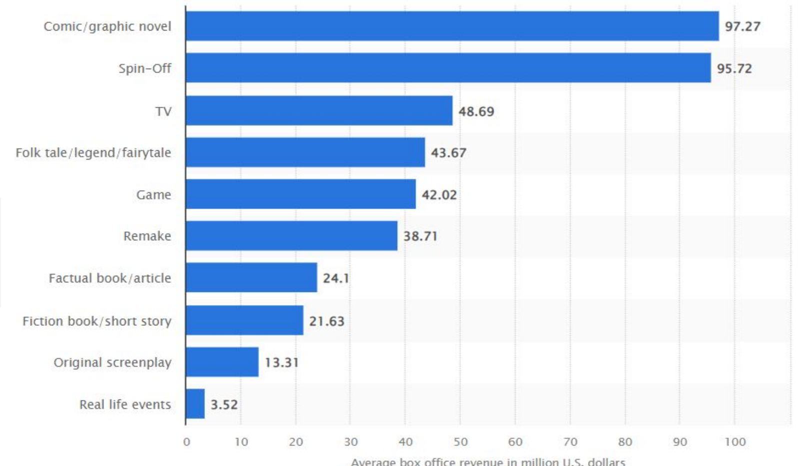
02 Comics transformed to movies, TV series, videogames and vice-versa

03 Sales dominated by print

04 Better search of suitable content can help average readers, artists, analysts



Source: Comic Book sales in USA[1]



Source: Average Movie Revenue (1995-2022)[2]

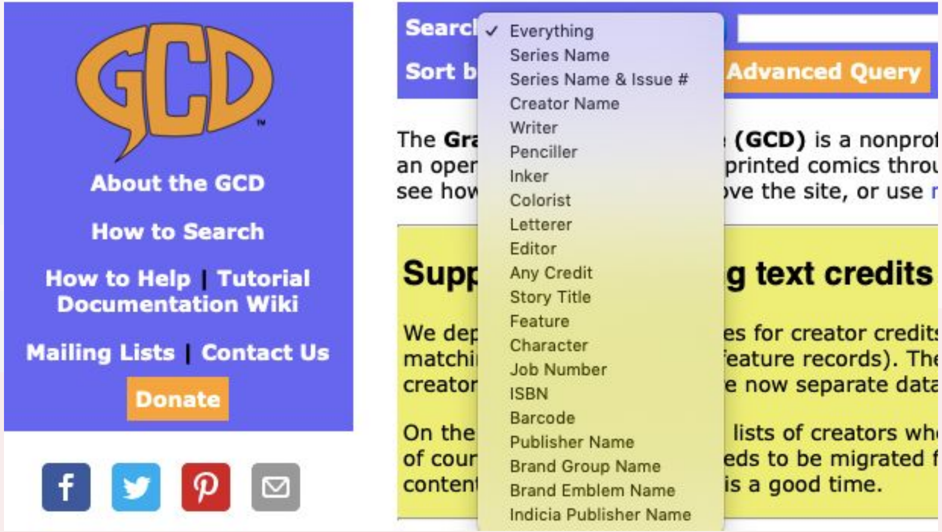
Motivation

01 Metadata or User based search engines

02 Differing user needs and new content discovery

03 Multi-Modality, Artistic Style, Geographies makes exploration difficult

04 Understand relation between results and query



Source: https://www.comics.org[3]

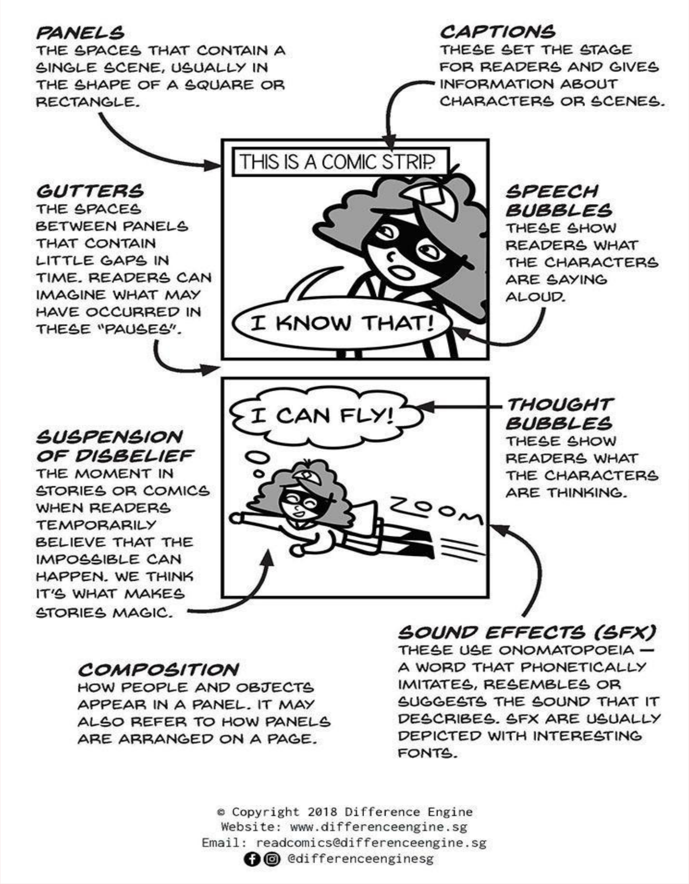
Background – Comic Elements

01 Panel forms the basic unit of comic book[5].

02 Sequencing of panels, shape and position of the panel controls time and renders artist vision

03 Dialogues are usually short and expressive. Caption signals scene change

04 Sound Effects capture objects action with its panel



Source: elements of comics[4]

Background – Adaptivity

01 **Distance based approaches try to find similarity between relevant and irrelevant books[6].**

- Metric learning[7]
- **Feature reweighting[8]**

02 Click based approaches track user activity on screen and tune search model[9].

03 Reinforcement Learning models search as exploration and exploitation[10].

Background – Explainability

01

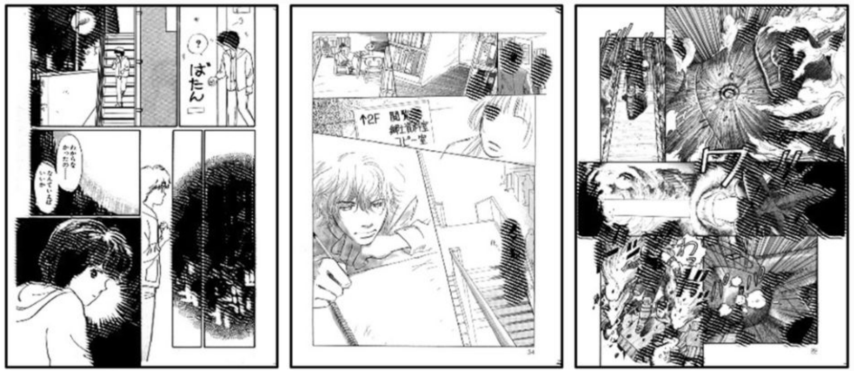
Explainable by Design[11]. No surrogate model.
- Trained from scratch to explain results.
- ILMART: Interpretable Learning to Rank model[12].

02

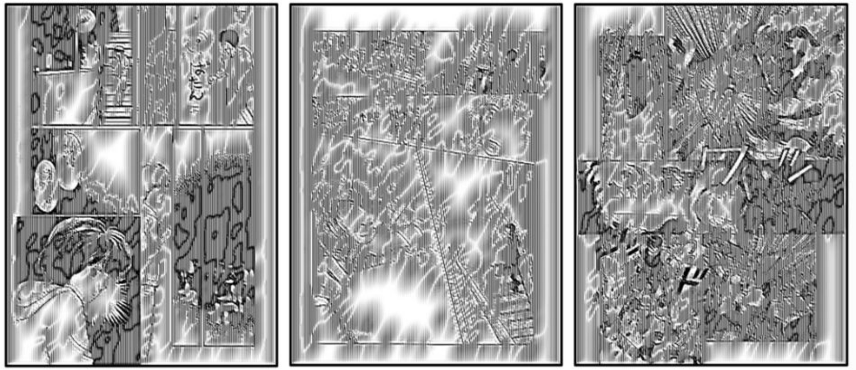
Post-Hoc
- Feature importance, **weights**, LIME[13], SHAP[14] ,LRP[15], and others are commonly used explanation techniques.
- X-Vision: Matching facets as explanation, Interpretable feature space with **reranking** [16].
- ExDocs: Interpretable feature space, provides keywords as explanation [17].
- SIMFIC: Feature importance of facets as explanation [18].

03

Comics
- Visualize features responsible for finding comic authors[19].
- Feature contribution towards visual stylometry[20].



Source: Feature visualization of higher layer neurons [19]

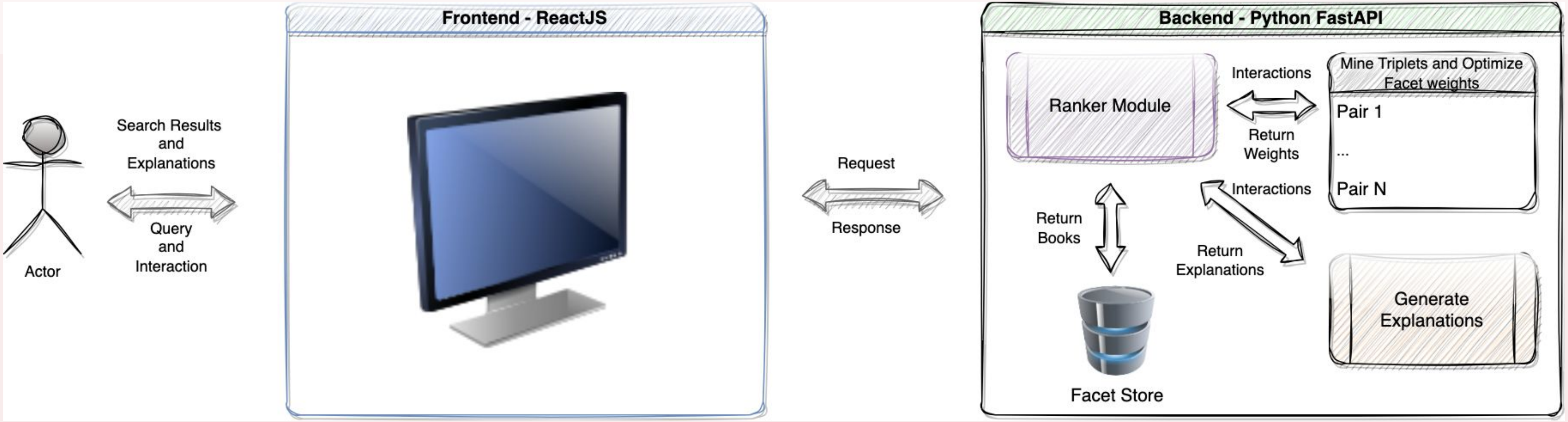


Source: Feature visualization of lower layer neurons [19]

Approach

01 Frontend captures user interaction. Displays search results and explanations.

02 Backend responsible for adapting to user preference and produce relevant explanations.



Datasets

01 Problems in creating and sharing datasets due to copyright issues

02 Available datasets

- eBDtheque[21]: 100 French comic pages annotated with balloons, text, panels
- Manga109[21]: 109 Manga comic books cutting across 12 genres. Only few volumes annotated
- COMICS[21]: 3948 Golden age copyright free comics. Panel, Text extracted by Machine Learning

03 Our dataset include

- 1000 comics/40 famous characters from internet, mostly from Golden Age and Silver age, in cbr/cbz/pdf format
- 710 books from COMICS dataset.
- **Toy dataset consisting of known books for experimentation**

Research Questions

01 What is the impact of domain based facets on search results quality?

02 What is the impact of user-system interaction on satisfying user's search context?

03 How does textual explanations generated from book cover help user understand personalization?

04 How does comparison table explanation help user discern between books from search results?

Facets

Facet Level	Facet Type	Facet Name	Dimensions	Rationale
Low Level	Visual	CLD	64	Represent average colors in book, thereby style
	Visual	EHD	128	Represent average texture in book, thereby style
	Visual	HOG	256	Represent average objects in book, thereby style
	Textual	Text Bag of Words	64	Represent entire text from book
Semantic	Visual	Book Cover Image	2048	Proxy for synopsis. Designed for attraction.
	Textual	Book Cover Prompt	384	Serves as explanation as well as Facet.
	Textual	Genre	16	Genre of the book
	Textual	Gender	3	Male oriented, Female oriented or Other
	Textual	Supersense	46	Coarse categories of ontologies, helpful for question answering
	Temporal	Story Pace	1	Average panels per page representing rough pace
	Temporal	Total Information in Book	1	Sum of text word count and panels in book

Implementation - Adaptivity

04

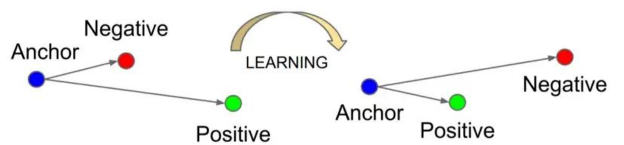
Can system understand users preferences?

- Triplet loss[22] tries to bring query and hovered book together and non-hovered book far apart by a margin.

- Loss updates facet weights by gradient descent

- 1000 epochs and learning rate 1e-3 .

- Local Minima, Quality Triplets, Correlated Facets



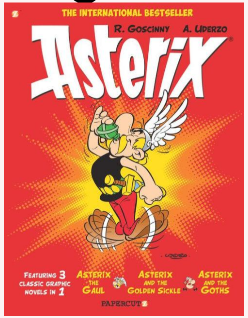
Source: Triplet Loss Illustration[23]

$$\sum_{i}^{N}\left[\|f(x_i^a)-f(x_i^p)\|_2^2-\|f(x_i^a)-f(x_i^n)\|_2^2+\alpha\right]$$

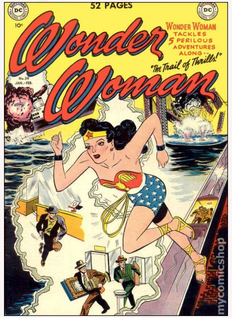
Source: Triplet Loss Formula[22]

Hovered

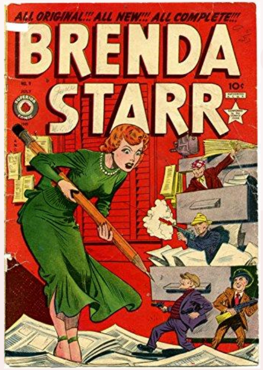
Not Hovered



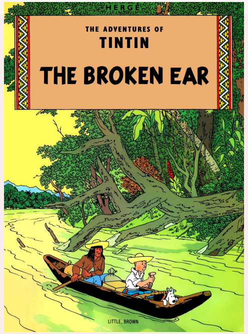
Source: Asterix[28]



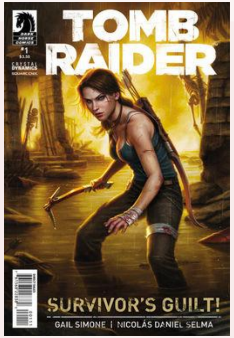
Source: Wonder Woman[29]



Source: Brenda Starr[26]

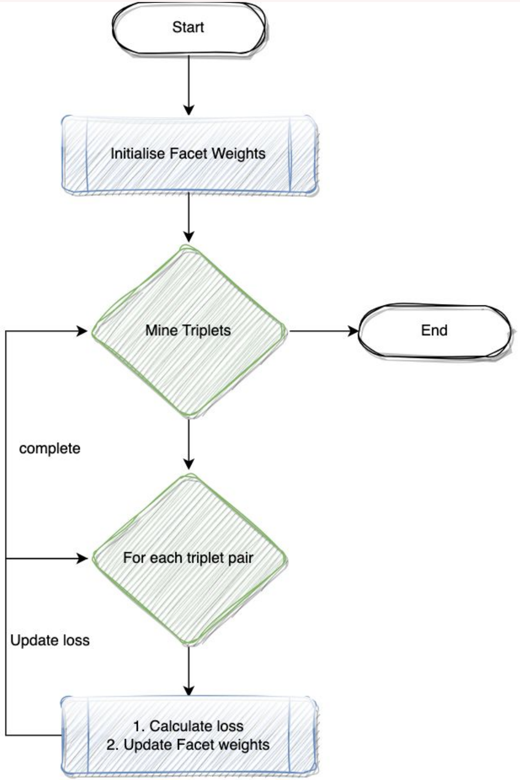


Source: Tin-Tin[27]

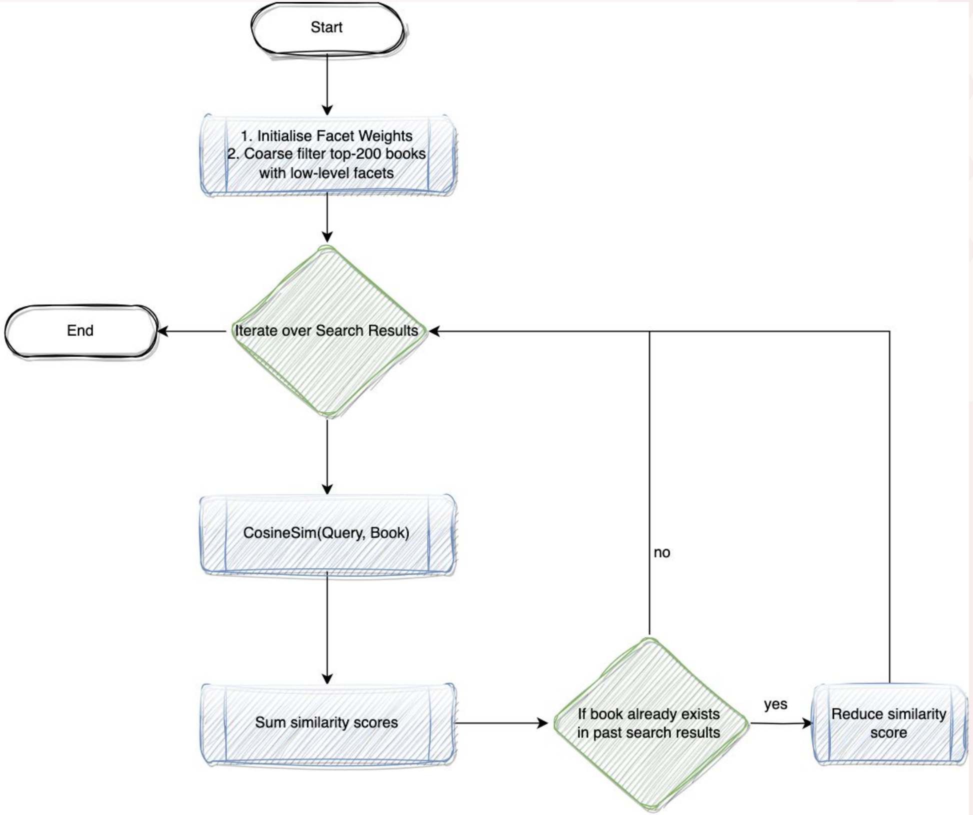


Source: Tomb raider[30]

Implementation - Reranking and Adaptation



Triplet Mining and Adaptation



Reranking

Implementation - Explainability

01

Global Explanation

- Sliders represent facet weights and also serves as explanation

02

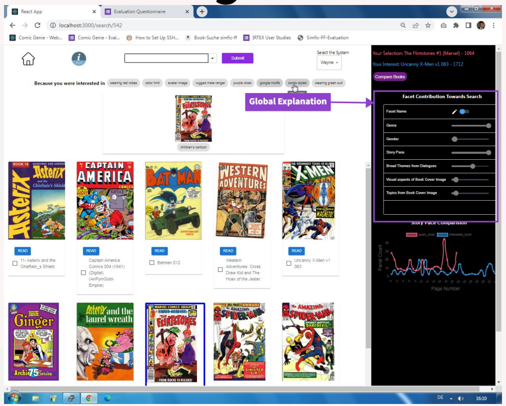
Personalization Explanation

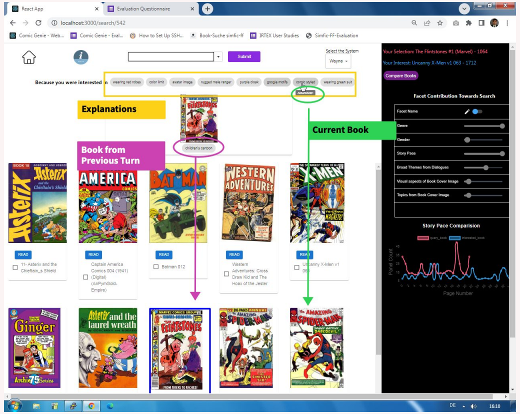
- Textual explanations from book cover that is most similar to previous search patterns

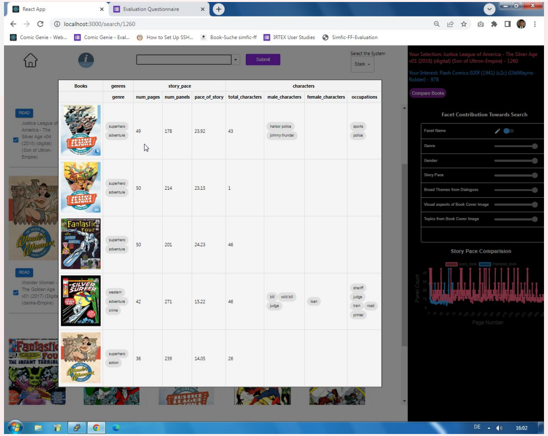
03

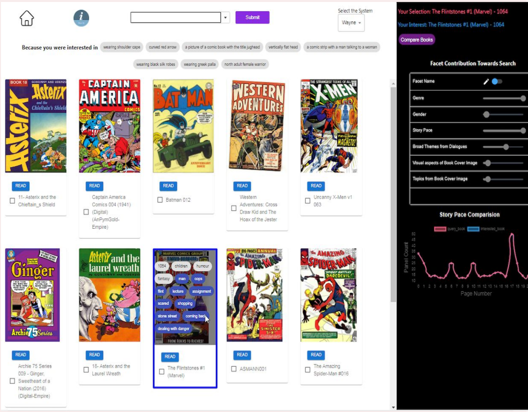
Local Explanation

- Comparison table to compare between search results and Keywords that acts as proxy for summary









Implementation - Personalization Explanation

01

Assumption

- if adaptation is done using hovered books from previous session, then there must be a common element and an explanation between books from previous session and current search results

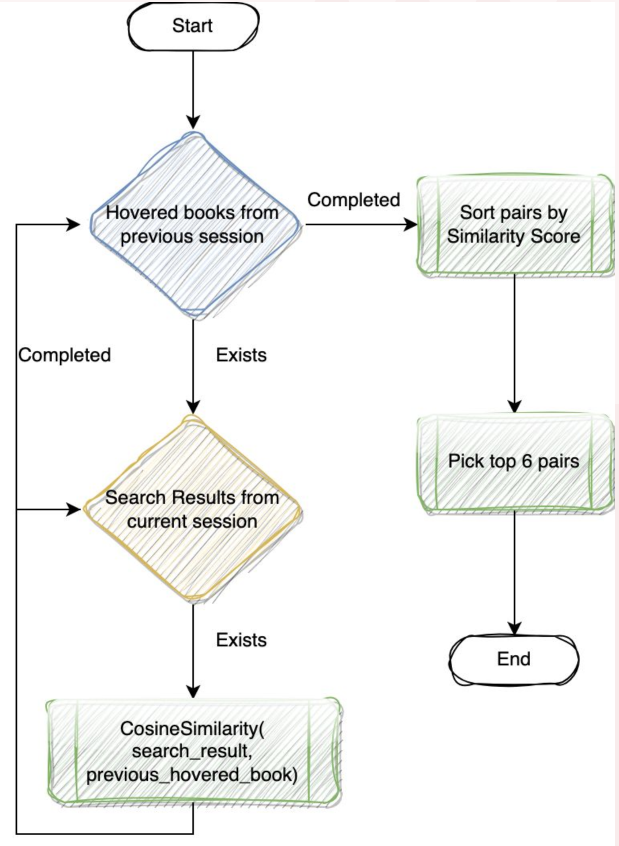
02

Find similarity for every pair of current search results and previous hovered books

- use book cover prompts as book cover can be easily verified by user.

03

Select top 6 similar pairs



User Evaluation

Usefulness metric[24] assesses the extent to which a search engine aligns with user intent and preferences, ultimately enhancing the their ability to find desired information.

User Evaluation facets[24]

01 Facet Usefulness: User satisfied with search results based on facets.

02 Interaction Usefulness: User can communicate their requirements with the system.

03 Explanation Usefulness: Explanations that help user understand search results better

User Evaluation

- Evaluation setup using Tobii Eye Tracker[25] to capture users cognition. 33 users participated in DKE Lab
- Greco-Latin square block with question order and task order randomly changed to reduce bias.

01 Interactive Precision

- Measures accuracy and relevance of user interactions. High precision means more relevant outcomes and less wasted effort.
- Interactive Precision = (Number of Relevant Documents Selected / Total Number of Documents Selected) * 100

02 Effectiveness

- Gauges system's success in achieving objectives. High effectiveness indicates reliable goal attainment.[6]
- Task Effectiveness = (Number of Relevant Artifacts Chosen by User / Total Relevant Artifacts) * 100

03 Efficiency

- Evaluates task completion with minimal resource consumption. Efficient systems save time and effort.
- Measured in time taken to complete task in minutes.

04 Satisfaction

- Reflects user contentment with system performance. High satisfaction leads to loyalty and engagement.
- Measured in likert scale between 1-7 with 1 being lowest and 7 being highest

RQ1: What is the impact of domain based facets on search results quality?

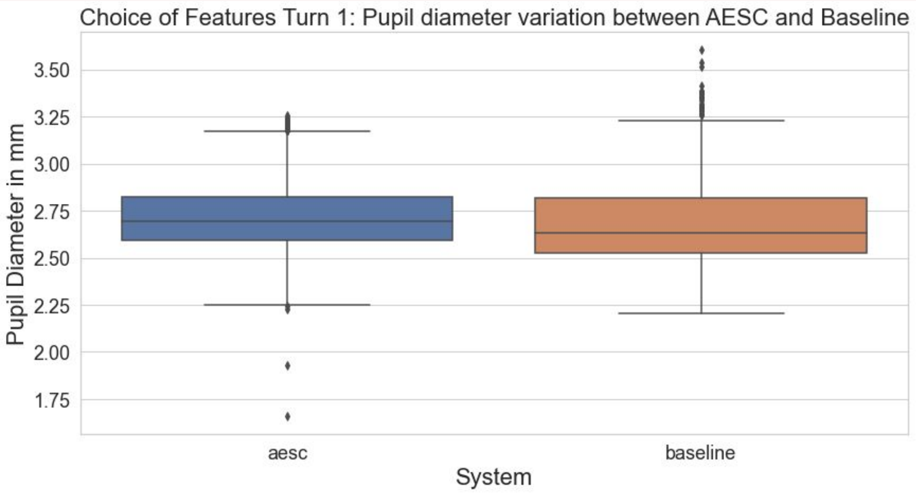
Assumption: AESC system will provide good quality results as compared to Baseline. Time taken to perform a task on AESC would be lesser than Baseline due to good quality.

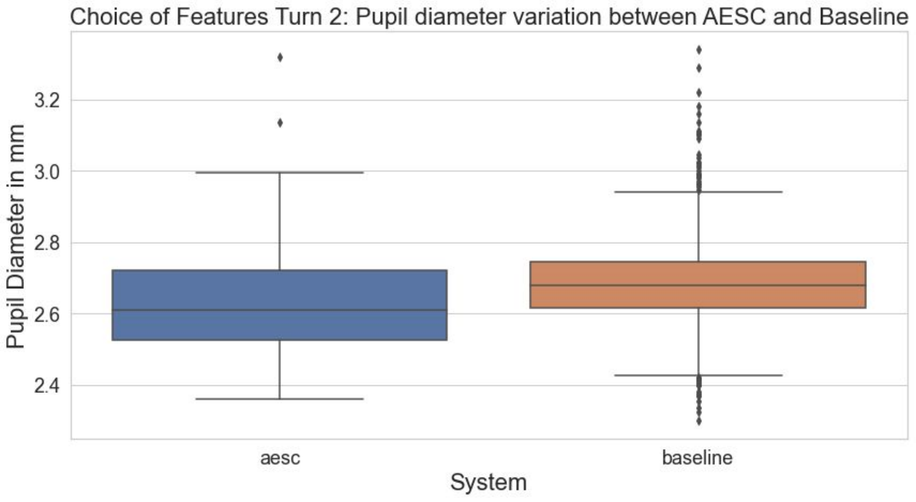
- AESC outperformed Baseline in Interactive Precision, Effectiveness, and Satisfaction, signifying improved search quality with domain-based facets. Users spent more time on AESC, possibly due to increased relevance in choices.

Metric	Turn 1 AESC	Turn 2 AESC	Overall AESC	Turn 1 Baseline	Turn 2 Baseline	Overall Baseline
Interactive Precision (in %)	**70.2 ± 41.71**	**84.26 ± 34.11**	**77.23 ± 37.91**	51.1 ± 50.1	53.2 ± 49.26	52.07 ± 49.66
Effectiveness (in %)	**39 ± 29.39**	**55.62 ± 36.53**	**47.5 ± 26.39**	20.02 ± 19.8	30.62 ± 30.04	25.31 ± 20.94
Efficiency (in min)	2.61 ± 0.63	2.33 ± 1.5	2.47 ± 1.5	**1.7 ± 0.46**	**1.46 ± 0.23**	**1.58 ± 0.34**
Satisfaction	-	-	**5.03 ± 1.51**	-	-	3.56 ± 2.05

RQ1: Impact of facets on search results quality?

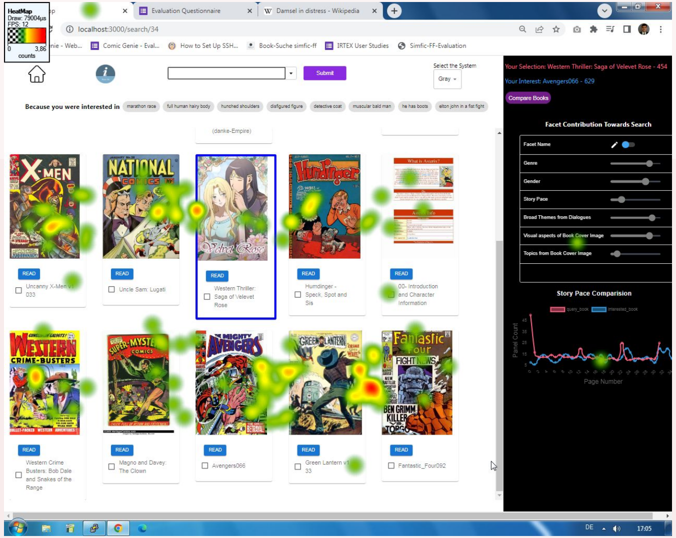
- Cognitive load remained similar during Turn 1, decreased for AESC in Turn 2, contrary to the expected increase with more relevant results.



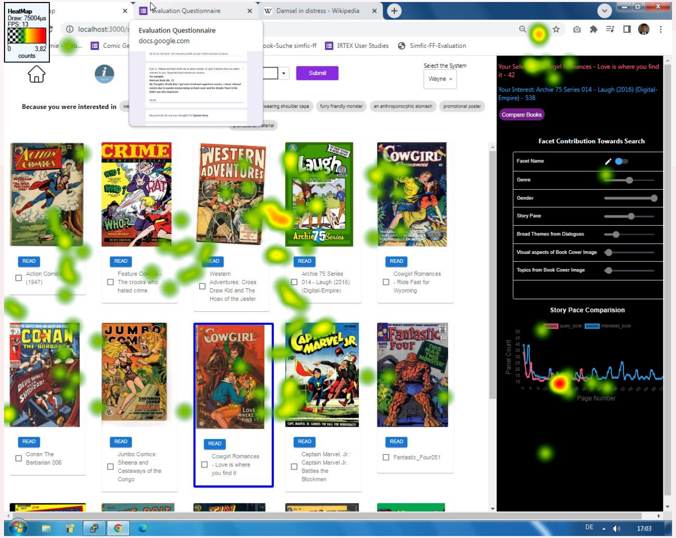


RQ1: What is the impact of domain based facets on search results quality?

- Heatmaps shows that cognitive load decreases a bit for AESC.



Heatmap of Baseline Turn 2



Heatmap of AESC Turn 2

RQ2: What is the impact of user-system interaction on satisfying user's search context?

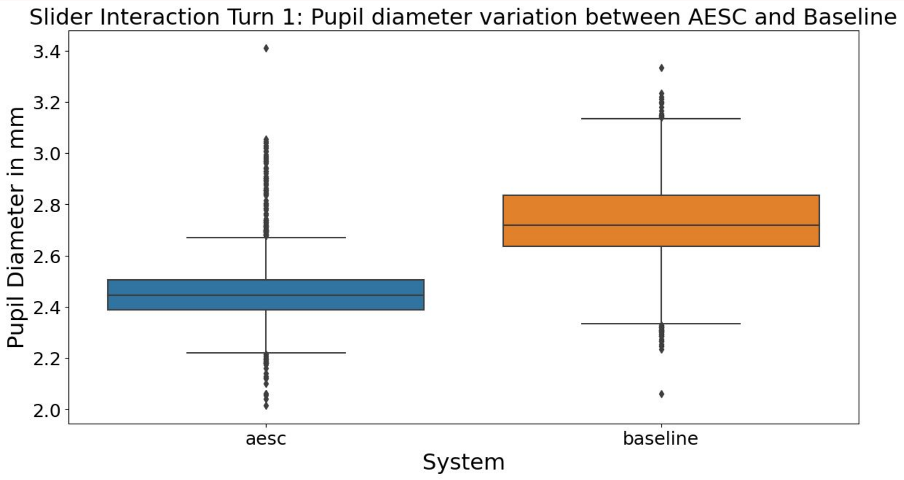
Assumption: Users would be able to interact with AESC system much better than baseline. Although the mode of communication would remain same, similarity model would result in better interaction.

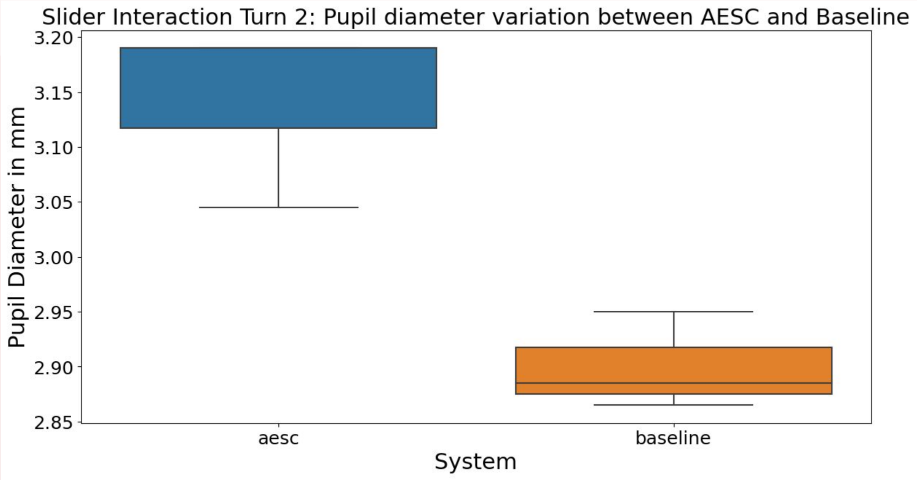
- AESC excels in Interactive Precision, Effectiveness, and Satisfaction compared to baseline, showcasing the influence of sliders in enhancing user interaction and system understanding. Users dedicated more time to AESC, possibly due to heightened engagement with sliders and book facets.

Metric	Turn 1 (AESC)	Turn 2 (AESC)	Overall (AESC)	Turn 1 (Baseline)	Turn 2 (Baseline)	Overall (Baseline)
Interactive Precision (%)	**63.72 ± 35.95**	**87.7 ± 26.92**	**75.71 ± 31.44**	59.9 ± 45.1	63.62 ± 42.26	61.75 ± 43.87
Effectiveness (%)	**68.12 ± 28.67**	**67.5 ± 32.82**	**67.81 ± 26**	68.12 ± 30.04	53.12 ± 31.15	60.62 ± 23.41
Efficiency (min)	6.4 ± 4.61	7.62 ± 6.65	7.01 ± 5.63	**2.5 ± 0.61**	**3.83 ± 1.81**	**3.16 ± 2.45**
Satisfaction	-	-	**5.41 ± 1.36**	-	-	4.43 ± 1.72

RQ2: What is the impact of user-system interaction on satisfying user's search context?

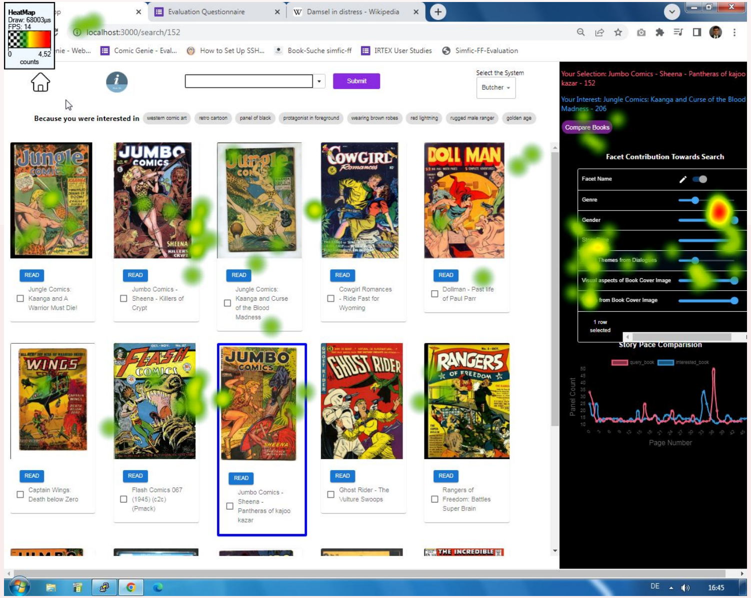
- During Turn 1 Cognitive load is less for AESC, but increases drastically during Turn 2.
- We thought that with better interaction it should be easy for users to find the books, but cognitive load increased more for turn 2. We can see from increased interaction on sliders, which could have bumped up load.



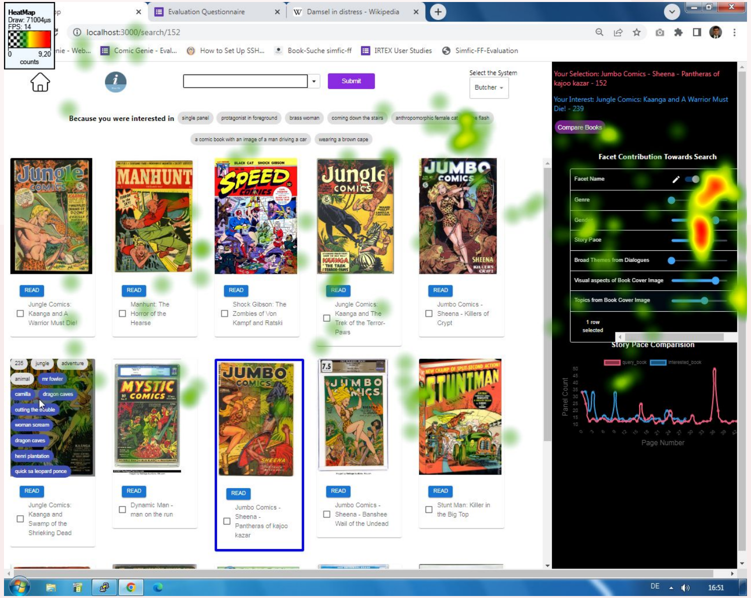


RQ2: What is the impact of user-system interaction on satisfying user's search context?

- Heatmaps shows that cognitive load increases for AESC, Slider interaction increased for AESC.



Heatmap of Baseline Turn 2



Heatmap of AESC Turn 2

RQ3: How does textual explanations generated from book cover help user understand personalization?

Assumption: Textual explanations derived from book cover text linked to previous search pattern would help explain personalization

- AESC demonstrates superior Effectiveness and Satisfaction, though not statistically significant, compared to baseline. Users invested more time in AESC, likely attributed to the preference for textual explanations, influencing the overall time spent.

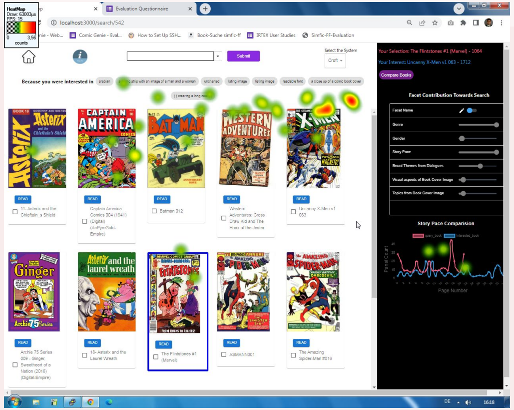
- Room for improvement in deriving textual explanations?

Metric	AESC	Baseline
Effectiveness (%)	**31.16 ± 14.5**	24.33 ± 16.83
Efficiency (min)	4.1 ± 1.74	**3.25 ± 1.69**
Satisfaction	**4.93 ± 1.61**	4.18 ± 1.92

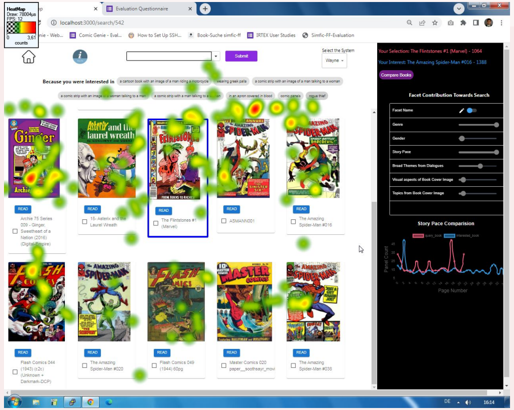
RQ3: How does textual explanations generated from book cover help user understand personalization?

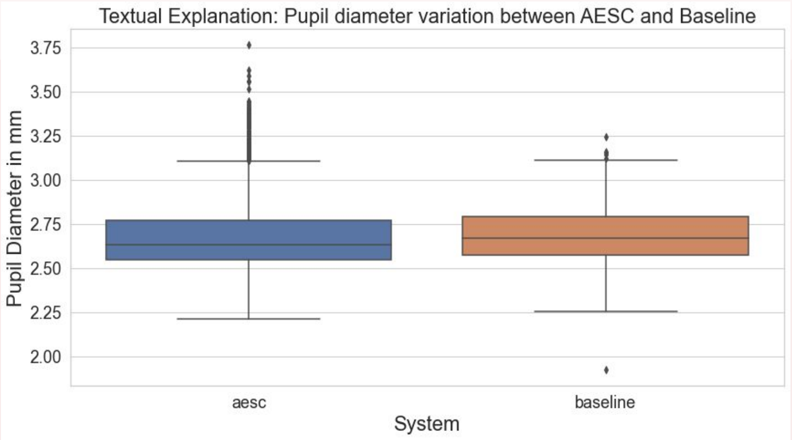
- Cognitive load remained almost same between two systems. As seen from heat maps, there was no considerable focus on textual explanations
- Lack of quality explanations could be the reason for drop in load for both systems.

Heatmap of Baseline



Heatmap of AESC





RQ4: How does comparison table explanation help user discern between books from search results?

Assumption: Comparison table should help compare between books in search results.

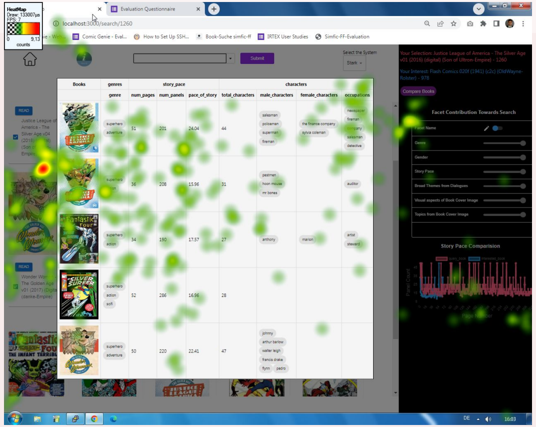
- AESC outperforms baseline significantly in Effectiveness, Efficiency, and Satisfaction, indicating enhanced user performance and efficiency in book selection.

Metric	AESC	Baseline
Effectiveness (%)	**84 ± 26.5**	59 ± 26.5
Efficiency (min)	**2.91 ± 1.82**	5.01 ± 4.59
Satisfaction	**5.87 ± 1.21**	5.31 ± 1.59

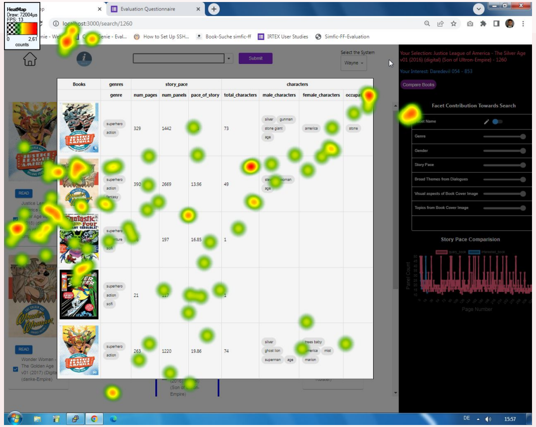
RQ4: How does comparison table explanation help user discern between books from search results?

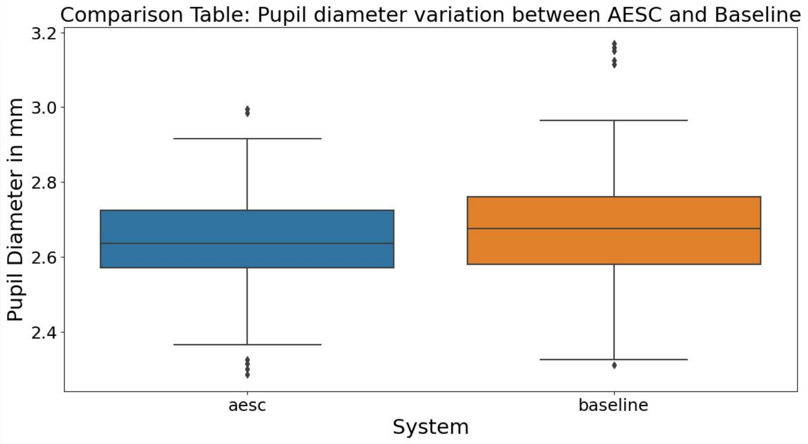
- Cognitive load remained almost same between two systems. As seen from heat maps, there was no considerable focus on comparison table.
- Comparison table looks the same in both systems which could have caused lower difference in loads.

Heatmap of Baseline



Heatmap of AESC





Conclusion

01 Domain based facets help retrieving relevant search results than random retriever

02 User interaction in form of feature weightage helps to communicate their preferences against a non-adaptive system.

03 Textual explanations derived from book covers and linked to previous search pattern provides flimsy explanation regarding personalization.

04 Comparison table serves as effective local explanation in comparing books from search results.

Future Work

01 Improvement in facets

02 Create benchmark dataset for measuring retrieval performance on comic book dataset.

03 Use better adaptive algorithms apart from simple gradient descent.

04 Improve personalization explanation

05 Include helpful information

References

1. Watson, A. (2022, July 5). *Comics sales revenue in North America by format 2021*. Statista. https://www.statista.com/statistics/369062/comics-revenue-print-digital/#:~:text=Revenue%20for%20printed%20comics%20and,dollars%20from%20the%20previous%20year.
2. Statista. (2023, January 5). *Average revenue of films in the U.S. & Canada 1995-2022, by selected source material*. https://www.statista.com/statistics/188689/movie-sources-in-north-america-by-average-box-office-revenue/.
3. *Grand Comics Database*. (n.d.). https://www.comics.org/.
4. Elements of Comics. https://differenceengine.sg/wp-content/uploads/2018/09/Adas-Adventure-Comics-For-Educators.jpg
5. Cohn, N. (2013). Navigating Comics: An Empirical and Theoretical Approach to Strategies of Reading Comic Page Layouts. *Frontiers in Psychology, 4*
6. Stober, S. (2011). Adaptive methods for user-centered organization of music collections.
7. Lee, J., Jin, R., & Jain, A.K. (2008). Rank-based distance metric learning: An application to image retrieval. *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 1-8.
8. Stober, S., & Nürnberger, A. (2011). An Experimental Comparison of Similarity Adaptation Approaches. *Adaptive Multimedia Retrieval*.
9. Joachims, T., Granka, L.A., Pan, B., Hembrooke, H., & Gay, G. (2005). Accurately interpreting clickthrough data as implicit feedback. *SIGIR Forum, 51*, 4-11.
10. Zhou, L. (2015). A Survey on Contextual Multi-armed Bandits. *ArXiv, abs/1508.03326*.
11. Anand, A., Lyu, L., Idahl, M., Wang, Y., Wallat, J., & Zhang, Z. (2022). Explainable Information Retrieval: A Survey. *ArXiv, abs/2211.02405*.
12. Lucchese, C., Nardini, F.M., Orlando, S., Perego, R., & Veneri, A. (2022). ILMART: Interpretable Ranking with Constrained LambdaMART. *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*.
13. Ribeiro, M., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
14. Lundberg, S.M., & Lee, S. (2017). A Unified Approach to Interpreting Model Predictions. *ArXiv, abs/1705.07874*.
15. Montavon, G., Binder, A., Lapuschkin, S., Samek, W., & Müller, K. (2019). Layer-Wise Relevance Propagation: An Overview. *Explainable AI*.
16. Polley, S., Mondal, S., Mannam, V.S., Kumar, K., Patra, S., & Nürnberger, A. (2022). X-Vision: Explainable Image Retrieval by Re-Ranking in Semantic Space. *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*.
17. Polley, S., Janki, A., Thiel, M., Hoebel-Mueller, J., & Nürnberger, A. (2021). ExDocS: Evidence based Explainable Document Search.

References

18. Polley, S., Ghosh, S., Thiel, M., Kotzyba, M., & Nürnberger, A. (2020). SIMFIC: An Explainable Book Search Companion. *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, 1-6.
19. Kim, Y. (2019). Feature visualization in comic artist classification using deep neural networks. *Journal of Big Data, 6*, 1-18.
20. Laubrock, J., & Dubray, D. (2018). Computational Analysis and Visual Stylometry of Comics using Convolutional Neural Networks. *International Conference on Digital Health*.
21. Augereau, O., Iwata, M., & Kise, K. (2018). A survey of comics research in computer science. *J. Imaging, 4*, 87.
22. Hoffer, E., & Ailon, N. (2014). Deep Metric Learning Using Triplet Network. *International Workshop on Similarity-Based Pattern Recognition*.
23. Weng, L. (2021, May 31). *Contrastive Representation Learning*. Lil'Log. https://lilianweng.github.io/posts/2021-05-31-contrastive/.
24. Kelly, D. (2009). Methods for Evaluating Interactive Information Retrieval Systems with Users. *Found. Trends Inf. Retr., 3*, 1-224.
25. Tobii T60 Eye Tracker.(n.d.). https://www.staff.universiteitleiden.nl/binaries/content/assets/sociale-wetenschappen/faculteitsbureau/solo/research-support-website/equipment/manual-tobii-t120_23122019.pdf
26. Brenda Starr (2012, Dec 12). *Digital Comic Museum*. https://digitalcomicmuseum.com/index.php?dlid=18306
27. Tin-Tin (2023, Feb 2). *Wikipedia*. https://en.wikipedia.org/wiki/The_Broken_Ear
28. Asterix (2023, Feb 2). *NYTimes*. https://www.nytimes.com/2019/11/07/arts/asterix-new-translated-editions.html
29. Wonder Woman (2011, Jun 12). https://comicvine.gamespot.com/wonder-woman-39-the-trail-of-thrills/4000-108840/
30. Tomb Raider (2014). https://en.wikipedia.org/wiki/Tomb_Raider_(comics)

Thank you!

Questions?