Board 97: Work-in-Progress: TextCraft: Automated Resource Recommendation for Custom Textbook Creation

Xinyuan Fan, University of Toronto

Xinyuan (Elva) Fan is currently pursuing a Master's degree in Electrical & Computer Engineering at the University of Toronto, following her Honours Bachelor in Computer Science from the University of Waterloo. At the University of Toronto, she worked on a research project focusing on web crawler-driven automated textbook creation. She can be reached at elva.fan@mail.utoronto.ca or elvafan625@gmail.com.

Dr. Hamid S Timorabadi P.Eng., University of Toronto

Hamid Timorabadi received his B.Sc, M.A.Sc, and Ph.D. degrees in Electrical Engineering from the University of Toronto. He has worked as a project, design, and test engineer as well as a consultant to industry. His research interests include the applicati

Prof. Salma Emara, University of Toronto

Salma Emara is an Assistant Professor, Teaching Stream in the Department of Electrical and Computer Engineering at the University of Toronto. She received her B.Sc. in Electronics and Communications Engineering from the American University in Cairo in 2018, and her Ph.D. in Computer Engineering from the University of Toronto in 2022. Her Ph.D. research focuses on improving reinforcement learning algorithms to solve problems in computer networking algorithms. Currently, she is interested in building software-tools for programming education and pedagogical practices that build testing and debugging skills for beginner programmers.

Work In Progress: TextCraft: Automated Resource Recommendation for Custom Textbook Creation

Abstract

Traditional textbooks are essential for teaching but sometimes need more flexibility to adapt to various courses and disciplines. Instructors have to change course content and align resources. Course instructors may recommend multiple textbooks to cover all topics, which can overwhelm and cost students. TextCraft is a web platform leveraging a database of 3 million book chapters to recommend relevant textbooks and assist instructors in creating customized textbooks, addressing the challenges of traditional textbooks. Users, whether instructors providing their course outlines or students specifying the topics they wish to learn, input their requirements into the application. Based on this input, the application recommends the ten most relevant textbooks. Users can preview and select from recommended content, using it as a reference to create tailor-made textbooks. TextCraft was evaluated through technical optimization of its BM25 search algorithm and user feedback surveys, which showed its effectiveness in suggesting resources and usability. The optimal search configuration achieved the highest Precision@10 and MRR scores, while user responses showed high satisfaction with TextCraft's relevance and usability. With an average relevance rating of 8.13 and strong future-use intent, the system effectively aligned educational content with course needs. These results emphasize TextCraft's value as a new educational technology tool, driving continuous enhancements. TextCraft also responsibly manages copyrighted material, incorporating copyright information into its database to ensure legal compliance and facilitate legal use in education, promoting content accessibility and lawful resource use in teaching and research.

Introduction

The educational resource landscape is currently experiencing significant shifts driven by technological advancements. Despite these changes, traditional textbooks remain prevalent, but they increasingly struggle to meet modern courses' diverse and dynamic needs. Traditional textbooks, typically organized to comprehensively cover a course's syllabus, need more flexibility to address specific course requirements. This inflexibility often results in a disconnect between provided content and the evolving curriculum demands. Additionally, the static nature of traditional textbooks limits their ability to incorporate current information, diminishing their effectiveness in rapidly advancing academic disciplines.

Research has highlighted the limitations of traditional textbooks in comparison to electronic formats. Rockinson-Szapkiw et al. [1] show that students using e-textbooks exhibit higher engagement and enjoyment in their learning process due to their interactive features. The

integration of hyperlinks and multimedia in e-textbooks meets diverse learning preferences, enhancing the educational experience. TextCraft simplifies the task for educators by streamlining the selection of teaching materials and expanding the range of resources available. It enables educators to identify and suggest content that aligns with specific learning objectives. This capability aids in providing customized and dynamic learning experiences that promote greater student involvement. Studies by Clark et al. [2], Kang et al. [3], Jamali et al. [4], Robinson [5], Shepperd et al. [6], and Woody et al. [7] have explored the preferences of faculty and students towards different e-textbook formats. Additionally, the economic aspect is critical, as acquiring numerous textbooks can be financially challenging for students. Hilton III et al. [8] highlight how open educational resources could alleviate the financial strain of textbooks, especially when costs prevent textbook purchases. Jenkins and Sánchez [9] investigate textbook affordability as a social justice issue, exploring how textbook costs affect student performance, emphasizing race/ethnicity and financial need.

TextCraft transforms the process of creating textbooks using its inventive web application, effectively adapting to the changing demands of modern educational settings. This flexible platform overcomes traditional textbooks' limitations by empowering instructors and students to tailor educational materials to their specific needs. With access to an extensive database of over 3 million book chapters, TextCraft allows users to input their course outlines or specific learning topics, efficiently suggesting the most recommended textbooks. The platform enhances the process of providing recommendations for resources and aiding users in effectively managing a personalized book collection aligned with their educational goals.

Importantly, TextCraft conscientiously manages copyrighted material. It incorporates copyright information within each book's database structure, ensuring legal compliance and enhancing user awareness. Instructors benefit from specialized tools designed to legally utilize copyrighted materials in line with fair use and educational allowances. This approach not only adheres to copyright laws but also promotes the accessibility of content for educational purposes. Users are well-informed about the copyright status of the materials they access, which supports legal and effective resource use in teaching and research.

In this paper, we develop TextCraft and investigate its applicability in customizing textbooks and recommending relevant resources. The main research question is: "How effective is TextCraft in aligning customized textbooks with course outlines and recommending suitable educational materials?" We hypothesize that TextCraft not only simplifies the textbook selection process but also precisely suggests resources matching specific course content and objectives. This research is driven by the goal of improving educational resource management through modern technology. TextCraft is designed to overcome the limitations of traditional textbooks by offering textbook customization and resource recommendations. The motivation behind this work is to

assess the impact of TextCraft in modern educational settings and its contribution to the evolving field of educational technology. By examining TextCraft's capabilities, this study aims to highlight its role in improving teaching and learning experiences through customized educational resources and effective material recommendations.

Overview of TextCraft

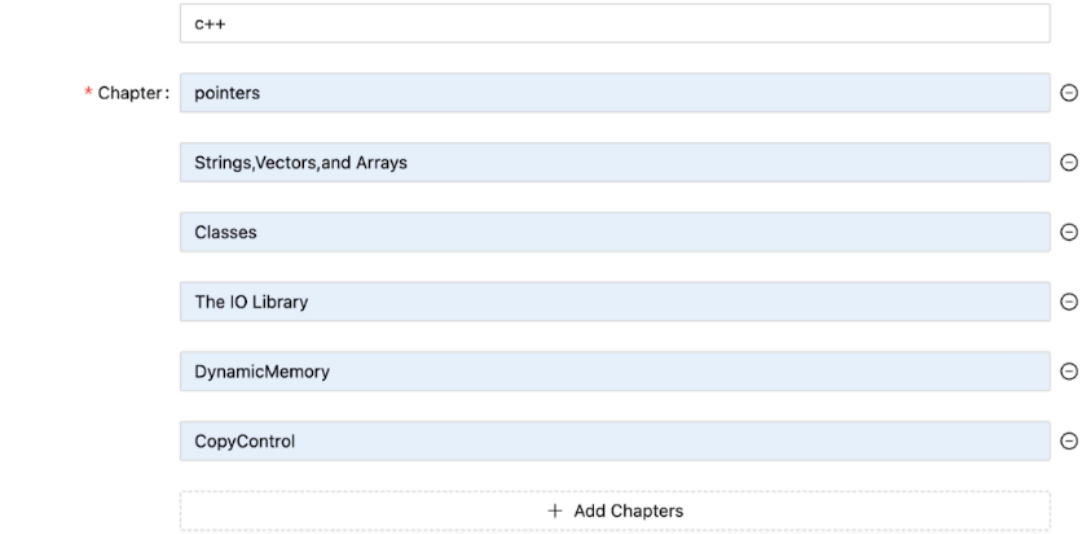
TextCraft's distinct advantages facilitate a more rapid book compilation process for instructors and offer recommendations from an expansive library of educational resources and advanced search functionalities. It provides instructors and students with the capability to obtain adapted educational textbook recommendations that align with their course syllabus and objectives. This customization provides significant advantages compared to traditional textbook models, which are more inflexible. It guarantees that the course content is relevant and matches the teaching objectives. In addition to its capabilities, TextCraft provides access to an extensive database from public libraries containing millions of book chapters. This extensive collection of resources ensures that instructors and students have a broad spectrum of current and diverse materials, facilitating the selection of the most appropriate material for their courses and learning goals.

Furthermore, TextCraft has an advanced search mechanism powered by a robust search engine that effectively recommends textbooks. This feature allows users to search by starting with broad topics and then narrowing the results by selecting more precise subtopics. This approach results in targeted and relevant search results, allowing users to efficiently locate the most relevant textbook materials for their specific educational needs. The process of receiving textbook recommendations through TextCraft, as shown in Fig. 1 and Fig. 2, starts when users log into the application and search for content pertinent to their courses, from broad subjects like "C++" to more specific subtopics such as "Pointers" or "Strings, Vectors, and Arrays."The system's advanced search mechanism retrieves a list of relevant books from the database. Users can interact with the search results by selecting books, viewing their contents, and previewing chapters for further review.

After selecting their preferred books, users are introduced to a comparison feature. This tool directly compares and evaluates chapters from different books, ensuring the choices align with educational objectives. Subsequently, an interactive module at the bottom of the screen helps users gather their selected chapters into a well-organized list. This module facilitates the organization of chapters and the inclusion of bibliographic information, such as titles, authors, and subjects. After completing the first stage of organization, TextCraft provides features that allow for the refinement of chosen chapters, simplifying the process of modifying content and integrating multimedia components. Once these modifications are made, the collection is saved and prepared for sharing or any additional changes in the future. TextCraft plays an essential part

in enhancing the learning and teaching environment by providing convenient access to educational materials.

1.Search



2. Review Search Results

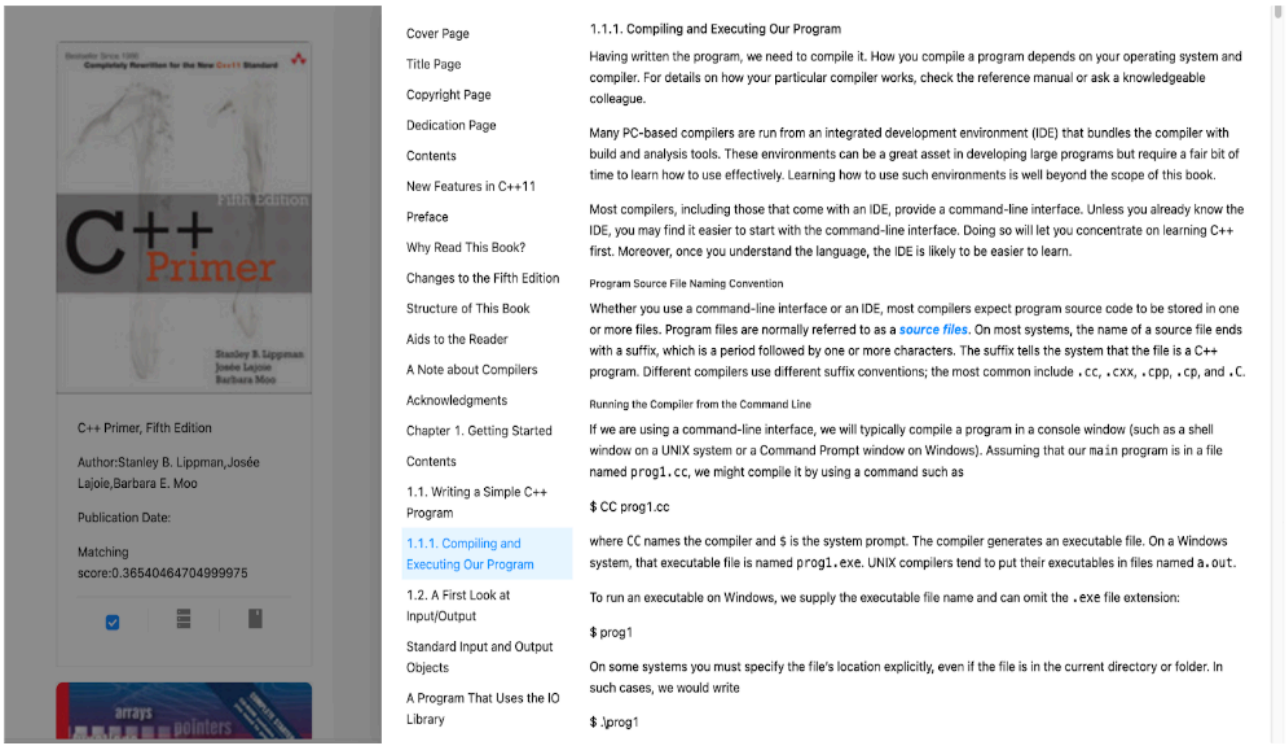


Figure 1: Procedure Part 1

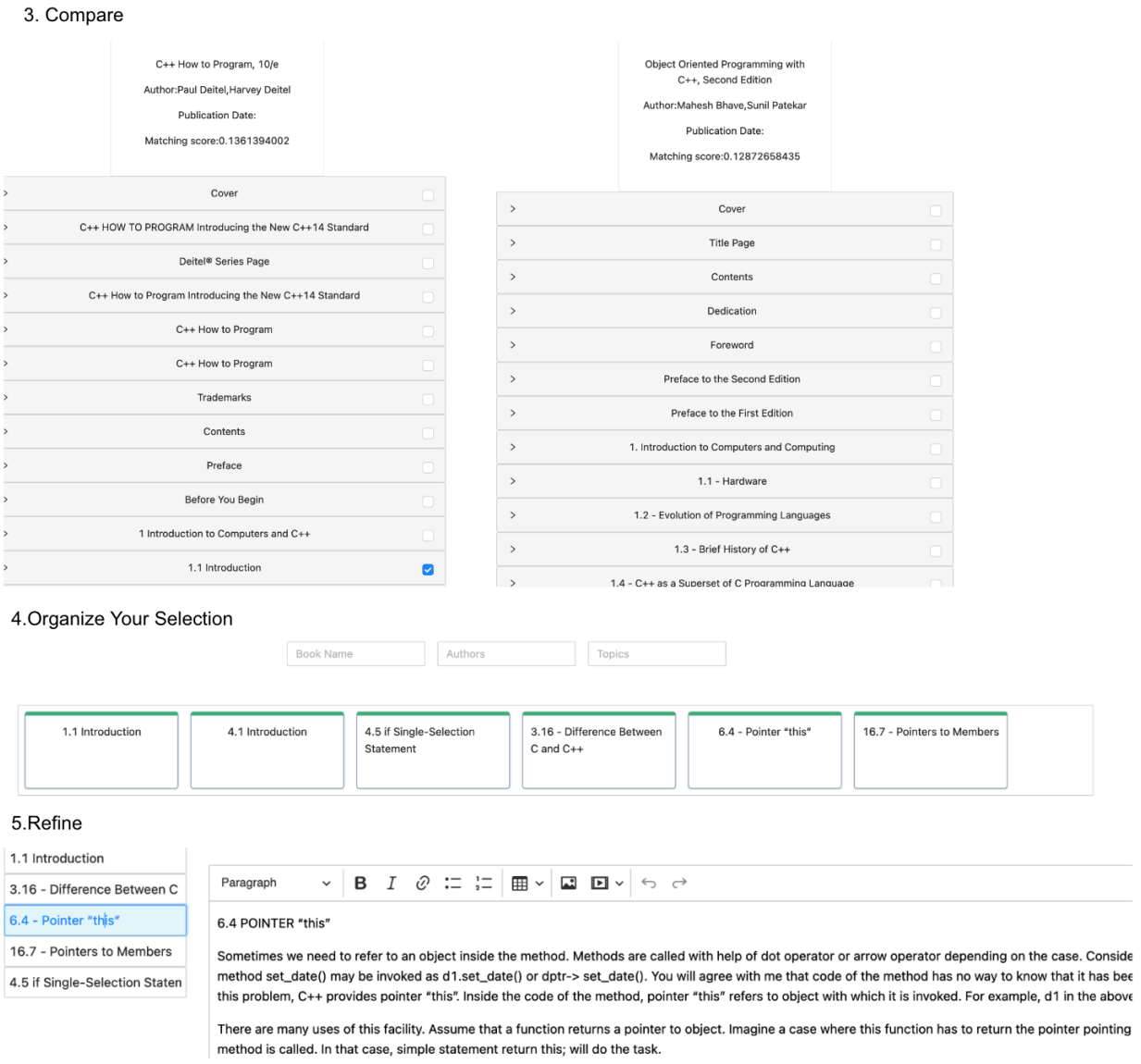


Figure 2: Procedure Part 2

Method

The evaluation of TextCraft involved a comprehensive approach, combining a technical assessment of the platform's search mechanism and a user feedback survey. The purpose of this dual-method strategy was to comprehensively evaluate the effectiveness of TextCraft in creating textbooks and recommending resources. This evaluation was conducted with the key research question, "How effective is TextCraft in textbook resource recommendation?" in consideration. This question was formulated to explore the utility and efficiency of TextCraft in meeting the diverse needs of educators and students within the educational technology field.

The technical assessment of TextCraft is designed to optimize its search mechanism using the BM25 algorithm essential to identifying appropriate educational resources. BM25 is a ranking function widely used in information retrieval to assess the relevance of documents against specific search queries[19]. It addresses term frequency (TF) and document length, which is important when calculating a document's relevance score. Unlike earlier models, BM25 introduces a novel approach to term frequency saturation, suggesting that term frequency's impact on relevance scores plateaus after a certain threshold, effectively reducing the influence of keyword stuffing. Additionally, it incorporates document length normalization and employs inverse document frequency (IDF) to highlight the significance of less common, more informative terms. As an enhancement of the TF-IDF model, BM25 adds parameters like k1 and b for fine-tuning term frequency saturation and document length normalization, ensuring a fair comparison across documents of varying lengths[20]. This mechanism ensures that search results are relevant and unbiased by document size, contributing to BM25's widespread use in search engines and information retrieval systems. Zhu et al. [10] highlight BM25's strength in term-frequency-based methods for dataset recommendation systems. Similarly, its use in digital libraries for text mining [11] demonstrates its capability to manage large textual datasets, affirming its suitability for TextCraft.

BM25's effectiveness hinges on the 'k1' and 'b' parameters. 'k1' affects term frequency normalization, influencing how terms are weighed in documents [12]. A higher k1 value allows the term frequency to influence the relevance score more significantly, potentially favoring documents with higher query term occurrences. Conversely, a lower k1 value reduces the impact of term frequency, which can be beneficial in avoiding the overvaluation of documents where a term appears repeatedly but may not necessarily be more relevant. Tuning k1 helps balance neglecting and overemphasizing the importance of term frequency in determining relevance. The 'b' parameter normalizes document length, impacting relevance calculations based on document size [12]. Setting b to 1 applies full normalization, making the document length significantly affect the relevance score, which can disadvantage longer documents or artificially boost shorter ones. A b value of 0 turns off length normalization, treating all documents as equal regardless of length, which can ignore the natural advantage of topical thoroughness typically found in longer documents. Adjusting b allows the algorithm to account for document length in a manner that aligns with the importance of length for a given search context or content type. Fine-tuning these parameters ensures the algorithm aligns with the complex nature of textbook content. Research indicates optimal 'b' values range from 0.3 to 0.9 [13], [14], [15], and 'k1' values are most effective between 0.5 and 2.0 [13], [14], [15]. These modifications are essential for refining BM25's performance in textbook searches. The assessment involved iterative tests with the BM25 algorithm's 'k1' parameter set from 0.6 to 2.0, increasing by 0.2, and 'b' from 0.3 to 0.9, increasing by 0.1. Testing used ten computer engineering topics, with each BM25 setting undergoing ten queries to simulate textbook search scenarios. We recorded the top ten results for

each query and configuration. The performance evaluation used two key metrics: Precision@10 [16] and Mean Reciprocal Rank (MRR) [17]. Precision@10 assesses how many of the top ten search results are relevant and crucial for quickly finding suitable textbooks. MRR measures how well the system ranks the first relevant result, emphasizing the importance of the top recommendation in saving time and resources. This method determined the most effective BM25 configuration for TextCraft's needs.

A comparative analysis using average Precision@10 and MRR identified the best relevant settings. The final BM25 configuration choice for TextCraft was based on these findings, enhancing its search accuracy and relevance.

In conjunction with the technical assessment, we conducted a survey targeting instructors and students to gather their feedback on TextCraft. The survey, "TextCraft: Automated Resource Recommendation for Custom Textbook Creation," was designed to capture both quantitative and qualitative data. Questions focused on the usability of TextCraft, the relevance of recommended resources, and overall satisfaction with the platform. The survey was distributed through academic networks to ensure a diverse and relevant pool of respondents.

Data analysis involved statistical tools for quantitative responses and thematic analysis for qualitative data. This approach allowed us to measure TextCraft's effectiveness quantitatively and qualitatively to understand user experiences, challenges, and suggested improvements. Comparative analysis was also conducted to contrast TextCraft with existing systems and identify areas for enhancement.

The survey also inquired about users' previous experiences with similar systems, allowing for comparative analysis and identification of TextCraft's unique features or potential improvements. The chapter concludes with a comprehensive summary of the findings, highlighting key insights derived from both instructors' and students' feedback, which are instrumental in guiding future enhancements of TextCraft.

Results

The experimental findings on the efficacy of various Elasticsearch BM25 configurations are visually represented through heat maps. Each cell within the heat maps is colored to indicate performance relative to Elasticsearch's default BM25 settings. Red hues signify configurations outperforming the default setting, while green suggests performance below the default. The default setting is represented by the mid-tone color and is characterized by a k1 value of 1.2 and a b value of 0.75, as per the default BM25 similarity configuration[18]. The vertical axis of the

heat maps is labelled with k1, denoting term frequency normalization, while the horizontal axis is labelled with b, indicating the extent of document length normalization. The heat maps contain cells representing a spectrum of k1 and b values, providing a comparative view of how each configuration fares against the others.

The Precision@10 heatmap in Table 1 illustrates the influence of different k1 and b parameters of BM25 on search precision. The heatmap uses darker shades of red to represent configurations that outperform the default BM25 settings. The most effective setting, marked by the deepest red, is found at k1=2.0 and b=0.8, where it delivers a peak precision score of 73. This optimal configuration underscores the significance of adjusting both term frequency normalization and document length normalization to enhance the algorithm's precision in identifying the top ten most relevant textbook materials. Such a combination of k1 and b values optimally balances term presence with document context, markedly improving the relevance of search results for TextCraft users. This finding highlights the critical role of fine-tuning k1 and b within the BM25 algorithm to significantly refine search outcomes, ensuring the recommended textbooks align closely with user queries.

Table 1: Heat Map of Percision@10 While Varying k1 and b

		b						
		0.3	0.4	0.5	0.6	0.7	0.8	0.9
K1	0.6	52	51	52	52	56	56	56
	0.8	52	55	57	53	57	58	58
	1	60	61	59	63	62	67	69
	1.2	56	60	63	64	62	67	68
	1.4	61	62	63	65	66	69	68
	1.6	59	62	63	66	67	72	70
	1.8	60	67	64	67	69	70	69
	2	62	66	66	66	67	73	72

The MRR heatmap analysis in Table 2 demonstrates that the most effective BM25 parameter configurations for ranking the most relevant document are identified by red cells, specifically when k1 values range between 1.4 and 1.7, with b values between 0.6 and 0.8. This observation indicates that fine-tuning k1 slightly higher and adjusting b slightly lower significantly enhances the algorithm's efficiency in ranking the most pertinent document first. The peak performance for MRR, achieving a score of 0.52083, is achieved with a k1 value of 2.0 and a b value of 0.7. This specific combination is highlighted as providing an optimal balance for ranking accuracy, effectively ensuring the most relevant document is prioritized in search results. Such fine-tuning underscores the critical importance of adjusting both term frequency normalization and

document length normalization within the BM25 framework to optimize search result relevance and precision.

Table 2: Heat Map of MRR While Varying k1 and b

		b						
		0.3	0.4	0.5	0.6	0.7	0.8	0.9
K1	0.6	0.25833	0.37833	0.40040	0.40278	0.40512	0.39429	0.41190
	0.8	0.35524	0.41583	0.41524	0.42262	0.43095	0.40690	0.41524
	1	0.33524	0.38262	0.46524	0.48429	0.48095	0.44524	0.44286
	1.2	0.34940	0.40750	0.47262	0.46762	0.48095	0.44286	0.43452
	1.4	0.32440	0.35190	0.42095	0.46667	0.43762	0.44774	0.43512
	1.6	0.35512	0.39440	0.46857	0.46167	0.45333	0.42595	0.43274
	1.8	0.33762	0.39302	0.45524	0.45667	0.45357	0.45357	0.50139
	2	0.31762	0.43107	0.38762	0.42833	0.52083	0.50040	0.50139

The analysis highlights the necessity of finely tuning the BM25 algorithm's k1 and b parameters to optimize search relevance and efficiency for TextCraft, with an ideal configuration identified at k1=2.0 and b=0.75. This specific adjustment has proven to yield the highest Precision@10 and Mean Reciprocal Rank (MRR), significantly enhancing the relevance and ranking of search results. Adopting this configuration for the survey lays a solid foundation for refining Elasticsearch configurations, which is crucial for information retrieval in educational contexts where search result precision is the highest priority. The exploration of these heat maps reveals the delicate balance required between term frequency and document length normalization within BM25, emphasizing the role of precise parameter adjustments in boosting the search algorithm's effectiveness. This strategy improves user experience by offering more relevant textbook suggestions and demonstrates the value of empirical testing in continuously improving educational technology tools.

The user survey on TextCraft yielded informative feedback on its proficiency in customizing textbooks and recommending resources, mainly collecting the views of instructors and students. The results demonstrated impressive satisfaction with the platform's ability to accurately align educational content with course outlines and suggest relevant textbooks. This indicates TextCraft's strong performance in meeting the specific needs of educational courses and improving resource selection. An average rating of 8.13 out of 10 for the relevance of books recommended in Fig. 3 by the system indicates a strong alignment between the system's suggestions and the users' specified topics. As shown in Figure 4, a significant 80% of survey participants agreed that the first book recommended by TextCraft closely aligned with their

search criteria, highlighting the system's adeptness at identifying key resources. However, a small fraction of dissenting opinions suggest an opportunity to further refine the recommendation algorithms to ensure top-ranking accuracy.

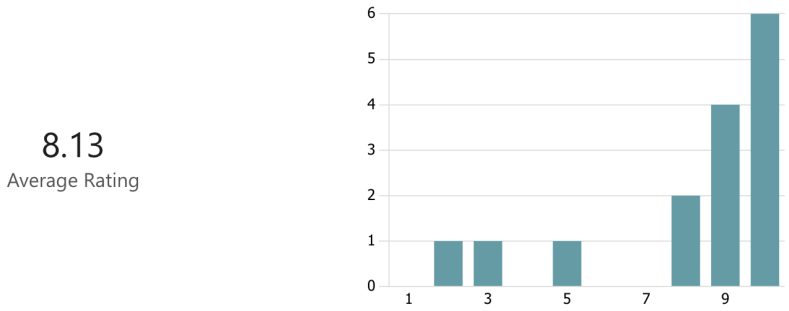


Figure 3: Relevance of Books Recommended by System

The survey demonstrated a strong intention for future use among participants, with approximately 93% indicating their plans to use TextCraft to search course materials. This reflects the application's perceived utility and potential impact on academic resource management. User feedback on the system's ease of use was overwhelmingly favorable, emphasizing TextCraft's user-friendly interface and streamlined process for textbook customization. Feedback on the system's effectiveness in meeting the need for finding relevant course materials was also largely positive, indicating that TextCraft successfully addresses one of the primary challenges faced by instructors and students. Users appreciated various features of TextCraft, including its user-friendly interface, the relevance of book recommendations, and the innovative approach to customizable textbook creation. The ability to compare and merge different textbook chapters stood out as a unique benefit, highlighting TextCraft's role in facilitating a more dynamic and customized learning environment.

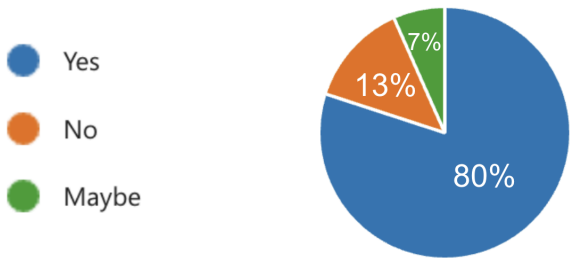


Figure 4: Is the First Book the Best Match

Furthermore, the majority of respondents had not previously used similar systems for textbook recommendations or academic resource customization, suggesting that TextCraft occupies a

special spot in the market. However, this also points to the potential for raising awareness and encouraging broader adoption among educators and students. Suggestions for improvement emphasized the need for enhancements in UI design and additional functionalities such as text highlighting and interactive features. These recommendations provide a roadmap for future development, ensuring that TextCraft continues to evolve in response to user feedback.

The survey results underscore the significance of TextCraft in the field of educational technology, illustrating its capacity to improve the approach to textbook customization and resource recommendation. By addressing the feedback and suggestions identified, TextCraft can further its contribution to the field of education, offering a solution that not only aligns educational content with course requirements but also improves the teaching and learning experience.

Discussion

The findings from our evaluation of TextCraft reveal an exciting potential for educational resource management and curriculum development. Notably, the optimization of the BM25 search algorithm demonstrates a substantial improvement in the relevance and precision of textbook recommendations. This outcome highlights the importance of fine-tuning search parameters to align with educators' and learners' specific needs. The combination of k1=2.0 and b=0.75, which results in the greatest Precision@10 and Mean Reciprocal Rank (MRR) scores, demonstrates the technical expertise required for customizing search features in educational environments. This precision in search results directly contributes to a more efficient and relevant discovery process for users, fundamentally enhancing the educational resource selection and customization process.

Furthermore, user feedback emphasizes the significant impact of TextCraft in simplifying the process of creating books and expanding the range of resources available to instructors. This aspect is particularly crucial in overcoming the limitations of traditional textbooks, which often need help to adapt to the diverse and evolving requirements of modern education. By providing a platform that not only speeds up the resource-gathering process but also ensures a high degree of relevance and adaptation, TextCraft addresses a challenge in education. It empowers educators to create learning experiences that are more aligned with their teaching objectives and responsive to students' learning needs.

Additionally, the study's results emphasize the importance of user experience in the development of educational technologies. The suggestions for UI improvements and the introduction of new functionalities, such as offline access and interactive features, point towards a continuous need for innovation in how educational content is delivered and engaged. These insights are

invaluable for guiding future enhancements of TextCraft, ensuring that the platform not only meets but exceeds its users' expectations.

This discussion, rooted in our study's outcomes, not only highlights TextCraft's contributions to educational technology but also sets the stage for further research and development in this field. By continuously refining its algorithms and user interface, TextCraft can support educators and learners in achieving their educational goals with greater efficiency and satisfaction.

Conclusion

The evaluation of TextCraft highlights its role in developing educational resources, showcasing a web application that improves textbook creation and recommendation. TextCraft overcomes the constraints of traditional textbooks by allowing instructors and students to customize texts to fit course outlines and learning objectives. Survey and technical assessments confirm TextCraft's effectiveness in aligning content with course needs and providing accurate textbook recommendations, demonstrating its potential as an educational tool. Despite its success, feedback indicates areas for improvement, such as enhancing user interfaces and expanding the resource database. These insights guide future development, ensuring TextCraft's constant development in response to user needs. The study suggests possibilities for further development, including advanced algorithms, mobile application development, and collaborative features, underscoring TextCraft's significant contribution to educational resource management.

In summary, TextCraft is an important solution for current educational challenges, improving the quality of teaching and learning experiences. The introduction of this technology represents a significant development in the field of education, providing an academic setting that is more flexible, inclusive, and efficient.

References

[1] A. J. Rockinson-Szapkiw, J. Courduff, K. Carter, et al., "Electronic versus traditional print textbooks: A comparison study on the influence of university students' learning," 2013.

[2] D. T. Clark, S. P. Goodwin, T. Samuelson, C. Coker, "A qualitative assessment of the Kindle e-Book reader: results from initial focus groups," Displays, 30.2 (2009): 49-52.

[3] Y. Y. Kang, M. J. Wang, R. Lin, "Usability evaluation of e-books," Displays, 30.2 (2009): 49-52.

[4] H. R. Jamali, D. Nicholas, I. Rowlands, "Scholarly e-books: the views of 16,000 academics," New Information Perspectives, 61 (1) (2010), pp. 33-47.

[5] S. Robinson, "Student use of a free online textbook," Academy of Educational Leadership Journal, 15 (3) (2011), pp. 1-10.

[6] J. A. Shepperd, J. L. Grace, E. J. Koch, "Evaluating the electronic textbook: is it time to dispense with the paper text?" Teaching of Psychology, 35 (1) (2008), pp. 2-5.

[7] W. Woody, D. Daniel, C. Baker, "E-books or textbooks: students prefer textbooks," Computers & Education, 55 (3) (2010), pp. 945-948.

[8] J. L. Hilton III, T. J. Robinson, D. Wiley, et al., "Cost-savings achieved in two semesters through the adoption of open educational resources," 2014.

[9] J. J. Jenkins, L. A. Sánchez, et al., "Textbook broke: Textbook affordability as a social justice issue," 2020.

[10] J. Zhu, B. G. Patra, and A. Yaseen, “Recommender system of scholarly papers using public datasets,” *AMIA ... Annual Symposium proceedings*, vol. 2021, pp. 672–679, 2021.

[11] G. A. Tinega, Prof. W. Mwangi, and Dr. R. Rimiru, “Text Mining in Digital Libraries using OKAPI BM25 Model,” *International Journal of Computer Applications Technology and Research*, vol. 7, no. 10, pp. 398–406, 2018, doi: 10.7753/IJCATR0710.1003.

[12] I. C. Mogotsi, “Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze: Introduction to information retrieval: Cambridge University Press, Cambridge, England, 2008, 482 pp, ISBN: 978-0-521-86571-5,” *Information Retrieval*, vol. 13, no. 2, pp. 192–195, 2010, doi: 10.1007/s10791-009-9115-y.

[13] A. Lipani, M. Lupu, A. Hanbury, and A. Aizawa, "Verboseness Fission for BM25 Document Length Normalization," Association for Computing Machinery, 2015.

[14] M. Taylor, H. Zaragoza, N. Craswell, S. Robertson, and C. Burges, "Optimisation methods for ranking functions with multiple parameters," Association for Computing Machinery, 2006.

[15] A. Trotman, A. Puurula, and B. Burgess, "Improvements to BM25 and Language Models Examined," Association for Computing Machinery, 2014.

[16] K. Jarvelin and J. Kekalainen, “IR evaluation methods for retrieving highly relevant documents,” SIGIR forum, pp. 41–48, 2000.

[17] D. R. Radev, H. Qi, H. Wu, and W. Fan, “Evaluating web-based question answering systems,” in Proceedings of the 3rd International Conference on Language Resources and Evaluation, LREC 2002, 2002, pp. 1153–1156.

[18] J. Lin and W. J. John, “PubMed related articles: A probabilistic topic-based model for content similarity,” *BMC bioinformatics*, vol. 8, no. 1, pp. 423–423, 2007, doi: 10.1186/1471-2105-8-423.

[19] C. D. Manning, Prabhakar. Raghavan, and Hinrich. Schütze, *Introduction to information retrieval*. Cambridge ; Cambridge University Press, 2008.

[20] A. I. Kadhim, “Term Weighting for Feature Extraction on Twitter: A Comparison Between BM25 and TF-IDF,” in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 2019, pp. 124–128. doi: 10.1109/ICOASE.2019.8723825.