Advancing the Art of Literature Review Visualization: A Journey Towards Enhanced Clarity and Insight

Mohammad Akif[*], Daud Khan[†], Waqas Ahmad[‡], Muhammad Idrees[§], Shahab Ul Islam[**]

Abstract

The Systematic Literature Review (SLR) is a type of academic paper, especially written on a specific research area. The SLR is the most important paper for researchers to get knowledge of certain area for further investigation. In SLR researchers normally dig and find the area of interest. The time and effort to collect, write, and report the statistics or figures found during this process is a cumbersome job. The same case is with the reader for reading and understanding other Scientific Literature Reviews (SLRs). It takes too much time to understand and read the SLR and conceptualize the statistics, figures, and other related numerals presentations. In the modern era of computing, researchers and ordinary readers want to visualize the literature in short, concrete, and meaningful ways. In this connection, this study attempts to visualize the literature in a specific domain, based on keywords and profile information. For this purpose, a mobile-based application is designed in this study to scrapes the data of a specific area based on a keyword. The user needs to give a keyword about any research area, i.e., Image Processing, etc. The application then scrapes all the parameter data from Google Scholar in a short time, and then some important items (based on keywords supplied) are visualized in graphs, charts, and numbers. Moreover, the scrapped data is downloadable in an Excel file for further analysis. The application may help scholars, researchers, and academicians to find and visualize the literature reviews more easily and efficiently.

Keywords*:* Systematic Literature Review; Mobile-based Application; Image Processing; Visualization.

Introduction

The Systematic Literature Review (SLR) is a crucial academic article used by researchers for digging and finding the area of interest. It contributes to the knowledge in the specific domain and requires significant time, effort, and energy to collect, organize, and present findings. It takes too much time to understand and read the SLR, and it

[*]Department of Computer Science, City University of Science and Technology, Peshawar 25000, Pakistan, akifkhan6556@gmail.com

[†]Department of Computer Science, Iqra National University, Peshawar 25100, Pakistan, daud.khan@inu.edu.pk

[‡]Corresponding Author: Department of Computer Science, Iqra National University, Peshawar 25100, Pakistan, waqas.ahmad@inu.edu.pk

[§]Department of Computer Science, City University of Science and Technology, Peshawar 25000, Pakistan, idreeskhanapple0334@gmail.com

[**]Department of Engineering, Parthenope University of Naples, Naples 80138, Italy, shahabul.islam001@studenti.uniparthenope.it

takes more time to conceptualize the statistics, figures, and other related numeric (Martins et al., 2023). Scientific Literature Reviews (SLRs) are searched by giving keywords in search engine such as Google Scholar to gather different research articles. Filters can be used limit the articles to the specific needs of the researchers (Valverde-Berrocoso et al., 2020). It may include or exclude an article by the title or internal content. Published research articles can also be selected based on citation and versions (Asadabadi, et al., 2021). The graphs and charts are normally used to make the SLR more visual.

Performing all these tasks manually take too much time for researchers, so there is a dire need to design an application to automate this whole process. This study aims to enhance the efficiency of writing and comprehension of Self-Portraits SLRs by utilizing real-time feedback and visual aids, ensuring accurate and insightful results for academics and the public. An Android and iOS based application is designed in this study in order to converts SLRs into Visual Literature Reviews (VLRs) by dynamically extracting data from Google Scholar, and presenting it visually, downloadable in Excel, and tabular (Booth et al., 2020). The application offers advanced data visualization through graphs, allowing users to filter data based on publication year, country, and web page scraping, enabling detailed graphs or tabular Excel file extraction.

The proposed application streamline the process of conducting and understanding SLRs by providing a user-friendly platform that enhances the accessibility and interpretability of academic research through dynamic, real-time visualizations, and comprehensive data representation. As a result, the program greatly advances our understanding of a variety of research topics by simplifying and making complex academic material more comprehensible.

Literature Review

When the researcher does the SLR it gathers, reads, and understands research papers related to the area where the researcher does research. Researchers search the SLR by giving keywords and then Google Scholar or by other ways all the research papers are gathered. In the researcher filters all the papers which papers should be included to read and which papers to be excluded (Valverde-Berrocoso et al., 2020). It may be included or excluded by the title of the paper or maybe the internal content of the paper because every paper should not be included in SLR. Paper selection depends on the area of research. After selecting the papers researcher reads and understands all the papers by using this way it writes his research, Papers are also selected by the citation of the paper how many people cited this paper as well as versions of the paper where the same

paper is published (Asadabadi, et al., 2021). The graphs and charts are made to make the SLR more visual. The graphs are made from different parameters of the research papers such as citation of paper, versions of papers, word count graph, and also the specific year the paper is published. The researcher does this whole process manually, so it takes too much time researchers, we think why not make an application that makes this work easy for the researcher, in which they do the SLR in minutes instead of taking a year or more? SLR paper publication is also tough work, it cannot be easily published by a publisher to make that simple and easy we made an application to save time and budget of the researchers. This application helps in the publication of SLR, it gives you the same result by giving keywords by scraping HTML tags techniques, the same result as Google Scholar. It searches and fetches the same papers that Google Scholar search for you. The system we made, researchers in their research SLR can use this System. Researchers have gotten the SLR by using specific keywords about the research area. The data about SLR has been scraped from Google Scholar and used by the researchers for their purposes.

The SLR aimed to explore key aspects of e-learning research, including topics, theories, modalities, and methodologies. Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol, 248 articles from top-tier journals in Educational Technology are analyzed. Three main nodes are identified: online students, online teachers, and curriculum-interactive learning environments. Massive Open Online Courses (MOOC) emerged as the most researched e-learning modality, with the Community of Inquiry and the Technological Acceptance Model being prominent theories (Neja et al., 2021).

Grant et al. (2018) explores different visualization techniques that can be applied to literature reviews. It discusses how visual representations can help researchers and readers grasp the evolution of research topics over time, understand key concepts, and identify gaps in existing knowledge. Jelier et al. (2021) focuses on the integration of text mining and visualization techniques specifically in the field of bioinformatics. It illustrates how these methods can enhance literature reviews by automating the extraction of relevant information and presenting it in a visually intuitive manner. Peterson et al. (2022) provides a structured taxonomy of visualization techniques for SLRs. It categorizes visualization approaches based on their purposes and effectiveness in supporting the review process. The paper also suggests future research directions to further advance the field. Lopes et al. (2020) presents a case study demonstrating how visualization techniques can reveal the evolution and structure of a specific research area, such as information retrieval. It

showcases practical applications of visualization in analyzing the growth of knowledge and identifying seminal works within a field. Shi et al. (2020) explores visual analytics approaches tailored for exploring large-scale scholarly data sets. It discusses techniques for visualizing complex relationships among research papers, authors, and citations, aiming to uncover hidden patterns and insights that traditional text-based analyses might overlook. Silva et al. (2021) provides a comprehensive review of visualization techniques used in SLRs within software engineering. It identifies common visualization methods and evaluates their effectiveness in synthesizing research findings and identifying trends in the field.

Existing studies reveal gaps in research, requiring innovative solutions. Automating SLRs, addressing institutional barriers, and integrating Cost Breakdown Structures can improve efficiency and outcomes. Engaging stakeholders and integrating Cost-benefit Analysis (CBA) of Disaster Risk Reduction (DRR) policies can mitigate communication issues. A SLR using PRISMA protocol can fill e-learning research gaps, stratification improves project planning reliability, and continuous software engineering in China can provide insights. Computer-mediated communication enhances second language acquisition, and bibliometric assessments improve software testing. Building Information Modeling (BIM), machine learning, and large language models can improve construction project risk management, flood control, software design accuracy, and software engineering tasks. These solutions can advance research and practice in these fields.

Methodology

The methodology for developing the mobile application to visualize literature reviews is presented. This includes a comprehensive description of the requirements analysis, participants, research instruments, and data collection processes. Figure 1 shows the steps followed in iterative development of the application.

Research Design

The study employed both descriptive and developmental research methodologies. The developmental approach guided the creation of both the hardware and software components of the system (Peterson, 2022). Descriptive research is used to evaluate the system's acceptability in terms of functionality, reliability, usability, efficiency, maintainability, and portability, aiming to advance the field of literature review visualization (Jelier, 2021). The system development followed an Iterative Development model within the Software Development Life Cycle (SDLC) framework.

Respondents of the Study

The system's acceptability is assessed through an evaluation conducted by a researcher with a group of Information Technology (IT) experts. These experts are selected using purposive sampling and included the researcher's supervisor as well as professionals from various IT roles: project manager, IT consultant, senior programmer, junior programmer, web developer, and full-stack developer.

Research Instrument

To define the project's business requirements, the researcher held brief meetings with key stakeholders of the VLR project, including professors conducting current research. These discussions helped identify the project's actual needs and business objectives.

Software Designing and Development

The study adopted an iterative development methodology for software creation. As described by Powell-Morse (2016), this approach involves starting with a basic version of the system and gradually enhancing it through successive iterations, adding complexity and features until the final product is complete.

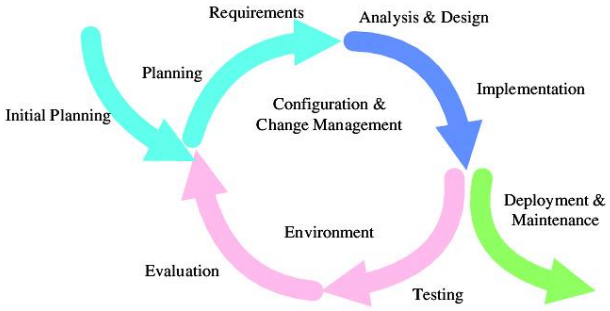

Figure 1: Iterative Development of the application.

The iterative development model is well-suited for today's rapidly changing environment due to its effectiveness in managing development risks. This approach involves creating the software in stages, with each iteration producing a functional product that serves as a foundation for the next, more advanced version. This cycle of refinement leads to continuous improvement in functionality and user interface.

Planning & Requirements

In this phase, the researchers began gathering data by reviewing existing literature and systems related to literature review visualization.

Potential functionalities and features are identified for inclusion in the system.

Analysis & Design

Following the completion of the planning phase, use case, system, and data flow diagrams are created to establish the system's foundation in terms of process flow and architecture. Additionally, a preliminary system design is developed to visualize the system's appearance and layout. These artifacts facilitated a smooth and efficient development process for advancing the art of literature review visualization.

Implementation

Once planning and analysis are completed, the team initiated the database implementation and coding phase for the literature review visualization system. Development adhered to a three-tier architecture, with system flow guided by the data flow diagram. Upon completing the system, it is uploaded to Google Play for user testing. In simpler words, Table 1 explains what the system does. It shows us what are the features of the system and whether the system meets the user's goals. The system must have the following functionalities mentioned in Table 1.

Table 1: System Functionalities

Requirement ID	Requirement Name	Requirement Description
FR-1	User registration	Users register themselves in the application
FR-2	Account verification	User verify itself using the email, clicking and verifying on link sent from the application during registration
FR-3	User login	Users login in the application
FR-4	Scraping data	Users scrap the data
FR-5	Download scraped data	Users download the scraped data in excel file
FR-6	Showing data on the application screen	Users show data with in-app in list form
FR-7	User registration	Users register themselves in the application
FR-8	FR-1 update profile	Users update the profile

Use cases can be used to provide a list of goals and this list can be used to establish the cost and complexity of the system. Each use case can be used to represent a sequence of simple steps, beginning with a user's goal and ending when that goal is fulfilled.

The Hierarchical Input Process Output (HIPO) technique is often used to plan or document a structured program. A variety of tools, including pseudocode and structured English, can be used to describe processes on an input-process output (IPO) chart. System flowcharting

symbols are sometimes used to identify physical input, output, and storage devices on an input process output (IPO) chart.

Figure 2 depicts the interactions between the USER and APP (application) for data scraping tasks. It features two actors: the USER, who initiates the process, and the APP, which facilitates it. The diagram illustrates several use cases: the USER searches for a keyword, prompting the APP to collect data based on the keyword, scrape the data, and then display the scraped data. The APP provides multiple options for presenting the data, including visual form, tabular form, and a downloadable Excel file. This diagram effectively outlines the steps and interactions involved in the data scraping module, highlighting the functionality provided by the APP to meet the user's data scraping needs.

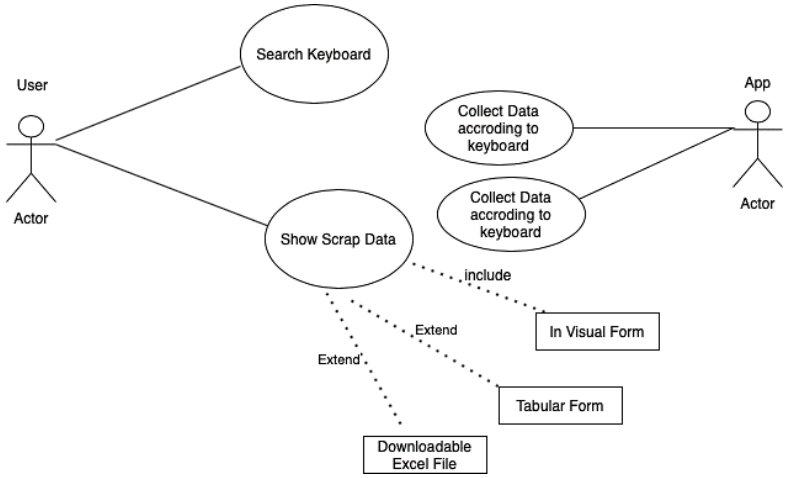

Figure 2: Data Scraping Module

Testing and Evaluation

Once the current iteration is coded and implemented, thorough testing of the system's frontend and backend functionalities is conducted. This process aimed to identify and rectify any bugs. Furthermore, this phase provided valuable insights for potential system improvements, especially in the dynamic search engine and collaborative researching modules. Upon completing all previous stages, domain experts are invited to evaluate the system. Data collected from their survey responses is analyzed using mean and standard deviation. This evaluation marks the end of one iteration in the iterative development model. Insights from this evaluation, along with the latest system version, are fed back into the planning and requirements phase, initiating a new development cycle.

Results and Discussion

The many components of the suggested system are depicted in Figure 3. Using a web scraping technology, users interact with the mobile application to manage their accounts and extract data from Google Scholar based on given keywords. Administrators keep an eye on every user and make sure the program runs well, handling any problems that may come up. Wide accessibility is made possible by the application's design, which is compatible with both the iPhone Operating System (iOS) and Android operating systems. The method uses information scraping from Google Scholar to obtain data.

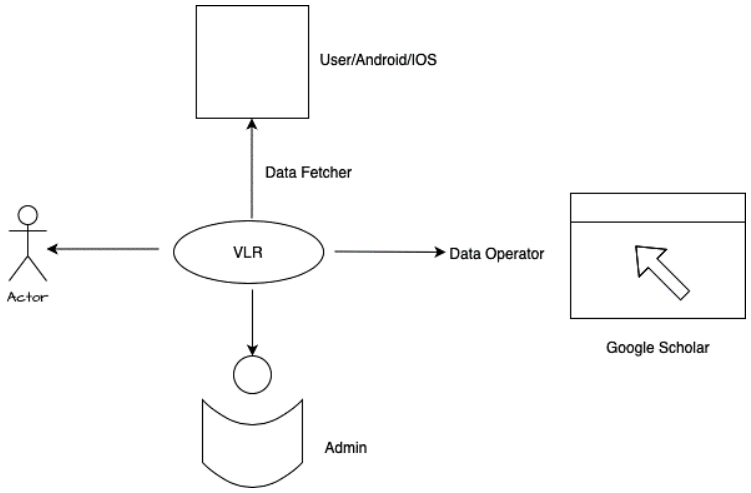

Figure 3: System Diagram

Scraped Data

Following the execution of a keyword search, the system retrieves and displays all relevant scraped data in a textual format. This process is visually represented in Figure 4, which illustrates how the data is organized and presented to the user.

Year-wise Publication Graph

The research papers are presented in a graph that illustrates the annual data extracted from Google Scholar. As depicted in Figure 5, this graph visualizes the frequency or quantity of research papers published each year, providing a clear representation of the trends and patterns over time. This comprehensive visualization offers valuable insights into the research landscape, highlighting the volume and distribution of academic publications sourced from a widely respected database.

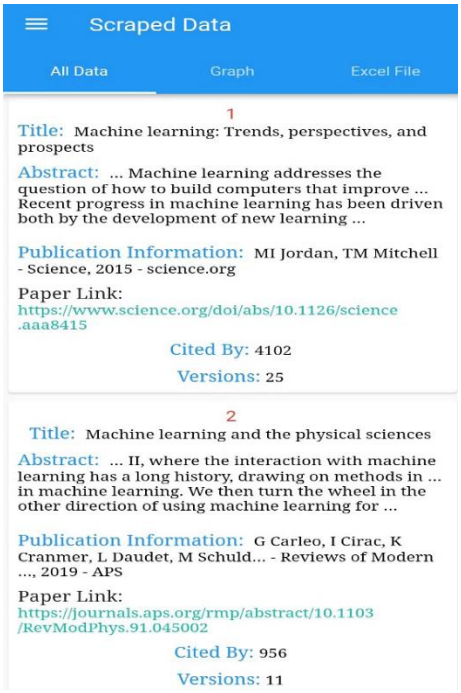

Figure 4: Scraped Data.

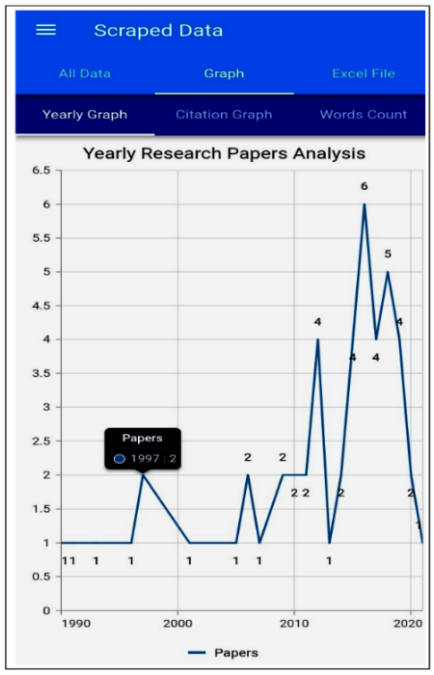

Figure 5: Year-wise Publication Graph.

Citation Graph

The graph visually represents research papers arranged according to the number of citations they have received. This organization helps to showcase the impact and influence of each paper within its academic context. The data used for this graph is sourced from Google Scholar, a widely utilized platform for academic literature, and the specific graph illustrating this information is referenced in the Figure 6.

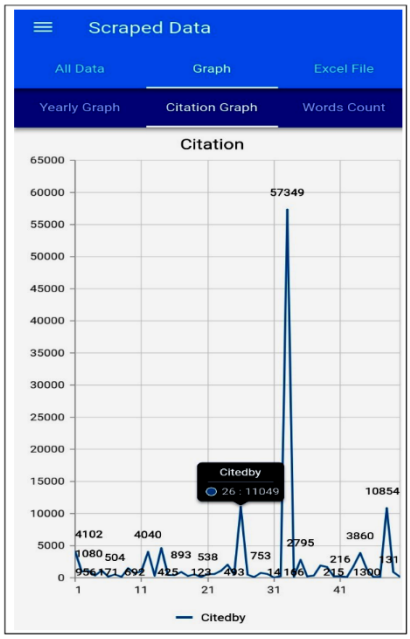

Figure 6: Citation Graph.

Word Count Graph

The figure below describes a graphical representation that visually presents the number of research papers and the respective word counts extracted from Google Scholar. This graph provides a clear overview of the distribution of research papers based on the length of their content. Google Scholar serves as the source for the data used in compiling this graph, emphasizing its credibility within academic research. Figure 7 shows the text serves to visually reinforce the information presented in the sentence.

Architecture Design

Figure 8 shows the structure and operation of the proposed application. This application can be accessible by both Android and iOS operating systems. It can be deployed on the Play Store and Apps Store to

be used by the researcher for both operating systems. The user credential data and personal information are stored on Firebase. The data is scraped from Google Scholar and that data is then managed in different forms in the application.

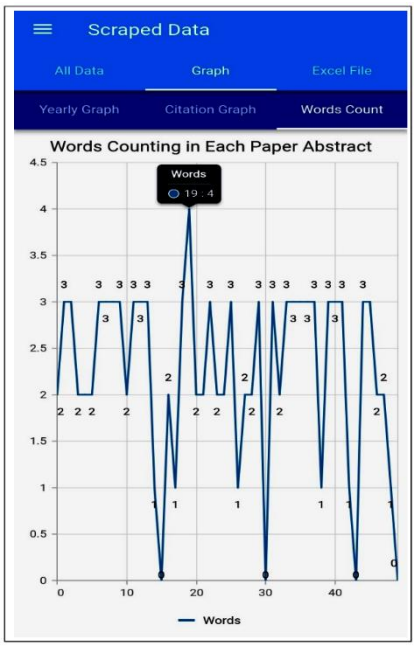

Figure 7: Word Count Graph.

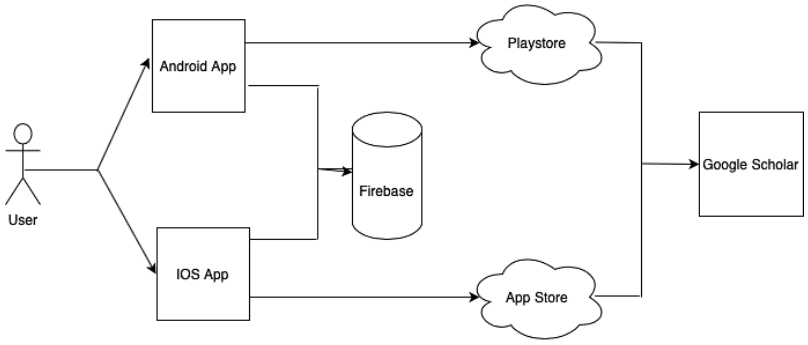

Figure 8: Architecture Diagram

Conclusion

Moving Towards Improved Visualization A mechanism called Literature Review is created to make it easier for researchers to acquire information. The SLR is a valuable tool for advancing knowledge in particular fields. However, gathering, combining, and reporting statistical

results during an SLR requires a significant amount of time, energy, and work. Similar to this, readers may need some time to understand and analyze scholarly articles. To improve comprehension, there is an increasing need in the modern day for real-time technological feedback and visualization tools like colors, graphs, and charts. The goal of this project is to design a system for visualizing SLRs utilizing a variety of current technologies to overcome these issues. To serve both academics and the general public, the produced visualizations strive to be accurate, timely, and insightful. It can be intimidating for researchers looking for a variety of articles in their field to observe current research trends. We have created an application for the Android and iOS platforms that allows users to convert SLRs into VLRs, thereby streamlining this procedure. Based on keywords entered by the user, such as "Image Processing," the application retrieves information pertinent to a particular field of study from databases like Google Scholar. The information is then displayed and accessible within the application as downloadable Excel files and tabular formats. Additionally, the program learns the user's tastes on its own and suggests papers depending on the user's search history. The goal of this customized recommendation system is to draw attention to certain user profiles' areas of interest.

References

Armitt, G., Slack, F., Green, S., & Beer, M. (2023, January). The development of deep learning during a synchronous collaborative on-line course. In Computer Support for Collaborative Learning (pp. 151-159). Routledge.

Asadabadi, M. R., & Zwikael, O. (2021). Integrating risk into estimations of project activities' time and cost: A stratified approach. *European journal of operational research*, *291*(2), 482-490.

Belt, E. S., & Lowenthal, P. R. (2023). Synchronous video-based communication and online learning: An exploration of instructors' perceptions and experiences. *Education and Information Technologies*, *28*(5), 4941-4964.

Cooper, C., Booth, A., Husk, K., Lovell, R., Frost, J., Schauberger, U., ... & Garside, R. (2024). A Tailored Approach: A model for literature searching in complex systematic reviews. *Journal of Information Science*, *50*(4), 1030-1062.

Hamer, S., Quesada-López, C., & Jenkins, M. How Have We Researched Developers' Contributions in Software Engineering? A Systematic Mapping Study. *A Systematic Mapping Study*.

Holanda, M., & Da Silva, D. (2021). Latin American women and computer science: A systematic literature mapping. *IEEE Transactions on Education*, *65*(3), 356-372.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering–a systematic literature review. *Information and software technology*, *51*(1), 7-15.

Kuss, D. J., Kristensen, A. M., & Lopez-Fernandez, O. (2021). Internet addictions outside of Europe: A systematic literature review. *Computers in Human Behavior*, *115*, 106621.

Laplante, P. A., & Kassab, M. (2022). *What every engineer should know about software engineering*. CRC Press.

Liu, S., Wang, X., Collins, C., Dou, W., Ouyang, F., El-Assady, M., ... & Keim, D. A. (2018). Bridging text visualization and mining: A task-driven survey. *IEEE transactions on visualization and computer graphics*, *25*(7), 2482-2504.

Martins, R. M., & Gresse Von Wangenheim, C. (2023). Findings on teaching machine learning in high school: A ten-year systematic literature review. *Informatics in Education*, *22*(3), 421-440.

Martínez-Fernández, S., Bogner, J., Franch, X., Oriol, M., Siebert, J., Trendowicz, A., ... & Wagner, S. (2022). Software engineering for AI-based systems: a survey. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, *31*(2), 1-59.

Nazar, N., Aleti, A., & Zheng, Y. (2022). Feature-based software design pattern detection. *Journal of Systems and Software*, *185*, 111179.

Qasaimeh, M., Al-Qassas, R. S., & Ababneh, M. (2021). Software design and experimental evaluation of a reduced aes for iot applications. *Future Internet*, *13*(11), 273.

Sami Ur Rehman, M., Thaheem, M. J., Nasir, A. R., & Khan, K. I. A. (2022). Project schedule risk management through building information modelling. *International Journal of Construction Management*, *22*(8), 1489-1499.

Shaturaev, J. (2023). Efficiency of investment project evaluation in the development of innovative industrial activities. *ASEAN Journal of Science and Engineering*, *3*(2), 147-162.

Wang, J., Huang, Y., Chen, C., Liu, Z., Wang, S., & Wang, Q. (2024). Software testing with large language models: Survey, landscape, and vision. *IEEE Transactions on Software Engineering*.

Zeinali Nejad, M., Golshan, M., & Naeimi, A. (2021). The effect of synchronous and asynchronous computer-mediated communication (CMC) on learners' pronunciation achievement. *Cogent Psychology*, *8*(1), 1872908.

Zhou, P., Ali Khan, A. A., Liang, P., & Badshah, S. (2021, June). System and software processes in practice: Insights from Chinese industry. In *Proceedings of the 25th International Conference on Evaluation and Assessment in Software Engineering* (pp. 394-401).