ENHANCING THE ACADEMIC RATING INDEX (ARI) VIA CROSS-PLATFORM DATA INTEGRATION: CONSOLIDATING SUPERVISORS' RESEARCH, SUPERVISION RECORDS AND EXPERTISE IN A CENTRAL ACADEMIC SEARCH AND RANKING ENGINE

Thapelo Mokole [1] and Marcia Mkansi [2]

[1] Department of Computer Science, University of South Africa, Pretoria, South Africa
[2] Department of Operations Management, University of South Africa, Pretoria, South Africa

ABSTRACT

This article looks at how the Academic Rating Index (ARI) can be integrated with a variety of academic platforms to improve data collection and increase access to supervisors' academic records and competency. The ARI aims to address challenges faced by postgraduate students in South Africa, such as an increasing student-to-supervisor ratio, insufficient assessment of supervisor effectiveness, and limited digital forums for student participation. The ARI uses cross-platform data collection and powerful search functions to help students locate and interact with distinguished supervisors who share their research interests. Supervisors can assert and amend their profiles using the site, ensuring that the information is accurate and up to date. This study builds on previous research and emphasizes the importance of a central directory for academic supervisory expertise.

KEYWORDS

Search engine, Postgraduate supervision, Student & Supervisor relationship

1. INTRODUCTION

Post apartheid South-Africa saw an increasing number of students (majority of whom were previously marginalized) seeking to further their education at colleges and universities, both at the undergraduate and postgraduate levels [1]. According to the department of higher education, between 830 and 982 thousand students enrolled in institutions of higher learning programs and between 57 and 84 thousand students enrolled in postgraduate programs from 2010 to 2021 [1,2]. The number of students increased not only at the undergraduate level, but also at the postgraduate level. The overwhelming influx of students saw an increasing imbalance of the ratio between students and supervisors, introducing a wide range of challenges. This study is primarily focused on the postgraduate level, scholars in South Africa identified a gap in the nation's low publication cohorts and contribution to the body of knowledge over the years in critical sectors such as math and science, stats and engineering, however, with increasing attention focus on social studies [3].

Bibhu Dash et al: EDUAI, CCSEAI - 2025
pp. 13-28, 2025. IJCI – 2025 DOI:10.5121/ijci.2025.140602

In an attempt to increase the contribution to the body of knowledge, in 2015 the South African department of higher education ushered in the Research Outputs Policy (ROP) to replace the Measurement of Research Outputs of Public Higher Education Institutions of 2003 to drive the hunger and an increase in quality research output by rewarding the contributors [1]. Postgraduate supervision strongly affects research quality in higher education. Institutions must ensure supervision meets quality standards and yields timely results, while supervisors support students with academic and personal challenges. Yet supervision often lacks systematic monitoring and evaluation tools, transparency, and clear allocation of responsibilities [7], [8], [9], [10]. Drawing on the challenges identified in the authors' earlier work [4], several persistent issues in supervisory relationships can be observed. In particular, power asymmetries, communication breakdowns, physical and temporal distance, misinterpretations, and inconsistent feedback are among the most frequently reported and influential factors that adversely affect the quality and effectiveness of these relationships. To address the above noted challenges the authors from the University of South Africa engaged in research to design and implement a digital solution named the ARI a proof of concept that serves as a digital supervision output model that combines or makes use of methods aligned with the fourth industrial revolution (4IR) based on a tried and tested theoretical framework in academia [4].

This paper is an extension on the aforementioned work by addressing one of the problems that came up when the platform was tested at several universities in South Africa. One recurring problem was that the platform needed to have search features so that students could look for supervisors who are highly rated in different areas of expertise and rate them, as well as check to see if they are open to taking on students with similar interests. The feedback provided that identified a lack for a centralized directory for experts in various dominance hierarchies or studies in particular that students can search when they need a supervisor [4]. To address the noted need, a recent feature was added to the ARI platform to build a database of supervisors' skills, expertise, publications, qualifications, specializations, professional positions and interests from an academic perspective so that students can search for them when they need a mentor who can help them and is available to take on any students.

2. LITERATURE

In the words of the noble Scottish-Irish physicist Lord Kelvin (William Thomson) of blessed memory whose famous quote "When you can measure what you are speaking about, and express it in numbers, you know something about it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts advanced to the stage of science." Sharing the same view is Peter Drucker, a renowned management consultant, educator, and author summing it all up in simple terms "you can't improve what you can't measure".

It is without doubt that an excellent higher education is necessary for building a knowledge-based economy in every country around the world and for dealing with the problems that arise within the nation as well as with keeping standards high when overseeing advanced (postgraduate) studies [5]. It is for that reason that the ARI seeks to quantify the sentiments of the relationship and by producing a statistical view that depicts the relationship and its strengths and places for development. As scholar [6] posits, one of the main goals of supervision is to make sure that the student follows the rules, has the right skills, and gets the help and support they need to complete the tasks at hand which will in turn prepare them for the marketplace. On the backdrop of the previous statement, we can conclude that the role of the supervisor is ensure they provide a good service, to ensure that the growth and progress exist in a given profession or any dominance hierarchy.

Throughout the course of this study, various problems have been identified and discussed in prior work by [4]. Table 1 below goes on to discuss the most prevalent issues that impede interactions between a student and a supervisor, which can have an influence on students. Given the difficulties illustrated in the table below, a clearer and more accurate understanding of the needs, roles, and expectations of all stakeholders (students and supervisors) or groups (schools or academic communities) is essential.

Table 1. Common relationship challenges

Challenges	**Impact**
Behavioural Issues and Perceptions	Behavioural difficulties and academic setbacks can strain student-supervisor relationships. Students from tough backgrounds frequently feel unsupported, which affects their performance and engagement [7]
Lack of Structural Support for Supervisors	A lack of tools to assist supervisors in addressing social expectations causes disagreements, misunderstandings, and stress. Difficult student relationships can cause supervisor stress and exhaustion [8]
Conflict and Dependency	Supervisor-student disputes can have a detrimental effect on pedagogical skill by fostering a reliance that makes teaching less effective [9]
Interpersonal Dynamics	An imbalance in behaviour dynamics, paired with interactions lacking complementarity and consistency, can be damaging to the student-supervisor relationship [10]
Perceptual Differences	Disparities in the way supervisors and students view each other can have an impact on learning outcomes, psychological growth, and comprehension of instructions [11].
Boundary Considerations	Relationship problems can arise from inappropriate intimacy, which includes socializing outside of work bounds, disclosing personal information, and showing partiality [12].
Supervisor Stress and Emotional Competence	Supervisor stress and lack of emotional competence can affect relationships, especially with disruptive students. Supervisors with higher emotional competence tend to have closer relationships with students [13–15]
Institutional Roles and Rituals	Ethnographic research has shown that institutional systems can force supervisors and students into roles that do not allow for fruitful relationships.
Classroom Instabilities	The interaction between students and supervisors can also be hampered by misunderstandings, arguments, and uneasiness brought on by unstable classroom conditions [16].
Economic and Structural Issues	The student-supervisor connection may be hampered by elements including disinterest, lack of incentives, absences, and reluctance to change. Inaccurate beliefs about teaching, economic degradation, and a lack of diversity all play important roles [17].

In addition to the challenges explored in (table 1 above), the student's impact and success is also dependent of the fact that every student at postgraduate level started out as an undergraduate. Because of this, some researchers have found that socioeconomic factors, as well as the quality of the schools and fields that undergraduates choose, are important because they affect how well they do and whether they get their postgraduate studies ([18]). Moreover, the students' motivation and perseverance are vital in determining their success when making early, crucial decisions about the selection of a course that aligns with their abilities and interests [19].

3. DESIGN SCIENCE RESEARCH METHODOLOGY

For the purposes of this paper, design science research (DSR) will be used, which is a problem-solving technique that involves the creation of new artifacts in the form of models, methods, and systems that help people develop and comprehend IT solutions [20,21]. Furthermore, researchers [22], using the work of [23], define DSR as "research that invents a new purposeful artefact to address a generalized type of problem and evaluates its utility for solving problems of that type".

3.1. Problem Relevance

The ARI was mainly developed as a tool to evaluate and control supervision quality, as well as to supplement the supervisor's citation index; nevertheless, since its deployment into the production environment, it has encountered additional obstacles and the need to improve specific aspects. This includes the op-opportunity to have a single source of truth for the supervisor or academic staff. Part of the initial research was to have a centralized place for rating or evaluating performance. To improve on that, we need to get the supervisors' details (publications, expertise, interests, and academic history) into a single place for ease of access and an overview of the track record.

3.2. Design Process

In the current scope of this study, the primary focus will be on data collection/sourcing, searching abilities, and mastery of data administration from the perspective of the supervisor. This data will be utilized to supplement student ratings, as well as to match and assign students to supervisors who share similar interests and research aims. With this information in hand, the institution will have a better idea of how to improve service quality and research output.

3.2.1. Existing Data Scraping and Integration Techniques

In order to collect the supervisor's data, this ARI system will need to link with external systems and databases in some form to pool valid and verified data. Data such as total citations and indexes, affiliation institutions, publications, qualifications, expertise, research interests, and supervisor projects can be pooled to demonstrate how knowledge is derived. This section provides an overview of existing strategies for automating the procedure. Web scraping, web crawling, Application-Programming-Interface (API), and Extract-Transform-Load (ETL) operations are some of the various methods for retrieving data from multiple sources [24]. In light of the aforementioned methods, ETL was utilized for data sourcing since it is a framework for processing data by taking it from various sources, transforming it into the appropriate style or structure, and then loading it into a target system, such as a data warehouse. Allows for the unification of data from multiple sources into a single version of the truth; yet it is labour intensive in the early phases to ensure adequate extraction and transformation of the source data [25].

3.2.2. Master Data Management (MDM) Approaches

MDM or its approaches can be defined as a way to aggregate data from various sources, check it for accuracy, and implement governance rules to ensure data quality over time, resulting in a single, consistent source of key business data across a company [26], to summarize, MDM can be defined as cutting-edge techniques to establish a single truth, which can be divided into four categories: registry, consolidation, coexistence, and centralized approaches. The Registry approach uses bi-directional data flow and matching algorithms to identify duplicates across multiple sources but requires unified data control for reliability [26–28]. Consolidation combines data from various sources into a single version of truth, offering high data quality and reduced duplicates, though it is less flexible and sensitive to changes [29–31]. Coexistence, similar to consolidation, allows phased implementation with bi-directional syncing but is complex and costly to maintain [31–33]. The Centralized approach stores data centrally, simplifying reporting and governance but at the cost of flexibility and high implementation expenses. Each method has its distinct pros and cons related to data flow, quality, and system adaptability [34–36]. For this study, the consolidation strategy was chosen because it combines master data from multiple sources in the hub to create the golden record, or one version of truth, in a unidirectional flow form [30].

3.2.3. Search Engine Technologies

A search engine is a software application that retrieves information from data collection according on the user's or client's request and delivers a list of records connected with the input [37]. A full text search engine, on the other hand, allows users to search across all of the text in documents, databases, or websites, rather than just specified sections or labels. In other words, it searches all of the text for information relevant to a search request [38], which is the topic of this work. In this section, we will look at tools like Elasticsearch, OpenSearch, Typesense, Apache Lucene, Apache Solr, and others for indexing and querying data using full text search capabilities. These platforms offer features like AI-driven results, scalability, full-text search, typo tolerance, and analytics. Each has its pros, like ease of use and robust performance, and cons, such as complexity in setup, resource intensity, and limited community support. Licensing varies from Apache 2.0 to MIT and GPL 3.0, with some offering cloud or self-hosting options [39–46]. After comparing the many open-source platforms available for consumption and usage, this study chose Apache Lucene, a powerful Java text search engine known for its adaptability, sophisticated full-text search capabilities, and extensive community support [38].

3.3. System Architecture and Design

System architecture and design are the organized and conceptual representations of the interactions and connections between various components and modules inside a system. Within the field of search engines, ETL methods and master data management serve as a framework to ensure the appropriate processing, management, and retrieval of data [38]. Figure 1 demonstrates how data is searched, cleaned, and indexed. The purpose is to provide a consolidated library of clean, rich, and valuable supervisor-related information. The ARI program enables supervisors to search and connect their online profiles across many academic platforms (ResearchGate, Semantic Scholar, and Google Scholar), which are then processed, edited, stored, and indexed for accessibility and system-broad visibility.

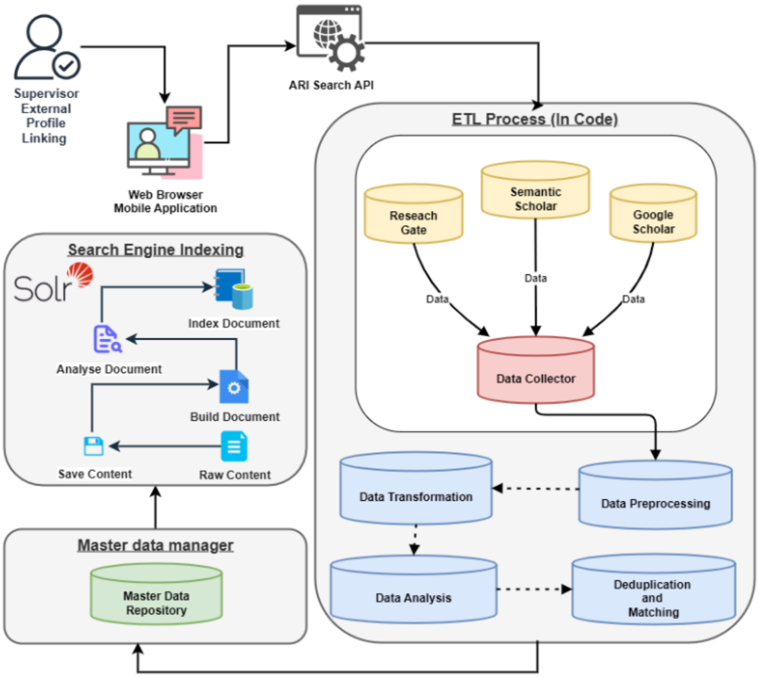


Figure 1. ARI Searching, ETL and Data management

ETL processes are critical in this architecture because they describe the procedures for extracting data from various sources, converting it to a usable format, and loading it into a database or search engine. This ensures that the data is golden, uniform, and ready for indexation. ETL operations are critical for maintaining data quality and integrity, which are necessary for accurate search results and optimal system performance [47]. MDM requires establishing a single source of truth by consolidating and managing an organization's critical data. In the context of search engines and ETL procedures, MDM ensures that the data indexed and searched is precise, consistent, and current. Figure 1 shows the ETL steps for data cleansing, de-duplication, and enrichment to provide a comprehensive and reliable dataset [47]. The system architecture and design in this context aim to create a cohesive and efficient ecosystem that supports the continuous flow of data from collection to indexing, allowing users to conduct effective searches and quickly get useful information.

3.4. Research Rigor

Software testing was key in this system as various forms of testing had to be invoked in order to achieve the objective. One way to make sure that software meets its quality, performance, and functionality standards is to test it in essence. Software testing makes sure that the quality, functionality, and speed of software meet the requirements [48]. There are several types of testing, including functional and non-functional. The former is used to validate the intended features or workings of the system based on what is expected of it, whereas the latter seeks to validate the solution's efficiency, performance, scalability, and portability in the given context, as stated by the authors of the well-known knowledge site [48].

- Functional poses the following question, what does the system do?
- Non-Functional poses the following question, how does the system do it?

The functional tests were confined to computer lab-based assessments and encompassed sanity, integration, and usability testing, which have been extensively detailed by other researchers [49]. The experiences of users, data integrity, operations, security, and performance were all confirmed through non-functional testing. The assessments were conducted on the primary domains of the ARI, specifically the supervisor search processing engine. The first assessments in this case focused solely on the search engine functionalities, rather than the complete system.

The first assessments in this case focused solely on the search engine functionalities, rather than the complete system. The capability to extract supervisor data from several sources, cleanse and format the data for storage in the ARI database and ensure it is indexed in Apache Solr. Figures 2, 3 and 4, presented below, illustrates the interface of the application in its web form as perceived from a supervisory standpoint.

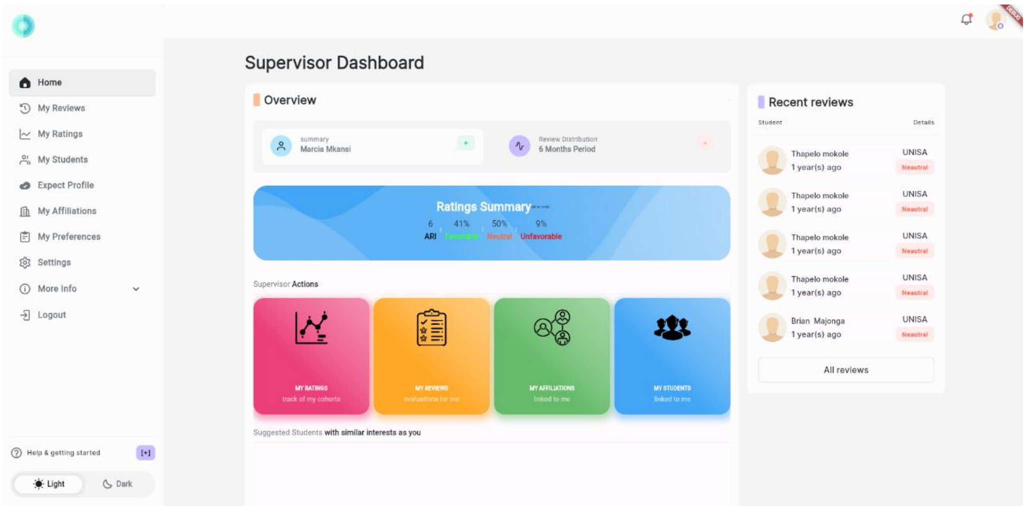


Figure 2. ARI Landing supervisor dashboard

The process of constructing a comprehensive profile for each supervisor can be achieved through the utilization of publicly accessible online information. This methodology enables the correlation of the supervisors to their respective publications, areas of interest, and fields of specialization. Such linkages are facilitated by the application's advanced deep internet crawling capabilities. Figure 3 showcases a streamlined landing page where users first link their profile before claiming a digital record—securing all key data associated with an academic personnel member in one place.

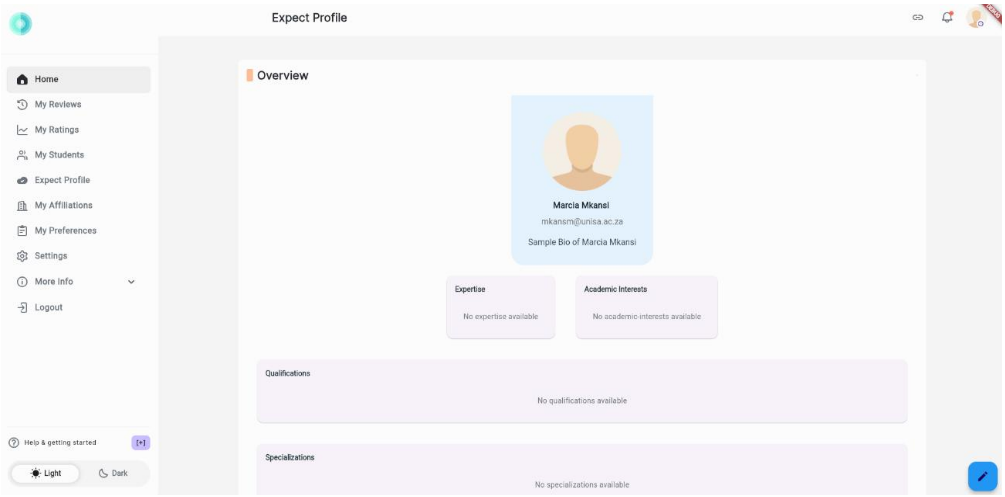


Figure 3. Supervisors expect profile view pre-linking view

In Figure 4, we see how a supervisor can perform a search based on individuals' names and their associated affiliations. The system then scans the virtual environment to identify and retrieve the most relevant information for the specific user.

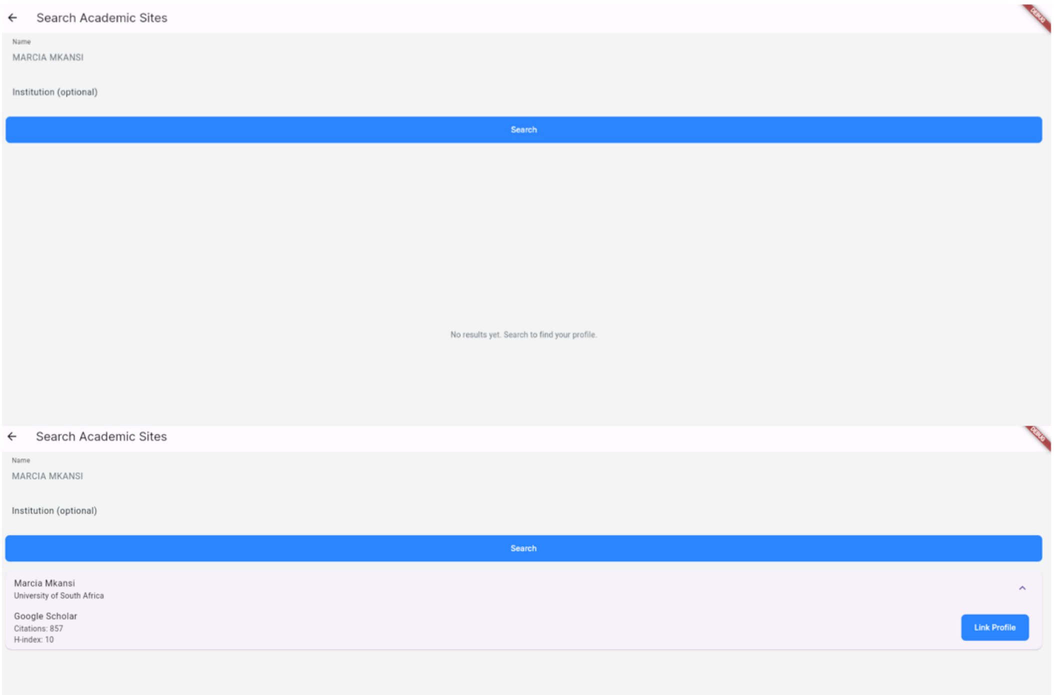


Figure 4. Supervisor searching for their profile online

Figure 5 below depicts how the citation and publication data of the supervisor and how it appears from the source (Google Scholar) using the harzings Publish or Perish application; highlighted are datasets copied from Google Scholar into the ARI using the search function. As previously stated, and shown in Figure 6, the citation index, publication history, and institution of affiliation will be extracted using the API, transformed and persisted using the ETL and MDM, and indexed using Apache Solr.

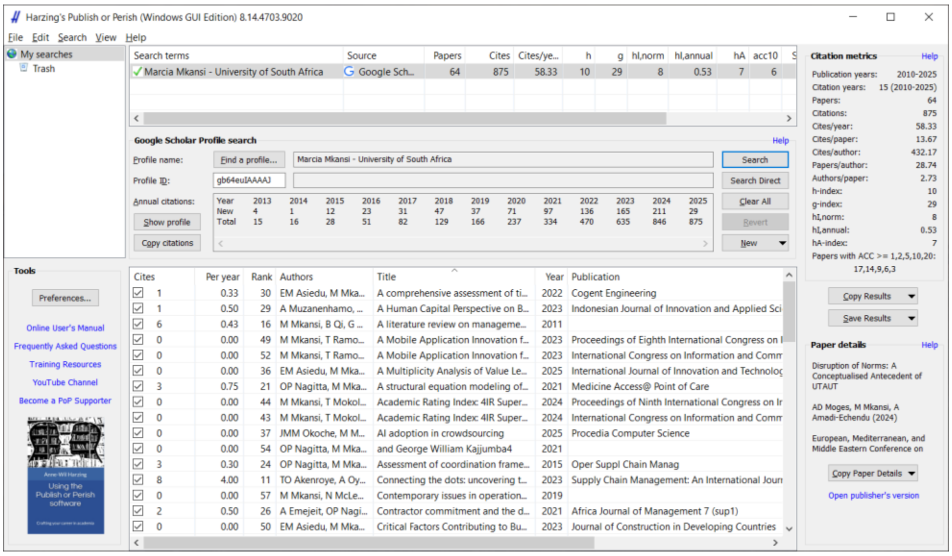


Figure 5. Sources of online profiles linked to a sample supervisor

The ability to retrieve supervisory data from multiple sources, subsequently purify and structure the data for integration into the ARI database, and ensure its indexing within Apache Solr is depicted in Figure 6.

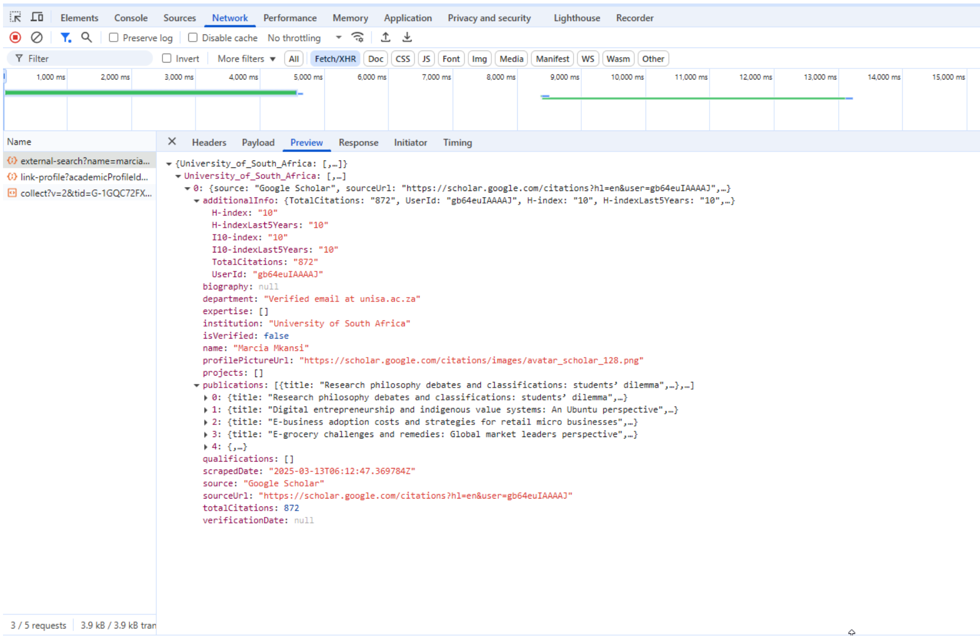


Figure 6. Typical response Json payload post gathering data from sources

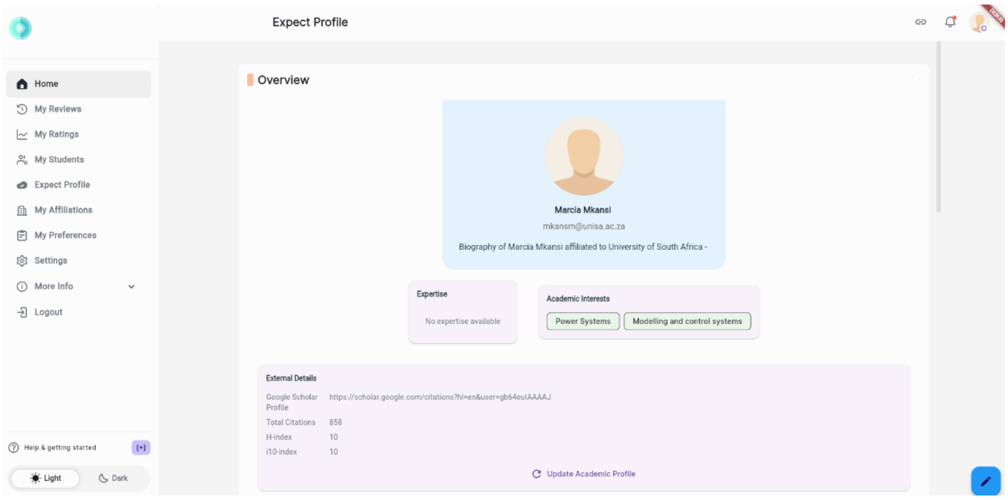


Figure 7. Extended view of post linking

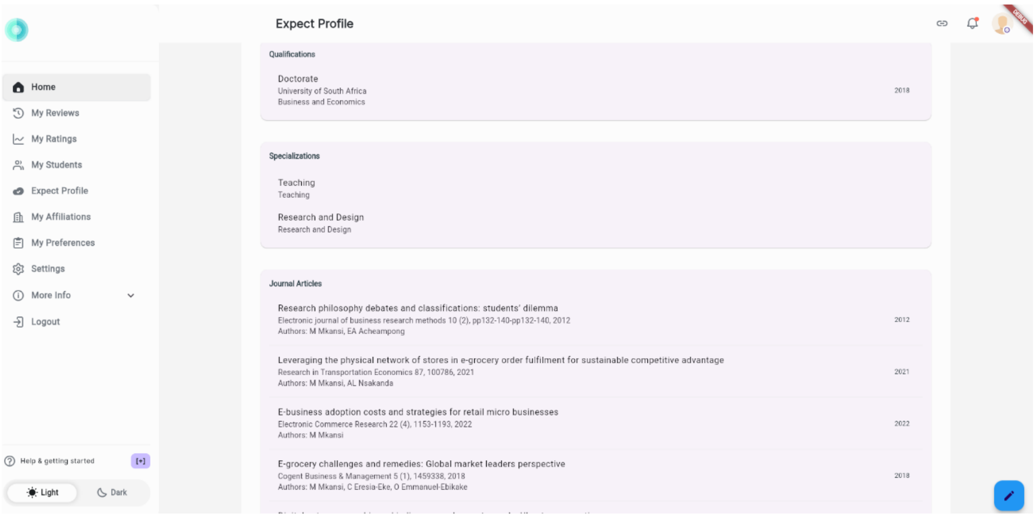


Figure 8. Extended view of post linking… continued

The researchers in this work developed integration capabilities to interface with the following data retrieval sites: Google Scholar, ResearchGate, and Se-mantic Scholar. The results from each data source are as follows. During the evaluation of online academic platforms, Google Scholar emerged particularly as the most dependable resource due to its publicly accessible data, which can be extracted. Conversely, data collection from ResearchGate was challenging due to the implementation of a Content Delivery Network (CDN) by the website administrators. Consequently, accessing the site's data became challenging because of the rate limiting measures designed to obstructs automated bots from retrieving or extracting information. Semantic Scholar utilizes API endpoints to provide data; however, subsequent to creation, the university restricted students from accessing API profiles. Although the data can be extracted and sourced, it constitutes a violation of university policies. Despite obtaining an API key from Sematic, it is ethically questionable to continue utilizing the service without the con-sent of the hosting university (UNISA).

4. Development Results

Swagger is used to expose the ARI's searching capabilities relevant to this study's focus, and endpoints are used to index the search engine data results set, which is then processed by the ARI application before being provided to the Solr instance. The endpoints, depicted in Figure 9 below provide search functionality from a variety of various websites and aggregate data using algorithmic ranking to provide a single version of truth. The program also has a feature that allows students to engage with it and request information about their supervisors and interests. The ARI has expanded its search functionality by incorporating additional data sources such as Google Scholar, ResearchGate, and Semantic Scholar. Google Scholar ranked as the most reliable source because to its easily accessible data, however ResearchGate created challenges due to its Content Delivery Network and bot rate restriction. While Semantic Scholar offers API endpoints, their utilization infringes institutional (UNISA) regulations.

The ARI application uses Swagger endpoints to index and sanitize search engine data before passing it to a Solr machine for further processing. These endpoints com-bine data from various sources into a single rendition of the truth. The ARI program also includes an instance of the Solr application, which can be accessible through a management interface and a Docker container. This design allows users to engage with the application and efficiently obtain information about supervisors and their interests. The following are the study's artifacts or outputs that can be used by both students and supervisors.

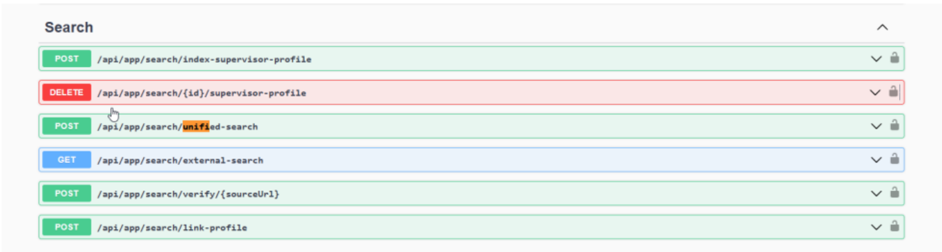


Figure 9. Swagger endpoints

The system also includes an active instance of the Solr application, as demonstrated in the sample view below. This configuration enables users to engage with the application and solicit information concerning supervisors and their research interests.

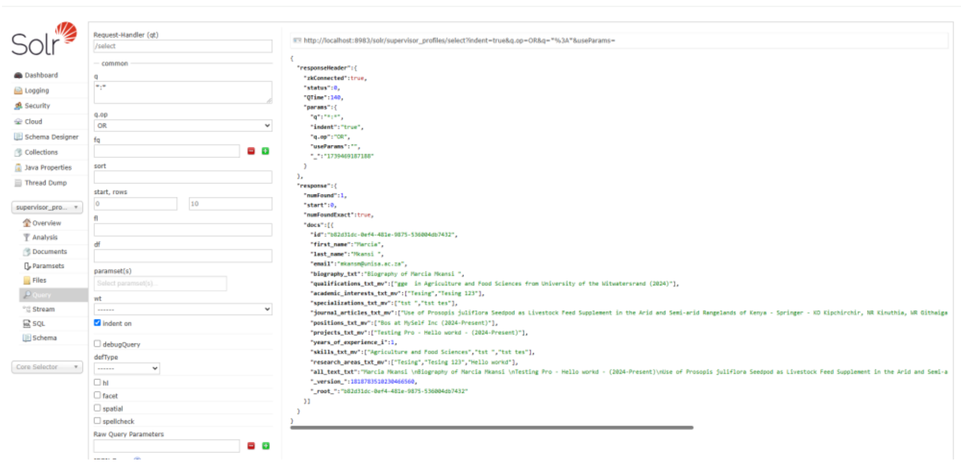


Figure 10. Solr management portal

The subsequent elements represent the artifacts or outputs of this research, which may be utilized by both students and supervisors.

- **Web**: https://ari-web-demo.web.app/
- **Android**: https://play.google.com/store/apps/details?id=za.ac.unisa.marcia.ari&hl=en_ZA
- **Apple**: https://apps.apple.com/app/id1556716168

5. DISCUSSION AND CONCLUSION

This section will delve into the findings and conclusions derived from our implementation and analysis of the Solr application. By providing a comprehensive overview of the application's functionalities, the challenges faced during deployment, and the potential benefits observed, we aim to demonstrate the value of Solr in managing and retrieving data efficiently. Our deployment of the Solr application revealed several key insights and areas of improvement. First, the application's ability to provide users with a robust interface for querying and retrieving information about supervisors and their interests proved highly effective. The integration with a Docker container simplified the deployment process, ensuring that the application remained scalable and easily accessible via the API.

Furthermore, the Solr management portal afforded us an intuitive means of overseeing the application's performance and optimizing search capabilities. This level of oversight is crucial for maintaining the efficiency and reliability of any large-scale data management system. In conclusion, the implementation of Solr within our infrastructure not only enhanced our data retrieval processes but also underscored the importance of utilizing advanced search technologies to manage complex datasets. Future work will focus on refining the application's functionalities, exploring additional use cases, and integrating new features to further leverage Solr's capabilities as well as building the front-end to expose the search engine on the application for smart searching purposes similar to that of google. which will have a heavy dependency on data in order to return satisfactory results.

6. FUTURE OF WORK

In a world saturated with data—constantly collected, quantified, and mined for sentiment and insight—this study sets out a bold and focused research agenda. The roadmap below charts the methods and technologies used so far and lays out the next steps in a clear, staged sequence, driving the project toward an increasingly comprehensive and nuanced understanding of the subject.

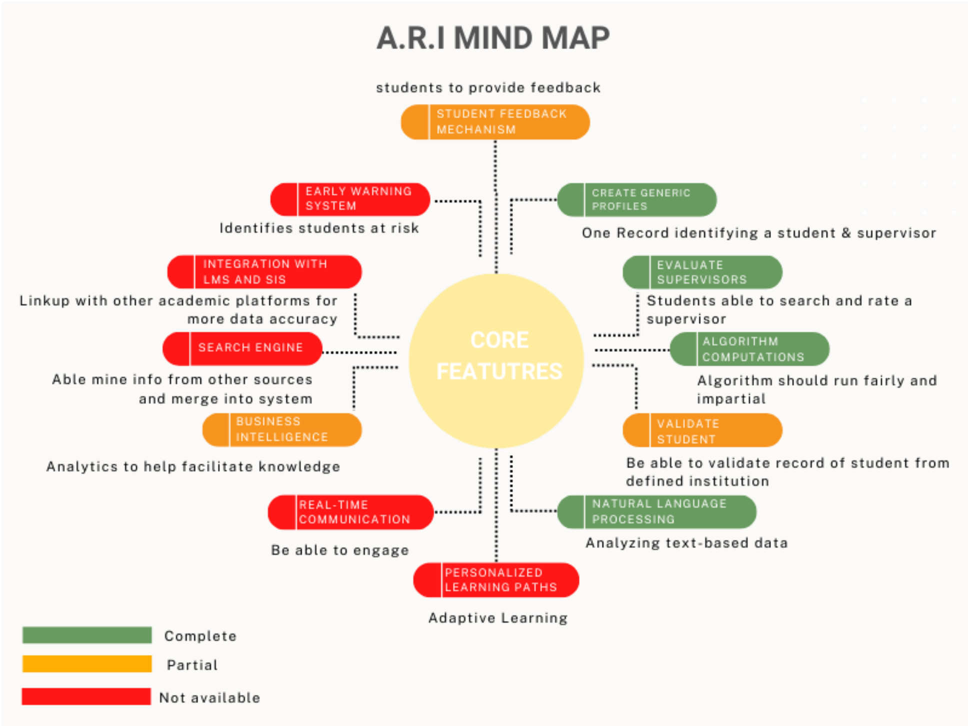


Figure 11. Academic rating index roadmap

Figure 11 above seeks to depict the various modules envisioned in the product development namely the inclusion of the following core modules: an interactive dashboard for individual and group performance, engagement, and progress; predictive analytics to flag at-risk students using historical and real-time data; adaptive learning to personalize paths; a feedback and analytics system for courses, instructors, and overall experience; progressive tutoring with real-time help, personalized feedback, and extra resources; virtual collaboration spaces for projects, discussions, and community building; integration for seamless LMS and SIS connectivity; networking features for peer insights, advice, and support; gamification (badges, rewards, challenges) to boost motivation; educator tools to monitor progress, find gaps, and adjust instruction in real time; a parent portal for student progress, attendance, and communication; attendance tracking to spot patterns and engagement issues; content delivery for hosting and streaming lectures, simulations, and virtual labs; integrated chat and video conferencing for real-time interaction; and career guidance with resources, recommendations, and analytics for academic and career decisions.

ACKNOWLEDGEMENTS

Praise and thanks to the great I am for the gift of life!

REFERENCES

[1] DHET, "Statistics on Post-School Education and Training in South Africa: 2021," 2023.
[2] DHET, "Statistics on post school education and training in south africa - Government Publications," 2010.
[3] M. P. Sebola, "South Africa's public higher education institutions, university research outputs, and contribution to national human capital," *Human Resource Development International*, vol. 26, no. 2, pp. 217–231, 2023, doi: 10.1080/13678868.2022.2047147.

[4] T. G. Mokole, “Academic-Staff Rating Index (ARI) System,” 2021, Accessed: Aug. 20, 2022. [Online]. Available: https://uir.unisa.ac.za/handle/10500/27704

[5] E. Kimani, “Challenges in Quality Control for Postgraduate Supervision,” *International Journal of Humanities Social Sciences and Education (IJHSSE)*, vol. 1, no. 9, p. 2349, 2014, Accessed: Oct. 02, 2022. [Online]. Available: www.arcjournals.org

[6] A. Saleh and M. Maher, “Challenges of Improving the Quality of Academic Supervision of Postgraduate Studies at the Faculty of Education, Damietta University,” vol. 7, no. 1, 2021, doi: 10.5296/jei.v7i1.18292.

[7] G. Espinosa and C. Beaudoin, “La relation enseignant-élève,” *La relation enseignant-élève*, Nov. 2022, doi: 10.1515/9782760557567.

[8] J. Wilkins, “An examination of the student and teacher behaviors that contribute to good student-teacher relationships in large urban high schools,” Thesis, State University of New York at Buffalo, 2006. Accessed: Jul. 25, 2024. [Online]. Available: https://www.researchgate.net/publication/36370361_An_examination_of_the_student_and_teacher_behaviors_that_contribute_to_good_student-teacher_relationships_in_large_urban_high_schools_electronic_resource

[9] Jeong-Nam, Bae and J. Han-ik, “The structural relationships among teacher-student relations, teacher emotions, teaching professionalism, and commitment perceived by secondary school teachers,” *Seoul Education Review*, vol. 67, no. 1, pp. 1–33, Mar. 2019, doi: 10.25152/SER.2019.67.1.1.

[10] H. J. M. Pennings, M. Brekelmans, P. Sadler, L. C. A. Claessens, A. C. van der Want, and J. van Tartwijk, “Interpersonal adaptation in teacher-student interaction,” *Learn Instr*, vol. 55, pp. 41–57, Jun. 2018, doi: 10.1016/j.learninstruc.2017.09.005.

[11] K. Jong Baeg and K. Jun-yeop, “Effects of Perceptual Differences in Student-Teacher Relationships Between Student and Teacher on the Level of Understanding Instructions: Relating with Teachers’ Job Satisfaction and Students’ Peer and Parental Trust,” *The Journal of Korean Teacher Education*, vol. 31, no. 3, pp. 245–267, Sep. 2014, doi: 10.24211/tjkte.2014.31.3.245.

[12] S. M. Plaut and D. Baker, “Teacher-student relationships in medical education: boundary considerations,” *Med Teach*, vol. 33, no. 10, pp. 828–833, Oct. 2011, doi: 10.3109/0142159X.2010.541536.

[13] K. F. McGrath and P. Van Bergen, “Attributions and emotional competence: why some teachers experience close relationships with disruptive students (and others don’t),” *Teachers and Teaching*, vol. 25, no. 3, pp. 334–357, Apr. 2019, doi: 10.1080/13540602.2019.1569511.

[14] P. Van Bergen, K. McGrath, and D. Quin, “Nurturing close student–teacher relationships,” in *Inclusive Education for the 21st Century*, Routledge, 2020, pp. 296–316. doi: 10.4324/9781003116073-16.

[15] P. Van Bergen, K. F. McGrath, D. Quin, and E. C. Burns, “Nurturing close student-teacher relationships,” in *Inclusive Education for the 21st Century*, London: Routledge, 2023, pp. 315–330. doi: 10.4324/9781003350897-19.

[16] D. Raufelder, W. M. Bukowski, and S. Mohr, “Thick Description of the Teacher-student Relationship in the Educational Context of School: Results of an Ethnographic Field Study,” *J Educ Train Stud*, vol. 1, no. 2, pp. 1–18, Apr. 2013, doi: 10.11114/JETS.V1I2.108.

[17] A. Becker Narvaes, “The teacher-student relationship: between and imaginaries, identifications and negotiations,” UFSM, Santa Maria, 2012.

[18] L. Zhang, “Advance to Graduate Education: The Effect of College Quality and Undergraduate Majors,” *Rev High Ed*, vol. 28, no. 3, pp. 313–338, 2005, doi: 10.1353/RHE.2005.0030.

[19] R. Ashraf, J. M. Godbey, M. M. Shrikhande, and T. A. Widman, “Student Motivation and Perseverance: Do They Explain College Graduation?,” *Journal of the Scholarship of Teaching and Learning*, vol. 18, no. 3, pp. 87–115, Oct. 2018, doi: 10.14434/josotl.v18i3.22649.

[20] J. Akoka, I. Comyn-Wattiau, and V. C. Storey, “Design Science Research: Progression, Schools of Thought and Research Themes,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13873 LNCS, pp. 235–249, 2023, doi: 10.1007/978-3-031-32808-4_15.

[21] P. Johannesson and E. Perjons, “An introduction to design science,” *An Introduction to Design Science*, vol. 9783319106328, pp. 1–197, Jul. 2014, doi: 10.1007/978-3-319-10632-8/COVER.

[22] A. M. Teperi, N. Gotcheva, and K. Aaltonen, "Design thinking perspective for developing safety management practices in nuclear industry," *Human Factors in the Nuclear Industry: A Systemic Approach to Safety*, pp. 309–326, Jan. 2020, doi: 10.1016/B978-0-08-102845-2.00016-8.

[23] J. Venable and R. Baskerville, "Eating our own Cooking: Toward a More Rigorous Design Science of Research Methods," *Electronic Journal of Business Research Methods*, vol. 10, no. 2, p. pp141-153-pp141-153, Dec. 2012, Accessed: Jan. 19, 2025. [Online]. Available: https://academic-publishing.org/index.php/ejbrm/article/view/1296

[24] A. Bodor, M. Hnida, and N. Daoudi, "Integration of web scraping, fine-tuning, and data enrichment in a continuous monitoring context via large language model operations," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 15, no. 1, pp. 1027–1037, Feb. 2025, doi: 10.11591/IJECE.V15I1.PP1027-1037.

[25] A. Walha, F. Ghozzi, and F. Gargouri, "Data integration from traditional to big data: main features and comparisons of ETL approaches," *Journal of Supercomputing*, vol. 80, no. 19, pp. 26687–26725, Dec. 2024, doi: 10.1007/S11227-024-06413-1/METRICS.

[26] A. Cleven and F. Wortmann, "Uncovering four strategies to approach master data management," *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2010, doi: 10.1109/HICSS.2010.488.

[27] A. Carr, "Data Management Implementation." Accessed: Jan. 24, 2025. [Online]. Available: https://www.google.com/search?q=4.2.2+Master+Data+Management+(MDM)+Approaches&rlz=1C1CHBF_enZA1001ZA1002&oq=4.2.2%09Master+Data+Management+(MDM)+Approaches&gs_lcrp=EgZjaHJvbWUyBggAEEUYOdIBBzM0M2owajeoAgCwAgA&sourceid=chrome&ie=UTF-8

[28] M. Khairi, "MASTER DATA MANAGEMENT MODEL EFFECTIVENESS IN INFORMATION TECHNOLOGY COMPANY," 2012.

[29] E. Tereshchenko, "Development and implementation of the data quality assurance subsystem for the MDM platform," 2021, Accessed: Jan. 26, 2025. [Online]. Available: https://lutpub.lut.fi/handle/10024/162973

[30] R. R. Pansara, "Master Data Management important for maintaining data accuracy, completeness & consistency," *Authorea Preprints*, Oct. 2023, doi: 10.36227/TECHRXIV.24053862.V1.

[31] G. Ali Shaykhian, M. A. Khairi, and J. Ziade, "Architectural Evaluation of Master Data Management (MDM): Literature Review," 2016.

[32] R. Ravjibhai Pansara, "Citation: Ronak Ravjibhai Pansara. Strategies for Master Data Management Strategies for Master Data Management Integration and Their Benefits," *Scholars Journal of Engineering and Technology Abbreviated Key Title: Sch J Eng Tech*, no. 2, pp. 40–47, 2024, doi: 10.36347/sjet.2024.v12i02.002.

[33] V. R. Rapolu, "Achieving MDM on data lakes," *International Journal of Management IT and Engineering*, vol. 10, no. 2, pp. 38–40, 2020, Accessed: Feb. 02, 2025. [Online]. Available: https://www.indianjournals.com/ijor.aspx?target=ijor:ijmie&volume=10&issue=2&article=006

[34] M. Heiskanen, "Data Quality in a Hybrid MDM Hub," May 2016, Accessed: Feb. 02, 2025. [Online]. Available: https://trepo.tuni.fi/handle/123456789/23907

[35] K. P. Kusuma Dewi, T. Fabrianti Kusumasari, and R. Andreswari, "Analysis and design of architecture master data management (MDM) tools for open source platform at PT XYZ," *Proceedings - 2019 5th International Conference on Science and Technology, ICST 2019*, Jul. 2019, doi: 10.1109/ICST47872.2019.9166255.

[36] R. Vilminko-Heikkinen, "Establishing a MDM function: First Steps in the Master Data Management Architecture Design," pp. 124–131, 2015, doi: 10.3233/978-1-61499-570-8-124.

[37] wwwsajscoza S. Afr J Sci, J. Edosomwan, and T. O. Edosomwan, "Comparative analysis of some searCh engines Affiliation," *S Afr J Sci*, vol. 106, no. 11, 2010, doi: 10.10520/EJC96996.

[38] Abhay Pai, "Search using Apache Solr in Enterprise." Accessed: Jan. 21, 2025. [Online]. Available: https://medium.com/@abhaypai/search-using-apache-solr-in-enterprise-or-crm-applications-a107d2d694e8

[39] OpenSearch, "opensearch-project/OpenSearch: 🔎 Open source distributed and RESTful search engine." Accessed: Feb. 09, 2025. [Online]. Available: https://github.com/opensearch-project/OpenSearch

[40] Typesense, “typesense/typesense: Open Source alternative to Algolia + Pinecone and an Easier-to-Use alternative to ElasticSearch ⚡ 🔍 ✨ Fast, typo tolerant, in-memory fuzzy Search Engine for building delightful search experiences.” Accessed: Feb. 09, 2025. [Online]. Available: https://github.com/typesense/typesense

[41] Meilisearch, “meilisearch/meilisearch: A lightning-fast search engine API bringing AI-powered hybrid search to your sites and applications.” Accessed: Feb. 09, 2025. [Online]. Available: https://github.com/meilisearch/meilisearch

[42] ElasticSearch, “elastic/elasticsearch: Free and Open Source, Distributed, RESTful Search Engine.” Accessed: Feb. 09, 2025. [Online]. Available: https://github.com/elastic/elasticsearch

[43] Lucene, “apache/lucene: Apache Lucene open-source search software.” Accessed: Feb. 09, 2025. [Online]. Available: https://github.com/apache/lucene

[44] Solr, “apache/solr: Apache Solr open-source search software.” Accessed: Feb. 09, 2025. [Online]. Available: https://github.com/apache/solr

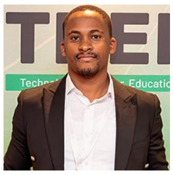
[45] Swirl, “swirlai/swirl-search: AI Search & RAG Without Moving Your Data. Get instant answers from your company’s knowledge across 100+ apps while keeping data secure. Deploy in minutes, not months.” Accessed: Feb. 09, 2025. [Online]. Available: https://github.com/swirlai/swirl-search

[46] MongoDB, “Full-Text Search: What Is It And How It Works | MongoDB | MongoDB.” Accessed: Feb. 09, 2025. [Online]. Available: https://www.mongodb.com/resources/basics/full-text-search

[47] J. Sreemathy, R. Brindha, M. Selva Nagalakshmi, N. Suvekha, N. Karthick Ragul, and M. Praveennandha, “Overview of ETL Tools and Talend-Data Integration,” *2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021*, pp. 1650–1654, Mar. 2021, doi: 10.1109/ICACCS51430.2021.9441984.

[48] A. Sethi, “A REVIEW PAPER ON LEVELS, TYPES& TECHNIQUES IN SOFTWARE TESTING,” *International Journal of Advanced Research in Computer Science*, vol. 8, no. 7, pp. 269–271, Aug. 2017, doi: 10.26483/ijarcs.v8i7.4236.

AUTHORS



Thapelo Mokole, a young African software engineer with nearly a decade of experience, works in the banking sector and has a broad career covering various industries. He holds a master’s degree in computing and is pursuing a PhD at Unisa. Through his start-up, Code Shandi's, he aims to build a skilled professional community to enhance Africa's contribution to the tech industry by incubating ideas and fostering knowledge sharing and connections.

Professor Marcia Mkansi holds a PhD in Supply Chain Management focused on the Grocery E-retail Sector and an MSc integrating supply chain management, E-Business, business information systems, and multimedia technologies, both from the University of Bolton, UK. She also has a Bachelor of Administration from the University of the North, South Africa, and an NRF Rating of C1. Her academic interests include promoting a triple helix approach involving industry, government, and academia to foster research-based innovation, societal value, economic transformation, and smart innovation development. Her specialization includes operations management, innovation and commercialization, and value chain management and sustainability.