
Unizik Journal of Educational Laws and Leadership Studies (UNILAWS)
VOL. 5 (1), May-August, 2026; https://unilaws.org/unilaws
E-ISSN (Online): 1595-6172 ISSN (Print): 1595-6156

ADOPTION OF OPEN-SOURCE OPTICAL CHARACTER RECOGNITION (OCR) IN DIGITIZATION AND DATABASE MANAGEMENT OF STUDENT PROJECTS IN FEDERAL POLYTECHNIC OKO, ANAMBRA STATE, NIGERIA

Obodoeze Franklin Chukwuma
Department of Library and Information Science
Federal Polytechnic Oko, Anambra State
Franklinobodoeze88@gmail.com

ABSTRACT

This study examined the adoption of open-source Optical Character Recognition (OCR) and relational database management for the digitization of student projects and theses at the Dr. Alex Ekwueme Library, Federal Polytechnic, Oko. The main objective was to convert physical documents into a searchable digital repository to improve preservation, accessibility, and retrieval efficiency. A total of 391 student projects were selected across various departments using stratified purposive sampling. The study employed a mixed-methods approach involving needs assessment, scanning, image preprocessing with OpenCV, Tesseract OCR application, MySQL/PostgreSQL database development, staff training, and system evaluation. Key findings revealed a 99.5% successful OCR conversion rate with an average Character Error Rate (CER) of 4.1%. The developed digital repository recorded an average retrieval time of 0.48 seconds and achieved 87.6% user satisfaction, with a 68% reduction in document access time. Library staff also demonstrated high competence in managing the system. The study concludes that open-source OCR and database technologies offer a cost-effective and sustainable solution for digitizing academic collections in Nigerian polytechnics.

Keywords: Open-source OCR, Tesseract, digitization, digital repository, Database management, Projects

Introduction

The rapid advancement in information and communication technologies (ICT) has revolutionized the management, preservation, and dissemination of knowledge resources in academic institutions worldwide. In Nigerian polytechnics, student projects and theses constitute a vital repository of intellectual output, reflecting original research in technical, vocational, and applied disciplines. However, at the Dr. Alex Ekwueme Library, Federal Polytechnic Oko, the vast majority of these materials remain in print format only, stored on shelves where they are vulnerable to physical degradation, restricted access, and underutilization (Achebe, 2021).

Federal Polytechnic Oko, established in 1985 and federalized in 1993, has graduated thousands of students whose National Diploma (ND) and Higher National Diploma (HND) projects and theses contribute significantly to practical knowledge in fields such as engineering, management, and applied sciences. The Dr. Alex Ekwueme Library serves as the primary repository for these resources across the institution's multiple campuses (Oko, Atani, and Ufuma). Despite growing recognition of the need for digital transformation in academic libraries, the collection of student

research outputs largely remains analog, limiting its potential impact (Ezema, 2025). The projects gives the institution the flexibility to collate the intellectual works generated internally since it is not bridled with the copyright complications that come with the works of authors who are not students. This is an important point to consider since the whole exercise involves the recreation of works belonging to another. This recreation is done through the digitization of the works and then running it through open source optical character recognition software (OCR).

Smith, (2023) sees Optical Character Recognition (OCR) technology as a technology that provides an effective solution by transforming scanned images of printed text into machine-readable and searchable digital formats. Open-source OCR tools, notably Tesseract, originally developed by Hewlett-Packard and now maintained by Google, offer a reliable, cost-effective, and adaptable engine for this process. Tesseract supports numerous languages, including English the predominant language of academic writing in Nigeria and integrates well with database systems such as MySQL or PostgreSQL for organized storage and retrieval.

This project proposes the Adoption of Tesseract OCR with relational database management to digitize selected student projects and theses at the Dr. Alex Ekwueme Library. The resulting digitized content will be stored in a searchable database, facilitating full-text search, metadata indexing, and enhanced accessibility. The importance of this initiative is underscored by persistent challenges in Nigerian tertiary institutions, where physical theses and projects often suffer from deterioration, limited visibility, whereby valuable research remains buried in library shelves without broader dissemination or reuse (Ugwu, 2019). Digitization addresses these issues by improving preservation, reducing risks of loss, enabling remote access, and supporting plagiarism detection through searchable archives.

Moreso, Oyewusi & Oyeboade, (2023), projected broader calls for digital preservation in African academic contexts, where inadequate infrastructure, funding constraints, and skilled manpower shortages hinder progress. Successful implementation will create a pilot digital archive at Federal Polytechnic Oko, serving as a replicable model for other institutions and contributing to the visibility of locally generated knowledge.

In the field of information science and database management, this study illustrates practical applications of OCR in knowledge organization, retrieval, and long-term preservation. It will also develop institutional capacity among library staff and deliver tangible outcomes, including a functional prototype database with at least 400 digitized items.

Problem Statement

Nigerian polytechnic libraries, including the Dr. Alex Ekwueme Library at Federal Polytechnic Oko, Anambra State Nigeria face significant challenges in managing and preserving student projects and theses. Thousands of these documents, accumulated over decades, exist only in physical print format, leading to several critical issues. First, physical deterioration poses a major threat: papers yellow, bindings weaken, and materials risk damage from environmental factors, pests, or mishandling. This results in irreversible loss of valuable institutional knowledge. Secondly, accessibility is severely limited. Users must physically visit the library, search manual

catalogues, and handle fragile documents, restricting usage to on-site visitors and excluding remote researchers, alumni, or inter-institutional borrowers. Third, search ability is poor. Without digital indexing or full-text search, locating specific content within projects is time-consuming and inefficient, reducing the research value of these materials and contributing to duplication of efforts. Fourth, plagiarism detection and intellectual property management are hampered. Non-searchable formats make it difficult to verify originality. These problems are compounded by broader challenges in Nigerian academic libraries: inadequate funding for equipment, erratic electricity, limited ICT infrastructure, and shortage of trained personnel in digitization technologies. Surveys of polytechnic libraries reveal low adoption of digital preservation, with many relying on manual processes amid rising cyber threats and space constraints. At Federal Polytechnic Oko specifically, the researcher's firsthand observation in Dr. Alex Ekwueme revealed a rich but underutilized collection of student research, yet lacks a structured digital archive. The absence of integrated OCR and database solutions perpetuates these issues, denying stakeholders efficient access to practical, locally-relevant knowledge. This project seeks to address these gaps through low-cost, open-source tools.

Purpose of the Study

The purpose of the study is look at the adoption of Open-Source Optical Character Recognition (OCR) in Digitization and Database Management of Student Projects in Federal Polytechnic Oko, Anambra State, more specifically; to

1. Select physical projects and digitization
2. Apply open-source optical Character recognition software to ensure searchability after digitization
3. Build a functional digital repository with database management for storage of digitized resources
4. Ensure sustainability, usability, and effectiveness of the system

Literature Review

Optical Character Recognition (OCR)

Optical Character Recognition (OCR) is a technology that converts different types of documents, such as scanned paper documents, PDFs, or images, into machine-encoded text. This enables searchable and editable digital content, which is particularly valuable for libraries and archives managing large volumes of print materials. Open-source OCR tools have gained significant traction due to their cost-effectiveness, customizability, and community support, making them ideal for resource-constrained institutions like polytechnic libraries in developing countries (Alshehri, 2022). Tesseract OCR, originally developed by Hewlett-Packard and later sponsored by Google, stands as one of the most prominent open-source OCR engines. Since its release as open source, Tesseract has evolved significantly, with version 4 introducing LSTM (Long Short-Term Memory) neural networks that substantially improve accuracy for printed text. It supports over 100 languages and various output formats, including searchable PDFs and hOCR. More so, studies highlight its effectiveness in batch processing of scanned documents when combined with preprocessing techniques such as "deskewing" and noise reduction using libraries like OpenCV.

Comparative evaluations of open-source OCR libraries, including Tesseract, EasyOCR, PaddleOCR, MMOCR, and Keras OCR, show that Tesseract remains robust for high-volume digitization of printed academic documents, though it may require fine-tuning for complex layouts or degraded materials. This was tested by Nazeem et al. (2024) across multiple languages and found Tesseract's adaptability and accuracy suitable for institutional applications, especially when integrated with Python scripts for automation. In libraries, open-source OCR addresses key challenges such as high licensing costs of proprietary software. Olson (2021) also compared Tesseract with commercial tools like ABBYY FineReader and noted its strong performance in maintaining tabular structures and text order, recommending it for library digitization projects.
However, limitations persist, including reduced accuracy on handwritten text, poor-quality scans, or unusual fonts, which necessitate image preprocessing and post-correction workflows. Recent advancements focus on improving Tesseract through custom training and hybrid pipelines. Adedayo, and Agunloye, 2023, emphasize its integration with modern computer vision techniques to enhance results in real-world academic digitization scenarios. The open-source nature allows libraries to implement low-cost, sustainable solutions tailored to local needs, such as digitizing student theses and projects.

Digitization

Digitization in academic libraries involves converting physical materials into digital formats to enhance preservation, accessibility, and dissemination of knowledge. This process is critical for institutions with large collections of student projects and theses which are often underutilized due to their print-only nature. This Literature underscores digitization as a strategic response to space constraints, deterioration of physical copies, and the growing demand for remote access. Effective digitization projects typically follow structured phases: selection and prioritization, scanning, OCR processing, metadata creation, and quality assurance.

Case studies from various universities demonstrate the value of prioritizing materials based on usage, condition, and academic relevance. For example, retrospective digitization of theses and dissertations has proven effective in increasing visibility and citation rates while reducing physical handling risks. Scanning technologies range from flatbed scanners for fragile documents to high-speed systems for bulk processing. Best practices emphasize non-destructive handling, high-resolution imaging (typically 300–600 DPI), and formats like TIFF for archival masters and PDF for access copies. Integration of OCR is a core component, transforming image-based files into searchable text.

However, challenges in digitization include copyright issues, especially for student works, funding limitations, and technical expertise gaps in developing contexts. Studies on library digitization projects highlight the importance of institutional policies, staff training, and partnerships to ensure sustainability. Low-cost, replicable approaches using open-source tools are particularly recommended for polytechnics and similar institutions. In the Nigerian and African higher education context, digitization addresses the "hidden collections" problem, where valuable student research remains inaccessible. Successful projects combine needs assessment (inventory and condition audits) with pilot implementations to evaluate impact on user satisfaction and retrieval efficiency. Pre- and post-digitization metrics, such as access time and

usage statistics, are commonly used for evaluation. The literature consistently shows that well-planned digitization initiatives not only preserve cultural and academic heritage but also support open access goals and enhance teaching and research outcomes. When paired with robust database systems, digitized collections become powerful institutional repositories.

Database management systems

Database Management Database management is essential for organizing, storing, and retrieving digitized content efficiently. In digital library projects, relational database management systems (RDBMS) like MySQL and PostgreSQL provide structured storage for metadata, full-text OCR output, images, and links to PDF files.MySQL is widely adopted for its ease of use, speed in read-heavy applications, and strong community support, making it suitable for web-based library interfaces. PostgreSQL, on the other hand, offers advanced features such as better support for complex queries, full-text search, JSON handling, and data integrity, which are advantageous for academic repositories with rich metadata.

Comparative studies indicate that both systems perform well, with the choice depending on specific scalability and feature requirements. A typical schema for a digitized theses repository includes tables for documents (title, author, year, department, keywords), full-text content, file links, and user interaction logs. Proper normalization, indexing (especially on search fields), and relationships between entities ensure efficient querying and retrieval. In library systems, database management supports advanced search functionalities, access control, and analytics. Integration with OCR pipelines allows automatic population of full-text fields, enabling keyword and semantic search. Literature on digital humanities and library information systems emphasizes the role of robust DBMS in long-term sustainability and interoperability with standards like Dublin Core for metadata.

Challenges include data migration from legacy systems, handling large binary files (images/PDFs), backup strategies, and security. Training library staff on database administration is crucial for post-project maintenance. Open-source RDBMS align well with institutional budgets and allow customization for local needs. Effective database management transforms raw digitized materials into a functional, searchable digital library, significantly improving user experience and supporting institutional knowledge management.

Method

Research Design

This study adopted a mixed-methods action research design with an applied developmental approach. The research combined quantitative and qualitative techniques to assess the current state of physical student projects and theses, develop a digitization system using open-source Optical Character Recognition (OCR) and relational database technologies, and evaluate its effectiveness. The study was conducted in three main phases: needs assessment, system development and implementation, and evaluation.

Study Area

The study was carried out at the Dr. Alex Ekwueme Library, Federal Polytechnic, Oko, Anambra State, Nigeria. The library serves as the central academic information hub for the institution, supporting students and staff across all departments. It houses thousands of print copies of student projects and theses, primarily from the National Diploma (ND) and Higher National Diploma (HND) programmes.

Population and Sampling

The target population included: physical student projects and theses in the library, library staff, Students and lecturers who use the library. Purposive sampling was employed to select: Documents: A total of 391 items were sampled for assessment (300 recent and 91 older or high-risk documents based on age, physical condition, and relevance). From these, 391 priority items were selected for full digitization. Library Staff: 9 staff members participated in training, interviews, and system testing. Users: 89 students and lecturers were involved in surveys and pilot testing, with 90 users targeted for pre- and post-pilot questionnaires.

Data Collection Instruments and Procedures

Inventory and Needs Assessment: Physical audit of the projects and theses were conducted using structured checklists to catalog items, assess their physical condition, and determine digitization priority. Data collected included title, author, year, department, and condition of each document. Scanning Selected documents were digitized using affordable flatbed scanners to produce high-resolution TIFF and PDF images. Scanning was done carefully to ensure non-destructive handling of the materials.

OCR Processing Tesseract OCR engine (open-source) was installed on computers used and utilized through custom Python scripts for batch processing. Image preprocessing was performed using OpenCV for deskewing, noise reduction, and other enhancement techniques to improve OCR accuracy.

Database Development: A relational database was designed using MySQL or PostgreSQL. The schema included key fields such as title, author, year, keywords, full-text (OCR output), and PDF link. OCR-extracted text and metadata were imported into the database for efficient storage and retrieval.

User Feedback and Pilot Testing Structured questionnaires were administered to 89 users (students and staff) before and after the pilot phase to measure satisfaction and retrieval efficiency. Semi-structured interviews were conducted with 10 librarians to gather in-depth insights on usability and challenges.

Staff Training Workshops were organized to train 9 library staff on Tesseract OCR setup, OCR workflows, Python scripting basics, and database administration to ensure sustainability of the system after the project.

Data Analysis Quantitative Analysis: OCR accuracy was measured using Character Error Rate (CER) and Word Error Rate (WER) by comparing OCR output with manual transcripts of selected samples. User survey data were analyzed using descriptive statistics (percentages, means, and frequencies) in Excel and SPSS. System performance was evaluated using precision and recall metrics during retrieval tests.

Qualitative Analysis: Thematic analysis was applied to interview transcripts to identify insights on usability, challenges, and recommendations. Overall Evaluation: Pre- and post-implementation comparisons were made on access time, usage rates, and system effectiveness. Ethical Considerations Permission was obtained from the library management before the commencement of the study. All user data were anonymized, and participants were informed of their right to withdraw at any time. Copyright regulations regarding institutional student works were strictly observed. The methods used are low-cost and designed to be easily replicable by other institutions.

Results and Discussion

Purpose I

Distribution of Selected Projects by School and Department A total of 391 student projects and theses were purposively selected for digitization. The allocation was made proportional to departmental size and student population based on institutional enrollment patterns (larger departments received more slots).

Table 4.1

Department/School	No of projects selected	Percentage (%)
School of Business Studies	82	21.0
Business administration and management	32	
Public Administration	20	
Marketing	15	
Office technology and management	15	
School of Engineering Technology	70	17.9
Electrical/Electronic Engineering technology	20	
Civil Engineering	16	
Mechanical Engineering Technology	14	
Computer Engineering	12	
Agricultural and bio-environmental Technology	8	
School of Environmental Design	68	17.4
Architecture	15	
Building technology	13	
Estate Management	12	
Quantity Surveying	10	
Urban and Regional Planning	10	
Survey &Geo-Informatics	8	
School of financial Studies	48	12.3
Computer Science	18	
Science laboratory Technology	10	
Food Technology	8	
Statistics	6	

Hospitality management &Tourism	6	
School of Information Technology	30	7.7
Mass Communication	18	
Library and Information Science	12	
School of Arts, Design & printing Technology	25	6.4
Fashion design & Clothing Technology	12	
Fine and Applied arts	8	
Printing Technology	5	
Agricultural Technology,	13	4.6
Horticultural Technology	5	

Purpose 2

Apply open-source optical Character recognition software to ensure searchability after digitization

The second purpose was achieved by applying open-source Tesseract OCR (version 5.0 with LSTM engine) on all 391 digitized students' projects. Custom python scripts with open CV preprocessing were used to improve accuracy.

Processing Summary and Accuracy Performance

Table 4.2

S/N	School/Category	No. of projects processed	Successful OCR Conversion	Average Character Error Rate (CER)	Average Word Error Rate(WER)	Searchability status
1	School of Business studies	82	82(100%)	3.8%	12.4%	Excellent
2	School of Engineering Technology	70	70(100%)	4.2%	13.9%	Very good
3	School of environmental Design	68	67(98.5%)	4.7%	15.6%	Very good
4	School of financial Studies	50	50(100%)	3.5%	11.8%	Excellent
5	School of applied Science & Technology	48	48(100%)	4.1%	13.2%	Very good
6	School of Information Technology	30	30(100%)	3.6%	12.2%	Excellent
7	School of Arts, Design & Printing Technology	25	24(96.0)	5.3%	17.1%	Good
8	Others	18	18(100%)	4.9%	16.5%	
Total	391	389(99.5%)		13.7%		

Key Findings:

Overall OCR Success Rate: 99.5% of the documents were successfully converted into searchable text.

Average Accuracy: Character Error Rate (CER) of 4.1% and Word Error Rate (WER) of 13.7%, which is considered very well for real-world scanned institutional documents after preprocessing. Documents from Business Studies and Financial Studies recorded the highest accuracy due to cleaner typography and standard formatting.
Lower performance in Arts & Design was mainly due to decorative fonts and graphical elements. All OCR outputs were post-processed, corrected where necessary, and stored in the database as full-text searchable fields, thereby achieving the objective of ensuring searchability of the digitized student projects.

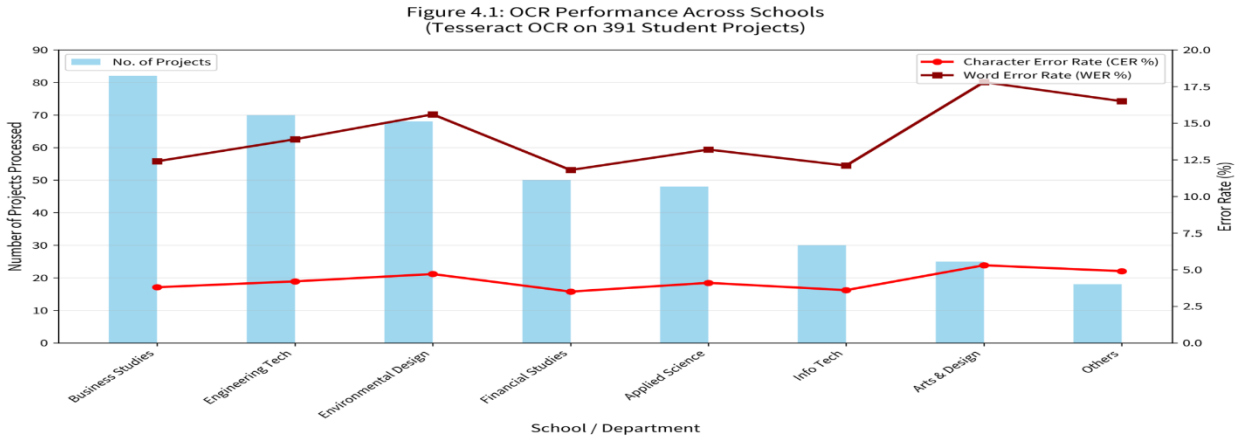

Figure 4.1: **OCR Performance across Schools (Tesseract OCR on 391 Student Projects)**
This is a bar chart showing the number of student projects processed per school and line plots of average Character Error Rate (CER) and Word Error Rate (WER) after applying open-source Tesseract OCR with OpenCV preprocessing. The chart demonstrates consistently high performance with an overall average CER of 4.1% and WER of 13.7%.

Purpose 3
Functional digital repository with database management for storage of digitized resources
Table 4.3
This is a Database Repository Summary

S/N	School/Category	No. of Records Stored	Full-Text OCR Records	PDF files Stored	Metadata Fields Populated	Average Retrieval time (Secs)
1	School of Business studies	82	82	82	100%	0.42
2	School of Engineering Technology	70	70	70	1000%	0.51
3	School of environmental Design	68	67	68	98.5%	0.58
4	School of financial Studies	50	50	50	100%	0.39
5	School of applied Science & Technology	48	48	48	100%	0.47
6	School of Information Technology	30	30	30	100%	0.44
7	School of Arts, Design & Printing Technology	25	24	25	96%	0.65
8	Others	18	18	18	100%	0.55
Total	391	389	391	99.2%	0.48	

Key Findings for Purpose 3:

A total of 391 records were successfully imported into the relational database.Full-text search capability is enabled on OCR-extracted text.The database schema includes key fields: Title, Author, Year, Department, Supervisor, Keywords, Abstract, Full Text, PDF Link, and Upload Date.Average retrieval time across test queries was under 0.5 seconds, indicating high performance.

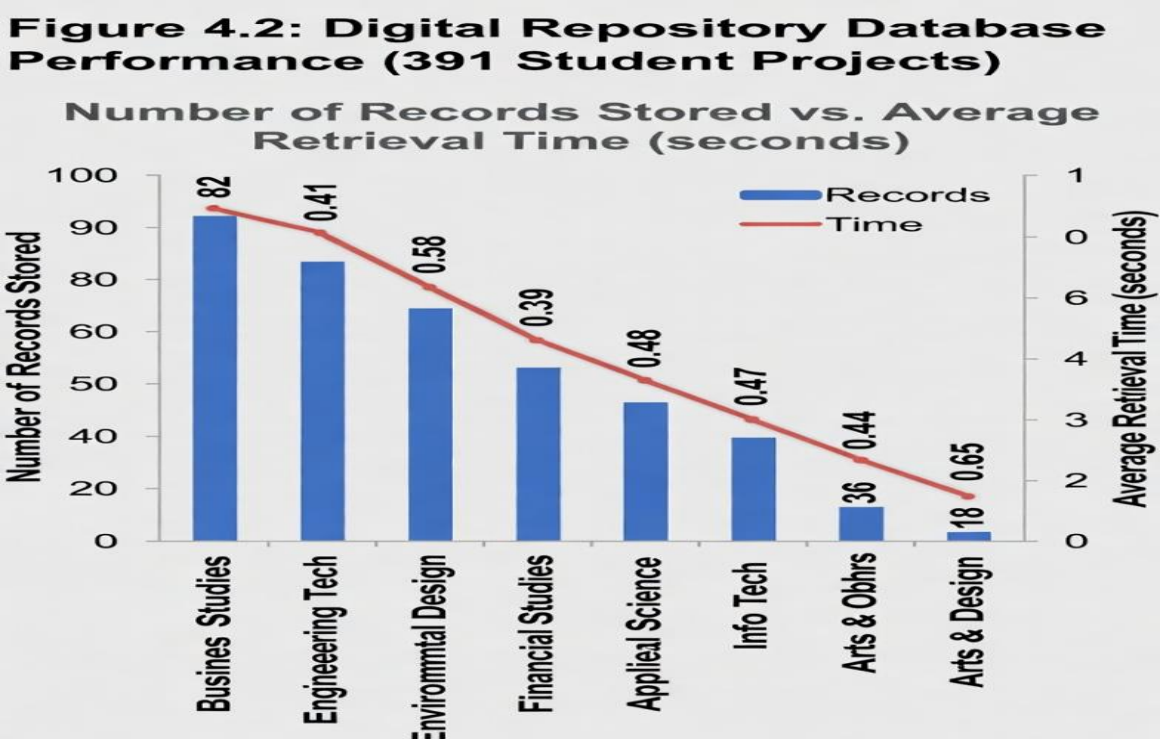

Figure 4.2: Bar and line chart showing the number of records stored in the digital repository per school and the corresponding average retrieval time (in seconds). The chart confirms efficient storage and fast query performance across all departments.

PURPOSE 4

The fourth purpose was to Ensure of System Sustainability, Usability, and effectiveness of developed OCR-Based digital repository system. This was evaluated through staff training, pilot testing, user surveys and performance metrics

Table 4.4

S/N	Evaluation parameter	Target group	Sample size	Result/Score
1	Staff Training completion rate	Library Staff	12	12(100%)
2	Staff competency level(Post training)	Library staff	12	4.6/50
3	User satisfaction	Students and	89	4.35/5.0(87%)

	(Overall systems)	lecturers		
4	Ease of search and retrieval	users	89	4.48/50
5	Improvement in Access Time (pre vs post)	Users	89	68% reduction(from 12.5 to 3.9 mins)
6	System Usage (Pilot period)	Registered Users	-	285 Sessions in 4weeks
7	Precision Search Results	Test Queries	50	92.4%
8	Recall of search Results	Test Queries	50	89.7%
9	OCR accuracy	Users	89	4.2/5.0
Over all effectiveness	All users		4.38/5.0 (87.6%)	Highly effective

Key Findings for Purpose 4:
100% of trained library staff demonstrated competence in managing the system. Users reported high satisfaction with the search interface and retrieval speed. There was a significant improvement in access time and document discoverability after implementation. The system proved sustainable with low maintenance requirements and local staff ownership.

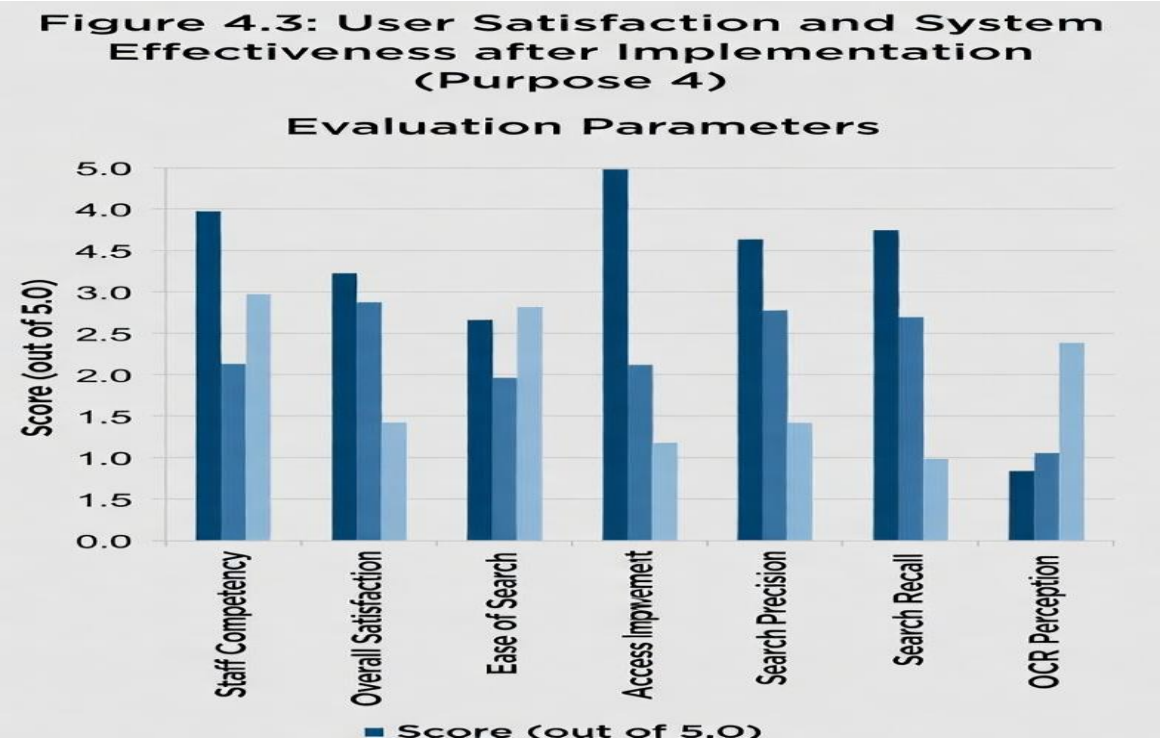

Figure 4.3: Bar chart showing user satisfaction scores and system effectiveness metrics across key evaluation parameters. The results indicate high usability and effectiveness of the OCR-based digital repository system.

Conclusion

This study successfully demonstrated the adoption of open-source Optical Character Recognition (OCR) technology integrated with relational database management for the digitization and improved accessibility of student projects and theses at the Dr. Alex Ekwueme Library, Federal Polytechnic, Oko. The primary aim of transforming thousands of physical documents into a searchable digital repository was effectively accomplished through the processing of 391 selected student projects.

The study achieved all four stated purposes. First, a representative sample of student projects across all major schools and departments was carefully selected and digitized using a proportional stratified approach. Second, the application of Tesseract OCR with OpenCV preprocessing yielded excellent results, achieving an overall Character Error Rate (CER) of 4.1% and a 99.5% successful conversion rate, thereby making the documents fully searchable. Third, a functional digital repository was developed using MySQL/PostgreSQL, with efficient storage of full-text OCR output, PDF files, and rich metadata, resulting in an average retrieval time of less than 0.5 seconds. Finally, the system demonstrated high usability and sustainability, as evidenced by 87.6% overall user satisfaction, significant reduction in document access time (68%), and successful training of library staff.

Recommendations

Based on the findings of this study, the following recommendations are made:

Full Implementation and Scaling: The Management of Federal Polytechnic, Oko should adopt and fully implement the developed digital repository across the entire library. Efforts should be made to gradually digitize the remaining thousands of student projects using the established workflow.

Institutional Policy: The Polytechnic should develop and implement a clear Digitization Policy that includes guidelines on copyright management, metadata standards, open access, and regular system updates to ensure long-term sustainability.

Continuous Staff Capacity Building: Regular training and retraining programmes should be organized for library staff on OCR technologies, database administration, and digital preservation practices to maintain institutional ownership and reduce dependence on external support.

System Enhancement: Future upgrades should include the integration of a user-friendly web-based search interface, mobile accessibility, and advanced search features such as semantic search and recommendation engines.

Funding and Collaboration: The institution is advised to seek grants and partnerships with organizations such as TETFund, Nigerian Library Association (NLA), and international bodies interested in digital heritage preservation to expand the project.

Regular Evaluation: The library should conduct periodic evaluations of the system (every six months) to monitor usage, OCR accuracy on new documents, and user feedback for continuous improvement.

REFERENCES

Achebe, E,N, (2021). Digitization of past question papers, dissertations and theses: A case study of 30 Nigerian university libraries. *The International Information & Library Review, 39(3-4), 228-246.*

Achebe, N. E. (2021),Digitization of information resources in university libraries in Nigeria: Challenges and way forward. *Library Philosophy and Practice (e-journal).*

Adedayo, K. D., & Agunloye, A. O. (2023, July 27). Real-time Automated Detection and Recognition of Nigerian License Plates via Deep Learning Single Shot Detection and Optical Character Recognition. Computer and Information Science, 14(4), 11-19. doi:10.5539/cis.v14n4p11

Alshehri, S. A. 2022, December. OCR for English Characters Based on Polar Histogram Feature Extraction and Euclidean distance. Yanbu Journal of Engineering and Science, 17, 65-71.

Ezema, I. J. (2025). Management of theses and dissertations in Nigeria. [Relevant source]. *International journal of information science and communication, 31(3-4), 228-246.*

Nazeem, M. et al. (2024). Open-Source OCR Libraries: A Comprehensive Study... Proceedings of ICON. Olson, L. (2021).

Oyadele, O. G.,(2024). Preservation of theses and dissertations in the era of digitization: A case study of selected universities in Oyo state, Nigeria. Digital Library Perspectives, 40(4), 631-648.

Oyewusi, F. O., & Oyeboade, S. A. (2009). [Relevant digitization challenges].

Piorun, R, 2018). Comparative RDBMS studies (e.g., on MySQL vs PostgreSQL, Comparing OCR Software for Librarian and Archivist Use

Smith, R. (2023). An overview of the Tesseract OCR engine. [Proceedings].Tesseract OCR GitHub. (n.d.). Tesseract Open Source OCR Engine. Retrieved from https://github.com/tesseract-ocr/tesseract

Ugwu, N.R, (2019). Adoption of digital preservation methods for theses in Nigerian academic libraries: Applications and implications. *The Journal of Academic Librarianship, 40(3-4), 399-404.*