HOUSE OF WISDOM: JOURNAL ON LIBRARY AND INFORMATION SCIENCES

ISSN 3089-6002. Published by Universitas Muhammadiyah Sidoarjo

Advancing Digital Archives via Machine Learning and Artificial Intelligence

Ahmed Nasser Lahmoud[1*], Mohsen Sajit Joudah[2]
Program Studi Ilmu Komunikasi, Universitas Muhammadiyah Sidoarjo, Indonesia
Email : [1]Zooba.987654321@gmail.com , [2]mohsenalzamili25@gmail.com

Abstract. General Background Modern digital repositories face severe operational constraints due to the exponential growth of born-digital materials and digital collections. Specific Background Traditional data organization systems struggle to handle massive digital volumes, leading to metadata deficiencies, processing backlogs, and compromised asset accessibility. Knowledge Gap Existing digital preservation strategies focus primarily on isolated automation tasks, lacking an integrated, multi-phase framework that combines literature landscape tracking, automated catalog error correction, and semantic linked data expansion. Aims This study develops a comprehensive computational framework leveraging machine learning, natural language processing, and knowledge schemas to optimize data workflows in multilingual institutional libraries. Results Utilizing a calibrated Random Forest Classifier paired with custom character-level n-gram frequency segmentation algorithms, the system achieved macro ROC-AUC scores of 0.958257, effectively isolating substandard catalog records and correcting phonetic transliteration errors in Cyrillic indexing systems. Novelty This framework introduces a unique dual-layer pipeline that resolves the physical contextual fragmentation of archives while simultaneously rectifying multilingual transmission flaws without requiring external spell-checking dictionaries. Implications The integration of intelligent automation structures provides memory institutions with scalable, reproducible asset tracking mechanisms that enhance metadata discoverability while preserving the essential contextual validation role of professional archivists.

Highlights:

1. Automated Random Forest pipelines achieved a 0.958 macro ROC-AUC efficiency in identifying corrupted catalog metadata records.
2. Character-level n-gram frequency mapping effectively isolates phonetic translation inconsistencies within multilingual indexing systems.
3. Cognitive graph modeling successfully reconstructs structural provenance patterns lost during analog-to-digital asset transformation.

Keywords: Computational Archiving, Machine Learning, Information Retrieval, Metadata Optimization, Digital Preservation

Introduction

Libraries, archives, and museums now operate in a completely different context due to the explosive expansion of digital information and digital originals. The acquisition, organization, preservation, and distribution of enormous volumes of digital content provide hitherto unheard-of difficulties for organizations in charge of cultural heritage preservation. The amount, complexity, and diversity of modern digital resources make traditional archiving and cataloging methods, which were first created for physical collections, increasingly insufficient. [1], [2]

As a result of these difficulties, computational archiving has become an interdisciplinary

ISSN 3089-6002. Published by Universitas Muhammadiyah Sidoarjo

topic that combines computer science, information science, archival science, and data analytics to create computational methods for resolving archival issues [3]. The potential of artificial intelligence, machine learning, natural language processing, and data mining approaches to support the evaluation, description, preservation, access, and management of archive material has been shown by computational archiving research. The potential for automating labor-intensive operations and enhancing information retrieval across large-scale collections has increased due to recent developments in large language models (LLMs), knowledge schemas, and semantic approaches [4].

The rise in original digital recordings is one of the most important effects of digital transformation. Original digital resources, in contrast to traditional paper collections, frequently exist in complex technological environments, have inconsistent or incomplete metadata, and lack a logical physical organization. As a result, archivists have to deal with the loss of contextual relationships that have historically aided in the interpretation and discovery of archives in addition to scale-related challenges. This situation, referred to as the "vanishing box problem," is the result of the physical organizational structures that upheld the source and original arrangement principles and represented significant links between data disappearing. [5, 6]

Institutions are still dealing with growing backlogs in processing, inconsistent metadata, and growing user demands for quick and advanced access to digital resources. By facilitating automated classification, information extraction, quality control, entity recognition, and semantic enrichment of archive resources, machine learning and natural language processing technologies present promising solutions [7, 8]. However, effective adoption necessitates strategies that support current professional workflows and uphold the crucial role that archivists and librarians play in guaranteeing contextual integrity, authenticity, and dependability.

The use of artificial intelligence (AI) methods to facilitate processes in digital libraries and archives is examined in this study. It focuses on three complementary areas: AI-assisted metadata enhancement through associated data and cognitive graphing techniques; machine learning approaches for automated error detection and correction in multilingual cataloging records; and computational analysis of archival literature to identify emerging trends in digital archives. The study shows how AI may improve professional practice, lessen processing loads, improve metadata quality, and make cultural heritage resources easier to access by testing these strategies in actual library and archive environments. These results offer useful advice for implementing intelligent technology in memory institutions and add to the expanding corpus of research on human-AI cooperation.

2. Related Works

Numerous studies into the use of computational techniques in libraries and archives have been sparked by the growing amount of digital and digital original resources. Initially, the main goals were to improve access to vast collections and deal with archiving backlogs. A "more product, less process" (MPLP) strategy was put forth by Green and Meissner [11], who focused on effective archival processing to lower material backlogs while preserving appropriate levels of description. Subsequent studies on scalable archive operations were greatly impacted by their work.

The groundwork for incorporating computational methods into archival procedures was established by computational archival science (CAS). According to Marciano et al. [2], CAS is an interdisciplinary field that addresses the difficulties posed by extensive digital collections by fusing computer science, information science, archive science, and data analytics. The potential of computational approaches in evaluating, characterizing, conserving, and retrieving archive materials has been shown by later research [3].

ISSN 3089-6002. Published by Universitas Muhammadiyah Sidoarjo

Another important field of study is digital preservation. The technological difficulties in maintaining digital information, such as media deterioration, software obsolescence, and dependence on technology, were emphasized by Thibodeau [4]. In a similar vein, Durante [5] stressed the significance of preserving electronic records' authenticity, dependability, integrity, and use throughout their existence. The theoretical underpinnings of contemporary digital preservation techniques were established by these works.

New difficulties that are essentially different from those related to digitized analog materials have been discovered through research on original digital records. In his analysis of the original arrangement principle's application to digital archiving, Zhang [6] made the case that documentary relationships are still extant but call for new kinds of representation. In a similar vein, Cato [7] talked about the necessity for creative methods to manage original digital collections in public sector organizations as well as the growing complexity of managing digital records.

In recent years, there has been a lot of interest in the use of machine learning and artificial intelligence in libraries and archives. Information retrieval, entity extraction, automatic classification, and metadata creation have all been accomplished with machine learning approaches. According to Higgins [8], computational methods that can handle enormous volumes of diverse data are becoming more and more important for digital preservation. In order to enhance the management of digital collections, Cucciolo [9] also stressed the significance of combining technology solutions with data management techniques.

Opportunities to enhance metadata and knowledge discovery have increased due to recent developments in natural language processing and semantic approaches. Richer semantic linkages between archival resources are made possible by knowledge maps and related data techniques, which improve contextual comprehension and discovery [3]. Large-scale linguistic model development has also created new opportunities for automated retrieval, summarization, and description in cultural heritage holdings.

Despite these developments, there are still a lot of gaps in the real-world application of AI technology in library and archival procedures. Few studies address full frameworks that integrate machine learning, natural language processing, and semantic techniques to concurrently handle various phases of archival and cataloging workflows; the majority of current research focuses on specific techniques or isolated jobs. By examining AI-powered methods for literature analysis, catalog quality control, and metadata enhancement in actual library and archive contexts, this study seeks to close this gap.

3. Research Approaches

In order to examine how artificial intelligence may enhance productivity in archives and libraries, this study used a mixed-methods research methodology that combined computational archival analysis, machine learning, statistical language modeling, and semantic methodologies. Three interconnected phases comprised the methodology: (1) computational analysis of archive material, (2) machine learning-based catalog quality monitoring, and (3) semantic metadata improvement.

3.1. Archival Literature Computational Analysis

Finding methodological and technological developments in the field of computational archaeology was the main goal of the first phase. Five analytical dimensions—archival function, collection type, computational approach, subject matter, and record lifecycle stage—were used to manually gather and classify a collection of publications in this discipline. Figure 2.1 shows examples of annotated papers from the 2023 CAS workshop,

ISSN 3089-6002. Published by Universitas Muhammadiyah Sidoarjo

demonstrating the application of the annotation system to recent CAS publications. To enable automatic analysis, the annotated recordings were subsequently transformed into structured textual descriptions (Figure 1).

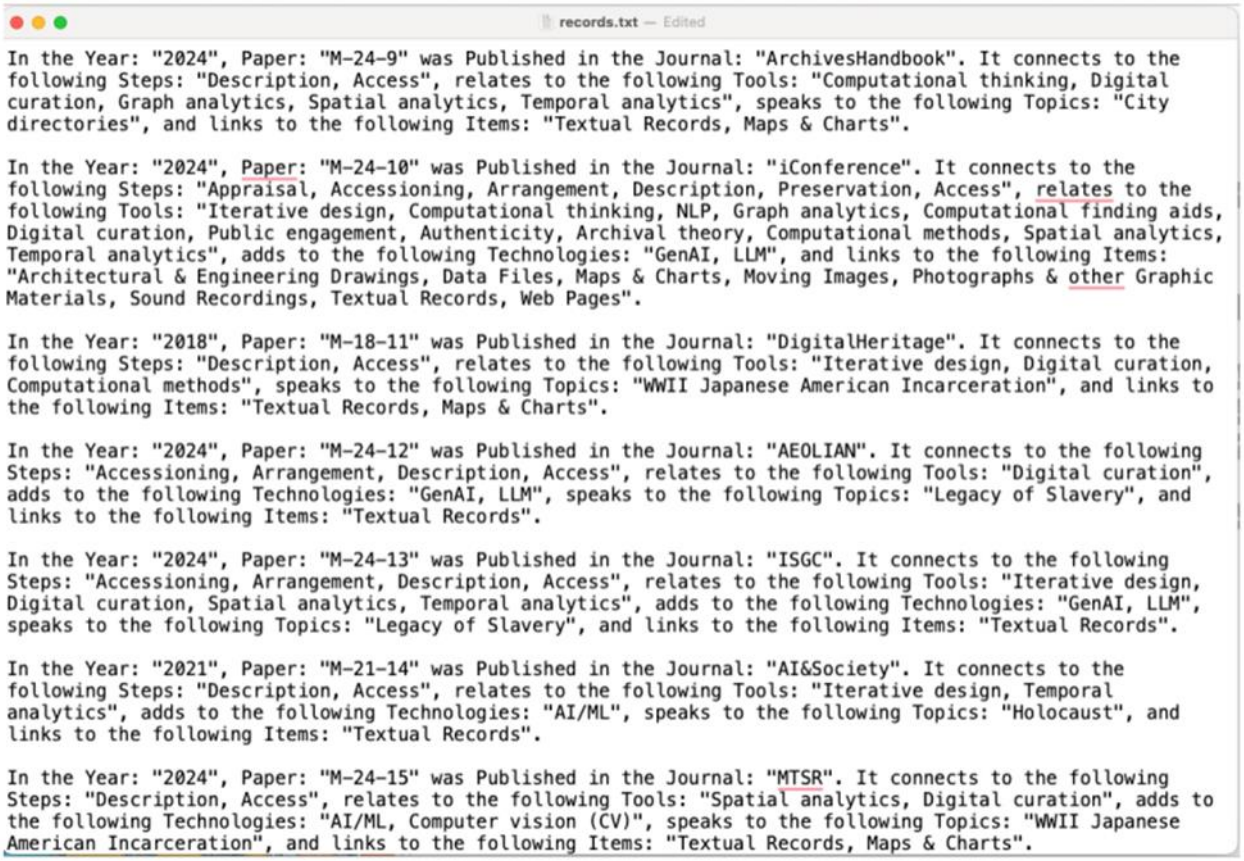


Figure 1 shows an example of seven research papers with annotations that were exported as plain English text.

Neo4j techniques were used to generate a knowledge graph (KG) to facilitate semantic exploration of the literature. From the annotated text set, entities and relationships were automatically retrieved and converted into a graphical representation (Figure 2). Interactive querying and computerized literature review with natural language coaching were made possible by the generated graph. Examples of graph-based searches and query outputs are displayed in Figures 3 and 4, and exemplary interactive query results are compiled in Table 1.

ISSN 3089-6002. Published by Universitas Muhammadiyah Sidoarjo

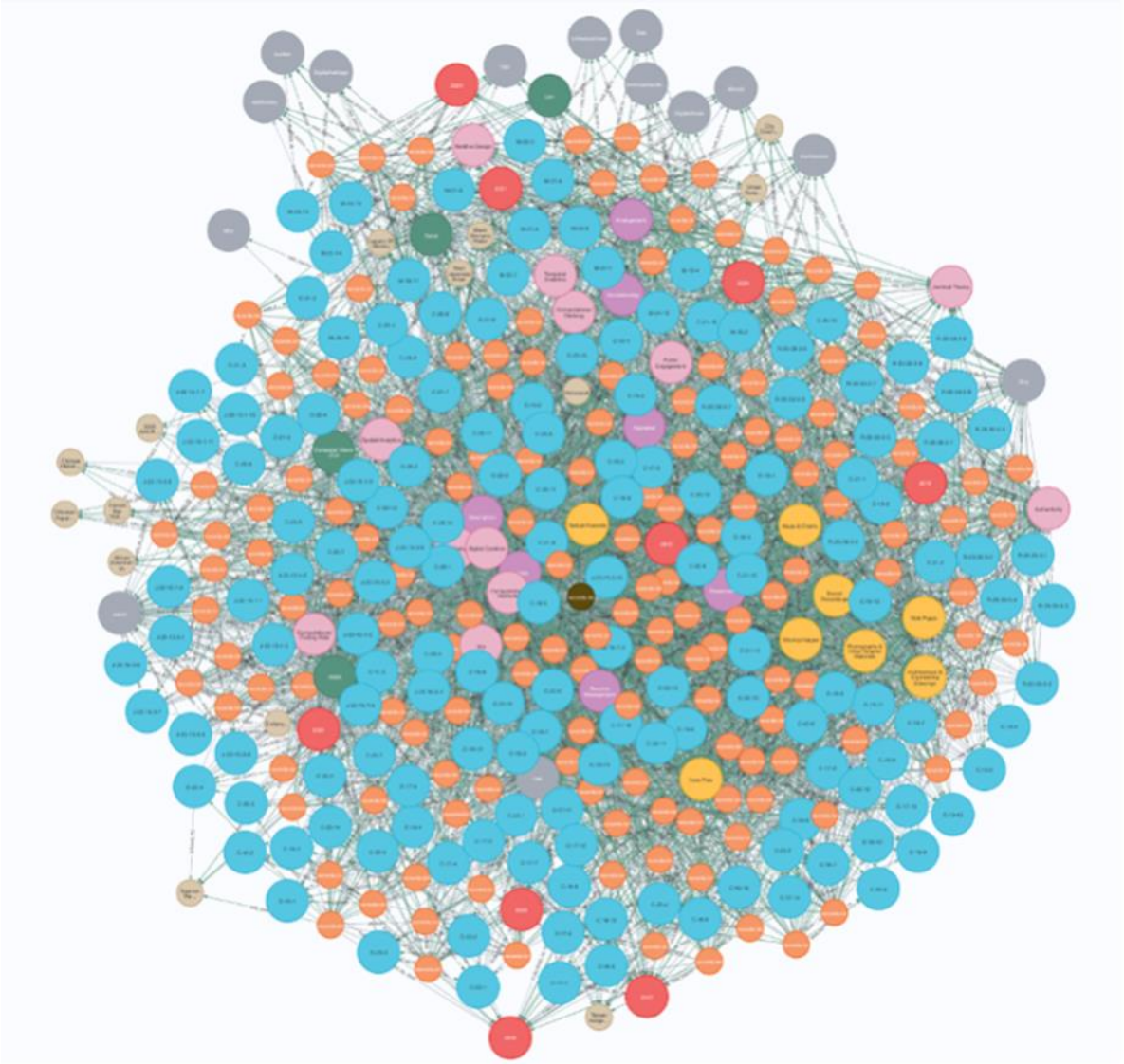


Figure No. 2:Entities and relationships are used to create a knowledge base (annotated papers output as plain text).

ISSN 3089-6002. Published by Universitas Muhammadiyah Sidoarjo

Figure No. 3: entity and relationship knowledge diagram.

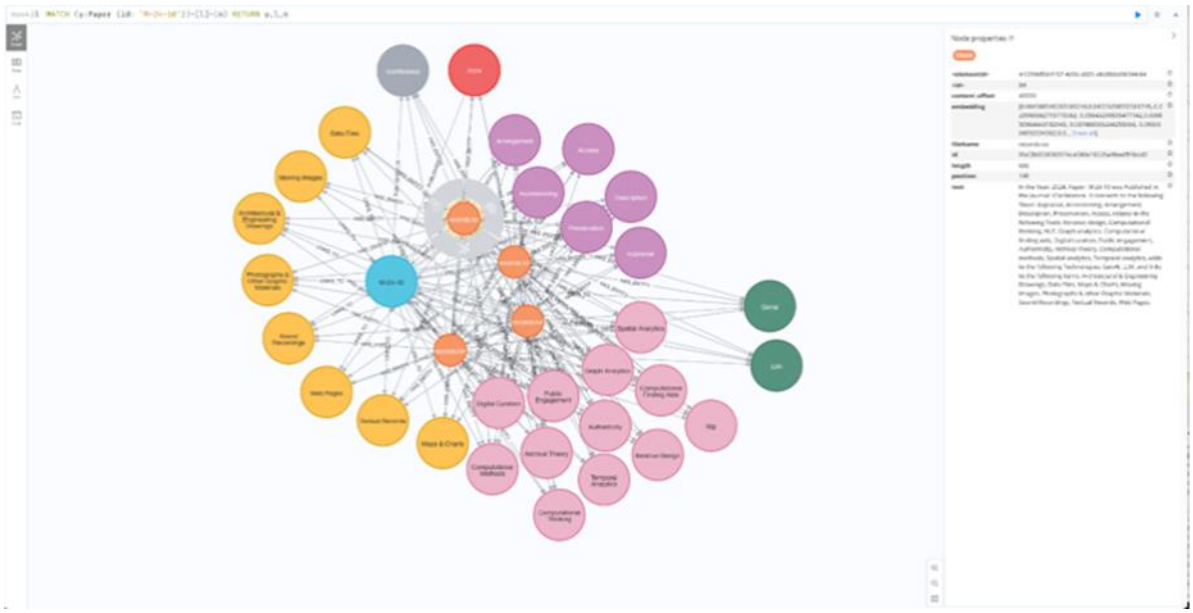


Figure No. 4: The knowledge map's search results.

ISSN 3089-6002. Published by Universitas Muhammadiyah Sidoarjo

Table 1: A mathematical review using dialogic queries.

Which paper IDs are related to 2024? The paper IDs related to the year 2024 are M-24-12, M-24-10, M-24-13, M-24-15, and M-29-9.
From these paper IDs which items are there? The paper IDs mentioned (M-24-10, M-24-12, M-24- 13, M-24-15, and M-29-9) link to the following items: • M-24-10: Architectural & Engineering Drawings, Data Files, Maps & Charts, Moving images, Photographs & Other Graphic Materials, Sound Recordings, Textual Records, Web Pages • M-24-12: Textual Records. • M-24-13: Textual Records. • M-24-15: Maps & Charts, Textual Records. • M-24-9: Maps & Charts: Textual Records.
From these paper IDs which Steps are there? The paper IDs you mentioned are associated with the following Steps: • M-24-10: Access, Accessioning, Appraisal, Arrangement, Description, Preservation • M-24-12: Access, Accessioning, Arrangement, Description • M-24-13: Access, Accessioning, Arrangement, Description • M-24-15: Access, Description • M-24-9: Access, Description
From these paper IDs which Tools are there? The paper IDs you mentioned are associated with the following Tools: • M-24-10: Computational Methods, Computational Thinking, Digital Curation, Graph Analytics • M-24-12: Digital Curation • M-24-13: Digital Curation • M-24-15: Digital Curation • M-24-9: Computational Thinking, Digital Curation, Graph Analytics

Tableau was used to create trend analyses and descriptive statistics to supplement the chart-based study. Key study subjects, archive functions, and computational techniques were identified using the generated charts (Figure 6). In order to evaluate the effect of digital transformation on archive practices, the distribution of both original and digital collections was also looked at (Figure 5).

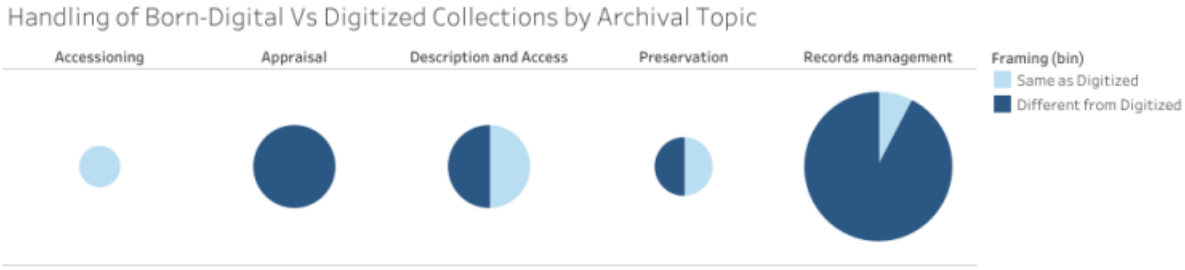


Figure No. (5): Original numerical sets versus numerical sets by archiving subject (Example of a knowledge graph query).

ISSN 3089-6002. Published by Universitas Muhammadiyah Sidoarjo

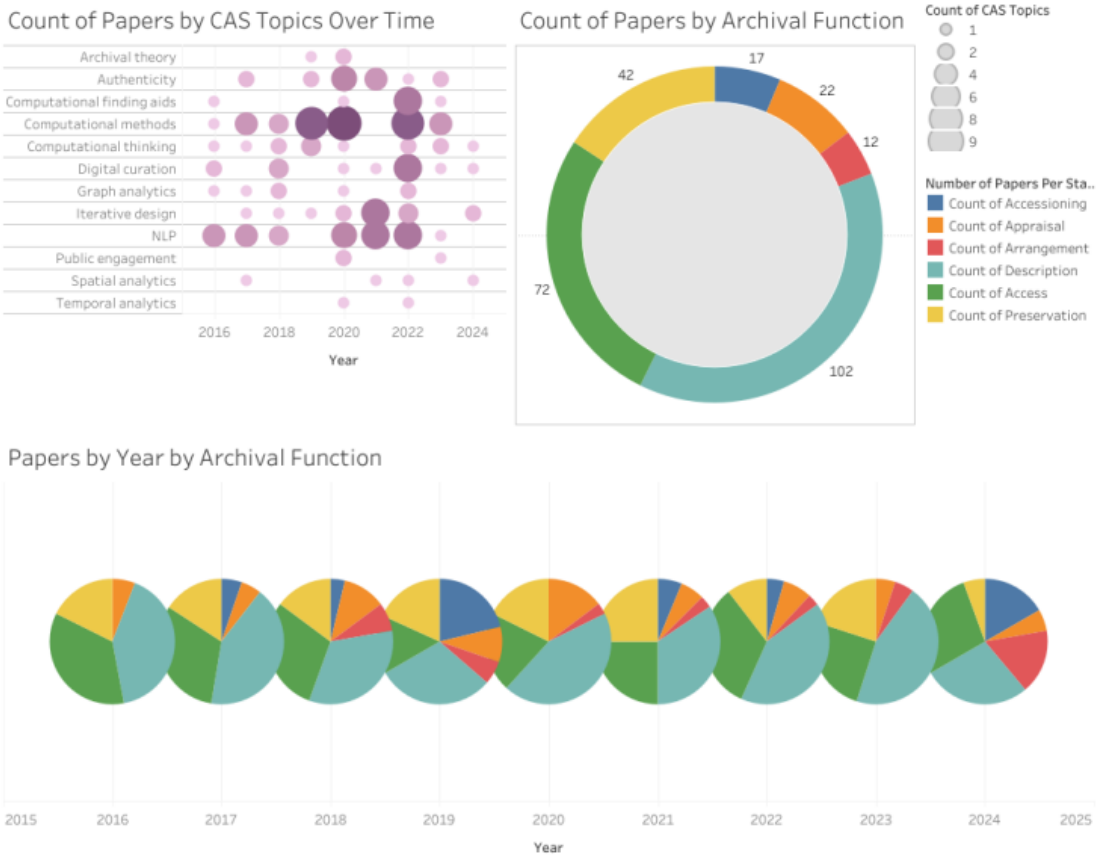


Figure (6): Tableau dashboard for computer review of CAS papers (original digital sets versus digital sets).

3.2 Machine Learning for Quality Assurance in Catalogs

The use of machine learning to automatically check the caliber of catalog records in multilingual libraries was the focus of the second phase. Millions of catalog records with transliterated and Cyrillic-written metadata fields made up the dataset. Following their extraction from a tab-separated dataset, the records underwent a thorough preprocessing procedure that included category coding, feature engineering, data cleaning, and normalization.

Feature segmentation algorithms were used to transform textual metadata information into numerical representations. Custom coding books and stereotype-based assignments were used to code nominal variables, such as language codes and country of publishing. Only a few information fields were kept for model building due to the computational constraints of large-scale text processing.

During the early trials, a number of machine learning methods were assessed. The Random Forest Classifier was chosen after comparative testing because of its classification performance, scalability, and resilience. To enhance probability estimation and assist human decision-making, the model was calibrated using a sigmoid calibration function.

Confusion matrices, classification accuracy, and ROC-AUC (Riot Characteristics Curve Area) were used to assess the model's performance. Figure 7, which displays the model's performance both before and after calibration, presents the findings of the comparative assessment of several feature configurations.

ISSN 3089-6002. Published by Universitas Muhammadiyah Sidoarjo

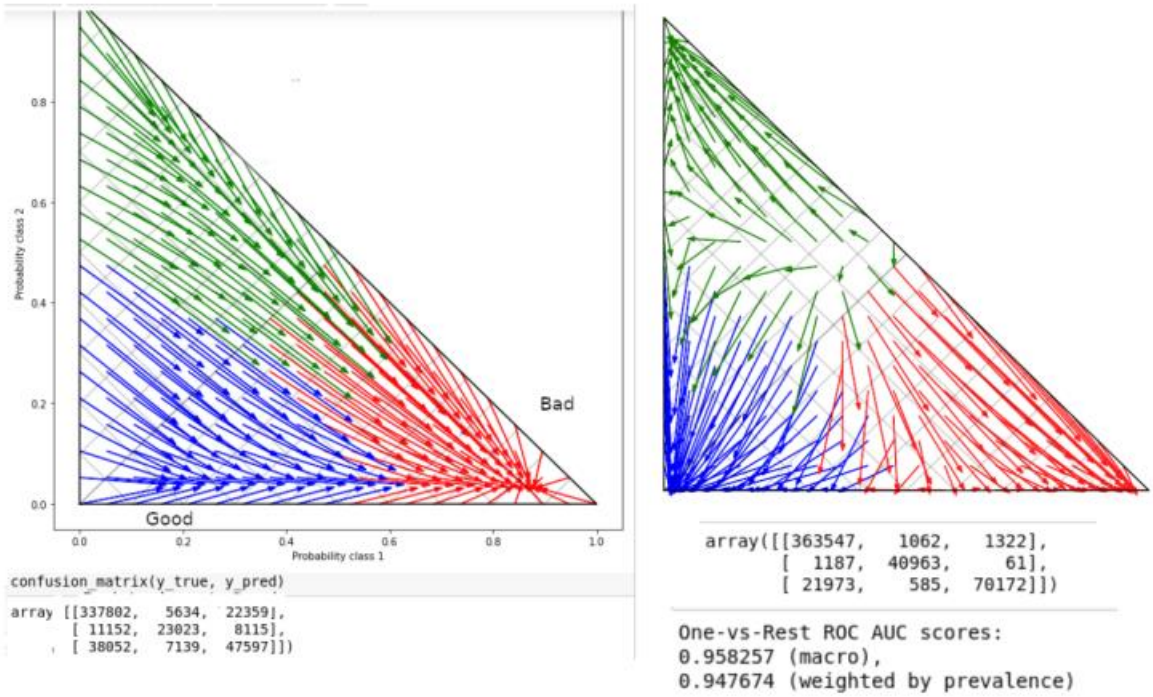


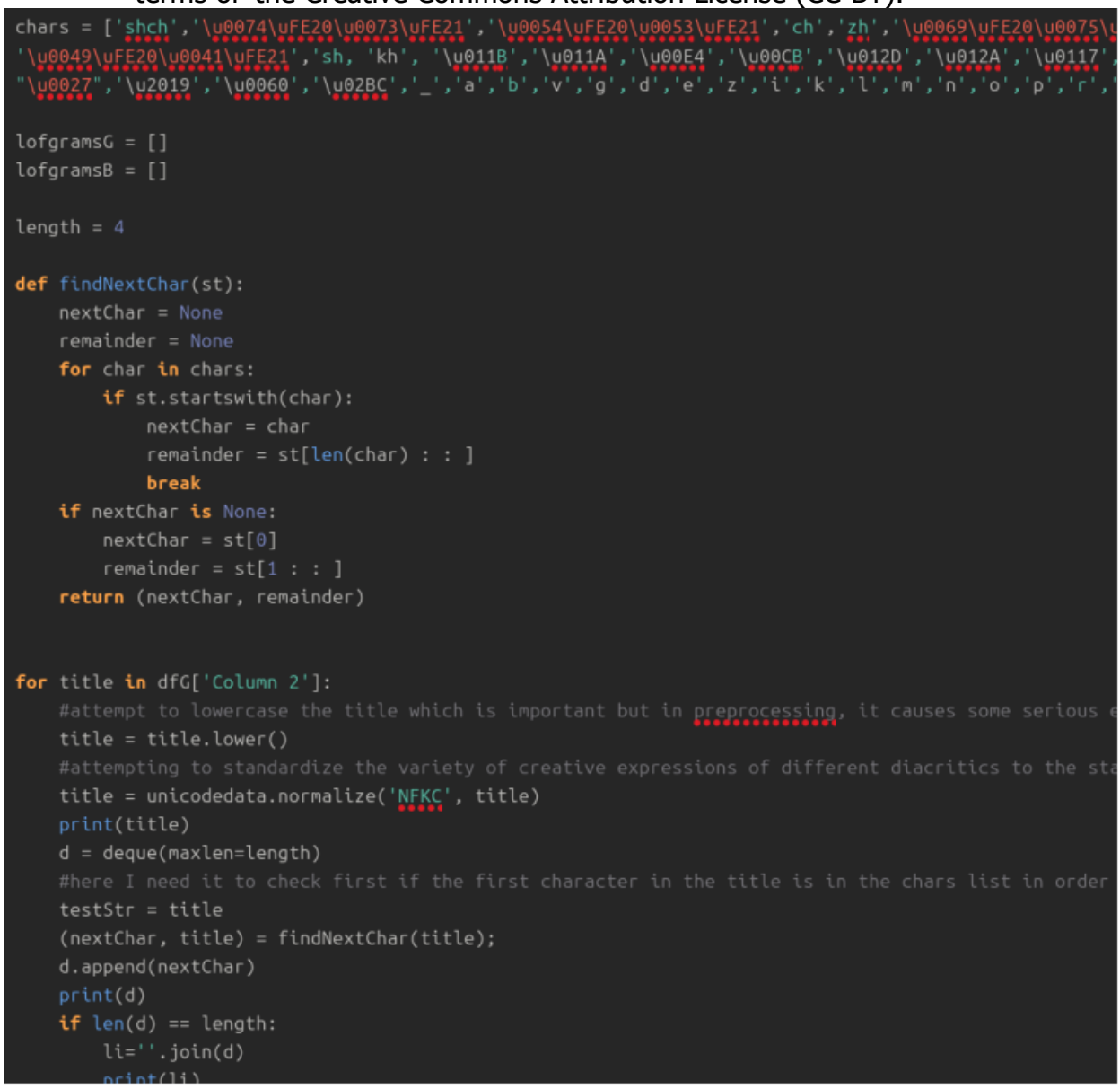
Figure No. (7): Comparison of values before and after calibration when using addresses (segmented 28) and 3 variables (left) versus addresses (segmented 23) and 8 variables (right) (results of machine learning model evaluation).

3.3 Statistical Language Modeling for Error Identification and Fixing

To find and fix transliteration problems in the Cyrillic indexing records, statistical language models were created after the records were classified. Since many transliteration systems lacked access to typical spell-checking tools, neogram frequency analysis at the character level was employed.

The transliterated text had to be divided into segments based on the values of the original Cyrillic letters as part of the preparation procedure. To recognize transliterated character sequences, such as Unicode diacritical marks and unique phonetic symbols, a unique method was created. The processing logic utilized to produce the right character sequences is shown in Figure 8.

ISSN 3089-6002. Published by Universitas Muhammadiyah Sidoarjo



```
chars = ['shch','\u0074\uFE20\u0073\uFE21','\u0054\uFE20\u0053\uFE21','ch','zh','\u0069\uFE20\u0075\u
'\u0049\uFE20\u0041\uFE21','sh', 'kh', '\u011B','\u011A','\u00E4','\u00CB','\u012D','\u012A','\u0117',
"\u0027",'\u2019','\u0060','\u02BC','_','a','b','v','g','d','e','z','i','k','l','m','n','o','p','r','

lofgramsG = []
lofgramsB = []

length = 4

def findNextChar(st):
    nextChar = None
    remainder = None
    for char in chars:
        if st.startswith(char):
            nextChar = char
            remainder = st[len(char) : : ]
            break
    if nextChar is None:
        nextChar = st[0]
        remainder = st[1 : : ]
    return (nextChar, remainder)


for title in dfG['Column 2']:
    #attempt to lowercase the title which is important but in preprocessing, it causes some serious e
    title = title.lower()
    #attempting to standardize the variety of creative expressions of different diacritics to the sta
    title = unicodedata.normalize('NFKC', title)
    print(title)
    d = deque(maxlen=length)
    #here I need it to check first if the first character in the title is in the chars list in order
    testStr = title
    (nextChar, title) = findNextChar(title);
    d.append(nextChar)
    print(d)
    if len(d) == length:
        li=''.join(d)
        print(li)
```

Figure No. (8): Code for splitting the translated text according to the values of the original characters n (translated text processing algorithm).

Then, for neograms of different lengths, character frequency tables were made. Based on substitution probabilities and Leuvenstein distance computations, these statistical models were used to identify unlikely character sequences and recommend corrections. This method preserved compatibility with current indexing procedures while enabling the automated detection of any phonetic transmission mistakes.

3.4 Improving Semantic Information

The third stage concentrated on applying semantic tools to enhance archival metadata. The metadata that already existed was converted into linked data structures that could depict the connections between archival entities. People, locations, organizations, events, and historical resources were connected under a single semantic framework using cognitive grap happroaches.

Contextual linkages between archival objects were automatically created by applying entity extraction and relationship identification techniques to metadata entries. The

ISSN 3089-6002. Published by Universitas Muhammadiyah Sidoarjo

resulting semantic structures strengthened contextual comprehension of archive materials, promoted exploratory exploration, and improved information retrieval.

3.5 Workflow for Research

There are five consecutive stages to the entire workflow:
• Archival items are gathered and categorized.
• Knowledge map creation and analysis.
• Machine learning-based index record preprocessing and classification.
• The statistical identification and rectification of faults in audio translation.
• Semantic enrichment of metadata using knowledge maps and associated data approaches.

Together, these stages offer a thorough framework for evaluating how artificial intelligence may improve metadata quality, decrease the backlog of archival records, and improve access to digital cultural heritage resources.

1. Findings and Conversation

The study's findings show how artificial intelligence (AI) technology can significantly enhance archival and library workflows through automated quality control, computational analysis, and semantic metadata enrichment. Together, the results of the three research phases show that AI can help manage large-scale digital and digital original collections while supporting professional archive procedures rather than taking their place.

A growing emphasis on using computational techniques for archival description, access, preservation, and record management was identified through computational analysis of the computational archiving literature. Digital preservation, information retrieval, and metadata augmentation were among the most often studied topics, according to an analysis of annotated papers. The knowledge map created at this time (Figure 2.3) effectively revealed the connections between computational methods, archive functions, and research themes, offering a more thorough grasp of advancements in the area. Additionally, the growing significance of original digital records as a primary focus of archive study was emphasized by the Tableau dashboard (Figure 2.7).

These results highlight the need for scalable solutions that can handle quickly expanding digital collections and represent the continuous digital transition of memory institutions.

The identification of the "vanishing box problem," a notion that describes the loss of contextual and organizational frameworks typically given by physical archival arrangements, was a major finding of the literature review. Preserving the contextual relationships that were previously communicated through boxes, folders, and file systems becomes more difficult as archives collect more original digital assets. The findings imply that knowledge maps and semantic approaches could offer useful tools for preserving and rebuilding these connections in digital settings, hence upholding archive concepts like source and original arrangement.

The study's machine learning component showed that large-scale index management can be efficiently supported by automated quality control systems. The random forest classifier produced the most dependable results after comparing other classification techniques. Model calibration enhanced prediction reliability and increased the agreement between expected and actual results, as evidenced by the comparative evaluation results displayed in Figure 3.2. The enhanced model effectively differentiated between records that met predetermined indexing criteria and those that needed quality control intervention. These findings show that by concentrating experts' attention on records that are more

ISSN 3089-6002. Published by Universitas Muhammadiyah Sidoarjo

likely to include errors, machine learning can greatly minimize the manual labor required to analyze millions of indexing entries.

A tiny fraction of records with quality problems were mistakenly categorized as acceptable, according to additional analysis of the model's performance. When these examples were analyzed, it became clear that many of them were from Russian-language libraries where voice transcriptions were either missing or recorded inaccurately. These records indicate an understandable source of ambiguity rather than a fundamental problem in the model because voice transcribe fields are frequently optional under cataloging requirements. The use of machine learning as a decision support tool that can boost productivity while preserving acceptable levels of accuracy is thus supported by the overall findings.

By addressing phonetic transliteration issues that are challenging to identify with conventional techniques, the statistical language modeling phase improved indexing quality control. Despite the inclusion of Unicode diacritical marks and multi-character translation rules, the translated text was effectively divided into meaningful character sequences by the proprietary preprocessing technique, as shown in Figure 3.3. Unusual character combinations and possible spelling mistakes were found using frequency-based analysis of these character sequences (n-grams). The system was able to recommend possible corrections by comparing the frequency distributions of correctly indexed data and those that needed intervention, which showed distinct statistical disparities. These findings show that language modeling approaches can be a useful tool for managing multilingual metadata, especially in communities that lack access to specialized linguistic knowledge.

5. Restrictions

Although the findings show how artificial intelligence might enhance archiving and library procedures, it is important to recognize some limitations.

Initially, multilingual cataloging datasets, mostly composed of transcribed and phonetically translated Cyrillic-based records, were used to create and assess the machine learning models. As a result, different languages, metadata standards, or institutional settings may result in varied model performance. Before the model is used in more extensive settings, it must undergo additional training and validation.

Second, n-gram analysis and character frequency distributions obtained from current indexing records are the foundation of statistical language models. This method may be less successful in resolving uncommon linguistic patterns, specialist terminology, or inconsistencies resulting from different phonetic translation standards, even though it is helpful in recognizing frequent phonetic translation problems. Additionally, model performance is significantly impacted by the quality and accessibility of training data.

Third, the consistency and completeness of current metadata are critical components of the metadata enrichment system. The accuracy of automated entity extraction and relationship identification can be lowered by incomplete records, unclear entity names, or lacking contextual information. As a result, human evaluation is still necessary to confirm semantic linkages and guarantee the quality of metadata.

Lastly, artificial intelligence-related ethical issues need to be considered. Inaccurate classifications, false links between archival things, and biases inherited from training data are all possible outcomes of AI systems. Transparency, explainability, and expert supervision are therefore still crucial components of any AI-powered archiving procedure.

6. Upcoming Projects

Future studies should examine how new AI technologies, such as Retrieval-Enhanced Information Generation (RAG) systems and Large Language Models (LLMs), can be

ISSN 3089-6002. Published by Universitas Muhammadiyah Sidoarjo

incorporated into archival and library operations. These technologies have the ability to enhance reference services, facilitate natural language interface with archival collections, and automate difficult descriptive jobs.

To assess the suggested framework over a larger range of languages, metadata schemes, and collection kinds, more research is required. Comparative studies including museums, libraries, archives, and other cultural heritage organizations would offer insightful information about how broadly applicable this strategy is.

The creation of explainable AI methods for archival applications should also be investigated in future studies. Enhancing automated decision-making's transparency will boost the confidence of information professionals and encourage the proper use of AI technologies. The accessibility and integration of digital cultural heritage resources can also be improved by research in automated source tracking, semantic consistency, and cross-institutional knowledge maps.

The creation of human-AI collaboration models that precisely outline the responsibilities of professional archivists and automated systems is another encouraging avenue. In digital environments, these models can help develop best practices for striking a balance between accountability, efficiency, and ethical governance.

7. Conclusions

In order to address current issues in libraries and archives, this study investigated the use of artificial intelligence (AI) approaches. The study showed how AI can help manage digital collections and large-scale digital original collections by combining computational literature analysis, machine learning-based cataloging quality control, statistical language modeling, and semantic metadata augmentation.

The findings demonstrated that computational techniques can boost metadata quality, increase the effectiveness of archival procedures, and enable efficient access to cultural heritage resources. A more comprehensive contextual representation of archive data was made possible by cognitive graph approaches, and machine learning models were successful in identifying records that needed quality control intervention. By automatically identifying and fixing translation problems, statistical language models also helped to increase the correctness of metadata.

The expanding significance of computational archiving as a framework for incorporating cutting-edge computational techniques into archival practice was also emphasized by the study. The results show that AI methods are especially useful in tackling the problems brought on by growing collection sizes, complicated metadata, and users' need for quick information access.

Despite these benefits, the study highlights that AI should be used in conjunction with professional knowledge rather than as a substitute. In order to ensure authenticity, maintain context, make moral decisions, and ensure quality, archivists and librarians are still crucial. As a result, the most successful archive environments of the future will be those that integrate AI's analytical powers with information workers' expertise.

To sum up, using artificial intelligence into archiving and library procedures is a viable way to lessen data overload, enhance metadata quality, and increase access to digital cultural property. Collaboration between humans and AI will become more crucial as digital collections expand in order to preserve and make the documentary record of society available to everyone.

ISSN 3089-6002. Published by Universitas Muhammadiyah Sidoarjo

Originality Statement

The author[s] declare that this article is their own work and to the best of their knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the published of and other published materials, except where due acknowledgement is made in the article. Any contribution made to the research by others, with whom author[s] have work, is explicitly acknowledged in the article.

Conflict of Interest Statement

The author[s] declare that this article was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

[1] A. S. Mohamed, "Artificial Intelligence Technologies in Library and Information Programs: An Analytical Study," Journal of the Humanities Sector, vol. 30, no. 1, pp. 991–1076, 2022, doi: 10.21608/JSH.2022.276724.

[2] Y. A. A. Hassan, "Artificial Intelligence: Foundations and Areas of Application in Libraries and Information Sciences," Arab International Journal of Information Technology and Data, vol. 2, no. 2, pp. 209–218, 2022.

[3] S. Al Jabri and A. B. S. Al Hinai, "Applications of Artificial Intelligence Technologies in Information Services in Libraries and Information Centers: Academic Libraries as a Model," Arab International Journal of Information Technology and Data, vol. 3, no. 3, pp. 45–58, 2023.

[4] S. A. Fahmy, "Chatbots and Their Uses in Information Institutions: An Exploratory Analytical Study," Scientific Journal of Libraries, Documents and Information, vol. 5, no. 15, pp. 269–310, 2023, doi: 10.21608/JSLMF.2022.172768.1147.

[5] L. S. S. Sami, "Analysis 'AI-Based Emotions as a Tool for Evaluating the Effectiveness of Information Services'," Journal of Library, Documentation and Information Science, vol. 4, no. 11, pp. 7–32, 2022, doi: 10.21608/JSLMF.2022.147720.1125.

[6] S. Dedication, "Applications of Artificial Intelligence Systems in Content Analysis and Indexing: An Applied Study of Natural Language Processing Systems," Journal of Libraries, Documents and Information, vol. 4, no. 11, pp. 89–120, 2022, doi: 10.21608/JSLMF.2022.131460.1111.

[7] M. F. A. al-Hadi, Digital Libraries, Cairo, Egypt: Egyptian-Lebanese House, 2019.

[8] M. F. A. al-Hadi, Introduction to Information and Library Science, Cairo, Egypt: Egyptian-Lebanese House, 2018.

ISSN 3089-6002. Published by Universitas Muhammadiyah Sidoarjo

[9] M. F. A. al-Hadi, Digital Libraries and Knowledge Management, Cairo, Egypt: Arab Group for Training and Publishing, 2020.

[10] A. M. al-Shami and H. al-Sayyid, Encyclopedic Dictionary of Library and Information Terms, Riyadh, Saudi Arabia: King Fahd National Library, 2017.

[11] K. A. al-Rahman, Electronic Archiving and Preservation of Digital Documents, Riyadh, Saudi Arabia: Al-Rushd Library, 2021.

[12] F. A. al-Jawad, The Shift Towards Electronic Archives in Government Institutions, Amman, Jordan: Dar Safaa, 2019.

[13] N. Mahmoud, Preserving and Providing Access to Digital Heritage in the Electronic Environment, Cairo, Egypt: University Knowledge House, 2022.

[14] A. al-Ali, The Semantic Web and the Organization of Digital Knowledge, Amman, Jordan: Dar al-Masirah, 2021.

[15] M. al-Salem, "The Role of Artificial Intelligence in Developing Digital Library Services," Journal of Digital Information Management, vol. 14, no. 3, pp. 112–125, 2023.

[16] N. A. al-Latif, "Metadata and Its Impact on Improvement Digital Information Retrieval," International Journal of Information Studies, vol. 9, no. 2, pp. 78–94, 2022.

[17] S. Al-Harbi, "Machine Learning Applications in Knowledge Organization and Information Retrieval," Arabian Journal for Science and Engineering, vol. 47, no. 4, pp. 3105–3118, 2022.

[18] M. Al-Qahtani, "Using Metadata in Building Institutional Digital Repositories," Journal of Information Science Theory and Practice, vol. 11, no. 1, pp. 55–69, 2023.

[19] A. Al-Jabri, "Big Data Management in Libraries and Information Centers," Journal of Enterprise Information Management, vol. 36, no. 5, pp. 1240–1258, 2023.

[20] H. Al-Alaq, "Smart Technologies in Managing Digital Documents and Archives," Computers in Human Behavior, vol. 139, p. 107520, 2023.