ENGENHARIAS, TECNOLOGIA, GESTÃO E TURISMO
ENGINEERING, TECHNOLOGY, MANAGEMENT AND TOURISM
INGENIERÍA, TECNOLOGÍA, ADMINISTRACIÓN Y TURISMO

millenium

Millenium, 2(Edição Especial Nº23)

en

REVOLUCIONAR A ORGANIZAÇÃO DO CONHECIMENTO COM APRENDIZAGEM AUTOMÁTICA PARA METADADOS, CLASSIFICAÇÃO E RECUPERAÇÃO EM BIBLIOTECAS

REVOLUTIONIZING KNOWLEDGE ORGANIZATION WITH MACHINE LEARNING FOR METADATA, CLASSIFICATION, AND RETRIEVAL IN LIBRARIES

REVOLUCIONANDO LA ORGANIZACIÓN DEL CONOCIMIENTO CON APRENDIZAJE AUTOMÁTICO PARA METADATOS, CLASIFICACIÓN Y RECUPERACIÓN EN BIBLIOTECAS

Rituraj Jain[1] *https://orcid.org/0000-0002-5532-1245*
Damodharan Palaniappan[1] *https://orcid.org/0009-0003-0721-3068*
Kumar Parmar[1] *https://orcid.org/0000-0002-2502-5680*
T Premavathi[1] *https://orcid.org/0009-0003-0172-2021*

[1] Marwadi University, Gujarat, India

Rituraj Jain – jainrituraj@yahoo.com | Damodharan Palaniappan – damomtpcse@gmail.com | Kumar Parmar – kumar4k@gmail.com | T Premavathi – prema.cse05@gmail.com

Corresponding Author:
Rituraj Jain
Marwadi University
360003 - Rajkot- India
jainrituraj@yahoo.com

RECEIVED: 27th September, 2024
REVIEWED: 20th May, 2026
ACCEPTED: 22nd May, 2026
PUBLISHED: 23rd June, 2026

DOI: https://doi.org/10.29352/mill0223.e37496

RESUMO

Introdução: O artigo irá rever como as transformações da aprendizagem automática levarão a modernização à abordagem dos sistemas de bibliotecas e contribuirão para métodos de organização do conhecimento para melhores processos de recuperação de informação.
Objectivo: Discutir como as técnicas de ML podem ser aplicadas no desenvolvimento das funções necessárias da biblioteca, como a criação de metadados, sistemas de classificação e recuperação de informação, de forma mais eficiente e precisa.
Métodos: As técnicas utilizadas envolvem um conhecimento profundo de diversos algoritmos de aprendizagem automática que incluem a PNL na geração de metadados e redes neuronais para classificação.
Resultados: Os estudos de caso provam de facto que estas técnicas foram aplicadas e implementadas em sistemas de bibliotecas reais, como revelado pelas principais bibliotecas que começaram a implementar tecnologias de ML. Estas implementações resultam em maior eficiência, precisão e experiência do utilizador. No entanto, esta pesquisa também aponta desafios como os dados de treino enviesados e a necessidade de transparência nas decisões algorítmicas.
Conclusão: São muitas as promessas do ML para o futuro das bibliotecas; a sua integração responsável necessita de uma abordagem criteriosamente equilibrada, com preocupação pelas questões éticas.

Palavras-chave: classificação; recuperação de informação; organização do conhecimento; aprendizagem automática; metadados

ABSTRACT

Introduction: This paper is going to explore how machine learning transformations will take library modernization a step further and help in knowledge organisation methods for better information retrieval processes.
Objective: To discuss how to make use of techniques in ML to develop the required library functions like metadata creation, classification systems, and information retrieval in an efficient and accurate manner.
Methods: The techniques used are based on a thorough knowledge of different machine learning algorithms, such as NLP for metadata generation and neural networks for classification.
Results: The case studies indeed prove that these techniques have been applied to and implemented in real library systems, as revealed by major libraries that have begun to implement ML technologies. These deployments lead to more efficiency, accuracy, and user experience. There are also issues highlighted in this research that could be problematic: biased training data, and transparency in algorithmic decisions.
Conclusion: There is a lot of promise in the future of libraries with the use of ML, and integration should be guided with a judicious balance, taking care of ethical issues.

Keywords: classification; information retrieval; knowledge organization; machine learning; metadata

RESUMEN

Introducción: El artículo analizará cómo las transformaciones del aprendizaje automático modernizarán el enfoque de los sistemas bibliotecarios y contribuirán a los métodos de organización del conocimiento para mejorar los procesos de recuperación de información.
Objetivo: Analizar cómo se pueden aplicar las técnicas de aprendizaje automático en el desarrollo de las funciones bibliotecarias requeridas, como la creación de metadatos, los sistemas de clasificación y la recuperación de información, de una manera más eficiente y precisa.
Métodos: Las técnicas utilizadas implican un profundo conocimiento de varios algoritmos de aprendizaje automático que incluyen el procesamiento del lenguaje natural en la generación de metadatos y las redes neuronales para la clasificación.
Resultados: Los estudios de caso demuestran que estas técnicas se han aplicado e implementado en sistemas bibliotecarios reales, como lo revelan las principales bibliotecas que han comenzado a implementar tecnologías de aprendizaje automático. Dichas implementaciones dan como resultado una mayor eficiencia, precisión y experiencia del usuario. Sin embargo, esta investigación también señala desafíos como los datos de entrenamiento sesgados y la necesidad de transparencia en las decisiones algorítmicas.
Conclusión: El aprendizaje automático es muy prometedor para el futuro de las bibliotecas; su integración responsable necesita un enfoque juiciosamente equilibrado que tenga en cuenta las cuestiones éticas.

Palabras clave: clasificación; recuperación de información; organización del conocimiento; aprendizaje automático; metadatos

DOI: https://doi.org/10.29352/mill0223.e37496

INTRODUCTION

Library science depends on knowledge organization for the arrangement and making huge informative resources more accessible. Description of the library resources with the help of metadata, classification into meaningful classes, and effective methods of retrieval are the hallmarks of knowledge organization. Metadata such as title, author, and topic matter, along with the publication date, depend on the description of library materials. It is this descriptive information that makes access to relevant materials easier. The classification schemes segregate resources into hierarchical groups and subgroups according to the shared attributes. Consequently, this makes browsing and discovery easier. Retrieval systems use metadata and classification as a means for the user to find material, normally using advanced search algorithms together with indexing approaches (Guerra-García et al., 2023). The recent development within Machine Learning has introduced powerful techniques that may potentially change the current ways of organizing knowledge in libraries. Machine Learning algorithms are distinctively superior at pattern finding, prediction, and automation of hard operations, making them ideal methods for the preparation of metadata, classification, and information retrieval frameworks. This paper presents how machine learning organizes and enriches library knowledge. It looks at how machine learning can automate metadata creation, enhance the accuracy and speed of classification, and provide more sophisticated and user-friendly retrieval systems. This present study applies machine learning in the enhancement of library knowledge organisation systems for greater productivity, effectiveness, and in meeting user demands. Figure 1 shows how library science knowledge was integrated with machine learning in enhancing metadata production, classification, and information retrieval. Thus, machine learning influences metadata, categorization, and retrieval systems in its first phase.

Machine learning automated the creation of metadata for publication date, author, and subject in the metadata pipeline. Metadata is important in the categorization of resources and is required for complete information on resources to be discovered or accessed. Machine learning enhances the paths used in the classification of resources. This approach arranges resources into hierarchies and subcategories in a taxonomy for convenience during browsing and discovery. Hence, consumers can access the relevant information more firmly (Rashid et al., 2022).

Machine learning in retrieval systems enhances the methods applied in searches. With state-of-the-art algorithms and indexing, these methods ensure that the search outcome is both efficient and relevant. Retrieval systems yield accurate and complete results since they understand user queries and index library resources efficiently (Xie, 2023). This is also a multimodal knowledge organization approach in library science that has proved that machine learning will be able to automate metadata development, categorization of resources, and retrieval of information. Machine learning approaches might indicate the potential of improving library information management throughout all steps of the knowledge organisation process.

1. LITERATURE REVIEW

1.1 Traditional Library Knowledge Organization

Libraries have always been involved in information organization and access through systematic methods. This includes, in the traditional sense, manual metadata creation, resource cataloging, and building information retrieval systems (Pessin et al., 2023). The librarian methodically creates metadata for every single library resource with great detail based on the set standards by MARC and Dublin Core. Such an approach identifies and embeds in the resource's descriptive features, subject headings, and controlled vocabulary phrases, so that resources are classified consistently with ease (Park & Lu, 2009).

The Dewey Decimal System and the Library of Congress Classification organize resources by subject to make sure resources are put in order through hierarchical schemes. A systematic way that encourages browsing and collocation for similar subjects, therefore causes ease of finding information related to a subject (H. R et al., 2019). Generally, library retrieval has adopted card catalogues and, recently, online public access catalogues. Appropriate resources are retrieved by means of keyword-based queries on metadata fields and on controlled vocabularies (Deolekar & Dangare, 2018)

1.2 Library Science Machine Learning: A New Era

Machine learning has opened vistas in library science, bringing revolutionary methods for the knowledge organization of resources. The machine-learning algorithm can extract useful data from large texts, photos, and library resources to create metadata. Natural language processing methods, such as Named Entity Recognition and Topic Modelling, may detect and capture key concepts, authors, and subject headings from texts with minimum human interaction (Wu et al., 2023)

More accurate classification: through machine learning algorithms, study the structure and pattern of existing classification systems in order to develop expertise in categorizing new resources. For instance, the support vector machines and Naive Bayes classifiers will classify the incoming documents into pre-specified categories by including knowledge from previous classifications, as shown in (Rashid et al., 2022). Knowledge Retrieval Improvement: Machine learning algorithms can learn user search habits, understand related phrases or synonyms, and personalize queries to improve search results. Other collaborative filtering or content-based recommendation systems can also offer suitable resources based on user preferences and past system interactions (Limbu et al., 2006).

DOI: https://doi.org/10.29352/mill0223.e37496

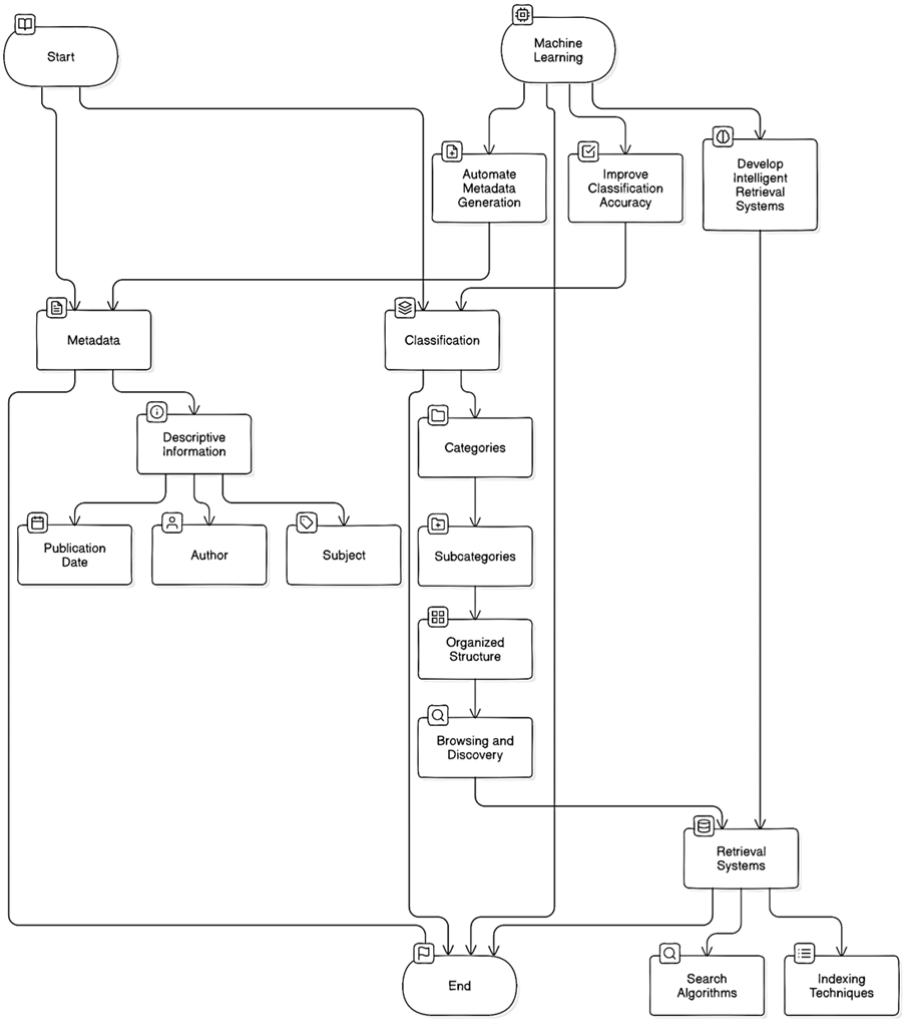

Figure 1 - Knowledge Organization in Library Science

1.3 Key ML-Powered Knowledge Organization Studies Results

Studies show machine learning can change knowledge organisations. AI and machine learning in libraries were thoroughly examined. These technologies increase metadata production, resource discovery, and user service, according to 32 scholarly studies (Das & Islam, 2021). Machine learning in libraries has been investigated for its history, development, and prominent researchers (L. Weinryb-Grohsgal, 2020). Automation, productivity, and user interactions are improved by machine learning. Qin says AI improves knowledge representation and organisation (Qin, 2020). Ontologies and semantic technologies in intelligent knowledge systems are investigated.

Machine learning helps novice researchers retrieve knowledge (Audeh et al., 2021). A paradigm for citation networks and third-party metadata outside content relevance is proposed in this research. More effective and personalised search results. AI and machine learning can semi-automate library subject indexing (Ahmed et al., 2023). The Annif platform and Library of Congress Subject Headings dataset are used to train the system on a massive bibliographic corpus. An AI/ML system can manage a growing library collection and reduce subject indexing, the study revealed.

The authors study how machine learning-enhanced knowledge management affects organisational innovation and industrial development. An ontology-based multilayer association and machine learning-inspired data mining method is presented and compared to existing algorithms. This study demonstrated a positive association between knowledge management levels, with acquisition and sharing being strongest. According to Zhou et al. (2020), the proposed knowledge management model boosts organisational creativity and industrial development and provides guidance on how to improve company knowledge management and innovation. Authors explain how AI makes libraries "smart". They claim AI can improve library content discovery and resource management. The book also argues that AI can personalise services and predict user needs using massive data sets. (Yu et al., 2019) acknowledge that infrastructure costs and a lack of qualified staff make AI deployment problematic, but they highlight AI's benefits in creating intelligent libraries.

DOI: https://doi.org/10.29352/mill0223.e37496

In (Hjørland, 2021), writers explore integrating information retrieval and knowledge organisation to enhance access to information. Knowledge organisation represents academic knowledge using classification systems, thesauri, and ontologies, unlike search engines like Google, which employ algorithms and user involvement. The essay claims that domain-specific knowledge and paradigmatic arguments necessitate a philosophical approach to knowledge organisation. It concludes that information retrieval and knowledge organisation systems cannot be impartial in such debates, supporting philosophical views in both domains. (Kalisdha, 2024) examines AI and ML in library and information science. It covers how AI and ML are transforming library cataloguing, recommendations, and digital preservation. NLP and other text analysis and information extraction applications are examined. In the digital age, libraries will manage and offer access to material, but this article discusses ethical difficulties and AI use pitfalls.

In, (Zamani et al., 2022) introduces REML, a novel paradigm that integrates machine learning algorithms with traditional information retrieval approaches. Since REML can use various knowledge bases, it overcomes the drawbacks of individual machine learning models. REML enhances generality, robustness, and business search and personal information management interpretability. In, (Soni, 2023) addresses AI's growing impact on information and library science. It explores how AI is affecting library operations like retrieval, knowledge organisation, user services, and decision-making. Improved search engines, customised recommendations, and automatic cataloguing are AI benefits. It tackles ethical issues relating to biases, privacy, and appropriate AI integration in information and library science without a wider vision. Citation graphs and document information are examined to improve scientific literature suggestions. In particular, integrating these characteristics into machine learning models enhances recommender systems for users inexperienced with terms or search strategies (Audeh et al., 2020).

These show machine learning's transformative power to solve knowledge organisation problems. Automation, accuracy, and user experience will transform digital library organisation, management, and access via machine learning.

2. METHODS

In modern library science, the use of machine learning is directed towards metadata production, categorization systems, and information retrieval. Case studies and empirical examples are used to explore machine learning techniques, algorithms, and their applications. The review compares and contrasts the advantages and disadvantages of machine learning as applied to these and makes an observation about the transformative nature of machine learning for digital knowledge organization.

2.1 Machine Learning for Enhancing Metadata Creation

Library metadata has been a manual process that is now transforming with machine learning. The use of natural language processing and supervised learning, enabling the machine to build its own metadata, is studied in detail. This is how NLP can assist computers in recognizing the meaningful information in dataset texts and consequently enhance the ability of the machines to recognize keywords, subjects in the header, and summarize the document. Rather, supervised learning algorithms are used to learn metadata from labelled data sets and make predictions on other unlabeled resources. (Miah et al., 2022). The metadata creation flowchart, with the help of machine learning, is shown in Figure 2.

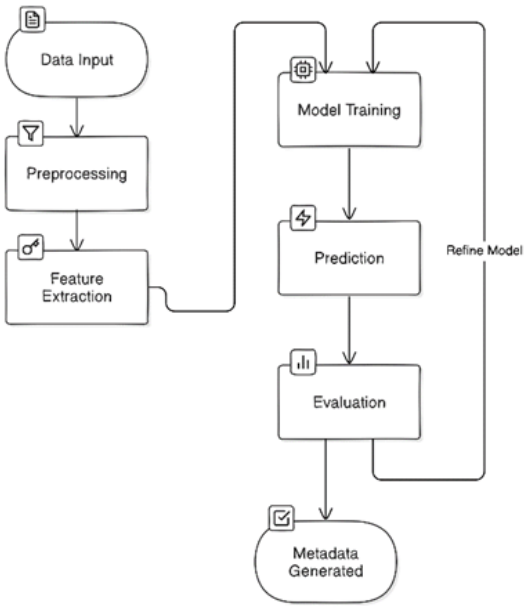

Figure 2 - Flowchart for ML Process in Metadata Creation

DOI: https://doi.org/10.29352/mill0223.e37496

Steps:

1. Collect textual data (e.g., books, articles, manuscripts).
2. Preprocessing: Data cleanup and formatting
3. NLP Feature Extraction: Extract keywords, headers, and summaries.
4: Train ML model using labelled data (supervised learning).
5. Prediction: Generate metadata for unlabeled resources using the model.
6. Evaluation: Determine accuracy and refine model (ongoing).

This section provides empirical examples of the use of machine learning. The National Library of the Netherlands uses machine learning to assign subject headings to digitized newspapers from the past. This considerably improves access (Rosa, 2024). Named entities in digitized texts are identified by the British Library by machine learning. It enhances data and research (Baker, 2015). While machine learning can enhance the efficiency, scalability, and consistency of metadata production, it also comes with challenges. Training data biases can drive metadata inequities and exacerbate the problem of lack of access to underprivileged areas. The machine learning models need to be validated and optimized for providing accurate and quality auto-generated metadata. The aims of this balanced view are to explore the advantages and disadvantages of using ML in metadata development.

2.2 Machine Learning in Classification Systems

Machine learning is replacing manual library classification with automated and sophisticated methods, according to the paper. There are several machine learning methods that could be used to automate classification tasks. Neural networks and decision trees are good supervised learning techniques to use for this task. Neural networks with human brain-like properties can identify complex patterns in data and categorize resources into classes with high accuracy after sufficient training (Guo et al., 2020). Decision trees can be analyzed with ease and can be used to simulate decisions and their impact (Rashid et al., 2022). Machine learning can classify real power for improving established classification schemes such as the Dewey Decimal System or the Library of Congress. These all-encompassing systems can become rigid and challenging to sustain consistency in the increasing number of information sources and their complexity. Machine learning algorithms can analyze classifications, uncover resource trends, and suggest refinements based on new knowledge. Such capabilities could tighten consistency and increase classification granularity for targeted information seekers. Classification-focused ML models must be evaluated for efficiency and efficacy. Common classification model performance measurements are accuracy, recall, and F1-score. These measures show whether a model efficiently classifies resources or misclassifies them. In a real-world library, an ML model must be assessed and updated to be accurate and efficient. At the National Library of Norway, an ML-based DDC number assignment system assigns DDC numbers to articles automatically, saving time in cataloguing. University of Illinois researchers constructed an ML model that accurately classifies research articles into academic fields, confirming ML's subject categorization potential.

2.3 Machine Learning in Information Retrieval

Library information retrieval is optimized for user access and relevancy using machine learning. Machine learning in library search is making search and retrieval systems smarter, more user-friendly, and Darwinian. Keyword-driven search methods sometimes fail to account for user intents and semantic relationships between topics. Machine learning algorithms, notably search and recommendation systems, have changed things. Machine learning analyses user behavior, search patterns, and content relationships to personalize search results, suggest relevant resources, and forecast future information demands. The review examines machine learning methods that improve information retrieval. The clustering technique organizes equivalent documents to help users navigate big collections and find relevant works they might have missed. Semantic search uses machine learning to comprehend search queries beyond keyword matching. Thus, they ensure accurate and relevant results and deliver the most authoritative resources faster, saving consumers time and effort. This is a user-centred review of machine learning-enhanced information retrieval methods. Machine learning personalizes search experiences by understanding user preferences, incorporating earlier engagements, and adapting to changing informational demands. Thus, users receive content tailored to their interests, study areas, and learning methods. User happiness, information delivery efficiency, and library resource interaction are greatly improved by this technique. In the review, specific examples and case studies show how machine learning has improved retrieval effectiveness, including measured increases in successful search results, user engagement with recommended materials, and user satisfaction with library digital interfaces. This paper discusses the potential benefits and practical consequences of integrating machine learning into library information retrieval systems. Table 1 lists significant knowledge organization ML approaches.

DOI: https://doi.org/10.29352/mill0223.e37496

Table 1 - Summarizing ML Techniques in Key Areas of Knowledge Organization

Key Area	Machine Learning Techniques	Case Studies / Examples	Benefits	Challenges
Metadata Creation	- Natural Language Processing (NLP) - Supervised Learning	- National Library of the Netherlands: Auto-generating subject headings - British Library: Named entity extraction from manuscripts	- Increased efficiency, scalability, consistency	- Biases in training data - Accuracy and quality control of auto-generated metadata
Classification Systems	- Neural Networks - Decision Trees - Supervised Learning	- National Library of Norway: Auto-assigning DDC numbers - University of Illinois: Classifying research articles into disciplines	- Improved accuracy, consistency, granularity of classification	- Bias in training data - Lack of transparency in "black box" models
Information Retrieval	- Clustering Algorithms - Semantic Search - Ranking Algorithms	- Enhanced user experience through personalized search results, proactive recommendations, and relevance predictions	- Higher engagement and user satisfaction with digital platforms	- Potential misinterpretation of user intent - Data privacy concerns in personalized systems

2.4 Architecture of Machine Learning for Knowledge Organization

Figure 3 shows a block diagram of generic components in a machine learning-enhanced knowledge organisation system. It lists metadata production, classification schemes, and information retrieval as areas where machine learning is improving library science. The web-based interface will represent the system's front end, where users interact. Library patrons, librarians seeking information, document submission, and resource retrieval will use the system. The infrastructure core is a machine-learning model based on enormous datasets, including library metadata like bibliographic entries, subject classifications, and user interactions.

The API Gateway processes user queries and sends them to the Metadata Service, Classification Service, and Retrieval Service. The Metadata Database stores and retrieves keywords, summaries, and subject headings created by the Metadata Service using machine learning algorithms. Machine-learning methods like neural networks and decision trees interact with the Classification Database to improve resource classification coherence and resolution. The Retrieval Service, on the other hand, improved information retrieval methods like clustering, semantic search, and ranking algorithms to provide relevant and focused search results and improve the Retrieval Database update with user queries. This architecture centres on the Machine Learning Service, which contains all the core machine learning models and predictions that support Metadata, Classification, and Retrieval. These services' training data sets, model parameters, and other resources will be stored in Shared Storage.

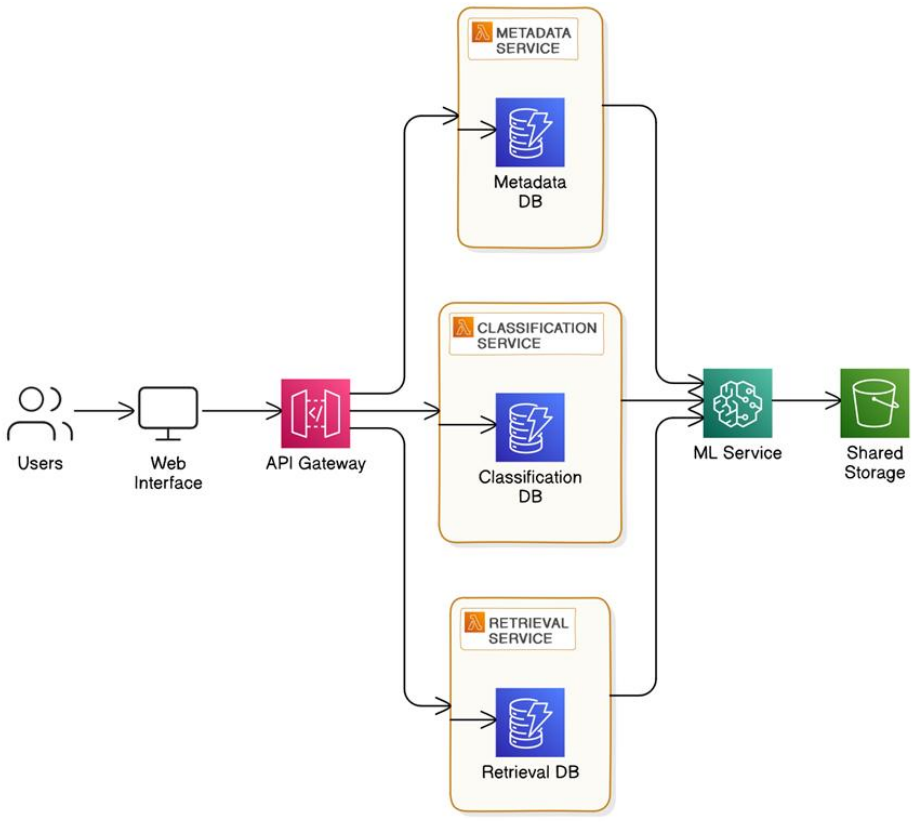

Figure 3 - Block Diagram of an Improved Knowledge Organization System Based on ML

DOI: https://doi.org/10.29352/mill0223.e37496

It collects textual and digital data from library archives, books, manuscripts, scholarly publications, and digital repositories. This initial step's raw data may have been pre-processed using data cleaning, normalisation, and feature extraction to train the machine learning model. After training, models are used to generate metadata, classify resources, and retrieve information in the library system, with subsequent analysis and improvement to reduce biases.

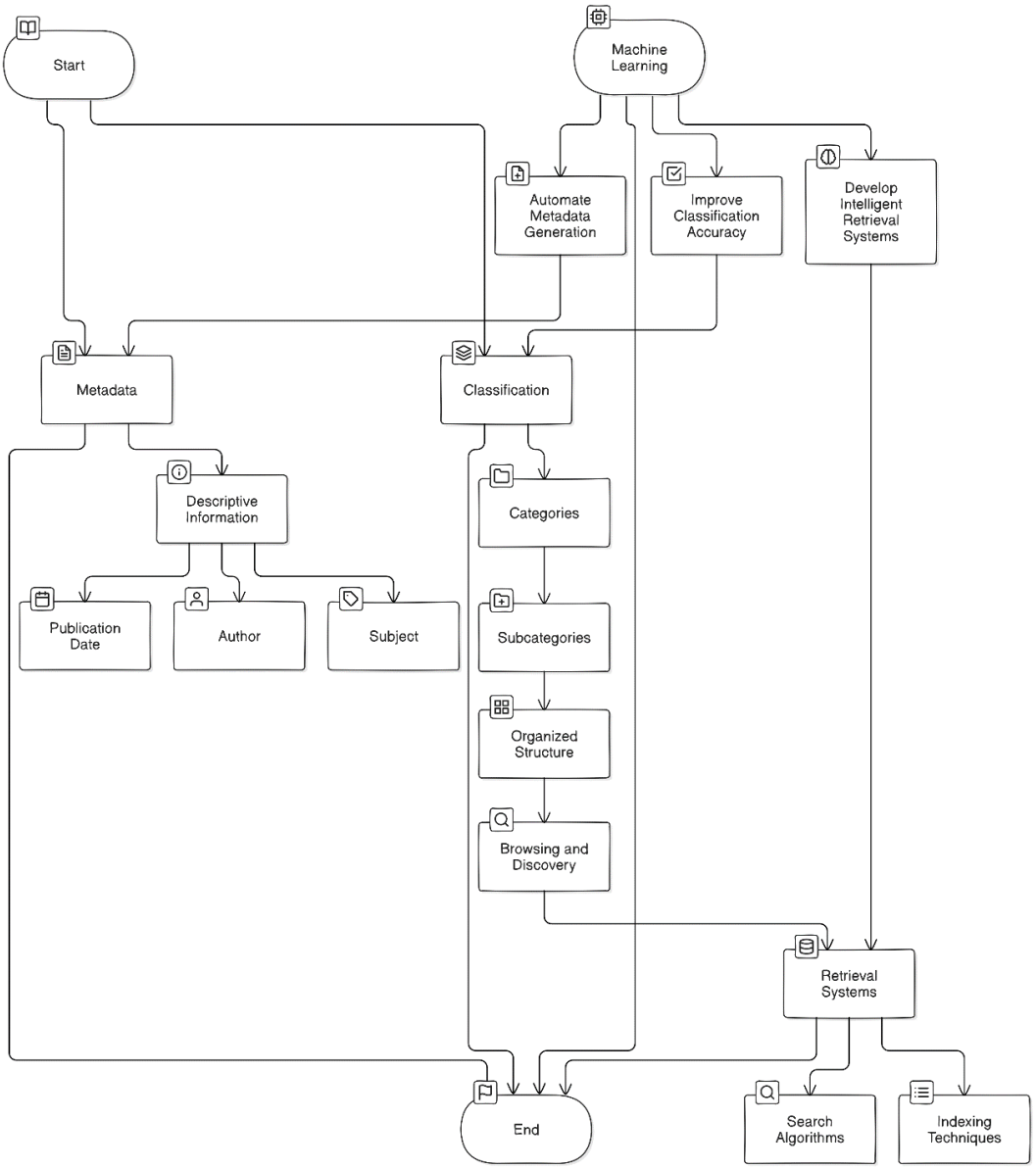

Figure 4 - Machine Learning Model in Knowledge Organization

After preprocessing, train machine learning models. Labelled datasets for metadata development, categorisation, and information retrieval will be used for training. NLP and supervised learning generate metadata by automatically creating subject headings, summaries, and named entities. Neural networks and decision trees are used in classification systems to categorise resources and allocate Dewey Decimal Classification numbers.

Clustering algorithms, semantic search, and ranking strategies personalise results by document/resource similarity, ordered by relevance to facilitate information retrieval. After training, these models are used to create metadata for digital libraries, classify new resources, and improve sophisticated search and recommendation systems to fulfil user needs.

DOI: https://doi.org/10.29352/mill0223.e37496

The suggested architecture allows library systems to integrate machine learning via a modular framework. This requires a scalable and efficient mechanism to improve metadata production, resource classification, and information retrieval for library knowledge organisation efficiency.

3. RESULTS

This section compares ML models and methods for library metadata production, classification, and information retrieval. The analysis shows the pros, cons, and hybrid ways that integrate ML with traditional methods for optimal performance.

3.1 Comparison Studies of ML Models and Techniques

Metadata Creation: Table 2 lists Natural Language Processing and Supervised Learning as the top metadata generation machine learning methods. NLP is highly effective at extracting keywords, subject headings, and summaries from massive textual corpora to help machines grasp context and linguistic structure. If a large labelled dataset is available, supervised learning lets a model learn metadata structure and apply it to unlabelled resources.

Classification Systems: Table 2 lists artificial neural networks and decision tree models for resource classification automation. Complex data patterns are better processed by artificial neural networks. They achieve great multi-class classification accuracy. Their big data application is excellent, but they require plenty of computational power. Decision trees are useful for interpretable classification system decision-making, but they degrade when utilised with complicated data sets.

Information Retrieval: Table 2 lists clustering methods, semantic search models, and ranking strategies that improve library information retrieval. Clustering methods allow users to conveniently manage big and complicated collections by grouping comparable documents, although performance depends on cluster quality. To offer relevant results, semantic search models interpret the user's query in context. These have helped library search algorithms improve relevancy and precision. Ranking algorithms rank resources by relevance, authority, and other variables to show the most relevant material first in search results.

Table 2 - ML techniques can be used for metadata creation, classification, and information retrieval

Area of Application	ML Technique	Effectiveness	Limitations	Best Use Cases
Metadata Creation	NLP	High for large unstructured text corpora	May struggle with non-standard or highly technical texts	Large digital libraries with extensive collections
	Supervised Learning	High if sufficient labelled data is available	Requires extensive training datasets	Libraries with established metadata templates
Classification	Neural Networks	High for large complex datasets	Computationally expensive; black-box nature	Automated large-scale classification tasks
	Decision Trees	Moderate for smaller, interpretable datasets	Performance declines with highly complex data	Smaller collections requiring transparency
	Clustering	Effective for document discovery	Requires well-defined similarity metrics	Large digital archives for document grouping
Information Retrieval	Semantic Search	High for improving query relevance	Computationally intensive; requires rich semantic data	Libraries with diverse user search behaviors
	Ranking Algorithms	Excellent for prioritizing high-value resources	Can suffer from biased training data	Libraries with a need for authority-based search

3.2 Effectiveness, Limitations, and Best-Use Scenarios

Effectiveness: Using Natural Language Processing and unsupervised learning approaches can effectively generate enormous amounts of metadata. This usage of NLP removes a lot of manual effort that was previously needed for this purpose and provides strong consistency, which is quite beneficial when processing big amounts of text. In classification systems, neural network models may outperform conventional methods in performance and accuracy. They can handle complicated data patterns that manual classification cannot. When decision transparency and interpretability are needed, decision tree models are useful. In information retrieval, machine learning-based semantic searches and rankings yield superior results over keyword-based systems. This model will improve user engagement and satisfaction by personalising information.

Limitations: One of the biggest drawbacks of utilising ML in many applications is the bias in the training data. For instance, biased training data in classification models may increase resource classification inequity. Neural networks are effective but often remain opaque, preventing end users from understanding their classification or retrieval results. Most advanced ML approaches, including neural networks and semantic search, need significant processing resources, making them problematic in low-resource contexts.

Best-Use Cases: Natural Language Processing is ideal for handling very large quantities of unstructured text input and automating the creation of subject headings or summaries. In the case of a library, an effective model of supervised learning requires good quantities of labelled data to train on, such as is achieved with well-structured metadata templates. An artificial neural network would be best for complicated classification cases that are highly accurate and voluminous, while decision tree models are for

DOI: https://doi.org/10.29352/mill0223.e37496

transparency and interpretability. Examples of such techniques include clustering algorithms and semantic search strategies to support users in finding and identifying, out of huge digital libraries, only that content that really matters.

3.3 Exploration of Hybrid Approaches

Combining machine learning with traditional methodologies to make use of the powers of each may turn out to be fruitful for library knowledge organisation.

Merging Old and New: Libraries may utilize NLP to generate metadata; human specialists then review and make updates where necessary. This would eliminate biased or incorrect metadata, and processing time/ labour costs. In the hybrid system, either decision trees or artificial neural networks classify resources that library staff can then manually review. Such systems weigh automation with openness. Human feedback may help improve information retrieval accuracy and relevance in machine learning models. Standard cataloguing systems with machine learning systems analyze user search trends to help improve the results of searches.

Hybrid Systems Advantages: Human intervention and machine learning of hybrid methods decrease the occurrence of bias and increase accuracy regarding metadata and classification. The intervention of a human in critical areas makes the models of machine learning appropriate and accurate. Hybrid systems enable libraries to automate various activities without affecting the data quality and reliability.

3.4 Investigation into Hybrid Directions

ML and traditional methods are merged in hybrid approaches to knowledge arrangement in libraries, ensuring better output. This is achieved by the integration of ML with the traditional method. Libraries may create metadata through NLP and then have the experts vet and refine it. This hybrid approach will reduce metadata bias and error; hence, it is efficient. Using this system, decision trees or artificial neural networks would suggest resource classification for the purpose of assessing it manually by the library staff. This would balance automation with openness. Humans refine this machine learning to correctly retrieve and retrieve full information. Libraries, through searching behavior of users and classic cataloguing, with the help of machine learning are able to present more relevant results. Benefits of Hybrid Systems Human interaction reduces bias in machine learning, therefore, making metadata and classification more inclusive and appropriate, hence the output from a machine learning model will probably be more appropriate and relevant. Hybrid systems support libraries with the objective of maximizing automation and preserving data quality.

4. DISCUSSION

This study shows how ML may improve library knowledge organization by automating metadata production, classification, and information search and retrieval. The study finds that ML can improve libraries in ways that conventional approaches cannot. Value ML, which comprises NLP and supervised learning, creates metadata, allowing libraries to standardize their various resources more accurately. Nevertheless, auto-generated metadata quality and accuracy remain concerns, especially in automated indexing of non-standard or specialized sources.

The taxonomies are based on neural networks and decision tree models for resource categorization. These models can also be used to deal with data patterns, allocating resources more accurately. There are several disadvantages to neural networks, such as a lack of transparency or the inability of users or the model to explain the decisions made when classifying. This brings up the question of trust and responsibility in library systems, which are mainly based on ML for decision-making. Further valuable results are observed with regard to the three-fold information retrieval enhancement. Using clustering algorithms and semantic search can yield more accurate results and enhance the search experience. These are ways of retrieving that can interpret the users' intentions incorrectly, and there is always a problem with data privacy in liberal arts computer systems.

The talk recognizes the ethical and infrastructural issues in ML. Libraries need to confront the biases in the generation of training data, which further exacerbate the injustices of how libraries are organized and the access to resources. Complex ML approaches consume a considerable amount of computing resources and are not easily accessible for libraries with limited resources or staffing. Fairness, transparency, and inclusiveness in library systems will require the help of ML and human intervention. Finally, while ML can enhance the experience of KO in libraries, ethical considerations, resource limitations, and user trust must all be taken into account when integrating ML into libraries.

CONCLUSION

The result of this study could help demonstrate the effectiveness of ML and how it can change the knowledge organization as it is conducted today in modern libraries in the creation of metadata, the construction of classification systems, and other aspects of information retrieval. The analysis highlighted the various approaches to ML, which can greatly enhance the efficiency, scalability, accuracy, and personalization in library services, including NLP, supervised learning, neural networks, clustering, and semantic search.

DOI: https://doi.org/10.29352/mill0223.e37496

It is also presented that ML systems are faster, more consistent, and more relevant to retrieval than many of the standard, manual processes. The discussion also highlighted the potential of library systems to benefit from ML for library automation in terms of generating metadata, classification of resources, and intelligent retrieval of resources to enhance user interaction. The research also outlined the main issues, including algorithmic bias, transparency, computational cost, and ethical issues, which provide an overall realistic view of the application of ML to libraries.

The main merit of this work is that it is a very comprehensive work that integrates theoretical discussions, practical case studies, and architectural designs for the implementation of ML-enhanced knowledge organization systems. The Study is significant as it will give a holistic framework covering all the areas of metadata production, classification, and retrieval, and integrates them to make them complete to comprise a holistic ecosystem based on ML. This study shows how several ML approaches are used to digitize libraries and to smart information management, not focusing on one specific activity of the library like the other studies. Another advantage of the research is that it focuses on the hybrid solutions in which humans are used in conjunction with machine learning. This will provide more transparency, equity, and trust in the automated results and keep the trusted role of librarians to verify and improve automated results. This is particularly beneficial to a multitude of libraries with limited resources.

In the future, increased explainability and openness of the ML models applied in libraries will be significant, as well as attempts to reduce bias in training data and to reinforce ethical governance systems of AI-based libraries. Finally, a long-term interaction study in real library environments is needed to evaluate the human–ML interaction. Moreover, future studies can be directed towards developing multilingual metadata, cross-domain interoperability, real-time recommendation systems, and implementing generative AI techniques within the digital library infrastructure. The guidelines can also help in a smart knowledge management system and ensure equitable access to information in the dynamic digital world.

AUTHORS´ CONTRIBUTION

Conceptualization, R.J., D.P., K.P. and P.T.; formal analysis, R.J., D.P., K.P. and P.T.; investigation, R.J., D.P., K.P. and P.T.; methodology, R.J., D.P., K.P. and P.T.; resources, R.J., D.P., K.P. and P.T.; visualization, R.J., D.P., K.P. and P.T.; writing – original draft, R.J., D.P., K.P. and P.T.; writing – review & editing, R.J., D.P., K.P. and P.T.

CONFLICT OF INTERESTS

The authors declare no conflict of interests.

REFERENCES

Ahmed, M., Mukhopadhyay, M., & Mukhopadhyay, P. (2023). Automated knowledge organization AI ML based subject indexing system for libraries. *DESIDOC Journal of Library & Information Technology, 43(*01), 45–54. https://doi.org/10.14429/djlit.43.01.18619

Audeh, B., Beigbeder, M., Largeron, C., & Ramírez-Cifuentes, D. (2020). Evaluating the usefulness of citation graph and document metadata in scientific document recommendation for neophytes. *Proceedings of the 35th Annual ACM Symposium on Applied Computing, 681–689.* https://doi.org/10.1145/3341105.3373886

Audeh, B., Beigbeder, M., Largeron, C., & Ramírez-Cifuentes, D. (2021). Improving exploratory information retrieval for neophytes. *ACM SIGAPP Applied Computing Review, 20*(4), 50–64. https://doi.org/10.1145/3447332.3447336

Baker, J. (2015). *The British Library machine learning experiment.* British Library Digital. https://shre.ink/3YZK

Das, R. K., & Islam, M. S. U. (2021). *Application of artificial intelligence and machine learning in libraries: A systematic review*. arXiv. https://arxiv.org/abs/2112.04573

Deolekar, R., & Dangare, A. (2018). Enterprise search: A new dimension in information retrieval. *2018 3rd International Conference for Convergence in Technology (I2CT)*, 1–6. https://doi.org/10.1109/I2CT.2018.8529602

Guerra-García, C., Pérez-González, H. G., Martínez-Pérez, F., Juárez-Ramírez, R., & Jiménez, S. (2023). Applying mechanisms of data profiling for assuring data quality in the software: A first approach. *2023 11th International Conference in Software Engineering Research and Innovation (CONISOFT),* 108–115. https://doi.org/10.1109/CONISOFT58849.2023.00023

Guo, Q., Jin, S., Li, M., Yang, Q., Xu, K., Ju, Y., Zhang, J., Xuan, J., Liu, J., Su, Y., Xu, Q., & Liu, Y. (2020). Application of deep learning in ecological resource research: Theories, methods, and challenges. *Science China Earth Sciences, 63*(10), 1457–1474. https://doi.org/10.1007/s11430-019-9584-9

H. R, S. Rashmi, B. S, Prof. Anisha, & Kumar. P, Dr. R. (2019). Smart document analysis using AI-ML. *International Journal of Innovative Research in Computer Science & Technology, 7(*3), 54–70. https://doi.org/10.21276/ijircst.2019.7.3.6

Hjørland, B. (2021). Information retrieval and knowledge organization: A perspective from the philosophy of science. *Information, 12*(3), 135. https://doi.org/10.3390/info12030135

DOI: https://doi.org/10.29352/mill0223.e37496

Kalisdha, A. (2024). The impact of artificial intelligence and machine learning in library and information science. *International Journal of Research in Library Science, 10*(1), 39–58. https://doi.org/10.26761/ijrls.10.1.2024.1733

Limbu, D. K., Connor, A., Pears, R., & MacDonell, S. (2006). Contextual relevance feedback in web information retrieval. *Proceedings of the 1st International Conference on Information Interaction in Context - IIiX,* 138. https://doi.org/10.1145/1164820.1164848

Miah, Md. S. U., Sulaiman, J., Sarwar, T. Bin, Naseer, A., Ashraf, F., Zamli, K. Z., & Jose, R. (2022). Sentence boundary extraction from scientific literature of electric double layer capacitor domain: Tools and techniques. *Applied Sciences, 12*(3), 1352. https://doi.org/10.3390/app12031352

Park, J., & Lu, C. (2009). Application of semi-automatic metadata generation in libraries: Types, tools, and techniques. *Library & Information Science Research, 31*(4), 225–231. https://doi.org/10.1016/j.lisr.2009.05.002

Pessin, V. Z., Santos, C. A. S., Yamane, L. H., Siman, R. R., Baldam, R. de L., & Júnior, V. L. (2023). A method of mapping process for scientific production using the Smart Bibliometrics. *MethodsX, 11*, 102367. https://doi.org/10.1016/j.mex.2023.102367

Qin, J. (2020). Knowledge organization and representation under the AI Lens. *Journal of Data and Information Science, 5*(1), 3–17. https://doi.org/10.2478/jdis-2020-0002

Rashid, F., Gargaare, S. M. A., Aden, A. H., & Abdi, A. (2022). Machine learning algorithms for document classification: Comparative analysis. *International Journal of Advanced Computer Science and Applications, 13*(4). https://doi.org/10.14569/IJACSA.2022.0130430

Rosa, J. D. L. (2024). *Artificial Intelligence at the National Library of Norway.* LUSTRE. https://lustre-network.net/artificial-intelligence-at-the-national-library-of-norway/

Soni, D. N. (2023). The role of artificial intelligence in information and library science: Opportunities and ethical considerations. *International Journal of Research in Humanities & Soc. Sciences, 11*(7), 41–43.

Weinryb-Grohsgal, L. (2020, July 22). *Machine learning + libraries: A report on the state of the field.* The Signal. The Library of Congress. https://shre.ink/3YF3

Wu, M., Brandhorst, H., Marinescu, M.-C., Lopez, J. M., Hlava, M., & Busch, J. (2023). Automated metadata annotation: What is and is not possible with machine learning. *Data Intelligence, 5*(1), 122–138. https://doi.org/10.1162/dint_a_00162

Xie, J. (2023). Research on information retrieval service innovation of university library. *SHS Web of Conferences, 169,* 01088. https://doi.org/10.1051/shsconf/202316901088

Yu, K., Gong, R., Sun, L., & Jiang, C. (2019). The application of artificial intelligence in smart library. *Proceedings of the 2019 International Conference on Organizational Innovation (ICOI 2019)*. https://doi.org/10.2991/icoi-19.2019.124

Zamani, H., Diaz, F., Dehghani, M., Metzler, D., & Bendersky, M. (2022). Retrieval-enhanced machine learning. *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2875–2886. https://doi.org/10.1145/3477495.3531722

Zhou, Z., Liu, Y., Yu, H., & Ren, L. (2020). The influence of machine learning-based knowledge management model on enterprise organizational capability innovation and industrial development. *PLOS ONE, 15*(12), e0242253. https://doi.org/10.1371/journal.pone.0242253