

# Digital Maktaba Project: Proposing a Metadata-Driven Framework for Arabic Library Digitization

Amina El Ganadi<sup>1,2,3,\*</sup>, Luca Gagliardelli<sup>1</sup>, Sania Aftar<sup>1</sup> and Federico Ruoizzi<sup>1,3</sup>

<sup>1</sup>University of Modena and Reggio Emilia, Modena, Italy

<sup>2</sup>University of Palermo, Palermo, Italy

<sup>3</sup>Fondazione per le Scienze Religiose (FSCIRE)

## Abstract

The rapid digitization of cultural heritage has underscored the critical need for robust digital libraries, particularly for underrepresented languages like Arabic and Persian. This paper describes the methodologies and challenges involved in developing a metadata-driven Arabic digital library, utilizing bibliographic metadata extracted from the Diamond catalogue. It explores advanced metadata schemas, such as Dublin Core, and integrates text recognition technologies and preservation strategies to address key concerns of accessibility, scholarly use, and the long-term preservation of Arabic-script texts. The paper delves into specific challenges of processing Arabic script, including handling calligraphy, diacritics, and ligatures, and introduces innovative solutions like the use of frontispiece images to train OCR systems. Furthermore, it discusses how integrated metadata could not only enhance text recognition but also improve user engagement by enabling refined search functionalities and better resource discovery. Finally, the paper outlines future directions for expanding metadata frameworks to ensure interoperability and the long-term preservation of cultural heritage.

## Keywords

Document Analysis, Arabic Digital Library, Bibliographic Metadata, Digitization, Digital Maktaba Project, OCR, Cultural Heritage, Natural Language Processing.

## 1. Introduction

In recent years, libraries have undergone a transformative shift toward digital environments, driven by advances in artificial intelligence (AI), machine learning, and digitization technologies. This transformation is not merely a technological upgrade; it represents a fundamental rethinking of how cultural heritage materials, scholarly works, and historical documents are preserved, accessed, and interpreted in digital formats. For institutions holding large collections of Arabic and Islamicate texts, this shift opens up new possibilities for enhanced discovery, interoperability, and user engagement.

Central to this digital evolution is the utilization of metadata as a core organizational principle. Metadata, from bibliographic information to structured descriptive tags, creates the backbone of any digital library system. Metadata provides key details about a resource such as author, title, publication date, and subject matter, playing a vital role in cataloguing, discovery, and resource management. Libraries have traditionally used standardized systems like MARC (Machine-Readable Cataloguing) and Dublin Core to maintain consistency and interoperability across institutions [1]. These frameworks form the foundation of efficient library operations, helping librarians and users navigate extensive collections of both physical and digital resources.

With proper standardization and adherence to established library protocols, metadata allows for scalable integration of diverse collections and seamless alignment with global digital library networks. As libraries migrate to digital repositories, metadata management becomes essential not only for

---

*IRCDL 2025: 21st Conference on Information and Research Sciences Connecting to Digital and Library Science, February 20-21 2025, Udine, Italy*

\*Corresponding author.

✉ amina.elganadi@unimore.it (A. El Ganadi); luca.gagliardelli@unimore.it (L. Gagliardelli); sania.aftar@unimore.it (S. Aftar); federico.ruozzi@unimore.it (F. Ruoizzi)

ORCID 0000-0002-8196-2628 (A. El Ganadi); 0000-0001-5977-1078 (L. Gagliardelli); 0000-0001-8151-8941 (S. Aftar); 0000-0003-2729-5016 (F. Ruoizzi)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

cataloguing but also for ensuring access, improving visibility, and fostering user engagement with digital resources [2].

Alongside these foundational elements, technologies like Optical Character Recognition (OCR) and AI-driven text analysis enable more accurate and accessible digital representations of physical materials. However, developing OCR technologies for Arabic texts presents unique challenges due to the script's cursive nature, variable letter shapes, and presence of diacritics. Mansoor Alghamdi and William Teahan [3] highlight that while handwritten Arabic text poses significant challenges for OCR systems, printed Arabic text is also notably difficult. Arabic-script OCR programs often struggle with accuracy, frequently failing to meet the high expectations set by their marketing claims [3].

Existing datasets, such as the Arabic Printed Text Image (APTI), have supported progress in OCR research, but efforts remain fragmented and often neglect specialized materials like ornate printed title pages. While advancements in deep learning and Vision-Language Models (VLMs) [4, 5] have improved recognition capabilities, the absence of targeted datasets for frontispiece images continues to limit the accuracy and efficiency of digital library cataloging systems.

This multifaceted approach, combining metadata extraction, standardized protocols, and cutting-edge OCR methods, ensures both the integrity of the original texts and the creation of rich, searchable interfaces that enhance the end-user experience. This article presents a methodology that prioritizes metadata extraction from physical library catalogs, digitization of frontispiece images, and subsequent enrichment using widely adopted standards. By referencing prominent case studies such as Europeana and WorldCat<sup>1</sup>, and highlighting platforms like eScriptorium<sup>2</sup>, we establish a framework that is both theoretically grounded and practically tested, ultimately advancing the integration of Arabic texts into the global digital information ecosystem.

Currently, this approach is being further demonstrated through a focused case study of the La Pira Library in Palermo<sup>3</sup>, showcasing its scalability and adaptability to other collections, including those containing Arabic and Persian texts. As this work progresses, the convergence of well-structured metadata, AI-assisted text analysis, and adherence to international standards continues to provide a promising pathway toward the creation of comprehensive and user-friendly digital libraries. Ongoing efforts aim to refine these methods, ensuring their effectiveness and broad applicability in diverse library settings.

## 2. Related Work

Many researchers have explored the Semantic Web's role in digital libraries, often focusing on theoretical aspects without addressing practical implementation [6, 7, 8, 9]. Others have studied bibliographic ontologies [10, 11, 12, 13] and AI-based document classification using similarity measures [14]. Some Arabic researchers have combined NLP with automatic classification and ontology creation [15].

The integration of artificial intelligence (AI) in the domain of library and information science has gained significant attention, with numerous studies investigating its applications, benefits, and associated challenges [16, 17, 18, 19, 20, 21]. Farag et al. [22] investigated AI adoption in Saudi academic libraries, finding limited staff understanding, with 69% not using AI. Despite its potential in indexing and user support, adoption is hindered by inadequate training, limited infrastructure and technical expertise. Hussain [23] examined AI integration in library services, noting its potential to enhance operations despite barriers like budget constraints, librarian perceptions, and limited technical skills. The study emphasizes affordable AI applications that can improve service delivery and drive library development. Brzustowicz, Richard [24] demonstrated ChatGPT's potential in automating library cataloging through MARC record generation, while Adetayo [25] highlighted its broader applications in academic libraries, including reference assistance and task automation, despite concerns about intellectual property, bias,

---

<sup>1</sup><https://search.worldcat.org/>

<sup>2</sup>eScriptorium is an advanced digital platform designed to streamline the transcription, annotation, and analysis of both printed and handwritten texts. Built on machine learning techniques, it provides a robust and flexible environment for processing historical and modern documents in a wide range of languages and scripts.

<sup>3</sup><https://lapira.diamonddrda.org/>

and job displacement. ChatGPT [26] exhibited notable proficiency in categorizing subjects within an Islamic digital library; however, issues like interpretability, generalization, and hallucinations remain key challenges.

Xu [27] explored AI applications in libraries, focusing on six technologies including OCR, NLP, and machine learning. The study analyzed their roles, implementation challenges, and potential impact on library development and reform. Kraken, an open-source OCR engine, achieves high-accuracy Arabic OCR on al-Abhath. The evaluation compares typeface-specific and general models, identifies error patterns, and recommends improving Arabic-script OCR through systematic training data, multilingual modeling, and better segmentation techniques [28]. As AI continues to evolve, libraries must adapt to emerging trends that will transform metadata creation, management, and utilization. Key developments include linked data adoption, predictive analytics, enhanced interoperability, and shifting librarian roles in metadata curation [29, 24, 1]. Moreover, several existing studies overlook the ethical and operational challenges associated with integrating AI into metadata management [30, 31, 17]. Key issues such as algorithmic bias, the risk of reduced human oversight, and the long-term sustainability of AI-driven metadata systems are frequently neglected.

Several projects have focused on digitizing and developing Arabic and Persian text corpora, making them highly relevant to our study within the digital domain. Notable initiatives include the Open Islamiate Text Initiative (OpenITI) [32], Shamela<sup>4</sup>, KITAB<sup>5</sup> and the Persian Digital Library<sup>6</sup> (PDL), which have significantly advanced research in this field.

### 3. Case Study of the La Pira Library

#### 3.1. La Pira Library: A Center for Islamic Scholarship

Established in 1953, the Foundation for Religious Sciences (FSCIRE) began as a specialized research center in the history of Christianity, housing the Dossetti Library in Bologna. Recognized in 2021 as a European research infrastructure for religious sciences, FSCIRE expanded its focus. In 2018, it opened the Giorgio La Pira Library in Palermo<sup>7</sup> as an extension of the Dossetti Library. Modeled on FSCIRE's extensive expertise, this new library serves as a dedicated facility for studying the history, doctrines, and theology of Islam. It boasts a vast collection of over 31,000 paper volumes and approximately 900,000 digital works, including significant manuscripts and texts in Western languages, Arabic, and Persian, as well as a unique 18th-century Koranic manuscript

#### 3.2. Integrated Overview of the Diamond Catalogue and La Pira Library's Classification Practices

The Diamond general catalogue<sup>8</sup>, developed by the Dominican Institute for Oriental Studies (IDEO), provides unified access to the collections of its partner institutions, including the La Pira Library. Supporting cataloguing in multiple languages, including Arabic and others, it broadens its applicability to diverse collections and user communities. Built upon the Resource Description and Access (RDA) format, the catalogue incorporates the Library Reference Model (LRM) developed by the International Federation of Library Associations and Institutions (IFLA). This model unifies previously separate frameworks—FRBR, FRAD, and FRSAD—into a single, cohesive system, enhancing the clarity, consistency, and interoperability of bibliographic records.

A central aspect of LRM-based cataloguing is the WEMI (Work, Expression, Manifestation, Item) model, which classifies:

---

<sup>4</sup><http://shamela.ws>

<sup>5</sup><https://kitab-project.org/>

<sup>6</sup><https://persdigumd.github.io>

<sup>7</sup><https://www.fscire.it/lang/eng/heritage/biblioteca-la-pira>

<sup>8</sup><https://ideo.diamonrda.org>

The screenshot shows the FSCIRE library catalog interface. On the left, the Dublin Core metadata for a monograph is displayed in XML format. On the right, the record is structured into sections: Access (Hard copy, Call number 9-4-70), Explore (Single work), and More publications in the same field. Green arrows indicate the mapping between the XML metadata and the structured view.

**Figure 1:** Monograph record from FSCIRE’s La Pira Library, extracted from the Diamond Catalogue and shown in the Dublin Core metadata format.

Work	Title	Author	Publisher	Call Number	ISBN	URI	Subject Headings
	ابن سينا وتلاميذه اللاتين	ابن سينا وتلاميذه اللاتين	مكتبة الخانجي، لآزيب محمود الضوري	9-609/6-7	978-977-505-019-9	<a href="https://lapira">https://lapira</a>	Philosophy and Sciences → Classical Islamic philosophy
	أداب الصبغة ونكر الصبغة والأخاداب الصبغة ونكر الصبغة	أداب الصبغة ونكر الصبغة ونكر الصبغة ونكر الصبغة	أداب الصبغة ونكر الصبغة ونكر الصبغة ونكر الصبغة	9-447-27	977-1600-62-1	<a href="https://lapira">https://lapira</a>	Theology / Kalām → Main theological issues → Ethics
	أصل العدل عند المعتزلة	أصل العدل عند المعتزلة	دار الفكر العربي، 1993 هـم إبراهيم يوسف، تقديم عاطف العراقي	9-727/1-44	978-0977-10-560-1	<a href="https://lapira">https://lapira</a>	Theology / Kalām → Sunni theology → mu'tazila

**Figure 2:** Overview of the bibliographic metadata extracted from the Diamond Catalogue.

- *Work*: The fundamental intellectual or artistic creation, identified by a title. This level may involve authors, compilers, or recipients.
- *Expression*: A specific realization of a Work, often distinguished by its language, edition, or format, involving roles such as intellectual editors, translators, or illustrators.
- *Manifestation*: The physical or digital embodiment of an Expression, typically identified by ISBN/ISSN, with responsibilities including publishers, printers, or copyright owners.
- *Item*: The individual physical or digital unit of a Manifestation, tracked by its call number and possibly linked to donors, binders, or dedicatees.

While many libraries utilize subject headings and intricate thematic schemes to facilitate discovery, the La Pira Library employs a distinct approach. Instead of traditional subject headings, it uses a topographic classification system rooted in the physical layout of materials within the library. This system is dynamic, allowing updates and expansions as the collection grows, and correlates each item’s classification with its physical location on the shelves. This approach categorizes materials into broad macro categories that are further subdivided into specific categories and micro-categories. Unlike systems relying on multiple conceptual descriptors, each book’s classification at the La Pira Library is directly tied to its placement within the library’s spatial configuration. This method, particularly effective given the Library’s wide thematic scope—from religious texts to historical works, supports an intuitive, location-based retrieval process that aligns with user navigation habits. The La Pira Library’s application of a physical, topographically based system alongside the advanced cataloguing standards of the Diamond catalogue exemplifies how libraries can adapt their cataloguing methods to best serve their collections and communities. This dual approach, balancing modern standards with practical,

user-centered design, highlights the library’s commitment to both preserving scholarly resources and enhancing accessibility.

## 4. Developing the Digital Maktaba prototype Using La Pira Library’s Bibliographic Metadata

As part of the Digital Maktaba project’s early efforts, the Fscire research team [33] initiated the development of a semi-automated tool aimed at cataloguing non-Latin texts, such as Arabic, Persian, and Azerbaijani. This phase employed EasyOCR<sup>9</sup> as OCR system to extract text and metadata from digitized PDFs. While this technology aimed to enhance the efficiency of the Digital Maktaba’s cataloguing processes, it faced several challenges. For instance, manual language specification was required before processing, which slowed down operations. Additionally, the metadata from EasyOCR was often inaccurate, with text boxes fragmented and incorrectly aligned due to the right-to-left structure of Arabic script. These inaccuracies resulted in significant errors in metadata indexing, adversely impacting retrieval accuracy. To enhance output quality, Google Docs was utilized to automatically infer the language of the documents. However, this approach failed to generate the necessary metadata, thereby limiting its utility for cataloguing. Additionally, the costs associated with API usage imposed further restrictions on its effectiveness.

Building on these foundational experiences, we are now pivoting the Digital Maktaba project towards a more robust integration of bibliographic metadata. In this next phase, we will leverage the extensive and already validated metadata from the La Pira Library, which includes over 15,000 catalogued entries. By employing metadata that has been rigorously validated by expert librarians, we aim to significantly enhance the accuracy and utility of our digital cataloguing efforts. Furthermore, this integration will involve collaboration with the La Pira Library’s IT team to ensure seamless data migration and system compatibility.

This strategic shift not only addresses the limitations encountered in the initial OCR workflows but also leverages the established cataloguing standards and practices of the La Pira Library. By incorporating validated metadata, we ensure that digital representations of texts are precise and truly reflective of their physical counterparts. This advancement aims to support more sophisticated retrieval and research capabilities, such as advanced search filters and cross-referencing features.

In the following section, we outline the specific methodology employed in leveraging La Pira Library’s extensive bibliographic metadata for the development of the Digital Maktaba prototype. This systematic approach is crucial for transforming raw data into a structured, accessible digital library format that serves both academic researchers and the general public. Our methodology underscores the pivotal role of bibliographic metadata as the backbone of a functional digital library. Rather than treating metadata as an afterthought, we place it at the center of the entire digitization, training, and integration pipeline.

### 4.1. Methodology

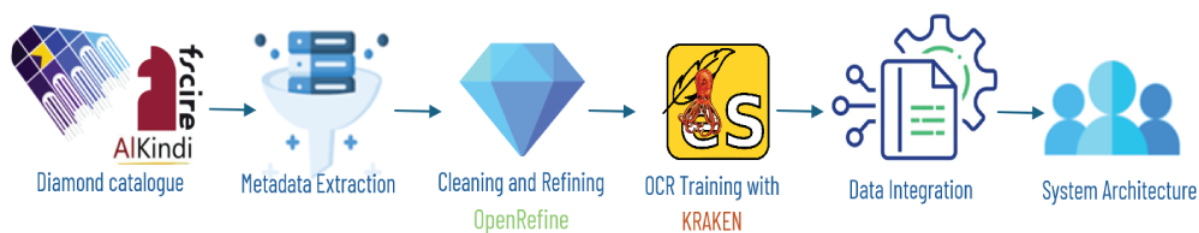
This section describes the various steps required for the development of Digital Maktaba prototype as illustrated in Figure 3.

As this project is ongoing, not all functionalities have been fully implemented yet. Currently, our efforts have been concentrated on the initial phase: metadata extraction. We have successfully extracted comprehensive metadata from the La Pira Library’s collections, and are now focusing on the critical task of labeling. Once the labeling work is completed, the data will undergo a thorough review by a linguistic expert specializing in Arabic to ensure its accuracy and reliability.

**Metadata Extraction.** The first step involves generating an annotated dataset of Arabic documents along with images of their frontispieces. We began with a dataset of 15,900 documents obtained from the Diamond catalogue, as shown in Figure 1. Each document is associated with a Uniform Resource Identifier (URI) (e.g., <https://lapira.diamondrda.org/manifestation/290168>) that links to its corresponding

---

<sup>9</sup><https://github.com/JaidedAI/EasyOCR>



**Figure 3:** Key Steps in the Workflow of the *Digital Maktaba* Prototype

page in the La Pira Library’s catalogue. Our primary focus is on Arabic-language documents that include an image of the frontispiece, as these are essential for training our Kraken OCR model. To identify and collect these resources, we developed a web crawler using the Scrapy library<sup>10</sup>. This crawler processes each document’s URI to extract metadata such as language, classification, and, when available, downloads the frontispiece image. From the initial dataset, 5,900 documents were identified as being written in Arabic, and 2,200 of these included an available frontispiece image.

For the Digital Maktaba project, we adapted La Pira’s topographic classification system<sup>11</sup> by integrating a subject headings section to the metadata extracted from the Diamond catalog (refer to Figure 2). The subjects are based on the topics provided by the library’s classification system (as shown in Figure 1), which categorizes the materials into thematic areas such as Philosophy and Sciences, Classical Islamic Philosophy, and Avicenna and Avicennism. This adaptation allowed us to structure the data in a way that improved resource discovery and metadata organization. Furthermore, the inclusion of subject headings has enriched the metadata, which will improve search and retrieval capabilities within the Digital Maktaba prototype.

**OCR Training with Kraken.** In this phase, we aim to construct an annotated dataset that links each document’s metadata (e.g., title, authors, publisher) with the corresponding text boxes extracted from its frontispiece. This process will be carried out using the eScriptorium interface, which utilizes the open-source Kraken OCR model. The Kraken model, provided to us by the eScriptorium development team, has an initial accuracy of 96.4% for Arabic script recognition and will serve as the foundation for further training on the frontispieces of Arabic books in our collection. The tool identifies and extracts different text boxes from the frontispiece images, retrieves the text from each box, and allows for manual correction to align with the metadata. An example of this process is illustrated in Figure 4. The completed annotated dataset will then be used to further train the Kraken OCR model, refining its performance specifically for this task and improving its accuracy on our dataset. In a subsequent phase, this trained model will be integrated into the Digital Maktaba prototype, replacing the EasyOCR model currently integrated into the existing prototype (see Figure 5). This integration is expected to enhance the overall accuracy and efficiency of text retrieval within the digital library system.

**System Architecture.** We will develop a modular system that includes a scalable metadata repository and an intuitive user interface for managing documents. This tool will enable authorized users to catalogue new documents using the pre-trained Kraken OCR model, as depicted in Figure 5, facilitating their addition to the library. Additionally, the system will offer an advanced querying interface, allowing users to efficiently search through the already catalogued documents.

**Data Integration.** We will also provide the capability to encode the generated metadata in XML formats following standards such as MARCXML and METS/ALTO. This approach ensures interoperability and scalability across our tool and other platforms, facilitating seamless data exchange and integration with existing library systems.

**Future Alignment with WorldCat and Europeana:** We plan to adopt the Europeana Data Model (EDM) to enrich semantic relationships and enhance multilingual support, essential for representing our diverse cultural heritage accurately. By integrating comprehensive bibliographic details such as

<sup>10</sup><https://scrapy.org>

<sup>11</sup><https://lapira.diamondrda.org/all/classifications>



Figure 4: The eScriptorium interface used for training the open-source Kraken OCR model.

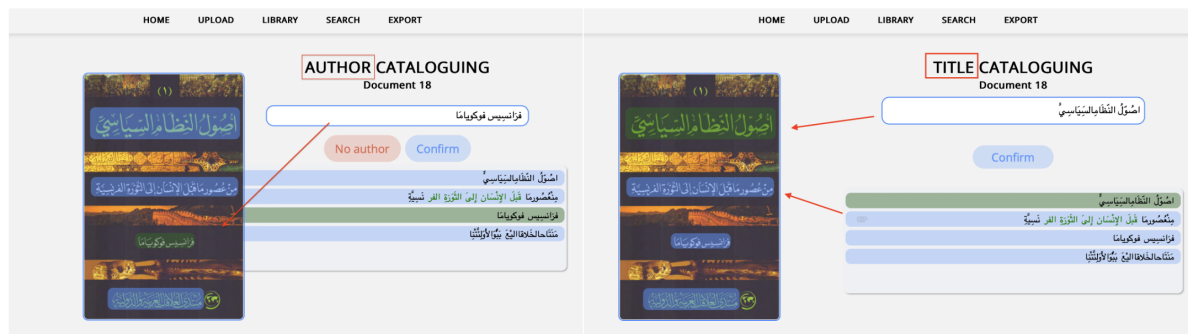


Figure 5: Screenshot of the Title and Author section of the Digital Maktaba prototype interface.

author information, publication data, and subjects, mirroring the structures used by WorldCat and the La Pira Library, we aim to improve interoperability with global library systems. This approach will facilitate smoother data exchange and integration, boosting the discoverability and accessibility of our collections.

## 5. Conclusion and Future Work

The intricate nature of Arabic script, characterized by ligatures, diacritics, and its right-to-left (RTL) orientation, poses significant challenges for OCR systems. Ongoing fine-tuning of Kraken models is a key component of efforts to enhance recognition accuracy. Preliminary endeavors, encompassing data preparation and initial testing, have demonstrated promising potential for these models to effectively process Arabic texts. Furthermore, this project addresses the variability in metadata formats by adopting universal standards such as Dublin Core and XML-based encoding. Plans to incorporate iterative user testing aim to refine the platform, ensuring it meets the diverse needs of its user base.

Initial results validate the feasibility of extracting and integrating bibliographic metadata from the Diamond catalogue, resulting in a prototype system that enables efficient metadata management and user access. These outcomes highlight the transformative impact of metadata-driven approaches in the design of digital libraries.

This study establishes a comprehensive framework for developing an Arabic digital library by tackling critical challenges and utilizing bibliographic metadata as a foundational element. Future work will focus on expanding the metadata repository, integrating advanced machine learning techniques to further enhance text recognition, and implementing user engagement strategies to improve the platform's usability and accessibility. This includes finalizing a comprehensive dataset of Arabic frontispiece images and collaborating with linguistic experts to ensure linguistic accuracy. By doing so, this effort will not only preserve and increase access to Arabic texts in digital libraries, supporting advanced cataloguing and research initiatives, but also pave the way for robust, user-friendly digital library solutions tailored to the unique complexities of Arabic script.

## Acknowledgments

This work was conducted within the PNRR project ITSEER - Italian Strengthening of the ESFRI RI RESILIENCE" (Avviso MUR 3264/2022) funded by EU – NextGenerationEU - Grant No IR0000014.

## References

- [1] A. Tella, O. Akanmu Odunola, L. WO, Cataloguing and classification in the era of artificial intelligence: Benefits, and challenges from the perspective of cataloguing librarians in oyo state, nigeria, *Vjesnik bibliotekara Hrvatske* 66 (2023) 159–176.
- [2] D. Oyighan, E. S. Ukubeyinje, B. T. David-West, B. D. Oladokun, The role of ai in transforming metadata management: Insights on challenges, opportunities, and emerging trends, *Asian Journal of Information Science and Technology* 14 (2024) 20–26.
- [3] M. Alghamdi, W. Teahan, Experimental evaluation of arabic ocr systems, *PSU Research Review* 1 (2017) 229–241.
- [4] J. Bai, S. Bai, S. Yang, S. Wang, S. Tan, P. Wang, J. Lin, C. Zhou, J. Zhou, Qwen-vl: A versatile vision-language model for understanding, localization, text reading, and beyond, 2023. URL: <https://api.semanticscholar.org/CorpusID:261101015>.
- [5] H. Liu, C. Li, Y. Li, Y. J. Lee, Improved baselines with visual instruction tuning, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 26296–26306.
- [6] N. I. A. El-Labban, The bibliographic ontologies and the bibliographic framework data model: comparative study, *Cybrarians Journal* (2016).
- [7] M. Z. Raza, N. F. Warraich, Impact of semantic web technologies on digital collections of libraries, in: *The 8th ALIEP Conference on the International Forum on Data, Information, and Knowledge for Digital Lives*, 2017.
- [8] B. Haslhofer, A. Isaac, R. Simon, Knowledge graphs in the libraries and digital humanities domain, *arXiv preprint arXiv:1803.03198* (2018).
- [9] Z. Raza, K. Mahmood, N. F. Warraich, Application of linked data technologies in digital libraries: a review of literature, *Library Hi Tech News* 36 (2019) 9–12.
- [10] Y. Hidalgo-Delgado, R. Estrada-Nelson, B. Xu, B. Villazon-Terrazas, A. Leiva-Mederos, A. Tello, Methodological guidelines for publishing library data as linked data, in: *2017 International Conference on Information Systems and Computer Science (INCISCOS)*, IEEE, 2017, pp. 241–246.
- [11] F. Nafis, A. Yahyaouy, B. Aghoutane, Ontologies for the classification of cultural heritage data, in: *2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)*, IEEE, 2019, pp. 1–7.
- [12] I. C. Dorobăț, V. Posea, Raising the interoperability of cultural datasets: the romanian cultural heritage case study, in: *Information Systems: 17th European, Mediterranean, and Middle Eastern Conference, EMCIS 2020, Dubai, United Arab Emirates, November 25–26, 2020, Proceedings 17*, Springer, 2020, pp. 35–48.
- [13] H. S. Patrício, M. I. Cordeiro, P. N. Ramos, From the web of bibliographic data to the web of bibliographic meaning: structuring, interlinking and validating ontologies on the semantic web, *International Journal of Metadata, Semantics and Ontologies* 14 (2020) 124–134.
- [14] T. Repke, R. Krestel, Interactive curation of semantic representations in digital libraries, in: *Towards Open and Trustworthy Digital Societies: 23rd International Conference on Asia-Pacific Digital Libraries, ICADL 2021, Virtual Event, December 1–3, 2021, Proceedings 23*, Springer, 2021, pp. 219–229.
- [15] S. E. Alamri, *Ontology Extraction from an Arabic Book*, Kent State University, 2020.
- [16] D. D. W. Praveenraj, K. Agarwal, B. Kim, V. Singh, Artificial intelligence applications in modern library services, *Library Progress International* 45 (2025) 1–11.
- [17] S. Priya, R. Ramya, Future trends and emerging technologies in ai and libraries, *Applications of Artificial Intelligence in Libraries* (2024) 245–271.



- [18] M. Bairagi, S. R. Lihitkar, Optimizing library services through the integration of artificial intelligence tools and techniques, in: Applications of Artificial Intelligence in Libraries, IGI Global, 2024, pp. 193–222.
- [19] M. Q. Affum, O. K. Dwomoh, Investigating the potential impact of artificial intelligence in librarianship, *Library Philosophy and Practice* (2023) 1–12.
- [20] M. Q. Affum, The role of artificial intelligence in library automation., *Library Philosophy & Practice* (2023).
- [21] L. Abudulsalami, A. K. Queeneth, S. S. NKapia, H. N. Ligola, E. L. Ovigie, B. O. Obande, M. Bilal, Artificial intelligence in academic libraries and its impact on library services and operations, *Omanarp International Journal of Library & Information Science* 1 (2024) 53–61.
- [22] H. A. Farag, S. N. Mahfouz, S. Alhajri, Artificial intelligence investing in academic libraries: Reality and challenges, *Library Philosophy and Practice* (2021) 1–34.
- [23] A. Hussain, Use of artificial intelligence in the library services: prospects and challenges, *Library Hi Tech News* 40 (2023) 15–17.
- [24] R. Brzustowicz, From chatgpt to catgpt: the implications of artificial intelligence on library cataloging, *Information Technology and Libraries* 42 (2023).
- [25] A. J. Adetayo, Artificial intelligence chatbots in academic libraries: the rise of chatgpt, *Library Hi Tech News* 40 (2023) 18–21.
- [26] A. El Ganadi, R. A. Vigliermo, L. Sala, M. Vanzini, F. Ruoizzi, S. Bergamaschi, et al., Bridging islamic knowledge and ai: Inquiring chatgpt on possible categorizations for an islamic digital library (full paper), in: CEUR Workshop Proceedings, volume 3536, 2023, pp. 21–33.
- [27] Z. Xu, Research on the application of artificial intelligence in the library sector, in: Third International Conference on Artificial Intelligence and Computer Engineering (ICAICE 2022), volume 12610, SPIE, 2023, pp. 1420–1429.
- [28] B. Kiessling, G. Kurin, M. T. Miller, K. Smail, Advances and limitations in open source arabic-script ocr: A case study, *arXiv preprint arXiv:2402.10943* (2024).
- [29] J. Yoon, J. E. Andrews, H. L. Ward, Perceptions on adopting artificial intelligence and related technologies in libraries: public and academic librarians in north america, *Library Hi Tech* 40 (2022) 1893–1915.
- [30] P. Chhetri, Analyzing the strengths, weaknesses, opportunities, and threats of ai in libraries., *Library Philosophy & Practice* (2023).
- [31] C. Mallikarjuna, An analysis of integrating artificial intelligence in academic libraries., *DESIDOC Journal of Library & Information Technology* 44 (2024).
- [32] M. T. Miller, M. G. Romanov, S. B. Savant, Digitizing the textual heritage of the premodern islamicate world: Principles and plans, *International Journal of Middle East Studies* 50 (2018) 103–109. doi:10.1017/S0020743817000964.
- [33] R. Martoglia, L. Sala, M. Vanzini, R. Vigliermo, et al., A tool for semiautomatic cataloguing of an islamic digital library: a use case from the digital maktaba project (short paper), in: CEUR WORKSHOP PROCEEDINGS, volume 3234, CEUR-WS, 2022.