International Journal of
Advanced Science Computing and Engineering

Journal Homepage www.ijasce.org/index.php/ijasce

Named Entity Recognition on Islamic Texts: A Systematic Review

Shasha Arzila Tarmizi [a,*], Saidah Saad [a]

[a] *Center for Artificial Intelligence Technology, The National University of Malaysia, Bangi, Selangor, Malaysia*

Corresponding author: [*]shasha.arzila@gmail.com

***Abstract*—This systematic literature review aims to comprehensively analyze Named Entity Recognition (NER) applications in Islamic texts, particularly the Quran and Hadith, across Arabic, Indonesian, English, and Malay languages. Materials comprised studies from major academic databases (2017-2024) implementing various NER approaches on Islamic textual datasets. The majority of studies reviewed focused on Hadith texts, with fewer examining Quranic texts and general Islamic literature. The methodology employed a PRISMA-based systematic review examining architectural components, diverse methodologies, comparative model performance, and extraction challenges in Islamic discourse. Traditional rule-based and statistical machine learning methods remain relevant, particularly in hybrid frameworks. However, the analysis reveals that transformer-based deep learning models consistently achieve superior performance, with the highest F1 Scores. Hadith datasets showed better NER performance than Quranic texts due to Hadith's structured and repetitive nature versus the Quran's greater linguistic diversity and complex syntactic structures. Most studies employed lexical and linguistic features to address distinctive characteristics of religious texts, with significant progress in handling specialized Islamic concepts and multilingual considerations. Despite these advancements, significant challenges persist, including the linguistic complexity of Classical Arabic, the scarcity of high-quality annotated corpora, and the difficulties of domain-specific entity identification. This review provides comprehensive guidance for researchers developing Islamic NER systems by identifying optimal methodological approaches and highlighting performance benchmarks across different text types, thereby enabling the development of more effective, culturally aware NLP systems for Islamic content.**

***Keywords*— Islamic texts; named entity recognition; Quran; Hadith.**

Manuscript received 11 Jun. 2025; revised 18 Aug. 2025; accepted 28 Oct. 2025. Date of publication 30 Dec. 2025.

I. INTRODUCTION

The explosion of textual data in this digital age has changed how information is created, shared, and consumed across all domains [1]-[20]. This enormous and continually growing corpus of text drives rapid advances in artificial intelligence (AI) and machine learning (ML). As a result, it is challenging to enable machines to comprehend, interpret, and produce human language, thereby making natural language processing (NLP) a vital field in AI research [21]-[45]. NLP relies on granular information extraction to fully capture the value embedded in textual data, a capability embodied by NER. Therefore, NER serves as a prerequisite for various downstream NLP tasks. NER primarily performs two functions in text information retrieval. First, it identifies certain information in unstructured data, known as entities. Then it classifies them into predefined categories. The majority of NLP applications, such as text summarization and question answering (QA) systems, are built on these functions. In textual data, the entities extracted are considered important pieces of information. For this reason, many NER-related studies focus on extracting specific entity types within a given domain [46]-[54]. For instance, in the biomedical field, NER is used to extract biomedical entities such as medications, proteins, and illnesses [55]. The development of NER models across different fields and languages is necessary because generic entity types, such as place or person names, may not always be useful for extracting information from domain-specific documents.

The domain of Islamic literature presents a steadily growing corpus of huge cultural, religious, and academic significance within this developing textual landscape [56]. The volume of Islamic literature is expanding globally, encompassing centuries-old manuscripts currently being digitized, modern scholarly works, online fatwa portals, and diverse digital content from Islamic institutions worldwide. The rise of digital availability motivates the growth of NLP applications in Islamic discourse to navigate, analyze, and extract insights, particularly from documents related to the Quran and Hadith. In Islam, both the Quran and the Hadith are important sources of knowledge and guidance. These two

sources, especially the Quran, which has remained unchanged over time, are quoted in most Islamic texts and literature. Consequently, much research in this area uses material from both scriptures as the primary sources for its dataset. The Quran, in particular, contains verses that mention numerous entities, from general ones to more specific ones, such as the name of God and the names of prophets. Therefore, the primary goal of the NER task in this field is to recognize and categorize specific terms denoting individuals, groups, places, and other important concepts in Islamic texts. The evolution of NER approaches in the Islamic domain reflects advances in NLP, from traditional systems built on handcrafted features and domain-specific rules to complex deep learning methods. In a study by [1], it was found that existing NLP libraries that perform well on general English texts struggle to handle the distinctive features of Islamic entities, resulting in poor performance. Specialized NER techniques are required due to the unique linguistic and cultural features of Islamic texts, including religious terms, proper names, and phrases [2]. Hadith, for instance, has a specific structure that includes the narrators [3]. These distinctions are often overlooked by generic models, underscoring the importance of NER techniques designed specifically for Islamic content.

The relevance of this research lies in its thorough examination of NER development in the critical yet underexplored domain of Islamic texts. This review is the first to synthesize NER methodologies applied to Islamic texts across multiple languages and sources from 2017 to 2024. It provides an understanding of the diverse methodologies, model components, and comparative performance within this specialized field. Therefore, this study examines four key aspects of NER in Islamic texts. First, we look at the main components of NER tasks, including the types of datasets used, the specific entity categories identified, and the features designed for entity extraction. Second, we review NER methods in this field, beginning with traditional approaches and progressing to hybrid architectures. Next, the efficacy of various approaches is evaluated through a comparative analysis of their performance. Lastly, we discuss the challenges of accurately extracting entities from Islamic texts, including linguistic nuances, semantic ambiguities, and the lack of high-quality annotated resources. With the aim of providing cross-linguistic insights and highlighting language-specific challenges, the selected papers reviewed in this study are not limited to English-language datasets; they also include datasets in other languages. Ultimately, this study contributes to the broader goal of developing domain-aware NLP technologies, bridging the gap between general-purpose NER models and the specific requirements of Islamic textual analysis.

This section provides context for the current systematic review by providing an overview of existing surveys and review articles on NER and its application in Islamic literature, which are relevant to this study. The objective is to highlight the current state of knowledge and the unique contribution of this study. A huge portion of the works analyzed are previous surveys of NER approaches. Alshammari et al [4] invented over 20 corpora of historical documents in various languages, including English and Arabic, to review the NER and classification methods used. The survey is relevant to our study since the Quran, Hadith, and numerous other Islamic literature are considered historical documents with distinct linguistic and semantic features [5]. While the aforementioned survey addressed NER challenges in general historical texts, such as adapting to language shifts, our research applies these insights specifically to the difficulties of handling a variety of Islamic texts. This distinct focus enables our study to serve as a reference, highlighting differences that may inform the development of NER models tailored to Islamic content.

This paper examines reviews of NLP specifically related to Hadith literature and to general historical documents. In one study, Azmi et al [7] did a survey on the computational and NLP-based research focusing on Hadith literature. The survey divided the studies into three categories: Hadith content-based studies, narration-based studies, and combined text-and-narration studies. Similarly, Binbeshr et al [8] discussed the approaches, datasets, evaluation metrics, and limitations of studies related to Hadith authentication and classification. In a more recent review, Sulistio et al [9] carried out a review focusing on studies that adopted the use of machine learning on Hadith texts, and analyzed the research trends, including languages, datasets, and algorithms employed. Although NLP in Hadith is covered in great detail in these works, their scope is broader than, or predates, a systematic review of NER approaches and their constituent parts in this field.

Complementing the surveys on Hadith, this study also examines surveys in Quranic research. A survey by [17] Analysis of NLP works related to the Quran by reviewing techniques, tools, and resources for Quranic NLP in the Arabic language. However, the survey does not address NER work on Quranic texts; it focuses primarily on NLP applications such as ontology, question-answering systems, translation, and speech processing, without addressing NER work in Quranic research. This research also looks at how Putra et al [10] review the text mining techniques specifically for the Indonesian translation of the Quran, along with the applications, challenges, and future research directions. Analytic techniques and tools (including NER and POS tagging) were employed to develop QA systems for the Quran. However, this study predates significant advances in NER methodologies, as it reviewed only papers published up to March 2017 [11].

Reviews of NER models applied to Arabic are also examined, given that Arabic is recognized as the original language of the Quran and Hadith. One of the earliest reviews by [12] compared the Arabic NER (ANER) models based on the machine learning approaches. The analysis discussed the linguistic resources used, the entity types extracted, the domains covered, the techniques employed, and the performance achieved by these models. A more recent and in-depth review on ANER models by [13] covered a broader spectrum of methodologies, from rule-based to deep learning and the latest pre-trained language models. This review highlighted the progression of ANER through a historical and developmental perspective. Unlike [12], this review also extended its scope to Classical and Dialectal Arabic in addition to Modern Standard Arabic (MSA). Another relevant survey on Arabic NLP applications by [14] focused on Arabic Optical Character Recognition (OCR) systems, examining the datasets and approaches used. Although these ANER reviews provide necessary linguistic context, they do not fully

elaborate on the unique conceptual and domain-specific entities in Islamic texts.

In addition, this study examines a more general review of NLP, like [15], which explored Word Sense Disambiguation (WSD) approaches in Islamic texts, along with the algorithms, datasets, and evaluation criteria. This study acknowledges the abundance of ambiguous words and phrases in Islamic literature in Persian and Arabic when implementing WSD systems. Moreover, Munshi et al [16] reviewed NLP techniques relevant to Islamic Fatwa automation systems, such as text classification and Question Answering/Chatbot systems. These systems depend heavily on the Quran and Hadith as primary sources of law and guidance. The aforementioned surveys address other critical NLP challenges in Islamic texts, but they do not provide a dedicated or systematic analysis, particularly for NER approaches. The existing reviews and surveys provide valuable insights into NER methodologies. However, the recent NER approaches applied to Islamic texts covering the crucial components (datasets, entity types, features), techniques, performance evaluations, and challenges, are yet to be systematically addressed and analyzed. This gap emphasizes the distinct contribution and necessity of the present systematic review.

II. Materials and Methods

This section describes the systematic methodology used to conduct a comprehensive review of the literature on NER in Islamic texts. The procedure involves formulating research questions and a review protocol in accordance with established systematic review guidelines. This protocol further specifies the search strategy, inclusion and exclusion criteria, and the statistical outcomes of the selection process.

A. *Research Questions Formulation*

This systematic review aims to summarize the relevant literature and provide a detailed overview of key findings on the advancement of NER for Islamic texts. To understand and improve NER systems in this field, five key research questions (RQs) underpin a systematic framework for the entire analysis and review process, as shown in Table 1.

TABLE I
Research Questions

RQ#	Question
RQ1	What kind of datasets are used in the works related to Islamic text?
RQ2	What are the features commonly used in handling entities from Islamic text?
RQ3	How are the entities extracted from Islamic texts?
RQ4	How are the performances of the applied approaches in identifying Islamic entities?
RQ5	What are the main challenges in the application of NER in the Islamic context?

The five RQs systematically examine five aspects of NER in Islamic texts. The first RQ examines the diversity of datasets selected for the studies, including aspects such as languages and sources. The second RQ examines the features commonly used to process entities in Islamic texts and their role in enhancing NER performance. The third aims to analyze approaches to extracting entities specific to the Islamic field. The fourth evaluates the efficiency of various techniques for accurately identifying Islamic entities and examines their strengths and limitations. The final research question aims to inform future advances in addressing the primary challenges of applying NER in the Islamic context.

B. *Review* Protocol

A rigid review protocol was established to ensure the careful curation of relevant studies for this systematic review. By focusing exclusively on NER studies related to the Islamic domain from 2017 to 2024, this protocol defined the precise scope of the literature search, regardless of the languages used in the datasets.

1) Search Strategy: An automated search strategy was developed and uniformly applied across the selected digital databases to identify suitable literature for this systematic review. This strategy involved utilizing each database's advanced search features and constructing a search string that included key keywords, their synonyms, and variant spellings. To ensure that no relevant studies were excluded during the search, Boolean operators ('AND', 'OR') were strategically employed in the search string. Although the core keywords remained consistent, the search string was adapted to accommodate the specific query formats of individual database search engines. The consistency of the search string across all databases is to uphold a standardized approach for a fair review. This entire search process complied to the PRISMA guidelines [18], with the resulting paper acquisition, screening, and inclusion detailed in the PRISMA flow diagram in Figure Fig.1.

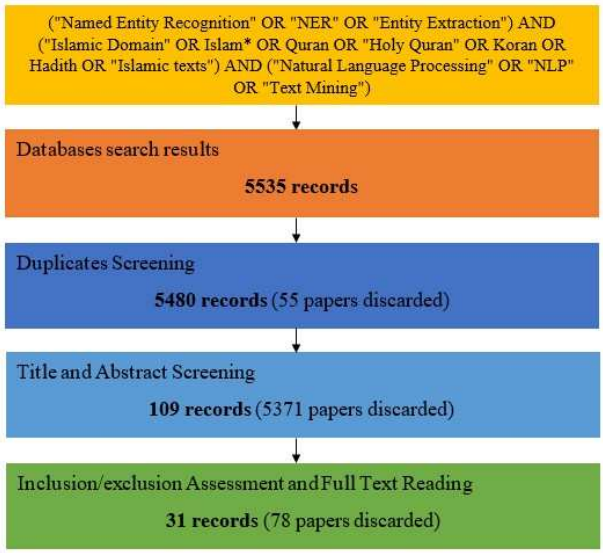

Fig. 1 PRISMA flow diagram for search strategy

2) Inclusion-Exclusion Criteria: Specific inclusion and exclusion criteria were also carefully designed and applied to the search results to ensure the relevance and quality of included literature. The initial screening process involved removing duplicates. Subsequently, the titles and abstracts are assessed to ensure alignment with the review's scope and objectives. The final selection was based on a full reading of the remaining papers.

The following is the list of the inclusion criteria:

- Articles published from 2017 to 2024.

- Articles written in English.
- Articles directly involving NER or any task requiring the recognition or extraction of specific entities.
- Evaluation and comparative study between models or methods.
- Articles focused on or directly related to the Islamic field, regardless of the dataset language used in the study.

The following is the list of the exclusion criteria:

- The type of extracted entity is not clearly defined.
- Study with Arabic language datasets that are not specific to the Islamic context (e.g., general news articles, unrelated social media texts).
- The methodology or approach used is not adequately described.
- Review articles or surveys (as this systematic review aims to synthesize primary research).

The PRISMA flow diagram in Figure 1 illustrates the careful selection of studies for this systematic review. Initially, the database search yielded 5,535 articles. Later in the first screening phase, 55 duplicates were found and removed, leaving 5,480 unique articles. Titles and abstracts are then screened to exclude articles outside the scope of the review, yielding 109 potentially relevant articles. The final stage involved a full-text review of these 109 articles against the defined inclusion and exclusion criteria. By the end of the screening process, 31 articles were selected for careful review in this study.

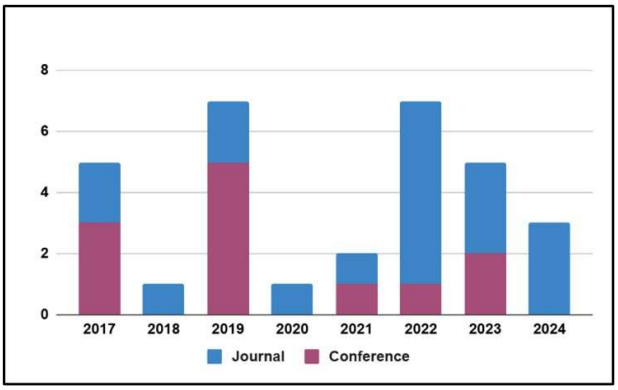

Fig. 2 Literature type distribution from 2017 to 2024

The distribution of these 31 selected papers by publication year and type (conference papers vs. journal articles) is presented in Figure 2. The annual volume of NER-related publications in the Islamic context remains relatively low, with fewer than ten studies identified per year. Journal publications appear to be more consistently present throughout the period, often exceeding conference publications in number. The years 2019 and 2022 recorded the highest combined output with 7 publications. In contrast, 2018 and 2020 had relatively few publications, with only one each.

III. Result and Discussion

Based on the rigid selection methodology specified in Section 3, this section presents a detailed analysis of the 31 selected studies on NER in Islamic texts. This analysis is structured around the research questions formulated in Section 3.1 with the aim of revealing the critical findings and providing an understanding of the current landscape. We first examine the datasets employed across these studies, focusing on the diversity of languages and sources. The features commonly selected for entity extraction are then analyzed to assess their role in improving NER performance. The discussion later turns to an in-depth examination of the approaches implemented for entity extraction in the Islamic domain. Next, this section presents comparisons of the performance of various techniques in accurately identifying Islamic entities. The final subsection addresses the final RQ regarding the primary challenges encountered in applying NER in this specialized context.

A. *What kind of datasets are used in the works related to Islamic text?*

A thorough examination of the selected articles reveals varied approaches to dataset construction, involving different languages, annotations, and sources. This analysis highlights similarities and critical gaps in current data availability, and subsequently explores the available resources for training and evaluating NER models in this specialized domain. It is important to understand the characteristics of these datasets, including their size, origin, and the specific entity types they annotate, in order to evaluate their role in the performance of the NER approaches employed.

1) Sources and Languages:

There is a growing collection of specialized datasets in Islamic text analysis, centered on two main categories: Hadith collections and Quranic texts. The Quran is the most authoritative source in terms of significance, while the Hadith serves as a supplementary source for religious laws and moral guidance [19]. These datasets show significant variations in size, language coverage, and annotation schemes. This review found that a majority of selected papers use Hadith datasets, followed closely by Quranic datasets in almost 40 percent of studies, as illustrated in Figure 3. In particular, a study by [32] is the only study that diverged from these primary sources, focusing on an Islamic article about the renowned Islamic scholar Ibn Abi Usaybi'ah.

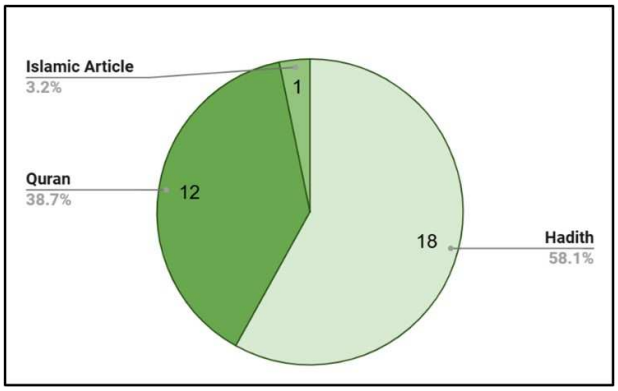

Fig. 3 Distribution of dataset sources

Hadith datasets. Despite serving as an additional source of Islamic texts, Hadith collections are used as the primary data source in most studies on extracting Islamic entities. Hadith,

which is a record of sayings and actions attributed to the Prophet Muhammad, possesses a unique structural characteristic [20]. One important feature is 'isnad' or chain of narrators, which documents the order of individuals through a Hadith is passed down. This feature is essential to prove its reliability and authenticity. The volume of Hadith literature is also significantly greater than that of the Quran, which contains fewer than 7,000 verses. This factor likely contributes to its prominence as a dataset in these studies. This preference for Hadith datasets suggests a rich resource for extracting entities associated with historical figures, theological concepts, and narrative structures.

Narrator identification is often the focus of NER tasks in Hadith texts. For instance, Emha et al [3] utilized 102 Hadith texts from the Bukhari Hadith Book, formatted in IOB format with B-Narrator and I-Narrator tags. Similarly, Rahman et al [21] employed 1,000 Malay-language Hadith texts, including an example from Sahih Bukhari. These smaller, focused datasets, often in translated languages such as Indonesian or Malay, demonstrate early attempts to apply NER to narrator recognition in particular linguistic variants of the Hadith. The manual annotation of these datasets in IOB or IOB-2 format, often with basic Person or Narrator tags, underscores the substantial effort required for initial data preparation in low-resource settings.

Other research addressing Arabic Hadith demonstrates a commitment to the original linguistic form. Alshammari et al [4] focused on Arabic Matn through the use of Sahih Al-Bukhari dataset from the Leeds and King Saud University (LK) Hadith corpus, which is a bilingual English-Arabic resource comprising 97 files and over 7,000 Hadiths. The use of established corpora, such as LK Hadith, represents significant progress toward standardized data sources. In addition, Balgasem and Zakaria [22] employed 235 Hadiths from Al-Bukhari's Book of Prayer, extracted from Sahih-Al-Bukhari online, containing 310 named entities of narrators. Furthermore, Najeeb [53] utilized 1,000 Arabic Hadiths from Sahih Muslim to build gazetteer lists from this source. These studies, typically smaller in scale than modern corpora, indicate early attempts to use structured datasets and external linguistic resources, such as gazetteers, to identify narrators in classical Arabic Hadith.

A pivotal development in classical Arabic NER research, particularly for Hadith, is the building of CANERCorpus [2]. This classical Arabic NER corpus was sourced from over 7,000 Hadiths from Sahih Al-Bukhari and manually annotated by experts. This corpus has now served as a foundational dataset for multiple subsequent studies, like studies by [51], [27], and [28]. Its importance lies in its extensive annotation, involving 21 distinct entity classifications (e.g., Person, Location, Organization, Book, God, Prophet). CANERCorpus, with approximately 258,264 words and 72,108 recognized entities, proves the critical need for high-quality datasets in this complex domain. Its availability has spurred further development and research on classical Islamic texts.

Other notable Hadith datasets and applications include the Hadith Dataset Repository, prepared for [23], which compiled data via web crawling from various authenticated sources (Sahih Muslim, Sahih Al-Bukhara, Sunnah Abu Dawood) in English, Urdu, and Arabic. In another study by [43] applied its methodology to six prominent Hadith collections, comprising a total of 34,458 Hadiths with Arabic text and Urdu/English translations, and also utilized CANERCorpus for entity extraction. This underscores the importance of larger, multilingual datasets for supporting more complex NLP tasks, such as knowledge graph construction. In 2017, a study by [46] introduced NoorCorp, a Classical Arabic corpus with 200,000 manually labelled words focusing on person, location, and organisation entities, alongside NoorGazet, which is an 18,000-name gazetteer compiled from Hadith books. The usage of Hadith datasets is beyond direct entity extraction, as shown by [54], who utilized prophetic Hadiths from multiple religious sources for classification. A study by [45] also design an architecture for question classification using Sahih Al-Bukhari. Additionally, the Gawāmi al-Kalim (GK) collection in [57] has digitized 828,841 Hadith reports and hyperlinked the names in the Isnads to a table containing biographical information on the narrators. The corpus was annotated by over 350 researchers and scholars and contained approximately 1400 sources, of which 900 were hadith collections, and the sources were related to Islamic law, theology, and literature. collection with 828,841 reports, including 447,205 Hadiths with hyperlinked narrator names. The GK collection is used by [34] in identifying narrators using Transformer-based models.

*Quranic datasets***.** Although Hadith collections constitute a significant portion of datasets in Islamic NER studies, Quranic texts are equally crucial, despite distinct characteristics and challenges in dataset construction. The Quran, as the central text of Islam, encompasses a wide range of topics for study and analysis across its 30 Juz (sections), 114 chapters, and 6236 verses. The Quran's rich textual structure contains numerous references to people, places, and events that are essential to understanding its message and historical context. Datasets derived from Quranic texts often focus on specific passages rather than covering the entire scripture. A common trend observed is the use of translated versions, particularly in Indonesian and English, due to the widespread readership and the complexity of direct Arabic annotation of the Quran.

Several studies focus on extracting general or person entities from Quranic translations. A study by [44] utilized English Quran verses from tanzil.net, comprising 19,473 tokens and 720 manually labelled entities for supervised learning. Similarly, a study by [30] employed a BIO tagging scheme on verses from Surahs Al-Fatiha, Al-Baqara, and Al-Imran using the same source (tanzil.net). For Indonesian translations, Mutia and Bijaksana [29] used BIO format for a dataset of 36,752 tokens and 3,139 human entities from Juz 1 to Juz 6 of the Tanzil Quran corpus. Likewise, a study by [31] labelled person entities (PER) in IOB-2 format on Juz 30 of the Indonesian Quran translation. These studies highlight the common adoption of IOB/BIO tagging schemes for Quranic translations. Most studies have employed relatively small datasets, often covering specific Juz or Surahs, owing to the linguistic intricacies of the Quranic text.

A recent contribution to standardized Quranic NER benchmarking is the IndQNER dataset for Indonesian Quran translation, proposed by [40]. This benchmark dataset contains 2,476 Named Entities (NEs) across 18 distinct classes. Quran and Tafseer experts are involved in carefully

examining a publicly available Arabic Quranic corpus (corpus.quran.com) and defining initial classes, such as Allah, Artifact, Angel, Person, and Location. The final annotated dataset comprises 3,117 sentences and 62,027 tokens, tagged in the BIO format and split into training, validation, and test sets. The introduction of IndQNER aims to foster further development and comparative analysis of Quranic NER, particularly in Indonesian-translated literature.

In addition to general entity extraction, several studies have focused on highly specialized entity types or specific translation methods. For instance, Hassan et al [35] uniquely targeted matching Allah's names across the entire original Arabic Quran. Tarmizi and Saad [26] utilized all 112 verses of Surah Al-Anbiya from the Hilali-Khan English translation to extract entities specific to the Quran. Tarmizi and Saad [1] used the same dataset to compare NER model performance. Moreover, some studies addressed more complex entity structures or broader NLP tasks in which NER could be a component. Dzidny et al [42] specifically addressed the challenge of nested people entities within Quranic texts. Its dataset demonstrates the complexity of annotation for nested structures by using tokens with POS tags and multi-level entity labels (Level-1, Level-2, and Joint-tag) in BIO format. A study by [48] created a manually annotated Quranic text for POS tagging by nearly 100 volunteers, using 43 POS tags and extracting proper noun locations. Although studies by [37] and [45] used large Quranic datasets from sources such as the Tanzil project for classification and similarity tasks; these comprehensive textual resources could serve as baselines for future NER endeavors.

In summary, Hadith datasets are often larger, containing over 100,000 tokens, whereas Quranic datasets are smaller and more focused. The small scale of many datasets, typically centered on specific chapters or Juz, is due to the labor-intensive nature of annotation. The emergence of benchmark datasets such as IndQNER and CANERCorpus, developed with expert input and comprehensive entity schemes, represents significant progress toward standardizing and advancing NER research in this domain. Future research could benefit from larger, more comprehensive datasets that capture a wider range of entity types beyond persons and divine names, enabling the development of more advanced NER systems.

2) Entity Types:

The Quran and Hadith convey guidance through extensive narratives and rich storytelling elements. They naturally cover numerous entities within their discourse, ranging from general ones to more specific religious concepts. The analysis reveals a spectrum of extracted entities unique to Islamic texts. This diversity shows the rich and nuanced information embedded within Islamic literature.

The most frequently targeted entity type across the reviewed studies is Person, particularly Narrator in the context of Hadith, due to the critical role of the 'isnad' (chain of narrators) in authenticating Hadith. Studies like [3], [21], [24], and [31] predominantly focus on identifying person names, often tagged as 'Narrator' or 'PER'. This emphasis extends to the original Arabic as Najeeb [53] identifying 'Narrator-Name' and 'Prophet-Name', and Balgasem and Zakaria [22] also targeting person names. Alkaoud and Syed [34] further highlights the importance of narrator identification as a core task.

In addition to the person entity, several studies also extract other standard NER categories, such as Location, Organisation, Book, and Temporal expressions (Date, Time). For example, Sajadi and Minaei [46] targets person, location, and organization entities. Tarmizi and Saad [1] also reports on common NER categories like Person, Location, Organisation, Time, Date, Number, and Money when applied to Quranic text. The comprehensive CANERCorpus, used in studies by [51], [28] and [27], includes these traditional categories (e.g., Loc, Org, Book, Date, Time) as part of its 21 distinct entity classifications. However, numerous studies have also attempted to extract specialized entities from Islamic texts. Study by [26], for instance, extracted more domain-specific entity types such as names of prophets and people (PER, PER-PROPHET, PER-GROUP), creations (CREATION), locations (LOC, LOC-AFTERLIFE), special time (STIME), and various names of God (GOD). Divine entities, such as God (Allah) and the Prophet, are also prominent in corpus like CANERCorpus and specifically addressed by [35] for matching Allah's names. The IndQNER dataset by [40] extends this to include Angel and Artifact (which includes religious artifacts). CANERCorpus also includes more abstract theological and conceptual entities like Sect, Religion, Heaven, and Hell as part of the 21 classes.

Furthermore, NER in Islamic texts often targets structural components unique to Hadith and Quranic discourse. In Hadith, this includes granular 'isnad' related entities such as 'Narrator-Name-Prefix', 'Received-Method', 'Received-Method-Prefix', 'Replacement', and 'Title' [53]. Study like [22] also delve into the 'semi-NEs' like "son of" and "father of" or "Termination Words" like glorification phrases within narrator chains. Aside from narrators, Alias et al [50] discusses the extraction of structural Hadith components such as Book Number, Book Name, Chapter Number, Chapter Name, Hadith Number, Hadith Sanad, and Matan. Some research addresses even more complex entities, such as nested entities. Dzidny et al [42] specifically attempted to identify multi-level 'PER' (person) entities within Quranic texts, utilizing Level-1, Level-2, and Joint-tag labels. Overall, the diversity of entity types extracted ranges from common nouns to more specialized theological concepts and structural elements. Additionally, it also emphasizes how NER applications in Islamic texts are becoming more sophisticated and reflects the particular informational requirements of this academic field.

B. What are the features commonly used in handling entities from Islamic texts?

The analysis of selected literature reveals a diversity of approaches, ranging from linguistically engineered features in traditional machine learning and rule-based approaches to learned representations, particularly through neural networks and pre-trained language models. These features can be broadly categorized into lexical and linguistic features and representation-based features (embeddings), which are typically used in hybrid combinations.

1) Lexical and Linguistic Features:

Early NER approaches and many traditional machine learning models in the Islamic domain heavily employ hand-crafted lexical and linguistic features. These features capture explicit textual and grammatical properties crucial for identifying entities. Part-of-Speech (POS) tagging, in particular, is widely used as a fundamental linguistic feature to identify named entities regardless of text type or language across multiple studies. In Hadith processing, POS tagging is used to recognize narrator names through religious keywords and pronoun positioning [22]. Meanwhile, for Arabic text processing, it is implemented with base phrase chunking to identify named entities [51]. For Indonesian translations, POS tagging facilitates proper noun detection in the Quran [48] and supports name indexing [24]. In addition, morphological analysis is incorporated, particularly for highly inflected languages such as Arabic. The features derived include word affixes (prefixes and suffixes) [46], [49] stem, root, and morphological ambiguity [51], [46], [34] and specific Arabic morphological cues like diptotes [51]. These features help disambiguate words and identify proper nouns. In addition to morphology, orthographic features like capitalization, digits, punctuation, and special characters are also commonly used as studied by [51] and [46] to provide basic visual cues about the nature of a word. In Arabic, the existence and importance of diacritics (Al-Tashkeel) are considered orthographic features [35]. Understanding the local context of a potential entity requires contextual word features, which are frequently obtained by examining words that precede and follow [28], [44], [46]. Furthermore, studies by [53] and [46] demonstrate the use of external knowledge sources, such as gazetteers and keyword lists, as features to enable a direct lookup mechanism for known entities.

2) Representation-based Features:

Strong representation-based features that automatically extract dense vector representations from text data have been made possible by the development of deep learning. Word embeddings, such as GloVe, enable models to capture word similarities and relationships by encoding semantic and syntactic information into continuous vectors, as shown by [4]. On the contrary, character embeddings, as used in studies by [28] and [4] provide a more granular level of representation, which is useful for handling out-of-vocabulary words and capturing morphological variation. The most notable of these features are contextual embeddings obtained from pre-trained language models (PLMs) such as BERT and RoBERTa. Compared with static embeddings, these models better capture polysemy and complex contextual dependencies by generating dynamic word representations that vary with the word's context. Research by [3], [27], [30] and [29] use these sophisticated embeddings as primary features and demonstrate their superior performance in capturing complex contextual information in Islamic texts.

3) Hybrid Features Combination:

Many effective NER systems in the Islamic domain employ hybrid approaches that combine multiple feature types and methodologies to leverage their complementary strengths. This approach often involves integrating traditional linguistic features with more advanced techniques. For instance, Balgasem and Zakaria [22] combined rule-based features (POS tags, keywords) with statistical methods that measure the "strengthen value" of candidates based on occurrence frequency. Similarly, Dzidny et al [42] utilized a rich set of features including word features, N-gram features, POS tags, lexical features (e.g., word shape, capitalization), and semantic features (dictionaries) to address the complexity of nested entities. Meanwhile, Hasan et al [45] combined POS tag features with statistical features based on the occurrences of nouns and verbs in the sentences.

Overall, the feature engineering in Islamic NER reflects how NLP progresses from manually crafted linguistic features to learned, dense embeddings. The continued use of gazetteers, even with advanced models [53], [46], highlights the value of external knowledge and curated lists as a feature type. This integrative approach proves that understanding the complex linguistic nature of Islamic texts requires a compound feature set to achieve accurate entity recognition.

C. How are the entities extracted from Islamic texts?

The approaches identified in the literature reflect an evolution from explicit rule-based systems and statistical machine learning models to advanced deep learning architectures and sophisticated hybrid methods. Each methodological paradigm showcases distinct advantages in processing the unique linguistic and structural properties of Islamic literature.

1) Rule-based approaches:

Rule-based approaches are foundational not only to Islamic NER but also to NER in general. This approach depends on meticulously crafted linguistic rules, patterns, and dictionaries derived from expert knowledge. These methods excel in precision for well-defined entities but can be labor-intensive to develop and less adaptable to linguistic variations.

A study by [28] proposed a novel rule-based system for Classical Arabic NER to identify entities like Person, Location, and Organization. This approach explicitly applied pattern matching and dictionary lookup, with rules based on word features, contextual features, word length, and frequency. This method is particularly suitable for Classical Arabic, given its relatively stable grammatical structure and the need for high precision in scholarly contexts. For highly specific entity types, Hassan et al. [35] developed a rule-based method tailored to match Allah’s names in the Holy Quran. The researcher built a dictionary of Allah's names and implemented rules based on orthographic features, N-grams (for approximate matching and similarity calculation), and affixes. Precision in this rule-based approach is important, as the accuracy of divine names is crucial in sensitive religious texts. In Hadith analysis, Rahman et al [21] employed rule-based heuristics covering keywords, character analysis, and morphological patterns to tag narrator names in Malay Hadith texts. This study demonstrated the efficacy of rule-based systems for identifying highly structured elements, such as narrators' names, particularly when supported by linguistic insights. Another study by [50] implemented a rule-based algorithm that used POS tags and contextual rules to identify narrators’ names and other general noun entities within Hadith 'sanad' parts. This approach emphasized a sequential process of word tokenization followed by rule application

based on predefined POS tags. Tarmizi and Saad [26] implemented the use of gazetteers to extract fixed types of entities, such as the names of the prophets and individuals in the Quran. Regular expressions are also part of the rules used to extract entities with a unique pattern, such as the name of God that often starts with prefixes such as 'Oft-', 'Most', and 'All-'. Compared with rule-based systems, statistical machine learning approaches can learn patterns from annotated data and generalize to unseen text. One of the earliest statistical methods, the Hidden Markov Model (HMM) and its Viterbi algorithm, is prominent in sequence labelling tasks within Islamic NER. In one study, a study by [25] applied HMM for name indexing in Indonesian Hadith translations by using the model's probabilistic framework to predict sequences of named entities. Similarly, a study by [31] used HMM with the Viterbi algorithm for person entity recognition in Indonesian Quranic translations, once again proving HMM's suitability for capturing sequential dependencies in text. In general, HMMs are well-suited to tasks in which the output sequence is probabilistic, and dependencies between successive tags are primarily local.

The Conditional Random Field (CRF) model has also been a popular choice due to its ability to capture dependencies among labels in a sequence without imposing strong independence assumptions on observations. Salah et al [51] extensively utilized CRF for classical Arabic NER by integrating a rich set of features, including lexical, morphological, syntactic, and semantic cues. This study highlights the ability of CRF in complex languages like Arabic by incorporating diverse, rich features and handling long-range dependencies. A study by [49] Also used CRF to recognize person entities in Indonesian Quran translations by leveraging features such as word affixes and chunking to enhance contextual understanding. In addition, the Support Vector Machine (SVM) has also been used to identify entities in English Quran versions, where its strength in high-dimensional feature spaces was utilized using word, character, and linguistic features [44]. Lastly, the Naive Bayes Classifier, known for its simplicity and efficiency, was experimented with to index names from Indonesian Hadith translations using features like POS tags and contextual information [24]. Although these statistical models require effective feature engineering, they still provide robust frameworks for learning from data.

2) Deep learning approaches:

have greatly advanced NER by automating feature learning and capturing intricate contextual relationships. Recurrent Neural Networks (RNNs), particularly their Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) variants, combined with CRF, have emerged as powerful architectures for sequence labelling. Alshammari et al [4] addressed the ability of BiLSTM-CRF and LSTM-CRF models to capture sequential information effectively when applied to Arabic Hadith by utilizing GloVe word embeddings and character embeddings. A study by [27] further explored complex deep learning architectures integrating word and character embeddings, including BiLSTM-CRF, CNN-CRF, and a hybrid BiLSTM-CNN-CRF. These models excel at processing sequences and learning hierarchical representations, making them suitable for the complex linguistic structure of classical Arabic.

More recently, Transformer-based models, such as BERT (Bidirectional Encoder Representations from Transformers) and RoBERTa, have set new benchmarks for contextual understanding. These models significantly improve their ability to capture semantic nuances and polysemy by generating highly contextualized word embeddings that consider the entire input sequence simultaneously. Emha et al [3] successfully fine-tuned the BERT model and utilized its pre-trained knowledge for automated Hadith narrator identification. Similarly, studies by [30] and [29] and effectively applied BERT and RoBERTa, respectively, for person entity recognition in Quranic translations. The application of these state-of-the-art PLMs enables models to achieve superior performance by automatically learning rich, distributed representations of words and their contexts.

3) Hybrid approaches:

often yield more robust and accurate NER by strategically combining different methods to take advantage of their complementary strengths. Najeeb [53] employed a sophisticated hybrid strategy combining HMM with gazetteer lists. An efficient cascade of knowledge-based and probabilistic techniques was demonstrated by first using the gazetteer lists to identify known entities (Points of Interest or POIs), and then using the HMM component to predict categories for unassigned or ambiguous Isnad-Phrases. Another noteworthy hybrid method by [22], combined rule-based techniques (using POS tags and keywords) with statistical measures like Log-likelihood Ratio (LLR) and Point-wise Mutual Information (PMI). The approach combined linguistic rules and statistical validation by assigning a "strengthen value" to the identified narrator candidates based on their frequency of occurrence.

Furthermore, several information-extraction tasks clearly demonstrate sophisticated hybrid frameworks. Alkaoud and Syed [34] combined rule-based techniques, heuristics, and machine learning models for narrator identification as well as utilising structured biographical data to enhance accuracy. Sajadi and Minaei [46] utilised an AdaBoost classifier, an ensemble machine learning method, which also functions as a hybrid through its strategic combination of multiple weak learners and integration of various orthographical, contextual, morphological, and lexical features. A study by [23] employed a complex hybrid approach combining rule-based methods (gazetteers), statistical methods (TF-IDF, cosine similarity), and machine learning (SVM) for comprehensive information retrieval and knowledge extraction from Hadith. Finally, a study by [43] proposed a hybrid strategy by combining rule-based techniques for specific pattern extraction with advanced machine learning methods, including word embeddings and deep learning, to generate knowledge graphs from Hadith. Hasan et al [45] also has successfully merged SVM classification with pattern-matching techniques to build a more robust system for question analysis in Hadith texts. This comprehensive application of hybrid methodologies emphasizes the need to integrate diverse strengths to address the multiple challenges of NER and information extraction in Islamic texts.

D. How are the performances of the applied approaches in identifying Islamic entities?

The performance of different NER models in the Islamic context varies significantly, depending on the underlying techniques and the text's unique linguistic characteristics. The performance metrics, typically F1-score, precision, recall, and accuracy, vary depending on the approach used, the specific entities targeted, and the characteristics of the datasets. Table 2 summarizes the performance of NER approaches on Hadith datasets, while Table 3 summarizes their performance on Quranic datasets.

Although they involve substantial manual labor, rule-based methods exhibit excellent accuracy for clearly defined entities in specific situations. A study by [28] demonstrated strong performance, achieving a 90.2% precision rate, an 89.3% recall level, and an F-measure of 89.5%. The study revealed that removing gazetteers drastically reduced the F-measure to 75.5%, proving their importance in recognizing most Arabic Named Entities. Furthermore, the absence of trigger words, patterns, or grammatical rules also led to performance declines (F-measures of 81.6%, 84.4%, and 88.4%, respectively), underscoring their critical role in identifying entities not covered by gazetteers. This proposed method clearly outperformed existing baseline tools such as GATE and Language Computer, which achieved lower F-measure values and failed to recognize certain named entities. Alias et al. [50] verified the effectiveness of their POS-tag and based rule algorithm in capturing relevant entities with a high precision of 96.72% and 93.64% for recall. Tarmizi and Saad [26] demonstrated a strong overall performance, achieving an F-score of 0.94, with a high precision of 0.98 and a recall of 0.91. This performance significantly outperformed general-purpose tools: a preliminary experiment using the general-purpose SpaCy NER library on the same dataset yielded an F-score of only 0.61. This result underscores the need for domain-specific customization in Quranic NER. Additionally, the same study reported perfect performance (F-score of 1.0) for specific entities such as PER-PROPHET (prophet names) and STIME (sacred times), largely attributable to the comprehensive gazetteers for these categories. Rule-based systems have been shown to achieve high precision on narrow, well-defined tasks where linguistic patterns are consistent.

In some studies, statistical machine learning models have been shown to outperform purely rule-based methods by learning patterns from data. Najeeb [53] reported that the Hidden Markov Model (HMM) method alone achieved an average accuracy of 83% for all Part-of-Isnads (POIs) predictions, which was higher than that of gazetteer lists alone (75% accuracy). This highlights HMM's capability in probabilistic sequence labelling. HMM achieves an impressive F1-score of 86% with the combination of POS-Tag, BIO Tag, contextual, and capitalization features when tested on Indonesian Hadith translation [25]. The F-1 score increased to 88\% as a result of further cross-validation refinement, especially when concentrating on state transition parameters with a larger variety of tokens. The system has also successfully indexed 84,052 names from the Hadiths. When evaluated on the Quranic dataset, a study by [31] reported that HMM achieved a best F1-score of 76% through the combination of preprocessing steps (including case folding, tokenizing, stopword removal, punctuation removal, stemming, and lemmatization) and POS-tagging features. Conversely, the system's performance drastically dropped to its lowest F1-score of 43.5% when only preprocessing was applied without POS-tagging. This contrast underscores the importance of rigorous data preprocessing and the substantial impact of POS tagging.

By using SVM, particularly the LinearSVC module, a study by [41] achieved a remarkable F1-score of 0.9 in indexing names from Indonesian Hadith translation through the effective combination of POS-Tag, Unigram, and Titlecase features. When using the same model on the Quranic dataset, a study by [44] found that a linear kernel combined with the Unigram feature produced the highest F-measure value of 0.75. While the linear kernel generally performed better across various features, the RBF kernel, although achieving the best precision, incurred a higher computational cost. An important finding was that the POS tagger alone performed poorly due to the Quranic text's distinct sentence structure compared with modern languages. On the contrary, the Unigram feature proved particularly effective, likely due to recurring patterns in entity structures within the Quran, such as common phrases like "those who are" or "the one who." The entity identification was facilitated by using Latin letters in the English translation, rather than in the original Arabic text. Unigram continues to be proven as essential in another study by [24] using a Naïve Bayes (NB) classifier. The combination of POS Tag, Unigram, and title-case features yielded the highest F1-score of 82.63%. The same study reported lower F1-scores for individual features (POS Tag: 76.75%; unigram: 71.14%). However, the study also identified areas for improvement by addressing errors, such as misclassifying non-person names beginning with a capital letter (e.g., "My father," "Judgment Day") and failing to recognize names immediately following punctuation.

Salah [51], who employed a Multinomial Naive Bayes classifier, reported an F-measure of 80% (86% precision and 75% recall) on the Classical Arabic dataset. When various features, specifically word-level, morphological, and knowledge-based features, were combined, the overall method attained an even higher F-measure of 0.805 (with 0.8750 precision and 0.7507 recall). This integrated approach consistently outperformed individual feature types across most categories. A study by [54] also showed that the Multinomial NB classifier achieved the highest accuracy (97.08%) in classifying Hadiths and outperformed LinearSVC (96.55%) and KNeighborsClassifier (96.44%). The model's exceptional accuracy stems from its strong ability to handle high-dimensional problems and to effectively identify key features during training. Furthermore, the study revealed that classification accuracy for Hadith categories based on their attribution to the Prophet Muhammad (Saying, Doing, Describing, Reporting) was significantly higher than that for classifications based on narrators' reliability and memory (Sahih, Hassan, Da'if, Maudu). In contrast, for classifying Quran verses, a study by [37] found that the kNN classifier outperformed SVM and NB classifiers using both CFS (Correlation-based Feature Selection) and IG (Information Gain) feature selection algorithms, without a fixed feature-selection threshold. The model achieved the highest F1-score of 0.773, with an Area Under the Curve (AUC) of 0.88. The study concludes that a lower selected

threshold for feature selection yields a more effective feature subset and improved classifier performance.

For CRF, a study by [49] reported an average F1-score of 0.77 when trained on 36,814 tokens from 954 verses. The best performance observed within this study was a precision of 0.86 with CRF and Prefix features, and an F1-score of 0.77 when combining CRF with both Suffix and Prefix features to extract the entity person. The results show that independent language features, such as affixes (suffixes and prefixes), significantly improved the model's performance. Interestingly, Dzidny et al [42] recorded an average F1-score of 0.741 (0.763 for Level-1 F1 and 0.719 for Level-2 F1) using the same classifier to extract the nested people entity.

In a comparative study of well-known general-purpose NLP tools on specialised domains, Tarmizi and Saad [1] revealed that while SpaCy, NLTK, and Stanford NER typically perform exceptionally well for general text, their performance significantly declines when used on Islamic text. SpaCy, for instance, extracted more entity types and achieved high F1 scores for PERSON (0.86), CARDINAL (0.92), and GPE (0.95), demonstrating strong performance on general text. However, there was a noticeable decline in performance for the Islamic text, especially for the person entity, whose precision fell to 0.44. Despite its strong recognition of prophet names, SpaCy misclassified an excessive number of non-person entities as persons. The same problem also occurred for NORP, GPE, and ORG, which had an F-score of 0 (i.e., no correct extractions) on Islamic text. NLTK exhibited a similar decline when processing Islamic text. It had very low precision (0.14) for person entities in this domain, resulting in an F-score of only 0.23. One of the main causes of this particular problem was NLTK's tendency to misclassify the entity ‘Allah’ as a person, resulting to serious errors. For general text, NLTK's F-score for PERSON was 0.56, but it drastically fell to 0.23 for Islamic text. The Stanford NER model consistently displayed precision scores significantly lower than recall scores on Islamic text, mirroring the performance problems of the other two models. Similar to NLTK, Stanford NER, a CRF-based model [58], frequently categorized ‘Allah’ as a person, contributing to its low precision (0.46) and F-score (0.57) for the person entity in Islamic text. It failed miserably to extract ORGANISATION entities from Islamic text (F-score 0), despite achieving respectable F-scores for PERSON (0.73) and ORGANISATION (0.74) in general text. Collectively, these results indicate a limitation: general-purpose NER models often perform poorly in highly specialized domains, such as Islamic texts, without substantial adaptation or domain-specific training, even though they perform well on heterogeneous data. These models misclassified and failed to recognize relevant entity types due to inherent biases or a lack of domain-specific knowledge, underscoring the need for customized solutions for these specialized applications.

Deep learning models, especially those that use pretrained language models, consistently demonstrate state-of-the-art performance by automatically learning rich contextual features. Emha et al [3] achieved an exceptionally high overall F1-score of 99.63% during its training phase by adding an additional feed-forward classifier into the last layer of a pretrained BERT model (specifically Cahya/bert-base-indonesian-1.5G). More remarkably, during evaluation on an additional forty New Hadith texts, the same model maintained its superior performance, achieving an F1 score of 99.27%, outperforming the general BERT model (F1 Score of 98.68%). Although the evaluation results were, on average, 0.28% lower than the training results, the consistent performance order of the pre-trained BERT models highlighted their robustness. While some minor tagging errors, such as "Abu" being incorrectly labelled as B-Narrator rather than I-Narrator, were noted, the overall performance represents a significant advance in accurately identifying narrators in Hadith texts. In another study by [30], the general BERT model only achieved an F1-score of 53% in extracting the people entity from the English Quranic dataset. Similarly, when a RoBERTa model was utilised for identifying human entities in the Indonesian Quran translation, it obtained an F-score of 52% [29]. In a direct comparison for this task, RoBERTa's F-score of 0.522 was marginally lower than BERT's 0.527, indicating that RoBERTa, despite generally exhibiting superior language processing capabilities due to its larger and more diverse training datasets, did not perform optimally. These findings suggest that even advanced transformer architectures face considerable hurdles in accurately extracting person entities from the Quran's unique linguistic and structural characteristics. For nested people entity extraction, a BiLSTM-CRF model integrated with FastText (skip-gram) features achieved the highest performance, boasting an average F1-score of 0.785 (0.81 for Level-1 F1 and 0.76 for Level-2 F1) [42]. IndoBERT, a pre-trained language model, also delivered strong results, achieving an average F1 Score of 0.755 (0.78 for Level-1 F1 and 0.73 for Level-2 F1).

For classical Arabic study by [27], the BERT-BGRU-CRF model demonstrated superior performance on the CANERCorpus, achieving the highest F-measure of 94.76% among all tested architectures. This model outperformed the BERT-BLSTM-CRF variant by 0.34 points in F-measure, indicating a slight advantage of the BGRU layer over the BLSTM in this context. The CRF algorithm's effectiveness in capturing tag dependencies was evident, as the BERT-CRF model achieved a 0.23-point improvement in F-measure over a BERT model with only a linear classification layer. Further enhancing the architecture by adding a BGRU layer yielded an additional 0.08-point increase in F-measure. Interestingly, the study found that concatenating CNN character-embedding representations reduced model performance, suggesting that BERT's inherent WordPiece tokeniser effectively addresses out-of-vocabulary (OOV) issues without requiring additional character-level features. Likewise, stacking additional deep learning layers did not improve performance. In another classical Arabic study by [34], ukhBERT achieved an F1 score of 96.15% for narration detection and 95.74% for narrator linking. The model's performance on the narrator linking task was particularly noteworthy, as it significantly surpassed BiLSTM-CNN by over 20% and the search-based approach by more than 45%. A pivotal contribution of this work is its demonstration of the effectiveness of transformer-based models in resolving entity linking problems within Arabic text. In an initial experiment by [4], the Stanza and Marefa-NER models achieved the best results on the LK Sahih Al-Bukhari dataset, with F1 scores of 0.826 and 0.807, respectively. Stanza, a pre-trained neural NLP model [59], also achieved the highest precision (0.787) and recall (0.869).

However, the landscape changed significantly when these models were tested on a new, corrected test set meticulously annotated by human experts using CANERCorpus guidelines. On this refined dataset, performance dropped considerably across the board, with Hatmimoha performing best, with an F1-score of 0.239. Marefa-NER followed at 0.227, and Stanza's F1-score fell to 0.191. This stark decline underscores a critical point: Islamic texts possess unique linguistic characteristics, including specific names of God and prophets, which differentiate them from MSA.

TABLE II
NER PERFORMANCES ON HADITH DATASETS

Ref.	Language	Methods	Best F1-score
[53]	Arabic	Hybrid (HMM, gazetteer)	0.86
[22]	Arabic	Hybrid (Rule-based, LLR)	0.82
[51]	Arabic	Statistical (MultinomialNB)	0.805
[28]	Arabic	Rule-based	0.895
[46]	Arabic	Hybrid (AdaBoost, gazetteer)	0.9604
[3]	Indonesian	Deep learning (Indonesian BERT)	0.9927
[27]	Arabic	Deep learning (BERT-BGRU-CRF)	0.9476
[54]	Arabic	Statistical (MultinomialNB)	0.97
[45]	Arabic	Hybrid (SVM and pattern matching)	0.8793
[4]	Arabic	Deep learning (Stanza)	0.826
[25]	Indonesian	Statistical (HMM)	0.88
[34]	Arabic	Deep learning (AraBERT)	0.9615
[24]	Indonesian	Statistical (NB)	0.8263
[41]	Indonesian	Statistical (SVM)	0.9
[23]	Arabic	Statistical (CRF)	0.9241
[43]	Arabic	Deep learning (custom SpaCy)	N/A
[50]	Malay	Rule-based	0.9516
[21]	Malay	Rule-based	N/A

Hybrid approaches strategically combine the strengths of different methodologies, often yielding improved, more robust performance by mitigating individual limitations. Najeeb [53] demonstrated this by achieving the highest average accuracy of 86% for all POI predictions when combining gazetteer lists with HMM, surpassing the individual performance of gazetteers (75%) and HMM alone (83%). Likewise, Salah and Zakaria [2] also demonstrated superior performance with a hybrid approach combining a rule-based method with statistical measures (LLR), achieving an F-measure of 0.82 (0.95 precision, 0.73 recall), which was better than the rule-based method alone (0.8 F-measure) or LLR alone (0.76 F-measure). In addition, Hasan et al [45] obtained the highest average F1-score of 87.93\% through the hybrid approach of SVM classifier and pattern matching in classifying Hadith questions.

Tables 2 and 3 present a comparison of NER performance between the Hadith and Quranic datasets. In addition to the approaches used, the tables also reported the languages of the datasets, providing a clear contrast in performance across specific languages. On average, the F1 Score of the best-performing models on Hadith datasets is higher than that on Quranic datasets.

TABLE III
NER PERFORMANCES ON QURANIC DATASETS

Ref.	Language	Methods	Best F1-score
[37]	English	Statistical (kNN)	0.775
[48]	Indonesian	Rule-based	0.968
[44]	English	Statistical (SVM)	0.75
[40]	Indonesian	Deep learning (BiLSTM-CRF)	0.96
[26]	English	Rule-based (gazetteer)	0.94
[30]	English	Deep learning (BERT)	0.53
[29]	Indonesian	Deep learning (RoBERTa)	0.522
[49]	Indonesian	Statistical (CRF)	0.77
[31]	Indonesian	Statistical (HMM)	0.76
[1]	Arabic	Rule-based	N/A
[42]	Indonesian	Deep learning (BiLSTM-CRF)	0.785

E. What are the main challenges in the application of NER in an Islamic context?

Islamic NER models face various challenges, particularly when applied to complex religious texts such as the Quran and Hadith. The complexities arise from the distinct linguistic characteristics of these texts, limitations in available data, and the inherent difficulties of existing methodologies, all of which point to important areas requiring further research.

1) Linguistic Complexities and Domain Specificity:

A primary challenge in applying NER to Islamic texts is the complex linguistic characteristics, especially for Classical Arabic. Salah et al [51] mentioned that issues like the lack of capitalisation, complex morphology, presence of multi-word expressions, and the significance of diacritics, complicate the process of entity identification. A study by [27] also stated that the complex morphology of Classical Arabic makes it prone to the Out-Of-Vocabulary (OOV) problem in NLP tasks. According to [22], recognizing Arabic names often requires treating them as multi-word expressions, which poses a specific challenge for NER systems. Studies by [24] and [25] both revealed a major linguistic limitation in indexing names in the Indonesian translation of the Hadith, namely the variation of a person's name when translated from Arabic texts. For example, the name 'Aisyah' can be spelled 'Aishah' or 'Aisha' in Indonesian texts. Furthermore, automatic and accurate entity extraction is challenging because Islamic texts often contain complex structural elements, including nested entities. For instance, the CRF model, a popular statistical method, struggles to identify nested people entities in Quranic verses [42]. This indicates that general NER solutions are often insufficient and require customization to account for the nuances of Islamic scholarly domains. The limitation is also evident in Arabic transliteration, where maintaining a consistent mapping between the Arabic script and other writing systems, such as English, is important for effective NER implementation [46]. Religious sensitivity and accuracy require an additional dimension in addressing the challenges. According to [35], the handling of religious terms demands extra attention because misclassification of terms like "Allah" can be considered offensive and compromise the credibility of the system. Therefore, careful consideration is necessary in the development and implementation of NER systems for Islamic texts. Alkaoud and Syed [34] emphasized the need for supervision from domain experts to ensure both linguistic and

contextual accuracy. A study by [43] also acknowledges the importance of expert validation, particularly when addressing the distinct Islamic terminology and Arabic linguistic features.

2) Data Scarcity and Annotation Burdens:

The lack of comprehensively annotated corpus is a major barrier to the advancement of NER in Islamic texts. Both studies by [21] and [50] mentioned the absence of an annotated corpus and existing systems to perform NER or develop POS tags for Malay Hadith texts. Alias et al [50] specifically stated that the development of annotated datasets is prone to errors and requires a lot of time and effort from skilled annotators. This manual burden stresses the need for automation in NER processes. Additionally, studies by [40] and [48] emphasized the limited availability of annotated data on Indonesian Quranic texts. In addition to the limited dataset, a study by [37] encountered difficulty identifying studies that classify Quranic verses based on English translation, as most studies focus on Arabic. Furthermore, inconsistencies coming from different preprocessing steps and the use of different Hadith corpus across studies make it challenging to compare diverse research studies related to Hadith science [53]. This makes it difficult to establish reliable benchmarks and conduct direct performance comparisons. A major constraint across multiple studies that impedes the development and evaluation of accurate NER models is the lack of annotated corpora specifically for Islamic texts. This limitation is evident when dealing with specialized entities such as Hadith narrators, as emphasized by [22] and [34], who require curated datasets to train domain-specific models. In addition, general-purpose NER tools are inadequate for Islamic texts due to their lack of training in the specific vocabulary, grammar, and contextual nuances unique to this domain [1].

3) Methodological Limitations and Gaps:

In addition to linguistic and data challenges, current NER methodologies possess specific limitations when applied to Islamic texts. Although gazetteer lists are useful, they usually do not cover the entire corpus and later suffer from ambiguity issues and subsequently yield lower prediction accuracy compared to other methods [53]. This indicates that gazetteer lists have limited applicability and are inefficient when used in isolation. Moreover, the technique developed by [24] could not yet accurately identify names following the punctuation '-' like 'Al-Khattab'. The system also requires further improvement to prevent misidentifying capitalized non-person names as person names. Besides, advanced models also struggle with certain complexities. For instance, as noted previously, CRF models struggle with nested entity extraction. Even deep learning models like IndoBERT and BiLSTM-CRF have not consistently extracted entities at very deep (e.g., level-4) nesting levels [42]. Gusmita et al [40] also found it difficult to implement transfer learning for NER tasks in this domain and suggested domain-specific fine-tuning for general models like IndoBERT. Furthermore, Najeeb [53] observed shortcomings in the exploration of certain advanced feature extractions, such as syntactic and semantic features, which could capture deeper linguistic relationships in these texts.

IV. Conclusion

In conclusion, this systematic literature review provides a comprehensive analysis of NER across diverse Islamic texts, primarily the Quran and Hadith, in languages including Arabic, Indonesian, English, and Malay. The majority of studies employed lexical and linguistic features, either in isolation or in hybrid combinations, to address the distinct linguistic elements of religious texts. The effort of many researchers in experimenting with various approaches to precisely extract important entities from Islamic texts is highly appreciated. The studies employed not only traditional rule-based and statistical machine learning methods but also advanced deep learning and hybrid architectures to further improve the NER system. The technological advancement of deep learning approaches, especially transformer-based models, has enabled them to perform impressively and consistently achieve the highest F1 Scores compared to other approaches. Nevertheless, rule-based and statistical methods remain useful, particularly when combined in hybrid frameworks. Interestingly, NER performance on Hadith datasets typically achieves slightly higher F1 scores than on Quranic datasets. The highest performance for Hadith models reached 0.9927, whereas Quranic models peaked at approximately 0.968. This is mainly because Hadith data are frequently more structured and repetitive, particularly for tasks such as narrator identification. Despite achieving high scores on specific, simpler entity types, Quranic texts exhibit slightly lower overall performance ceilings due to their greater linguistic diversity, complex syntactic structures, and the presence of difficult nested entities.

There are still many challenges to overcome in implementing NER in Islamic texts, despite recent advances. The limitations observed include the linguistic complexity of Classical Arabic and translated Quranic texts, the lack of a high-quality corpus, and the technical challenges of successfully identifying domain-specific entities. These challenges and limitations emphasize the need for continued research and development in Islamic NER models. To fully unlock the potential of NER in this crucial area, future research must prioritize the development of large-scale, well-annotated datasets, conduct deeper analyses of sophisticated linguistic features, and develop more robust models for complex entity structures. This review also suggests experimenting with advanced models from a closely related domain or building an entirely different architecture to tackle the issues mentioned. Substantial work remains to address the challenges encountered in implementing NER in Islamic texts. This systematic review offers a thorough analysis to guide future researchers in improving NLP applications in this specific field, with the broader aim of integrating with the current wave of AI innovation. Such specialized advancements can enhance the capabilities of modern NLP systems to understand culturally significant content in Islamic contexts.

Acknowledgment

This research study was financially supported by the Fundamental Research Grant Scheme under the grant FRGS/1/2024/ICT02/UKM/02/8 provided by the Ministry of Higher Education, Malaysia.

REFERENCES

[1] S. A. Tarmizi and S. Saad, "Review on NER tools for extracting the literary document in Islamic domain," in *Proc. E-Proc. Theme Impact Influence Technol. Comput. (E-TECH)*, 2022, p. 194.

[2] R. E. Salah and L. Q. Zakaria, "Building the classical Arabic named entity recognition corpus (CANERCorpus)," in *Proc. 4th Int. Conf. Inf. Retr. Knowl. Manage. (CAMP)*, Kota Kinabalu, Malaysia, Mar. 2018, pp. 1–8, doi: 10.1109/INFRKM.2018.8464820.

[3] T. L. Emha, Z. I. M. Yusoh, and B. M. Aboobaider, "BERT based named entity recognition for automated Hadith narrator identification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 1, 2022, doi:10.14569/ijacsa.2022.0130173.

[4] I. K. Alshammari, E. Atwell, and M. A. Alsalka, "Evaluation of Arabic named entity recognition models on Sahih Al-Bukhari text," *Int. J. Islamic Appl. Comput. Sci. Technol.*, vol. 11, no. 1, pp. 1–8, 2023.

[5] M. Shahiditabar, M. A. Mozaheb, M. Mohseni, A. Babaii, A. Dehchali, and S. M. Mousavi, "The place of language in the Holy Quran as a linguistic heritage," *J. Appl. Linguist. Lang. Res.*, vol. 4, no. 4, pp. 291–300, 2017.

[6] G. Inoue, B. Alhafni, N. Baimukan, H. Bouamor, and N. Habash, "The interplay of variant and size and task type in Arabic pre-trained language models," *arXiv*, 2021. [Online]. Available: https://arxiv.org/abs/2103.11478.

[7] A. M. Azmi, A. O. Al-Qabbany, and A. Hussain, "Computational and natural language processing based studies of hadith literature: A survey," *Artif. Intell. Rev.*, vol. 52, pp. 1369–1414, 2019, doi:10.1007/s10462-019-09692-w.

[8] F. Binbeshr, A. Kamsin, and M. Mohammed, "A systematic review on hadith authentication and classification methods," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 20, no. 2, pp. 1–17, 2021, doi:10.1145/3434236.

[9] B. Sulistio *et al.*, "The utilization of machine learning on studying Hadith in Islam: A systematic literature review," *Educ. Inf. Technol.*, vol. 29, no. 5, pp. 5381–5419, 2024, doi: 10.1007/s10639-023-12008-9.

[10] S. J. Putra, T. Mantoro, and M. N. Gunawan, "Text mining for Indonesian translation of the Quran: A systematic review," in *Proc. Int. Conf. Comput., Eng., Des. (ICCED)*, Kuala Lumpur, Malaysia, Nov. 2017, pp. 1–5, doi: 10.1109/CED.2017.8308122.

[11] F. S. Utomo, N. Suryana, and M. S. Azmi, "Question answering systems on holy Quran: A review of existing frameworks, approaches, algorithms and research issues," *J. Phys., Conf. Ser.*, vol. 1501, no. 1, p. 012022, 2020, doi: 10.1088/1742-6596/1501/1/012022.

[12] R. Salah and L. Q. Zakaria, "A comparative review of machine learning for Arabic named entity recognition," *Int. J. Adv. Sci., Eng., Inf. Technol.*, vol. 7, no. 2, pp. 670–678, Apr. 2017, doi:10.18517/ijaseit.7.2.1810.

[13] X. Qu *et al.*, "A survey on Arabic named entity recognition: Past, recent advances, and future trends," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 3, pp. 943–959, Mar. 2024, doi:10.1109/TKDE.2023.3303136.

[14] I. Alsheikh, M. Mohd, and L. Warlina, "A review of Arabic text recognition dataset," *Asia-Pac. J. Inf. Technol. Multimed.*, vol. 9, no. 1, pp. 69–81, 2020, doi: 10.17576/apjitm-2020-0901-06.

[15] Y. V. Farahani, B. Janfada, and B. M. Bidgoli, "A review of algorithms, datasets, and criteria in word sense disambiguation with a view to its use in Islamic texts," in *Proc. 8th Iran. Joint Congr. Fuzzy Intell. Syst. (CFIS)*, Mashhad, Iran, 2020, pp. 172–179, doi:10.1109/CFIS49607.2020.9238679.

[16] A. Munshi *et al.*, "Towards an automated Islamic Fatwa system: Survey, dataset and benchmarks," *Int. J. Comput. Sci. Mobile Comput.*, vol. 10, no. 4, pp. 118–131, Apr. 2021, doi:10.47760/ijcsmc.2021.v10i04.017.

[17] M. H. Bashir *et al.*, "Arabic natural language processing for Qur'anic research: A systematic review," *Artif. Intell. Rev.*, vol. 56, no. 7, pp. 6801–6854, Jul. 2023, doi: 10.1007/s10462-022-10313-2.

[18] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *Int. J. Surg.*, vol. 8, no. 5, pp. 336–341, 2010, doi: 10.1016/j.ijsu.2010.02.007.

[19] J. Sulaiman, "Hadith as a source of reference for creative thinking and problem solving among Muslim students in tertiary education," in *Proc. UnIPSAS Conf.*, 2021, vol. 1, pp. 78–88.

[20] N. A. S. Ilahi, N. Nurwahidin, and M. Izdiyan, "The concept of Hadith, meaning, and position of Hadith, implementation and comparison in Hadith learning application: Hadith Encyclopedia and Hadith Collection," *J. Middle East Islamic Stud.*, vol. 10, no. 1, p. 6, 2023, doi:10.7454/meis.v10i1.160.

[21] N. A. Rahman *et al.*, "Tagging narrator's names in Hadith text," *J. Fundam. Appl. Sci.*, vol. 9, no. 5S, pp. 295–306, 2018, doi:10.4314/jfas.v9i5S.21.

[22] S. S. Balgasem and L. Q. Zakaria, "A hybrid method of rule-based approach and statistical measures for recognizing narrators name in hadith," in *Proc. 6th Int. Conf. Electr. Eng. Informat. (ICEEI)*, Langkawi, Malaysia, Nov. 2017, pp. 1–5, doi:10.1109/ICEEI.2017.8312417.

[23] A. Mahmood, H. Khan, W. Rehman, and Z. Khan, "Query based information retrieval and knowledge extraction using Hadith datasets," in *Proc. 13th Int. Conf. Emerg. Technol. (ICET)*, Islamabad, Pakistan, Dec. 2017, pp. 1–6, doi: 10.1109/ICET.2017.8281714.

[24] F. Azalia, M. A. Bijaksana, and A. Huda, "Name indexing in Indonesian translation of Hadith using named entity recognition with Naïve Bayes classifier," *Procedia Comput. Sci.*, vol. 157, pp. 142–149, 2019, doi: 10.1016/j.procs.2019.08.151.

[25] W. P. Sari, M. A. Bijaksana, and A. F. Huda, "Indexing name in Hadith translation using hidden Markov model (HMM)," in *Proc. 7th Int. Conf. Inf. Commun. Technol. (ICoICT)*, Kuala Lumpur, Malaysia, Jul. 2019, pp. 1–5, doi: 10.1109/ICoICT.2019.8835296.

[26] S. A. Tarmizi and S. Saad, "Named entity recognition for Quranic text using rule based approaches," *Asia-Pac. J. Inf. Technol. Multimed.*, vol. 11, no. 1, pp. 112–122, Dec. 2022, doi: 10.17576/apjitm-2022-0101-08.

[27] N. Alsaaran and M. Alrabiah, "Classical Arabic named entity recognition using variant deep neural network architectures and BERT," *IEEE Access*, vol. 9, pp. 91537–91547, 2021, doi:10.1109/ACCESS.2021.3092261.

[28] R. Salah, M. Mukred, L. Q. Zakaria, R. Ahmed, and H. Sari, "A new rule-based approach for classical Arabic in natural language processing," *J. Math.*, vol. 2022, p. 7164254, 2022, doi:10.1155/2022/7164254.

[29] A. Mutia and M. A. Bijaksana, "People entity recognition in Indonesian Alquran translation using Roberta," *J. Inf. Syst. Res. (JOSH)*, vol. 5, no. 2, pp. 648–656, Jan. 2024, doi:10.47065/josh.v5i2.4838.

[30] R. Ningtias and M. A. Bijaksana, "People entity recognition for the English Quran translation using BERT," *J. Media Informat. Budidarma*, vol. 7, no. 1, pp. 536–544, Feb. 2023, doi:10.30865/mib.v7i1.5586.

[31] M. A. K. R. M. S., M. A. Bijaksana, and A. Huda, "Person entity recognition for the Indonesian Qur'an translation with the approach hidden Markov model-Viterbi," *Procedia Comput. Sci.*, vol. 157, pp. 214–220, 2019, doi: 10.1016/j.procs.2019.08.160.

[32] N. P. Gibson, *Labeling Religious Affiliation in Ibn Abī Usaybia's History of Physicians: A Quest*. Vienna, Austria: Verlag der Österreichischen Akademie der Wissenschaften, 2023, doi:10.1553/medievalworlds_no18_2023s246.

[33] Z. Adnan, "Methodology study Hadith by Imam Bukhari," *THARWAH: J. Islamic Civilization Thought*, vol. 3, no. 1, pp. 15–32, 2023.

[34] M. Alkaoud and M. Syed, "Learning to identify narrators in classical Arabic texts," *Procedia Comput. Sci.*, vol. 189, pp. 335–342, 2021, doi:10.1016/j.procs.2021.05.109.

[35] M. M. Hassan, D. AL-Nasrawi, R. J. Hassan, and N. T. Mahdi, "Rule based method of name entity recognition for matching Allah's finest names in Holy Quran," *J. Eng. Appl. Sci.*, vol. 13, no. 13, pp. 3618–3623, 2018.

[36] M. T. Al-Hilali and M. M. Khan, *The Noble Qur'an: English Translation of the Meanings and Commentary*. Madinah, Saudi Arabia: King Fahd Complex for the Printing of the Holy Qur'an, 1997.

[37] A. Adeleke, N. Samsudin, A. Mustapha, and N. M. Nawi, "Comparative analysis of text classification algorithms for automated labelling of Quranic verses," *Int. J. Adv. Sci., Eng., Inf. Technol.*, vol. 7, no. 4, pp. 1419–1427, Aug. 2017, doi: 10.18517/ijaseit.7.4.2198.

[38] M. H. Shakir, *The Quran with Text and Translation*. New York, NY, USA: Createspace Independent Publishing, 2016.

[39] M. M. Pickthall, *The Meaning of the Glorious Quran*. London, U.K.: Islamic Cultural Centre, 2001.

[40] R. H. Gusmita, A. F. Firmansyah, D. Moussallem, and A. Ngonga Ngomo, "IndQNER: Named entity recognition benchmark dataset from the Indonesian translation of the Quran," in *Natural Language Processing and Information Systems*. Cham, Switzerland: Springer, 2023, pp. 170–185, doi: 10.1007/978-3-031-35320-8_12.

[41] F. Yusup, M. A. Bijaksana, and A. Huda, "Narrator's name recognition with support vector machine for indexing Indonesian Hadith

translations," *Procedia Comput. Sci.*, vol. 157, pp. 191–198, 2019, doi:10.1016/j.procs.2019.08.157.

[42] D. Dzidny, M. A. Bijaksana, and K. Lhaksmana, "Supervised learning approaches for nested people entity extraction in Indonesian translated Quran," *Building Informat., Technol. Sci. (BITS)*, vol. 4, no. 1, pp. 272–278, Jun. 2022, doi: 10.47065/bits.v4i1.1758.

[43] A. Kamran, A. Abro, and B. Basharat, "SemanticHadith: An ontology-driven knowledge graph for the hadith corpus," *J. Web Semant.*, vol. 78, p. 100797, Jun. 2023, doi: 10.1016/j.websem.2023.100797.

[44] M. A. Maulana, M. A. Bijaksana, and A. F. Huda, "Entity recognition for Quran English version with supervised learning approach," *Indones. J. Comput. (Indo-JC)*, vol. 4, no. 3, pp. 77–86, 2019.

[45] A. M. Hasan, T. H. Rassem, and M. N. Noorhuzaimi, "Combined support vector machine and pattern matching for Arabic Islamic Hadith question classification system," in *Recent Trends in Data Science and Soft Computing*. Cham, Switzerland: Springer, 2019, pp. 278–290, doi: 10.1007/978-3-319-99007-1_27.

[46] M. B. Sajadi and B. Minaei, "Arabic named entity recognition using boosting method," in *Proc. Artif. Intell. Signal Process. Conf. (AISP)*, Shiraz, Iran, Oct. 2017, pp. 281–288, doi:10.1109/AISP.2017.8324098.

[47] A. Abdelmegied *et al.*, "A modified version of AlQuAnS: An Arabic language question answering system," in *Knowledge Discovery, Knowledge Engineering and Knowledge Management*. Cham, Switzerland: Springer, 2019, pp. 184–199, doi: 10.1007/978-3-030-15640-4_10.

[48] S. Raharjo, R. Wardoyo, and A. E. Putra, "Detecting proper nouns in Indonesian-language translation of the Quran using a guided method," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 32, no. 5, pp. 583–591, Jun. 2020, doi: 10.1016/j.jksuci.2018.06.009.

[49] F. D. Arvianto, M. A. Bijaksana, and A. F. Huda, "People entity recognition in Indonesian Quran translation with conditional random field approach," in *Proc. 7th Int. Conf. Inf. Commun. Technol. (ICoICT)*, Kuala Lumpur, Malaysia, Jul. 2019, pp. 1–5, doi:10.1109/ICoICT.2019.8835266.

[50] N. Alias *et al.*, "Tagging algorithm and POS tags for narrator's name in Hadith document," in *Proc. 4th Int. Conf. Artif. Intell. Data Sci. (AiDAS)*, Ipoh, Malaysia, Sep. 2023, pp. 126–130, doi:10.1109/AiDAS60501.2023.10284594.

[51] R. Salah, M. Mukred, L. Q. Zakaria, and F. A. M. Al-Yarimi, "A machine learning approach for named entity recognition in classical Arabic natural language processing," *KSII Trans. Internet Inf. Syst.*, vol. 18, no. 10, pp. 2895–2919, Oct. 2024, doi:10.3837/tiis.2024.10.005.

[52] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, Minneapolis, MN, USA, Jun. 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.

[53] M. M. A. Najeeb, "A hidden Markov model-based tagging approach for Arabic Isnads of Hadiths," *Math. Problems Eng.*, vol. 2022, p. 7160509, 2022, doi: 10.1155/2022/7160509.

[54] H. M. AbdElaal *et al.*, "Classifications of Hadiths based on supervised learning techniques," *Int. J. Comput. Sci. Netw. Secur.*, vol. 22, no. 11, pp. 1–10, Nov. 2022.

[55] N. Goyal and N. Singh, "Named entity recognition and relationship extraction for biomedical text: A comprehensive survey, recent advancements, and future research directions," *Neurocomputing*, vol. 618, p. 129171, 2025, doi: 10.1016/j.neucom.2024.129171.

[56] H. I. Yilmaz, M. C. Izgi, E. E. Erbay, and Y. Ozdemir, "Studying early Islam in the third millennium: A bibliometric analysis," *Humanit. Soc. Sci. Commun.*, vol. 11, p. 1521, 2024, doi: 10.1057/s41599-024-04058-2.

[57] *Gawami al-kalim*. Doha, Qatar: Islamweb, 2021.

[58] P. Qi *et al.*, "Stanza: A Python natural language processing toolkit for many human languages," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics, Syst. Demonstrations*, 2020, pp. 101–108, doi:10.18653/v1/2020.acl-demos.14.

[59] J. R. Finkel, T. Grenager, and C. Manning, "Incorporating non-local information into information extraction systems by Gibbs sampling," in *Proc. 43rd Annu. Meeting Assoc. Comput. Linguistics (ACL)*, Ann Arbor, MI, USA, Jun. 2005, pp. 363–370, doi:10.3115/1219840.1219885.