SemanticTafsir: Building a Cultural Heritage Ontology and Knowledge Graph from the Quranic Exegesis of al-Tabari

Amna Basharat [a,*], Amna Binte Kamran [a] and Misbahur Rehman [b]

[a] *Department of Computer Science, FAST National University of Computer and Emerging Sciences, Islamabad, Pakistan*

[b] *Goethe University Frankfurt, Germany*

E-mails: amna.basharat@nu.edu.pk, amna.kamran@nu.edu.pk, rehman@em.uni-frankfurt.de

Abstract. Tafsir, the classical exegesis of the Quran, represents a cornerstone of Islamic intellectual and literary tradition. Rooted in the teachings of the Prophet Muhammad and elaborated by early scholars, tafsir provides interpretive insights into Quranic verses through historical, linguistic, theological, and jurisprudential lenses. Among the most authoritative and influential works in this tradition is Tafsir al-Tabari, a comprehensive commentary compiled by Muhammad Ibn Jarir al-Tabari in the 9th century CE. Despite the foundational role of such works in the Islamic heritage, they remain largely underrepresented in structured, semantically annotated digital forms. This paper introduces *SemanticTafsir*, an OWL ontology and an RDF-based knowledge graph designed to semantically model Tafsir al-Tabari and support its exploration as a rich cultural and intellectual resource. The ontology captures the structural, thematic, and referential components of the text, including Quranic verses, layered commentary, embedded hadith, narrator chains, and interpretive themes. Developed using established ontology engineering methodologies, *SemanticTafsir* reuses and aligns with external vocabularies including SemanticHadith, Schema.org, and DBpedia to ensure semantic coherence and interoperability within the broader Linked Data ecosystem. Our core contribution lies in automating the semantic transformation of TEI-encoded tafsir manuscripts into a knowledge graph that preserves both the literary structure and scholarly nuance of the original work. The pipeline produces RDF representations that support advanced querying, cross-referencing, and thematic exploration, enabling users to navigate complex exegetical relationships at scale. We evaluate the ontology in terms of logical consistency, ability to resolve competency questions, and representational fidelity. The resulting knowledge graph is accessible via SPARQL endpoint and supports multilingual and semantically rich querying for scholars in Islamic studies, cultural heritage research, and digital humanities. By bridging classical Islamic exegesis with Semantic Web technologies, *SemanticTafsir* contributes to the digital preservation, accessibility, and scholarly engagement with a core component of global cultural heritage. The ontology and knowledge graph are openly available at: https://github.com/A-Kamran/SemanticTafsir

Keywords: Knowledge Graph, Ontology Design, Cultural Heritage, Semantic Web, Digital Humanities, Quran, Tafsir, Islamic Knowledge

1. Introduction

Islamic intellectual heritage encompasses a vast body of interpretive literature, among which *tafsir*, the exegesis of the Quran, holds a central place. Tafsir texts provide layered interpretations of Quranic verses, informed by

*Corresponding author. E-mail: amna.basharat@nu.edu.pk.

historical context, linguistic nuance, and jurisprudential reasoning. Among the most influential and enduring works in this tradition is Tafsir al-Tabari, composed in the 9th century by the scholar Muhammad Ibn Jarir al-Tabari. This work is not only a religious commentary, but a record of scholarly interpretation, narrative transmission, and linguistic analysis - representing a core component of Islamic cultural memory and intangible heritage. Despite its historical and interpretive richness, tafsir literature remains largely absent from structured digital representation efforts. Existing platforms and digital repositories provide access to primary texts, but often lack semantic structure, contextual annotation, and interoperability. As a result, these critical sources of cultural and religious knowledge remain difficult to explore, interlink, or query in meaningful ways.

Recent advances in Semantic Web technologies and Linked Open Data (LOD) offer promising tools for the preservation, representation, and dissemination of cultural heritage. These approaches promote best practices for using standardised web technologies (e.g., RDF, URIs, SPARQL) to interconnect and enrich data from diverse sources [1, 2]. Sectors such as education [3, 4], scientific research [5], medicine [6–9], libraries [10, 11], urban planning [12], and cultural heritage [13] have leveraged these principles to build applications that incorporate structured, semantically rich data.

Within this broader ecosystem, Knowledge Graphs (KGs) have emerged as a powerful paradigm for organising and querying interconnected knowledge. Despite their relevance, the religious domain—particularly Islamic knowledge—remains underexplored in terms of applying these technologies to support data integration, retrieval, and semantic enrichment. Historical documentation covers events, literature, music, and religious texts, leading to advanced ontologies and knowledge graphs. Key projects include the Enslaved Ontology for African slave trade data [14], virtual archives of lost Irish records [15], and a semantic portal for Finnish Civil War victims [16]. Other examples are ArCo for Italian Cultural Heritage [17], semantically-encoded biographies [18], the Archive Dynamics Ontology [19], and ontologies for historical architectures and graffiti [20–22]. Additionally, knowledge graphs have been created for Dante Alighieri's works [23–25], Buddhism [26], and Greek mythology [27].

Within the Islamic domain, a growing body of work has explored ontology-based representations of the Quran and hadith [28–30], but the complex interpretive structure of tafsir has yet to receive comparable attention. Hulliyah et al. and Khazani et al. emphasised the importance of contextualising Quranic text to enhance comprehension, proposing methods for summarising and knowledge representation [31, 32]. Daud et al. further developed this by integrating the Quran, hadith, and tafsir into an ontology for semantic search on Zakat, demonstrating a growing interest in digital methodologies for Islamic texts [28]. Ahmed et al. provided a comprehensive review of methodologies and tools used in Quranic ontology development [33]. Farooqui et al. suggested a highly granular ontology design for Al-Quran tafsir, emphasising the need for detailed and precise knowledge representation [29]. Ta'a et al. underscored the importance of Islamic knowledge management, proposing methods to effectively manage and utilise Islamic knowledge resources. Bashir et al. surveyed the field of Quranic natural language processing, highlighting computational techniques to aid in the study and understanding of the Quran [34].

Despite these advancements, there remains a critical gap in the systematic digital representation of tafsir literature. While foundational studies have focused on morphological annotation of the Quran [35], ontology modelling [36–41], and Arabic natural language processing [34, 38], these efforts have not fully addressed the complexity of tafsir literature. Repositories such as QuranComplex and quran.com provide valuable resources, yet their integration and interoperability remain limited, hindering comprehensive exploration and analysis of Islamic knowledge sources [42]. The digital representation of tafsir literature requires advanced semantic models capable of capturing its nuanced interpretations and historical context effectively. Several ontologies have been developed to represent key concepts, entities, and relationships in the Quran [40, 41]. These ontologies enable semantic annotation and retrieval of Quranic knowledge. Kamran et al. presents the design, development, and publishing of the hadith corpus as a knowledge graph, aiming to enhance interlinking and knowledge discovery in the Islamic domain [30]. The SemanticHadith ontology describes and relates core structural concepts from the hadith, with prominent collections published as an RDF-based knowledge graph.

Integrating tafsir with other Islamic knowledge sources through Semantic Web technologies offers a transformative path for enhancing scholarly access and cultural preservation. As interpretive texts that span centuries, tafsir works encode not only theological insights but also linguistic heritage, intellectual discourse, and narrative traditions central to Islamic civilization. Their digital representation requires semantic frameworks capable of capturing this interpretive richness while facilitating discovery, reuse, and contextual analysis.

This paper presents the design and implementation of *SemanticTafsir*, a cultural heritage ontology and RDF-based knowledge graph based on Tafsir al-Tabari. Developed using TEI-encoded manuscripts, the ontology models verse-level commentary, narrator references, cited hadith, thematic annotations, and the hierarchical discourse structure of tafsir literature. *SemanticTafsir* supports semantic interoperability through reuse of established vocabularies, including Schema.org, DBpedia, and the previously developed SemanticHadith ontology. Our approach addresses the dual challenge of intellectual preservation and semantic accessibility. By automating the transformation of TEI-encoded tafsir into RDF, we enable both humanistic and computational exploration of classical Islamic thought. The knowledge graph supports SPARQL querying, thematic navigation, and linked data interconnection, thereby facilitating new modes of access and engagement for scholars, educators, and cultural heritage institutions. *SemanticTafsir* represents a step toward the long-term digital preservation of interpretive Islamic texts. It contributes not only to the growing body of Islamic digital humanities but also to broader discussions around the semantic modelling of interpretive traditions and the encoding of intangible religious and literary heritage.

2. Background Context and Motivation

The Quran, revealed to the Prophet Muhammad between 610 and 632 CE, lies at the heart of Islamic religious and cultural life. Its verses were initially transmitted orally and later compiled into a written corpus during the Prophet's lifetime and standardised in a uniform codex under the third Caliph, Uthman ibn Affan. Despite this standardisation, linguistic variations persisted due to its oral transmission, which started to be standardised in the 10th century. The interpretation of the Quranic text posed challenges to subsequent generations of scholars, leading to the development of tafsir, or Quranic exegesis. The transmission and interpretation of the Quran has, since its inception, been accompanied by a body of scholarly commentary - tafsir - that contextualizes its verses through theological, linguistic, and historical lenses. Tafsir aimed to deepen understanding and disseminate insights through teaching circles, drawing upon the sayings of the Prophet, interpretations by his companions, and subsequent scholars' analyses.

Over time, tafsir evolved into a distinct scholarly tradition, shaped by teaching circles, oral transmission, and manuscript production. These works encode the cumulative insights of early Islamic scholars, preserving intellectual debates, interpretive principles, and exegetical methods. By the 8th century CE, comprehensive tafsir compilations began to appear as independent literary works, marking a key moment in the documentation of Islamic scholarly heritage.

2.1. Tafsir as Intangible Cultural Heritage

Tafsir literature plays a foundational role in preserving and transmitting Islamic interpretive knowledge. More than theological commentary, tafsir reflects the evolution of intellectual traditions, linguistic analysis, and jurisprudential reasoning within Islamic societies - making it indispensable for understanding the Quran beyond its literal text. Through tafsir, generations of scholars have contextualized divine revelation, offering guidance on matters of faith, ethics, and law while responding to the evolving needs of Muslim communities.

Tafsir represents a cumulative intellectual tradition, reflecting diverse methodologies and perspectives across time and geography. These works preserve the interpretive voices of early Islamic scholars and continue to inform contemporary discourse, making them a cornerstone of Islamic scholarly heritage. As dynamic engagements between scripture and scholarship, tafsir texts embody an ongoing dialogue that is both historically situated and spiritually resonant.

In this sense, tafsir constitutes a form of intangible cultural heritage - transmitted through teaching, writing, and commentary, and integral to religious identity and intellectual continuity. In the digital age, preserving and modelling this literature is essential to ensuring its continued relevance and accessibility not only for safeguarding its cultural and intellectual value but also for enabling new forms of scholarly engagement. Semantic modelling provides a means to structure and interconnect this complex body of knowledge while respecting its historical and interpretive depth.

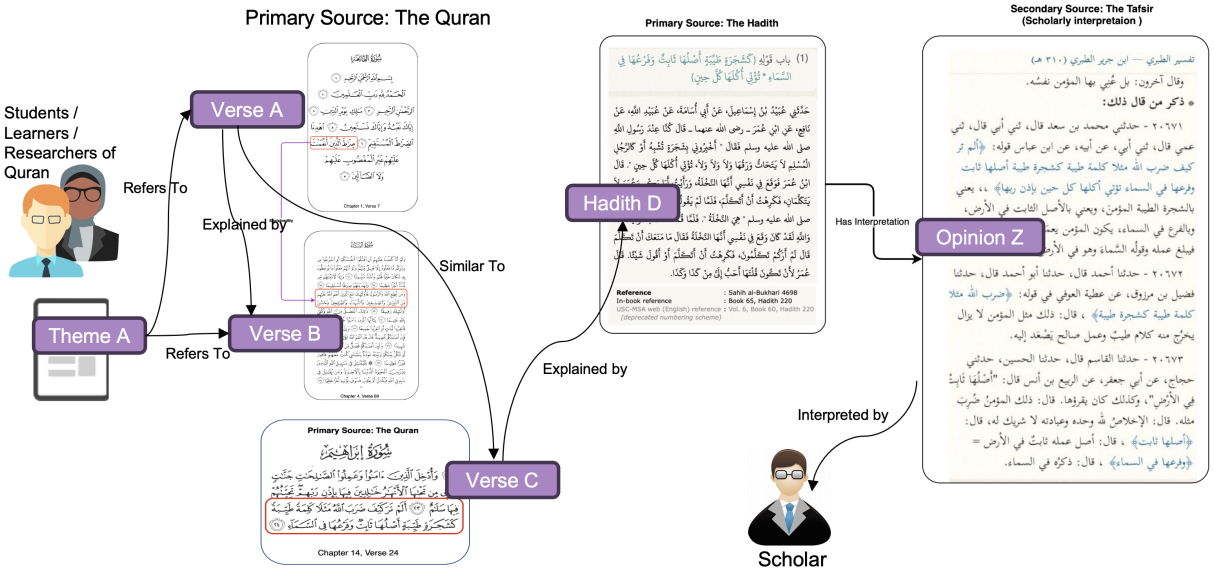

Fig. 1. Motivational Scenario - A typical knowledge seeker's path

2.2. *Muhammad Ibn Jarir al-Tabari and His Tafsir*

Muhammad Ibn Jarir al-Tabari (839–923 CE), born in Tabaristan (modern-day Iran) and active in Baghdad, is one of the most influential figures in Islamic scholarship. A polymath historian and jurist, al-Tabari's magnum opus - Tafsir al-Tabari - remains a cornerstone of Sunni exegetical tradition. His work synthesizes early oral interpretations, narrations from companions, and linguistic analysis, providing a rich source for understanding both the Quran and early Islamic thought.

Beyond its doctrinal value, Tafsir al-Tabari is a document of immense cultural and intellectual heritage. It preserves the transmission of scholarly discourse, narrative traditions, and interpretive authority from the formative period of Islam. Modelling this work semantically offers a means to preserve and make accessible a seminal text in the history of Islamic interpretation.

2.3. *Motivation: Semantic Modelling for Preservation and Access*

The vast corpus of tafsir literature presents unique opportunities and challenges for digital heritage preservation. These texts are rich in intertextual references, interpretive strategies, and scholarly voices—making them critical to understanding Islamic intellectual history. Yet their complexity poses challenges to traditional digitization efforts: varied terminology, nested commentary structures, and cross-references to hadith and jurisprudence demand sophisticated, semantically rich models.

Semantic modelling offers a powerful tool for capturing the interpretive logic and structural features of tafsir. It enables the encoding of themes, references, narrators, and scholarly perspectives in ways that support both humanistic inquiry and computational analysis. Figure 1 illustrates a typical scholarly use case, where researchers engage with Quranic verses and seek interpretive depth from sources like Tafsir al-Tabari.

By creating an ontology and knowledge graph tailored to the structure and discourse of tafsir, we enable structured access, semantic search, and contextual exploration—enhancing both scholarly engagement and long-term preservation. These efforts support the integration of tafsir into broader digital Islamic heritage frameworks, bridging manuscript culture with semantic technologies.

2.4. Challenges in Modelling Interpretive Islamic Texts

Although growing efforts have targeted semantic modelling of the Quran and Hadith, tafsir literature introduces a higher degree of complexity. Existing ontologies address thematic tagging or verse-level annotation, but few offer comprehensive, interoperable models tailored to tafsir's layered discourse structure. Projects such as Farooqui et al. [29] and Ahmed et al. [43] offer valuable precedents but remain limited in scope or coverage.

Tafsir texts integrate multiple interpretive layers—legal, historical, theological—alongside references to narrators, traditions, and linguistic analysis. Modelling such texts semantically requires not only advanced NLP techniques, as noted by Bashir et al. [34], but also domain-informed ontological design. Challenges include extracting and formalizing entities, managing textual variants, modeling interpretive relationships, and ensuring alignment across diverse scholarly traditions.

Despite advances, there remains a scarcity of publicly accessible linked data resources specifically tailored to tafsir. Addressing this gap is essential to ensure the digital preservation, semantic integration, and scholarly utility of these foundational works of Islamic knowledge.

3. Design and Development of *SemanticTafsir* Ontology

In recognition of the absence of a standardised ontology for Quranic exegesis, we introduce *SemanticTafsir*, an OWL ontology specifically designed to semantically represent the structural and interpretive features of classical tafsir literature. The design process is grounded in established ontology engineering principles and tailored to support the cultural heritage goals of preservation, semantic access, and contextual exploration of Islamic texts. The following subsections outline the structure of Tafsir al-Tabari and detail the conceptual and technical steps undertaken in the ontology's development.

3.1. Tafsir Structure

Tafsir al-Tabari exhibits a hierarchical structure common to classical exegesis. It consists of collections corresponding to Quranic surahs (chapters), each subdivided into interpretive units aligned with full verses or specific verse fragments. These are annotated with the author's commentary, which integrates thematic exposition, linguistic interpretation, and references to hadith. Hadith citations include chains of narrators (sanad), culminating in a source (the matn). These narrations function within the tafsir as interpretive tools rather than isolated traditions. The text also references named entities such as persons, locations, time periods, and institutions, which support contextual analysis. This layered structure—linking Quranic text, commentary, narrative authority, and literary devices—presents both an opportunity and a challenge for semantic modelling, particularly for the purposes of digital preservation and cultural heritage scholarship.

3.2. Conceptual Knowledge Modelling

The design of the *SemanticTafsir* ontology follows the seven-step methodology proposed by Noy and McGuinness in Ontology Development 101 [44]. These steps include defining the ontology's scope, identifying key terms, reusing existing ontologies, developing class hierarchies and properties, defining property constraints, and creating instances.

The ontology was developed using Protégé (v5.5.0), a widely adopted ontology editor that supports OWL and integrates reasoning tools such as Fact++ and HermiT. UTF-8 compatibility and plug-in extensibility made Protégé suitable for modelling Arabic-language data from TEI-encoded sources. The ontology uses the prefix `tafsir:` and is published under the namespace `http://www.semantictafsir.com/ontology/`, with a persistent dereference-able URI available at https://purl.org/semantictafsir. The RDF/Turtle serialisation is hosted on GitHub Pages for long-term accessibility and public reuse.[1]

[1] https://purl.org/semantictafsir

To ensure semantic interoperability and reduce modelling redundancy, *SemanticTafsir* incorporates terms and alignments from existing vocabularies such as Schema.org [45], Dublin Core [46], DBpedia [47], and Wikidata [48]. We adopted top-level vocabularies with minimal ontological commitment (e.g., `schema`, `dc-terms`) to maximize reusability and alignment with other cultural heritage datasets, following the classification of Partridge et al. [49]. In addition, hadith-related entities and properties are reused directly from the SemanticHadith ontology [30], ensuring consistency and interoperability across Islamic knowledge graphs.

This reuse strategy supports integration with broader semantic infrastructures while preserving the interpretive nuance and literary structure that characterise tafsir texts.

3.3. Scope of the Ontology – Competency Questions

The *SemanticTafsir* ontology s designed to document, preserve, and semantically represent interpretive structures found in Quranic exegesis. The ontology captures Quranic verses, thematic commentary, hadith citations, and hierarchical sectioning, enabling contextualized search, advanced knowledge discovery, question answering systems and semantic exploration. Beyond preservation, *SemanticTafsir* supports applications in information retrieval systems and question-answering (QNA) systems, offering a robust foundation for further scholarly and computational analyses. Its intended audience includes scholars of Islamic Studies, Arabic literature, and Digital Humanities, as well as technical users building semantic access systems.

To define and validate the ontology's functional scope, we constructed a set of competency questions (CQs), shown in Table 1, based on the archetype patterns proposed by Ren et al. [50]. These CQs guide ontology design and assess its ability to support knowledge discovery, textual alignment, and interpretive analysis. Many of these questions reflect tafsir-specific use cases, including: thematic annotations of Quranic verses, verse fragments, poetic references, the hierarchical structure of tafsir chapters and sections, and the referential patterns across narrators and traditions — use cases that are unique to exegetical texts.

Some of the CQ structures, particularly those related to narrators and transmission chains, are adapted from prior work on *SemanticHadith* [30]. However, in *SemanticTafsir*, these competency questions are recontextualized: hadith are not stand-alone entities but appear as embedded, cited components within Quranic commentary. This shift alters both their ontological representation and their functional role in scholarly exploration. Where *SemanticHadith* focuses on canonical hadith collections (e.g., the *Sihah Sitta*), *SemanticTafsir* integrates hadith as interpretive tools within the structure of commentary. This distinction alters both the scope and function of the reused patterns, aligning them with the exegetical and thematic concerns of the tafsir corpus.

To reflect this emphasis, Table 1 begins with tafsir-specific questions, followed by hadith-related CQs adapted to the exegetical domain. This organisation underscores the centrality of interpretive structure in the ontology's design, while highlighting its ability to interoperate with other semantic representations of Islamic knowledge.

3.4. Reused Ontologies

As part of our goal to construct a semantically rich and interoperable ontology for tafsir literature, we integrate concepts from existing ontologies wherever appropriate. This strategy ensures alignment with established semantic standards and facilitates broader reusability across cultural heritage and religious knowledge graphs. We began by compiling a comprehensive list of key terms derived from a detailed analysis of *Tafsir al-Tabari*, supplemented by a high-level review of additional Tafsir sources such as those by Ibn Kathir and Al-Jaza'iri (see Section 4.1). These terms were modelled as classes and properties, with their semantics formalized using OWL axioms. Informed by prior research in Islamic knowledge modelling [29, 40, 51–56], we reused and extended relevant terms from publicly accessible ontologies. Notably, we incorporated the SemanticHadith ontology [30] to represent hadith elements embedded in tafsir texts. These reused classes and properties are referenced with the prefix `sh:`, maintaining consistency across related Islamic knowledge representations.

In addition to domain-specific resources, we reused widely adopted vocabularies such as Schema.org, DBpedia, Wikidata, QuranOntology, and DCMI Metadata Terms (Dublin Core) [41, 45, 47, 48, 57]. Where applicable, we established semantic alignments using `owl:equivalentClass` and `owl:equivalentProperty`, enabling interoperability and enhancing discoverability within the Linked Open Data cloud.

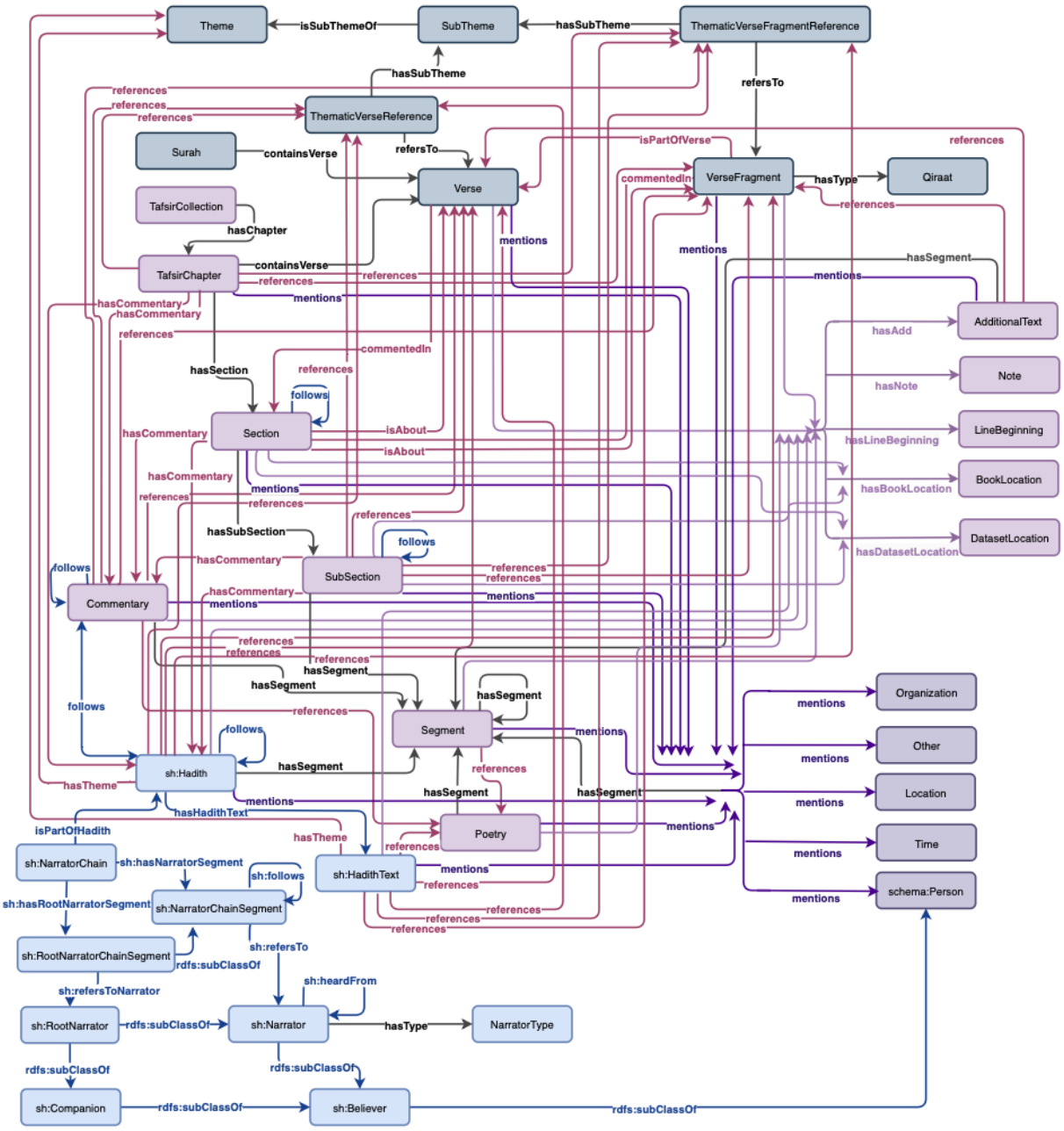

Fig. 2. Conceptual model of the *SemanticTafsir* ontology.

To ensure semantic clarity and extensibility, we structured these reused terms within our ontology through subclassing and subproperty modelling. For instance, `schema:Person` serves as a superclass for entities such as `hadith:HadithNarrator`, while `schema:hasPart` and `schema:partOf` serve as super-properties for internal relationships like `hasCommentary`, `hasSection`, and `isPartOfVerse`. These alignments allow us to preserve the domain-specific integrity of tafsir concepts while embedding them within a larger web of semantically linked resources. We also use DCMI Metadata Terms (Dublin Core) [57], a standard ontology for representing metadata, to describe the metadata of the *SemanticTafsir* ontology.

This reuse-oriented approach not only ensures robust coverage of interpretive Islamic concepts but also strength-

Competency Questions	Patterns
Find all Poetry in the tafsir that mentions a Person X.	Which [CE1] [OPE] [CE2]?
List all VerseFragments discussed in Chapter X.	What is the [DP] for a particular [CE]?
List all Themes associated with Verse X in the tafsir.	What are the types of [CE]?
Which Chapter hasSection mentioning Person Y?	Which [CE1] [OPE] [CE2]?
How many themes are mentioned in Chapter X?	How many [CE1] [OPE] [CE2]?
Which themes are discussed in multiple chapters?	Which are [CE]?
What Section do I need to examine to find Verse Y?	What [CE1] do I need to [OPE] [CE2]?
What are the types of HadithNarrators?	What are the types of [CE]?
Which verses are most frequently referenced in the tafsir?	What [CE] has the [NM] [DP]?
Which Commentary mentions Location B?	Which [CE1] [OPE] [CE2]?
Do all sections mention multiple persons?	Do [CE1] have [QM] values of [DP]?
Where can I find sections about a specific Verse X?	Where do I [OPE] [CE]?
List all verse numbers where an entity of type "Other" is mentioned.	What [CE] has the [NM] [DP]?
Which Commentary mentions both a Person and an Organisation?	Which [CE1] [OPE] [CE2]?
Search a Hadith where NarratorChain has Narrator A and Narrator B but not Narrator C and HadithText includes Theme A and Location B	Does [CE1] [OPE] [CE2]?
All the Hadith narrated from Narrator A	Find [CE1] with [CE2].
How many Hadith narrated by Narrator A	How many [CE1][OPE] [CE2]?
How many Hadith narrated by Narrator A from Narrator B	How many [CE1][OPE] [CE2]?
List of narrators by the number of their narrations	What [CE] has the [NM] [DP]?
Which Narrator narrated most Hadith about Theme A	Which [CE1] [OPE] [QM] [CE2]?
Most narrated Theme by Narrator A	What is the [NM] [CE1] to [OPE][CE2]?
Number of Hadith by Theme narrated by Narrator A	How many [CE1][OPE] [CE2]?
What is the frequency of a specific chain or part of a chain	How many [CE1][OPE] [CE2]?
Any NarratorChain that is repeated more than 10 times	How many [CE1][OPE] [CE2]?
Frequency of partial NarratorChain repeating at least ten times	How many [CE1][OPE] [CE2]?
Search Hadith 'mauquf' from Narrator A	What type of [CE] is [I]?
Search Hadith that references ayah 11:11 (or surah 11 i.e. any ayah of surah 11)	Which [CE1] [OPE] [CE2]?

Table 1

Competency Questions Mapped to CQ Archetypes/Patterns as identified by [50] (CE = class expression, OPE = object property expression, DP = data type property, I = individual, PE = property expression, NM = numeric modifier, QM = quantity modifier).

ens the ontology's role as a bridge between classical scholarly texts and contemporary digital heritage infrastructures. It supports both current analytical needs and future expansions within the broader domain of semantic knowledge representation.

3.5. Ontology Design

The ontology design process was guided by the structural and interpretive features of classical Tafsir literature. Our goal was to preserve the literary hierarchy and capture meaningful semantic relationships across verses, commentary, and referenced traditions. The conceptual model for the *SemanticTafsir* ontology reflects this emphasis, as shown in Figure 2.

Key entities and their relationships are modelled as follows:

- **Chapters and Sections:** *Tafsir al-Tabari* is organized by surahs (chapters), each containing multiple interpretive sections. These are modelled using distinct classes, with Sections containing commentary on individual verses.
- **Verses and Verse Fragments:** Quranic verses are represented as Verse entities, while partial references within the commentary are modelled as VerseFragment. Sections are semantically connected to these through the isAbout property.

- **Commentaries and Hadith:** Interpretive text units authored by al-Tabari are modelled as `Commentary` instances, often incorporating referenced `Hadith`. These are linked via `containsCommentary` property preserving their contextual placement in the exegetical structure.
- **Named Entities and Annotations:** Referenced `Person`, `Location`, `TimePeriod`, and `Organisation` entities are modelled separately and connected using the `mentions` property. This supports entity-based querying and cultural contextualisation.

This modular and hierarchical modelling strategy allows for the representation of layered interpretive content, facilitates semantic disambiguation, and supports the preservation of the original structure of the tafsir manuscript. The design is intentionally extensible, allowing for future incorporation of additional sources or refinements.

3.6. Classes, Hierarchies, Properties, and Facets

Drawing on the domain analysis outlined above, we developed a class hierarchy grounded in the structural semantics of the tafsir corpus. Classes were created for all terms corresponding to entities with distinct identity or function. Using Protégé, we finalized 32 classes, along with 37 object properties and 18 data properties. High-level domain classes include `TafsirCollection`, `TafsirChapter`, `Verse`, `Hadith`, and `Commentary`. These are further refined by subclasses such as `Section`, `Subsection`, `VerseFragment`, `Theme`, `NarratorType`, and `Poetry`. This top-down approach ensures semantic clarity and reflects the internal structure of tafsir. Object properties define relationships between entities (e.g., `hasCommentary`, `isPartOfVerse`), while data properties encode literal values such as labels, dates, or identifiers. Property facets include datatype constraints, cardinality, and domain/range specifications to maintain data consistency across instances. These are summarised in Tables 1–3 in the Supplementary Information.

3.7. Modelling Decisions

Our ontology leverages established OWL design patterns to address common modelling challenges in literary and exegetical corpora [58]. Key patterns applied include part-whole relations, value sets, and n-ary relations.

Part-whole relations: To model hierarchical textual structure - e.g., chapters, sections, subsections, and commentary—we implement part-whole relationships using sub-properties of `hasPart` and `isPartOf`, as recommended in W3C best practices [59]. Sub-properties such as `isPartOfVerse`, and `isPartOfHadith` reflect document composition without semantic ambiguity.

Value sets (enumerated individuals): We adopt a `value set` modelling pattern [60] to represent types of narrators. We define a class `NarratorType`, which enumerates the individuals `sahabi`, `rawi`, `shaykh`, `unknown_shaykh`, and `unknown_rawi`. These individuals represent the origin categories of narrators referenced in the Tafsir al-Tabari, capturing whether a narrator was a companion of the Prophet, a later transmitter, a teacher, or of unknown status.

This approach enables controlled vocabulary enforcement via object properties (e.g., `hasNarratorType`) and supports validation and filtering within SPARQL queries. While these values are modelled as individuals rather than subclasses, this design remains fully compatible with taxonomic classification: value assertions can coexist with subclass-based reasoning if needed. An alternative modelling approach—representing these types as subclasses of `Narrator` (e.g., `Sahabi`, `Rawi`, etc.) and typing each narrator as an instance of the appropriate subclass—is also valid and commonly used in OWL ontologies. Our choice reflects both a pragmatic understanding of the source data, where narrator roles are often treated as attributes rather than rigid types, and a desire for easier integration with systems that favour value-based classification. The implementation of this modelling choice, including enumerated individuals and disjoint axioms, is illustrated in Figure **??**, as rendered in Protégé. Future versions of the ontology may explore both patterns to evaluate their effects on reasoning, query performance, and interoperability.

N-ary relations: Complex thematic references are modelled using OWL n-ary design patterns [61]. In cases where a text segment (X) is annotated with multiple themes in varying contexts of a verse or verse fragment, binary property assertions are insufficient. A particular challenge arises when multiple references to the same verse occur with different themes across various contexts. This complexity makes it difficult to trace which thematic relationship

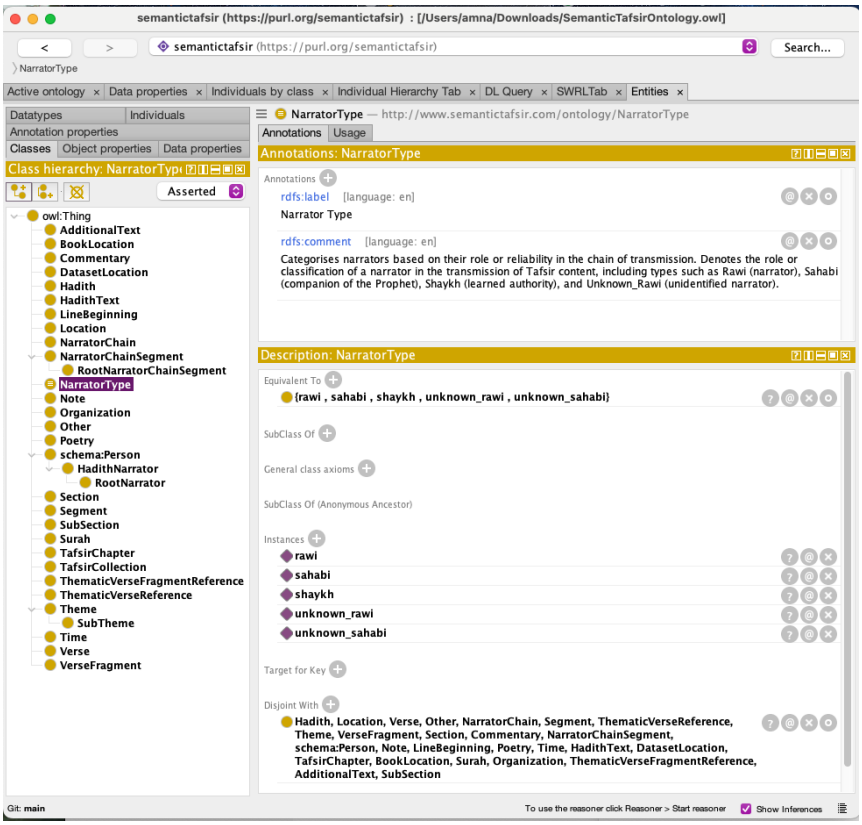

Fig. 3. Protégé interface displaying the 'NarratorType' class and its enumerated individuals in the *SemanticTafsir* ontology. The value-set pattern captures narrator roles as enumerated values, supporting value-based classification and controlled vocabulary for reasoning and query filtering.

applies in each specific instance. For example, X references verse Y, and verse Y hasTheme themeA. This binary relationship approach fails when there are multiple thematic references to the same verse in different contexts. To address this, we implement the OWL N-ary pattern [61]. This pattern allows us to model complex relationships involving more than two entities by introducing an auxiliary node, *ThematicVerseReference*, which captures the relationship between the theme and the verse, ensuring precise tracking of thematic references. The revised model includes:

- X references *ThematicVerseReference*
- *ThematicVerseReference* refersTo verse Y
- *ThematicVerseReference* hasTheme themeA

Similarly, this N-ary pattern is applied to verse fragments. For cases where a text segment (X) references a fragment of a verse with a specific theme, we introduce another auxiliary node, *ThematicVerseFragmentReference*, to manage these relationships. The model includes:

- X references *ThematicVerseFragmentReference*
- *ThematicVerseFragmentReference* refersTo verseFragment
- *ThematicVerseFragmentReference* hasTheme themeA

These auxiliary nodes (*ThematicVerseReference* and *ThematicVerseFragmentReference*) serve as intermediaries to record the context of each thematic reference, maintaining clarity and traceability of thematic relationships across different segments and contexts. By employing the OWL N-ary pattern, we can effectively manage and query complex thematic annotations within our ontology, ensuring accuracy and comprehensiveness in our data representation.

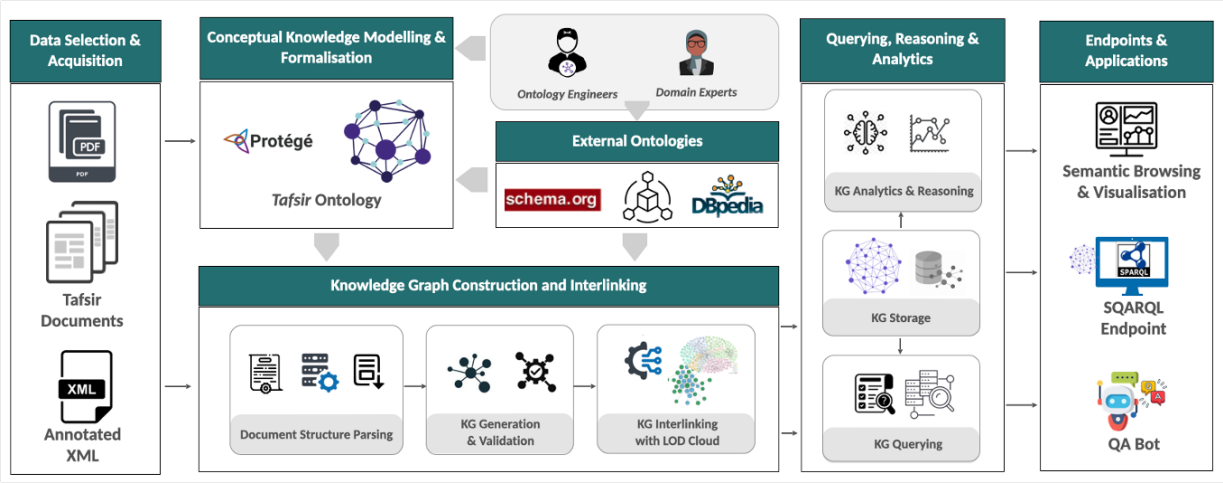

Fig. 4. Overview of the *SemanticTafsir* knowledge graph construction framework. The key stages of the framework include Data Selection and Acquisition, Conceptual Knowledge Modelling and Formalisation, *SemanticTafsir* Knowledge Graph Construction and Interlinking, Querying, Reasoning and Analytics and Endpoints and Applications.

The final step in our design pipeline involves populating the ontology with individuals. This is done via our custom KG-Generator, which extracts data from TEI-encoded Tafsir manuscripts, instantiates relevant classes, assigns property values, and links entities using defined object properties. This automation ensures scalable, reproducible generation of a semantically rich knowledge graph.

4. Methodology for Tafsir al-Tabari Knowledge Graph Construction

The construction of the *SemanticTafsir* knowledge graph follows a reusable pipeline for transforming TEI-encoded exegetical texts into semantically rich, interlinked data. This process not only supports the scholarly exploration of interpretive Islamic texts but also contributes to the broader goal of integrating classical Islamic knowledge into the Web of Data [42]. Figure 4 outlines this process, which includes data acquisition, RDF conversion, enrichment, and deployment. The following subsections detail each stage of the implementation.

4.1. Data Selection and Acquisition

To construct the *SemanticTafsir* knowledge graph, we began with a comprehensive review of several major tafsir collections. This analysis ensured that the ontology design could accommodate the varied structures and interpretive styles found across classical Islamic exegesis. In particular, we examined works by Ibn Kathir, Al-Jaza'iri, and most prominently, al-Tabari.

For this study, we focus on the complete edition of *Jāmi al-Bayān an Tawīl Āy al-Qurān* by Abu Jafar Muhammad ibn Jarīr al-áabarī (d. 923 CE), one of the earliest and most widely referenced Sunni exegeses. Al-áabarī's work is notable for its comprehensive citation of narrations, early linguistic analysis, and influence on subsequent Quranic commentary traditions. The decision to use Tafsir al-Tabari was based on both its scholarly significance and the availability of a complete, TEI-annotated XML version, which was essential for structured data extraction. This digital edition—annotated and curated as part of a named entity recognition (NER) and topic modelling project [43]—served as the primary source for knowledge graph construction.

The version used is the Turki Edition of Tafsir al-Tabari, spanning 26 volumes and over 18,500 pages. This TEI-encoded corpus includes rich annotations aligned verse-by-verse across 51,704 sentences. Notable annotation counts include:

– 176,105 person references

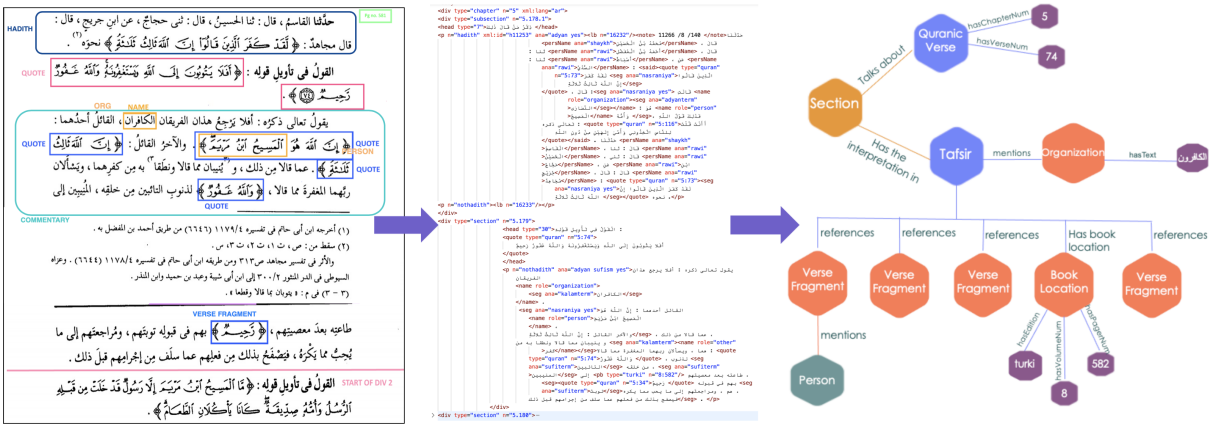

Fig. 5. Structure of the XML/TEI Dataset

- 5,583 locations
- 22,026 organisations
- 4,160 temporal references
- 12,453 miscellaneous entities

These annotations form the basis of our RDF transformation pipeline. They enable entity extraction, semantic alignment, and ontology population, and serve as the foundation for linking interpretive discourse with broader semantic infrastructures. Figure 5 illustrates a representative TEI snippet and the annotation structure.

4.2. *Conceptual Knowledge Modelling and Formalisation*

This phase focused on structuring the ontology to represent the interpretive content of classical tafsir while ensuring semantic clarity, extensibility, and alignment with existing knowledge graphs. As described in Section 3, the ontology design followed the Ontology 101 methodology [44], supported by competency questions to guide modelling scope and reasoning requirements.

Key modelling tasks included:

- **Formalising Competency Questions:** Derived from use cases in Quranic exegesis, these CQs define the ontology's expressive needs and guided class/property selection.
- **Entity and Relation Modelling:** Core concepts such as verse, commentary, theme, hadith, narrator, and verse fragment were represented using OWL classes. Their relationships—e.g., `hasTheme`, `containsCommentary`, `mentions` - were modelled using object properties.
- **Ontology Vocabulary Reuse:** We integrated terms from Schema.org, Dublin Core, DBpedia, Wikidata, and the SemanticHadith ontology, using `owl:equivalentClass` and `owl:equivalentProperty` to preserve interoperability.
- **Use of Ontology Design Patterns:** Patterns for part-whole hierarchies, value sets, and n-ary relations (see Section 3.7) were applied to ensure clarity, consistency, and extensibility.
- **Implementation and Testing:** The ontology was implemented in OWL 2 using Protégé and validated via reasoning and CQ resolution. It was subsequently used to populate the *SemanticTafsir* knowledge graph.

4.3. *Knowledge Graph Construction and Interlinking*

The *SemanticTafsir* knowledge graph was constructed through a semi-automated pipeline designed to process TEI-encoded Tafsir texts and generate RDF triples aligned with the ontology. This process ensures semantic fidelity to the original manuscript structure while enabling computational access and interlinking with external data sources.

4.3.1. Data Parsing and Entity Extraction

We began by analysing the structure and annotation tags of the TEI-encoded dataset (Tafsir al-Tabari, Turki edition). A custom parser was developed using a recursive depth-first traversal to read the XML tree, preserving both the hierarchical and sequential order of annotated entities.

Each tag (e.g., `person`, `place`, `time`, `verse`) was mapped to a corresponding data class. These classes encapsulate attributes and relationships, preparing the content for ontology-based instantiation. Maintaining annotation order was critical for aligning interpretive structure with the TEI source.

4.3.2. Knowledge Graph Generation

Parsed entities were transformed into OWL individuals using an object-oriented pipeline implemented in Python. Each ontology class (e.g., `Verse`, `Commentary`, `Narrator`) corresponds to a Python object responsible for instantiating individuals, assigning literal values (data properties), and establishing links (object properties). RDF triples were dynamically generated using the Owlready2 library [62], which supports in-memory OWL ontology manipulation. PyArabic [63] was used to handle morphological processing and tokenisation of Arabic text. The resulting knowledge graph is fully OWL-compliant and contains semantically aligned instances that preserve the literary hierarchy of the Tafsir. The pipeline is modular, supporting future ingestion of additional TEI-based sources with minimal configuration changes.

During the RDF transformation process, we encountered a number of structural anomalies in the TEI-encoded source, including mis-nested tags, redundant divisions, and unfamiliar annotations (e.g., `<add type='parenthesis'>`, misused `<persName>` attributes). These issues, while technically routine, required targeted preprocessing to ensure semantic alignment with the ontology schema. We applied rule-based strategies to preserve interpretive content (e.g., parenthetical notes) while maintaining RDF validity. These adjustments were crucial to maintaining data fidelity, but do not affect the conceptual structure of the knowledge graph. Implementation details are documented in the project repository[2].

4.3.3. Knowledge Graph Interlinking with LOD Cloud

To integrate *SemanticTafsir* into the broader Linked Open Data ecosystem, we established links with external knowledge graphs. Our primary focus was on aligning narrator entities across DBpedia [47], Wikidata [48], and our previously published SemanticHadith knowledge graph. The linking process involved using OpenRefine [64] for automated reconciliation based on name similarity and class alignment, matching Arabic names of narrators with instances of `Companion`, `Muhaddith`, or similar types in Wikidata and DBpedia, and establishing `owl:sameAs` and `rdfs:seeAlso` links between matching instances to enhance entity resolution.

Due to the linguistic complexity of Arabic names and inconsistency in transliteration, expert curation and validation were essential. Annotators verified matches manually, resolved ambiguities, and corrected false positives. This hybrid process ensured high precision and domain relevance for the resulting links.

As a result, we successfully aligned a large subset of narrator entities with external resources. These linkages significantly enhance the semantic context and enable cross-graph query federation (see Table 2 in Section 5.1).

4.4. Endpoints and Applications

To promote accessibility, reproducibility, and future research, the ontology and resulting knowledge graph are publicly hosted across multiple platforms:

- **Ontology Access:** RDF serialization is available at https://purl.org/semantictafsir, and browsable at https://a-kamran.github.io/SemanticTafsir/ontology.ttl.
- **SPARQL Endpoint:** The deployed graph is available via a public SPARQL endpoint at http://www.semantictafsir.iknex.com/sparql/, hosted on GraphDB.
- **Code and Reference Implementation:** The full source code, scripts, and configuration files are available on GitHub[3], supporting community contributions and issue tracking.

[2] https://github.com/A-Kamran/SemanticTafsir/blob/main/TEI_ConversionNotes.md
[3] https://github.com/A-Kamran/SemanticTafsir

These interfaces enable both programmatic querying and human exploration of Tafsir data. Future applications include multilingual retrieval interfaces, knowledge-enhanced reading environments, and integration with Islamic manuscript repositories and digital humanities platforms.

5. Results and Discussion

The development of the *SemanticTafsir* ontology and corresponding knowledge graph represents a significant step toward the semantic preservation of Quranic exegetical knowledge. The ontology captures interpretive structures, linguistic commentaries, and cited narrations—providing a semantic infrastructure for exploring classical Islamic knowledge as linked data.

By reusing and aligning with the SemanticHadith ontology [30], we ensured interoperability and semantic coherence between referenced Hadith and exegetical context. This alignment enables federated analysis across Islamic textual traditions, enhancing interpretive research within the broader cultural heritage landscape.

5.1. Ontology and Knowledge Graph Evaluation

The evaluation of the *SemanticTafsir* ontology and knowledge graph followed a multi-dimensional strategy:

- **Logical Consistency:** Using reasoners such as HermiT [65], Pellet [66], and FaCT++ [67], we validated OWL semantics and inferencing capability.
- **Ontology Quality:** The OOPS! tool [68] detected minor pitfalls—e.g., missing labels, inverse properties, and inconsistent naming—that were corrected post-evaluation.
- **MIRO Reporting:** We applied the MIRO ontology evaluation framework [69], documenting our design decisions, reuse justifications, and intended use-cases in a machine-readable format[4].
- **Competency Question Resolution:** To evaluate the knowledge graph constructed from the *SemanticTafsir* ontology, we focused on its ability to answer competency questions derived from the ontology's scope and requirements. The process involved running SPARQL queries against the knowledge graph and verifying the results for accuracy and completeness. The competency questions covered various aspects of the knowledge represented in the graph, such as the relationships between Quranic verses, hadith, and scholarly commentary. In particular, the competency questions identified in Section 3.3 are successfully handled. The results of these queries, along with additional resources and detailed query results, are hosted on GitHub for transparency and reproducibility[5]. In Figure 6 We also present an example SPARQL query and its corresponding results addressing a representative competency question related to the interpretation of Quranic verses and their contextual connections.

Table 2 summarises the knowledge graph statistics, including class distributions, individual counts, and linked entity totals. Figure 7 presents an illustrative graph excerpt visualising the layered structure of a tafsir commentary, its linked verse, and referenced hadith.

Overall, the *SemanticTafsir* framework demonstrates robustness in both its ontological design and data instantiation. It satisfies key design goals—logical soundness, semantic richness, reuse of existing vocabularies, and query capability—while supporting long-term digital preservation and interpretive continuity of Islamic heritage texts.

5.2. Analytical Capabilities of the SemanticTafsir Knowledge Graph

The *SemanticTafsir* Knowledge Graph functions not only as a structured repository of Quranic exegesis but also as an analytical lens into the interpretive traditions of Islamic scholarship. By semantically encoding themes, narrator roles, and exegetical layers, it enables users to uncover patterns that span linguistic, theological, and legal discourses.

[4] https://github.com/A-Kamran/SemanticTafsir/blob/main/MIRO.md
[5] https://github.com/A-Kamran/SemanticTafsir/blob/main/CompetencyQuestionsAndSPARQLQueries.md

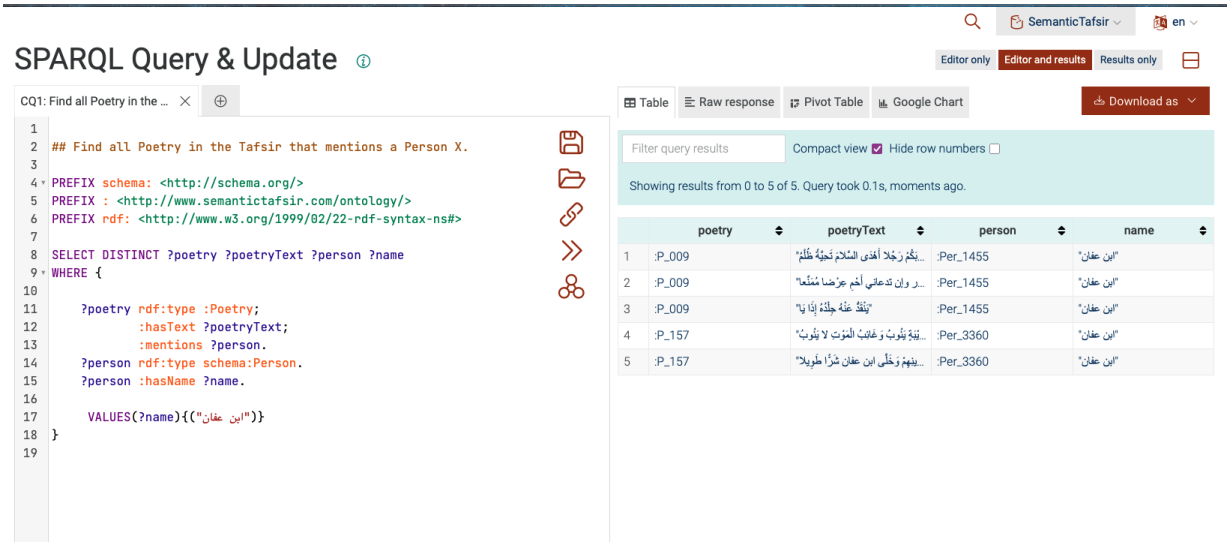

Fig. 6. SPARQL query execution and results showing all Poetry in the Tafsir that mentions a Person X in GraphDB

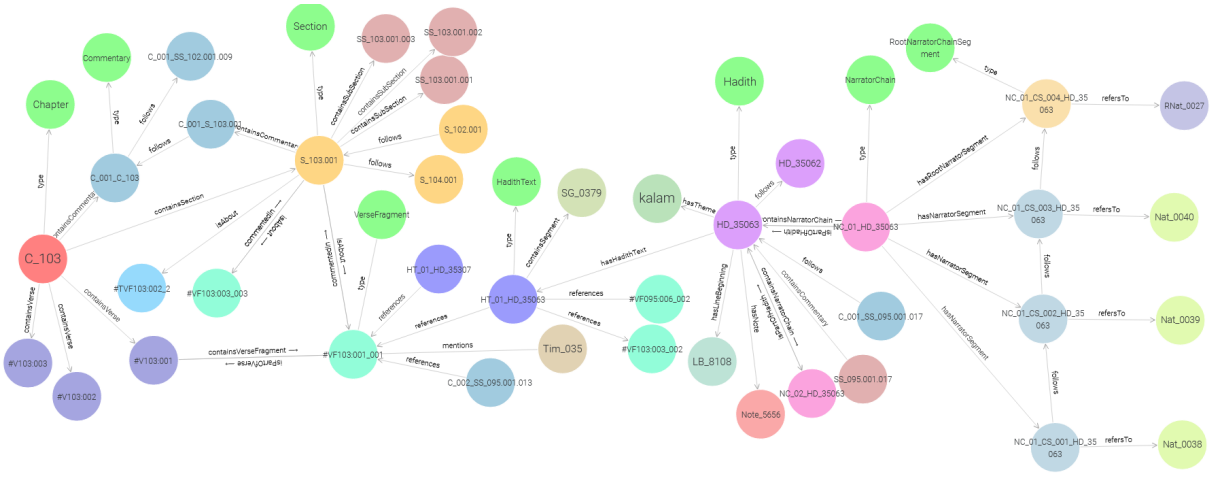

Fig. 7. Visualisation from the *SemanticTafsir* Knowledge Graph. (Visualisation created using GraphDB by OntoText.)

The knowledge graph facilitates thematic analysis by allowing the querying and visualisation of major interpretive themes—such as *fiqh* (jurisprudence), *lugha* (language), and *kalam* (theology)—across Quranic verses and Tafsir sections. Figure 8 illustrates the frequency and distribution of major themes—such as *fiqh* (jurisprudence), *lugha* (language), and *kalam* (theology)—across chapters, allowing researchers to identify thematic concentrations and their exegetical significance. Figure 9 highlights relationships between narrators and the themes they frequently address, revealing implicit roles of transmitters in shaping interpretive focus. These visualisations are backed by SPARQL queries that trace thematic and structural connections across verses, commentaries, and cited hadith.

By supporting such multidimensional queries and exploration, the graph supports new modes of scholarly inquiry into Tafsir literature—uncovering interpretive trajectories, exegetical dependencies, and thematic overlap—thus advancing both Islamic studies and semantic cultural heritage analysis.

Table 2
Statistics of the *SemanticTafsir* ontology and the *SemanticTafsir* knowledge graph.

	Variables	Number
Structure & Ontology	Ontology Classes	32
	Object Properties	37
	Data Properties	18
Knowledge Graph	Total Individuals	591,751
	Hadith	36,515
	Hadith Narrators	7870
	Persons	2800
	Locations	477
	Organizations	1153
	Temporal Events	218
	Themes	50
External Links to *SemanticHadith*	`owl:sameAs`, Narrator	2890
	`owl:sameAs`, Verse	338
External Links to Wikidata &/or DBPedia	`owl:sameAs`, Places	34
	`owl:sameAs`, Narrators	370
	`owl:sameAs`, Person	634
	`owl:sameAs`, Prophet	23
External Links to Quran Ontology	`owl:sameAs`, Verse	6236
	`owl:sameAs`, Surah	114
	`owl:sameAs`, Prophet	23

5.3. Intended Usage

The *SemanticTafsir* ontology and knowledge graph are designed to serve a broad community across digital humanities, Islamic studies, and semantic web research. For researchers and scholars, the graph offers a means to conduct fine-grained investigations into Tafsir content, tracing verse-level commentaries, thematic patterns, and citation structures. By supporting semantic querying and interlinking of concepts, it facilitates comparative studies and deepens scholarly understanding of exegetical traditions. In educational settings, the structured nature of the ontology enables students and instructors to interact with Quranic exegesis in an engaging and analytical manner. Visualisations and structured queries make it possible to explore thematic relevance, historical context, and intertextual links, thus enriching pedagogical experiences in Islamic studies and related disciplines. The ontology formalises the interpretive structure of classical texts, preserving their intellectual and cultural significance in machine-readable form. By aligning with linked data standards and vocabularies, it supports sustainable reuse and integration in wider digital heritage infrastructures. Beyond academic applications, the project supports digital preservation by encoding interpretive knowledge in a semantically rich, machine-readable format. This contributes to long-term accessibility and sustainability of Islamic intellectual heritage, ensuring that Tafsir literature remains available for future generations of scholars, educators, and cultural institutions. The ontology is designed with interoperability in mind, using widely adopted vocabularies such as Schema.org and DBpedia to facilitate integration with other linked data resources. This semantic compatibility promotes knowledge enrichment and cross-domain exploration, supporting broader goals of data reuse and heritage interconnection. In addition, the open-source nature of the knowledge graph pipeline invites developers to adapt the framework for other religious, literary, or historical corpora—enabling the creation of search interfaces, recommendation systems, or semantic tools tailored to specific interpretive needs.

5.4. Future Directions

Looking ahead, several directions for expansion and enhancement of the *SemanticTafsir* project are envisioned. One immediate goal involves integrating additional Tafsir collections to broaden the interpretive coverage. By incorporating classical and modern texts beyond Tafsir al-Tabari, the graph will enable comparative analysis across

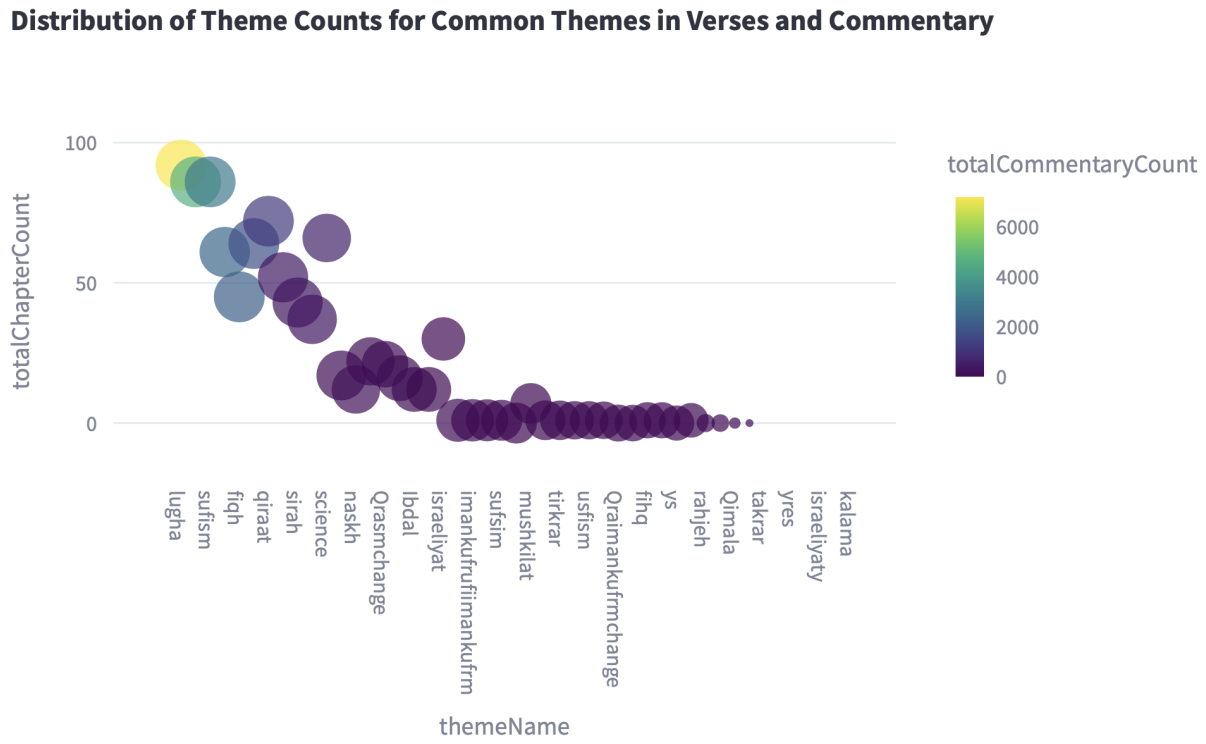

Fig. 8. **Distribution of Theme Counts for Common Themes in Verses and Commentary**

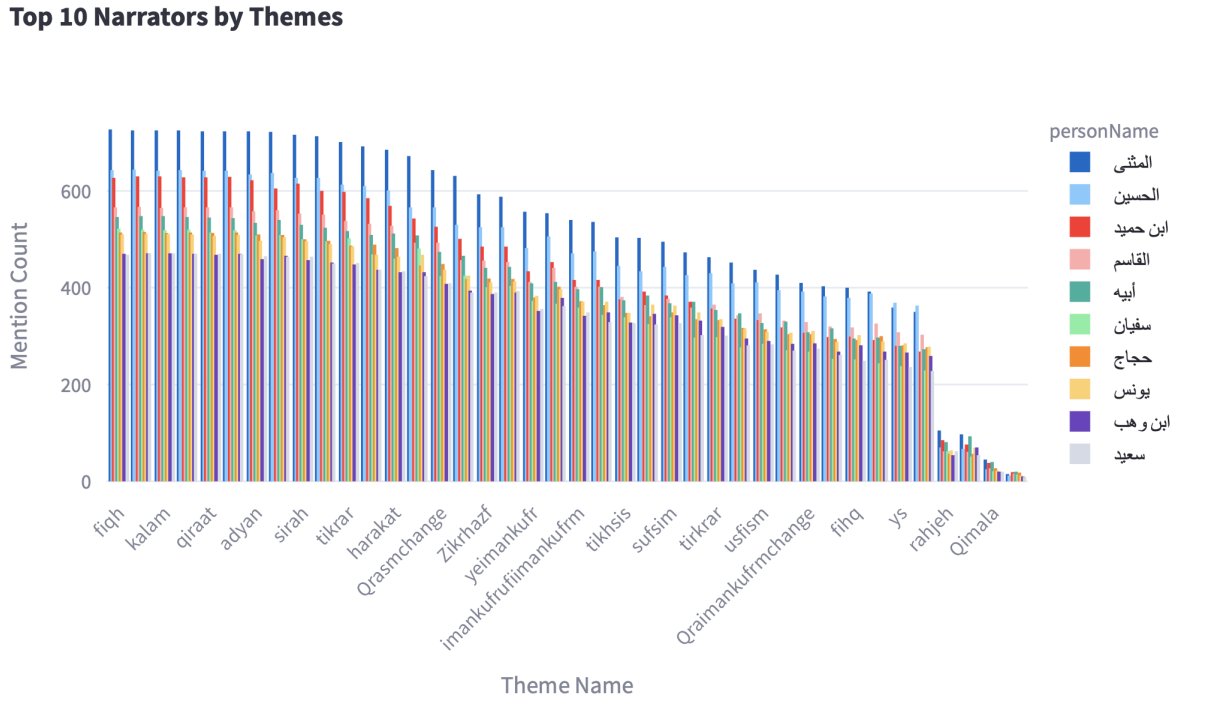

Fig. 9. Relationship between (top ten) narrators and themes, revealing each narrator's contribution to thematic discussions within Tafsir Al-Tabari.

centuries, authors, and schools of thought, thereby deepening scholarly engagement with Quranic interpretation. Another key direction is the semantic linking of Tafsir with other Islamic knowledge domains, including jurisprudence (fiqh), hadith, and theological treatises. This will enable cross-referencing and exploration of how exegetical insights draw from, and contribute to, broader religious discourses. Such interlinking will strengthen the role of *SemanticTafsir* as a bridge between textual traditions in the Islamic intellectual ecosystem.

To improve accessibility for non-technical users, development of a natural language interface and a graphical SPARQL query builder is also planned. These interfaces would allow scholars, educators, and the general public to interact with the knowledge graph without requiring expertise in RDF or query languages, thus promoting inclusive engagement with Islamic heritage.

The graph's annotated content also presents opportunities for machine learning applications. In future work, NLP and annotation models could be trained on the graph's data to automate the processing of additional Tafsir texts, accelerating ontology population and discovery of latent thematic or narrative patterns. In the broader context of digital humanities, the structured modelling of Tafsir as a layered discourse resource opens new pathways for manuscript studies and textual historiography. Future efforts will explore how this model can be extended to represent manuscript variants, marginalia, and the evolution of exegetical commentaries across time and geography. Ultimately, the project remains committed to openness, extensibility, and scholarly collaboration. By supporting community contributions and promoting reuse, the *SemanticTafsir* ontology aspires to become a foundational infrastructure for semantic Islamic knowledge representation, with continued relevance in both academic and heritage-driven domains.

Together, these future directions support the project's broader aim: to serve as a dynamic infrastructure for the semantic preservation, interpretation, and engagement with the intellectual heritage of Islamic exegesis.

6. Conclusion

This paper has presented the design, development, and evaluation of the *SemanticTafsir* ontology and knowledge graph, developed to semantically model and preserve the rich intellectual tradition of Quranic exegesis, with a focus on the Tafsir of al-Tabari. By formalising the interpretive structure of classical texts using linked data principles, we offer a semantic framework that supports the preservation of traditions while enabling digital exploration and scholarly reuse.

The ontology captures the layered structure of Tafsir literature, including Quranic verses, commentaries, hadith, and thematic annotations. Leveraging existing vocabularies such as Schema.org, Dublin Core, and the previously developed SemanticHadith ontology, we ensured semantic interoperability and alignment with the broader Linked Open Data ecosystem. Through a combination of ontology design patterns, modular implementation, and automated RDF generation from TEI-encoded sources, we constructed a reusable pipeline for digital knowledge representation in the Islamic humanities. Evaluation of the ontology confirmed logical consistency, modelling coherence, and its capacity to answer competency questions relevant to Islamic studies. The resulting knowledge graph supports SPARQL-based querying, thematic exploration, and narrative analysis, providing new opportunities for scholarly engagement, educational use, and the digital preservation of Islamic interpretive heritage.

As a contribution to the semantic web and cultural heritage communities, the *SemanticTafsir* project demonstrates how ontological methods can encode intangible religious knowledge in structured, machine-readable form—supporting its transmission, reuse, and scholarly re-interpretation in the digital age. Future directions include the integration of additional Tafsir collections, multilingual access, and broader linkage with Islamic jurisprudence and historical corpora. By fostering interoperability, extensibility, and community participation, we hope to advance the long-term preservation and meaningful accessibility of Islamic intellectual traditions.

Additional Information

Supplementary Information accompanies this paper.

Data Availability

Ontology, Knowledge Graph, ontology documentation, SPARQL Queries corresponding to Competency Questions, MIRO report are available at https://github.com/A-Kamran/SemanticTafsir.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. Bizer, T. Heath and T. Berners-Lee, Linked data-the story so far, *Semantic Services, Interoperability and Web Applications: Emerging Concepts* (2009), 205–227.
[2] T. Heath and C. Bizer, Linked data: Evolving the web into a global data space, *Synthesis lectures on the semantic web: theory and technology* **1**(1) (2011), 1–136.
[3] S. Softic, B. Taraghi, M. Ebner, L. De Vocht, E. Mannens and R. Van de Walle, Mining and visualizing usage of educational systems using linked data, in: *Immersive Education: 4th European Summit, EiED 2014, Vienna, Austria, November 24-26, 2014, Revised Selected Papers 4*, Springer, 2015, pp. 17–26.
[4] N. Li, Curriculum data association organization and knowledge management method for unstructured learning resources, *International Journal of Emerging Technologies in Learning (Online)* **15**(6) (2020), 79.
[5] T.K. Attwood, D.B. Kell, P. McDermott, J. Marsh, S. Pettifer and D. Thorne, Utopia documents: linking scholarly literature with research data, *Bioinformatics* **26**(18) (2010), i568–i574.
[6] X. Chen, S. Jia and Y. Xiang, A review: Knowledge reasoning over knowledge graph, *Expert Systems with Applications* **141** (2020), 112948.
[7] Y. Shang, Y. Tian, M. Zhou, T. Zhou, K. Lyu, Z. Wang, R. Xin, T. Liang, S. Zhu and J. Li, EHR-Oriented Knowledge Graph System: Toward Efficient Utilization of Non-Used Information Buried in Routine Clinical Practice, *IEEE Journal of Biomedical and Health Informatics* **25**(7) (2021), 2463–2475.
[8] S. Ali, I. Zada, Z. Mehmood, A. Ullah, H. Ali, M. Ullah et al., Publishing and interlinking covid-19 data using linked open data principles: toward effective healthcare planning and decision-making, *Mathematical Problems in Engineering* **2022** (2022).
[9] E. Rajabi, R. Midha and J.F. de Souza, Constructing a Knowledge Graph from OpenStatistical Data: The Case of Nova ScotiaDisease Datasets (2022).
[10] E. Miller and M. Westfall, Linked data and libraries, *The Serials Librarian* **60**(1–4) (2011), 17–22.
[11] L.C. Howarth, FRBR and linked data: connecting FRBR and linked data, *Cataloging & Classification Quarterly* **50**(5–7) (2012), 763–776.
[12] A. Yazdizadeh and B. Farooq, Smart mobility ontology: Current trends and future directions, *Handbook of smart cities* (2020), 1–36.
[13] J. Marden, C. Li-Madeo, N. Whysel and J. Edelstein, Linked open data for cultural heritage: evolution of an information technology, in: *Proceedings of the 31st ACM international conference on Design of communication*, 2013, pp. 107–112.
[14] C. Shimizu, P. Hitzler, Q. Hirt, D. Rehberger, S.G. Estrecha, C. Foley, A.M. Sheill, W. Hawthorne, J. Mixter, E. Watrall et al., The enslaved ontology: Peoples of the historic slave trade, *Journal of Web Semantics* **63** (2020), 100567.
[15] C. Debruyne, G. Munnelly, L. Kilgallon, D. O'Sullivan and P. Crooks, Creating a knowledge graph for ireland's lost history: knowledge engineering and curation in the beyond 2022 project, *ACM Journal on Computing and Cultural Heritage (JOCCH)* **15**(2) (2022), 1–25.
[16] H. Rantala, I. Jokipii, E. Ikkala and E. Hyvönen, WarVictimSampo 1914–1922: A national war memorial on the Semantic Web for digital humanities research and applications, *ACM Journal on Computing and Cultural Heritage (JOCCH)* **15**(1) (2022), 1–18.
[17] V.A. Carriero, A. Gangemi, M.L. Mancinelli, L. Marinucci, A.G. Nuzzolese, V. Presutti and C. Veninata, ArCo: The Italian cultural heritage knowledge graph, in: *The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II 18*, Springer, 2019, pp. 36–52.
[18] M.A. Stranisci, V. Basile, R. Damiano, V. Patti et al., Mapping Biographical events to ODPs through Lexico-Semantic Patterns?, in: *CEUR WORKSHOP PROCEEDINGS*, Vol. 3011, CEUR-WS, 2021, pp. 1–12.

[19] O. Vsesviatska, T. Tietz, F. Hoppe, M. Sprau, N. Meyer, D. Dessì and H. Sack, ArDO: an ontology to describe the dynamics of multimedia archival records, in: *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, 2021, pp. 1855–1863.
[20] D. Mikhaylova and D. Metilli, Extending RiC-O to model historical architectural archives: The ITDT ontology, *ACM Journal on Computing and Cultural Heritage* **16**(4) (2023), 1–15.
[21] A. Goy, D. Colla, D. Magro, C. Accornero, F. Loreto and D.P. Radicioni, Building semantic metadata for historical archives through an ontology-driven user interface, *Journal on Computing and Cultural Heritage (JOCCH)* **13**(3) (2020), 1–36.
[22] M. Trentin and A. Felicetti, Between Written and Visual Communication: CIDOC CRM Ontology for Medieval and Early Modern European Graffiti, *ACM Journal on Computing and Cultural Heritage* (2023).
[23] V. Bartalesi and C. Meghini, Using an ontology for representing the knowledge on literary texts: The Dante Alighieri case study, *Semantic Web* **8**(3) (2017), 385–394.
[24] A. De Domenico, D. Cantone and M.N. Asmundo, An Ontology for the Iconographic Sources of La Divina Commedia., in: *CILC*, 2020, pp. 157–168.
[25] R. Damiano, V. Lombardo and A. Pizzo, The ontology of drama, *Applied Ontology* **14**(1) (2019), 79–118.
[26] T. Wu, C. Gao, G. Qi, L. Zhang, C. Dong, H. Liu and D. Zhang, KG-Buddhism: the Chinese knowledge graph on Buddhism, in: *Semantic Technology: 7th Joint International Conference, JIST 2017, Gold Coast, QLD, Australia, November 10-12, 2017, Proceedings 7*, Springer, 2017, pp. 259–267.
[27] J. Pastor-Sánchez, E. Kontopoulos, T. Saorín, T. Bebis and S. Darányi, Greek mythology as a knowledge graph: from chaos to zeus and beyond, *Seman. Web J. Rev* (2021).
[28] A. Daud, M.H. Ullah, A.R. Banjar and A.A. Alshdadi, Ontological Modeling and Semantic Search in Quran, *IJCSNS* **22**(5) (2022), 771.
[29] N. Farooqui, M.F. Noordin and R. Othman, Ontology matching: A case of English translation of Al-Quran Tafsir, in: *2018 International Conference on Information and Communication Technology for the Muslim World (ICT4M)*, IEEE, 2018, pp. 1–6.
[30] A.B. Kamran, B. Abro and A. Basharat, SemanticHadith: An ontology-driven knowledge graph for the hadith corpus, *Journal of Web Semantics* **78** (2023), 100797.
[31] K. Hulliyah and S. Rosmaria, The Overview of Summary Interpretation Verses In The Holy Qur'an Based On Knowledge Graph Method, in: *Proceedings of the 2nd International Conference on Quran and Hadith Studies Information Technology and Media in Conjunction with the 1st International Conference on Islam, Science and Technology, ICONQUHAS & ICONIST, Bandung, October 2-4, 2018, Indonesia*, 2020.
[32] M.M.M. Khazani, H. Mohamed, T.M.T. Sembok, N.M.M. Yusop, S. Wani, Y. Gulzar, M.H.M. Halip, S. Marzukhi and Z. Yunos, Semantic graph knowledge representation for Al-Quran verses based on word dependencies, *Malaysian Journal of Computer Science* (2021), 132–153.
[33] R.I. Ahmed, M.H. Sayed and T.M. Wahbi, Quran Ontology: Review on Recent Research Issues (2018). https://api.semanticscholar.org/CorpusID:219569813.
[34] M.H. Bashir, A.M. Azmi, H. Nawaz, W. Zaghouani, M. Diab, A. Al-Fuqaha and J. Qadir, Arabic natural language processing for Qur'anic research: a systematic review, *Artificial Intelligence Review* **56**(7) (2023), 6801–6854.
[35] K. Dukes and N. Habash, Morphological Annotation of Quranic Arabic, in: *LREC*, 2010.
[36] H.S. Al-Khalifa, M. Al-Yahya, A. Bahanshal, I. Al-Odah and N. Al-Helwah, An approach to compare two ontological models for representing quranic words, in: *Proceedings of the 12th International Conference on Information Integration and Web-based Applications & Services*, 2010, pp. 674–678.
[37] S. Baqai, A. Basharat, H. Khalid, A. Hassan and S. Zafar, Leveraging semantic web technologies for standardized knowledge modeling and retrieval from the Holy Qur'an and religious texts, in: *Proceedings of the 7th International Conference on Frontiers of Information Technology*, 2009, pp. 1–6.
[38] A. Farghaly and K. Shaalan, Arabic natural language processing: Challenges and solutions, *ACM Transactions on Asian Language Information Processing (TALIP)* **8**(4) (2009), 1–22.
[39] A.R. Yauri, R.A. Kadir, A. Azman and M.A.A. Murad, Ontology semantic approach to extraction of knowledge from Holy Quran, in: *2013 5th International Conference on Computer Science and Information Technology*, IEEE, 2013, pp. 1–5.
[40] M.A. Sherif and A.-C. Ngonga Ngomo, Semantic Quran, *Semantic Web* **6**(4) (2015), 339–345.
[41] A. Hakkoum and S. Raghay, Ontological approach for semantic modeling and querying the Qur'an, in: *Proceedings of the International Conference on Islamic Applications in Computer Science And Technology*, 2015.
[42] A. Basharat, B. Abro, I.B. Arpinar and K. Rasheed, Semantic Hadith: Leveraging Linked Data Opportunities for Islamic Knowledge., in: *LDOW@ WWW*, 2016.
[43] S. Ahmed, R. Van Der Goot, M. Rehman, C. Kruse, Ö. Özsoy, A. Mehler and G. Roig, Tafsir dataset: A novel multi-task benchmark for named entity recognition and topic modeling in classical arabic literature, in: *Proceedings of the 29th International Conference on Computational Linguistics*, 2022, pp. 3753–3768.
[44] N.F. Noy, D.L. McGuinness et al., Ontology development 101: A guide to creating your first ontology, Stanford knowledge systems laboratory technical report KSL-01-05 and . . . , 2001.
[45] R.V. Guha, D. Brickley and S. Macbeth, Schema. org: evolution of structured data on the web, *Communications of the ACM* **59**(2) (2016), 44–51.
[46] D.C.M. Initiative et al., Dublin core metadata element set, version 1.1, Dublin Core Metadata Initiative, 2012, [Online; accessed 20. Aug. 2022].
[47] C. Becker and C. Bizer, DBpedia mobile-a location-aware semantic web client, *Proceedings of the Semantic Web Challenge* (2008), 13–16.

[48] D. Vrandečić and M. Krötzsch, Wikidata: A Free Collaborative Knowledgebase, in: *Proceedings of the 2014 ACM conference on Web science*, ACM, 2014, pp. 106–107.
[49] C. Partridge, A. Mitchell, A. Cook, J. Sullivan and M. West, A Survey of Top-Level Ontologies-to inform the ontological choices for a Foundation Data Model (2020).
[50] Y. Ren, A. Parvizi, C. Mellish, J.Z. Pan, K. Van Deemter and R. Stevens, Towards competency question-driven ontology authoring, in: *European Semantic Web Conference*, Springer, 2014, pp. 752–767.
[51] H. Abdelnasser, M. Ragab, R. Mohamed, A. Mohamed, B. Farouk, N.M. El-Makky and M. Torki, Al-Bayan: an Arabic question answering system for the Holy Quran, in: *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)*, 2014, pp. 57–64.
[52] Q.A. Abed, Ontology-based approach for retrieving knowledge in Al-Quran, PhD thesis, Universiti Utara Malaysia, 2015.
[53] K. Dukes, Statistical parsing by machine learning from a classical Arabic treebank, *arXiv preprint arXiv:1510.07193* (2015).
[54] M. Alqahtani and E. Atwell, Aligning and merging ontology in al-Quran domain (2016).
[55] M.M. Alqahtani and E. Atwell, Developing bilingual arabic-english ontologies of al-quran, in: *2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR)*, IEEE, 2018, pp. 96–101.
[56] R. Ahmed and E. Atwell, Developing an ontology of concepts in the Qur'an, *International Journal on Islamic Applications in Computer Science and Technology* **4**(4) (2016), 1–8.
[57] D.U. Board, DCMI Metadata Terms.
[58] A. Gangemi and V. Presutti, Ontology design patterns, in: *Handbook on ontologies*, Springer, 2009, pp. 221–243.
[59] A.L. Rector, Simple Part-Whole Relations in OWL Ontologies, 2005, Accessed: 2025-05-12.
[60] A.L. Rector, Representing Specified Values in OWL: "Value Partitions" and "Value Sets", 2017, Accessed: 2025-05-12.
[61] N. Noy, A. Rector, P. Hayes, C. Welty, H. Thompson, D. Solomon and D. McGuinness, Defining N-ary Relations on the Semantic Web, 2006, W3C Working Group Note, accessed: 2025-05-12.
[62] J.-B. Lamy, Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies, *Artificial intelligence in medicine* **80** (2017), 11–28.
[63] T. Zerrouki, pyarabic, An Arabic language library for Python, 2010. https://pypi.python.org/pypi/pyarabic.
[64] K. Ham, OpenRefine (version 2.5). http://openrefine. org. Free, open-source tool for cleaning and transforming data, *Journal of the Medical Library Association: JMLA* **101**(3) (2013), 233.
[65] R.D. Shearer, B. Motik and I. Horrocks, Hermit: A highly-efficient OWL reasoner., in: *Owled*, Vol. 432, 2008, p. 91.
[66] B. Parsia and E. Sirin, Pellet: An owl dl reasoner, in: *Third international semantic web conference-poster*, Vol. 18, Citeseer, 2004, p. 13.
[67] D. Tsarkov and I. Horrocks, FaCT++ description logic reasoner: System description, in: *International joint conference on automated reasoning*, Springer, 2006, pp. 292–297.
[68] M. Poveda-Villalón, M.C. Suárez-Figueroa and A. Gómez-Pérez, Validating ontologies with oops!, in: *International conference on knowledge engineering and knowledge management*, Springer, 2012, pp. 267–281.
[69] N. Matentzoglu, J. Malone, C. Mungall and R. Stevens, MIRO: guidelines for minimum information for the reporting of an ontology, *Journal of biomedical semantics* **9**(1) (2018), 1–13.