

Structuring Women's Familial Roles in *Sahih al-Bukhari*: A Semantic Field Theory Model for Thematic Hadith Extraction

Norzulaili Mohd Ghazali¹, Norfarhana Ahmad Ghafar²,
Hishomudin Ahmad³, Mawaddah Abd Manan⁴

^{1,4}Fakulti Pengajian Quran dan Sunnah, Universiti Sains Islam Malaysia, Malaysia, ^{2,3}Fakulti Pengajian Bahasa Utama, Universiti Sains Islam Malaysia, Malaysia

*Email: norfarhana@usim.edu.my

DOI Link: <http://dx.doi.org/10.6007/IJARBSS/v15-i9/26416>

Published Date: 14 September 2025

Abstract

Women's familial roles—as mothers, wives, and daughters—are central to Islamic ethical and social thought, yet narrations addressing these roles in *Ṣaḥīḥ al-Bukhārī* are dispersed across books and chapters, limiting their accessibility and pedagogical value. Recent scholarship affirms the importance of these roles for questions of authority, leadership, and family well-being, but highlights the absence of systematic frameworks to organize them thematically. This study introduces the Semantic Field Theory–Hadith Extraction Model (SFT-HEM), a conceptual framework for semantically clustering narrations on women's familial roles. Grounded in Semantic Field Theory, the model employs lexical relations such as synonymy, hyponymy, and meronymy to move beyond keyword searches and rigid RDF taxonomies that miss semantic nuance. From an initial corpus of 1,463 narrations, manual curation and validation through classical commentaries refined the set to 60 role-specific narrations: 45% on mothers, 40% on wives, and 15% on daughters. By integrating classical semantics with computational tools such as Arabic WordNet and morphosyntactic parsers, SFT-HEM offers a linguistically grounded, theologically respectful approach that benefits educators, students, and digital humanists, while also laying the foundation for future automation in Arabic NLP and ontology-based retrieval.

Keywords: Semantic Field Theory, Hadith Data Extraction, Arabic Lexical Semantics, *Ṣaḥīḥ al-Bukhārī*, Digital Islamic Text Analysis

Introduction

Women's familial roles—as mothers, wives, and daughters—hold a central place in Islamic ethical and social thought, with prophetic narrations presenting these roles as foundations of responsibility, moral education, and communal continuity. Contemporary scholarship highlights their enduring importance by demonstrating that women's narratives in sacred

texts remain integral to religious authority and leadership (Erwani & Siregar, 2025), while transmitters such as ‘Ā’ishah (RA) contributed decisively to hadith validation (Mehfooz, 2021). Yet in *Ṣaḥīḥ al-Bukhārī* these narrations are dispersed across books and chapters, limiting accessibility, weakening pedagogical application, and obscuring continuity in women’s religious agency (Stowasser, 2011).

This fragmentation generates methodological challenges that are further intensified by pedagogical and technological constraints. Traditional approaches emphasize naqlī (transmitted) memorization rather than thematic or semantic integration (Junaedi et al., 2025), while curricula in Islamic education often struggle to address the plural realities of contemporary Muslim societies (Selsuk, 2021). Digital repositories have expanded access but typically rely on rigid RDF schemas or keyword-based searches that cannot capture synonymic variation or contextual nuance (Bendjamaa & Taleb, 2024), resulting in retrieval systems that risk flattening the rhetorical and theological texture of prophetic discourse. These limitations underscore the urgent need for methods that are semantically precise yet theologically respectful.

In response, this study introduces the Semantic Field Theory–Hadith Extraction Model (SFT-HEM), a framework that clusters narrations on women’s familial roles in *Ṣaḥīḥ al-Bukhārī* through lexical relations such as synonymy, hyponymy, and meronymy. Unlike keyword-driven or ontology-first approaches, SFT-HEM enables corpus navigation without doctrinal inference, thereby preserving the rhetorical and theological integrity of prophetic diction. By integrating classical semantic insights with computational tools such as Arabic WordNet and morphosyntactic parsers, the framework advances a linguistically grounded and pedagogically relevant approach to digital hadith scholarship. It offers thematic clarity for education and research while also opening new directions for natural language processing and ontology-based text analysis.

Literature Review

The evolution of hadith data extraction—from manual annotation to computational modelling—reflects a broader epistemic shift in Islamic studies. This transition marks a critical juncture at which traditional modes of textual engagement intersect with the methodological affordances of digital scholarship. Early efforts, such as thematic indexing and curated compilations, laid essential groundwork but remained heavily reliant on expert intuition rather than replicable semantic logic.

Contemporary digital repositories like *Maktabah Shamela* have extended this legacy by transforming classical collections into searchable, digitized formats. These platforms offer rapid, keyword-based access to authoritative texts. However, their retrieval mechanisms remain lexically constrained; they are unable to detect conceptual overlap, root-based morphological variation, or deeper semantic proximity across narrations.

Although *Maktabah Shāmilah* lacks semantic tagging capabilities, its comprehensive indexing of *Ṣaḥīḥ al-Bukhārī* and widespread use in traditional and academic circles rendered it an effective corpus for lexeme-driven extraction in the initial stage of thematic structuring.

More recent computational initiatives—such as HadithEnc and SemanticHadith (Kamran et al., 2023)—mark a significant methodological advance. By employing machine-readable structures and RDF-based ontologies, these projects enable scalable querying and interlinking of hadith corpora. In doing so, they help usher hadith studies into a digitally mediated phase, where intertextual analysis becomes increasingly feasible. Yet these gains come with notable limitations. RDF models, shaped by Western taxonomic logics, often impose rigid subject–predicate–object schemas that are less attuned to the context-sensitive, fluid, and rhetorically layered nature of prophetic discourse. As a result, they frequently fall short in capturing the dynamic semantic nuances embedded in hadith texts.

Theoretical Foundations of Semantic Field Theory

Semantic Field Theory (SFT) emerged within structuralist linguistics as a framework for understanding how words derive meaning not in isolation but through their relationships within structured lexical domains. Jost Trier introduced the concept in 1931, and it was later refined by Adrienne Lehrer (1974) and D. A. Cruse (1986). At its core, SFT posits that meaning arises through semantic networks—fields of conceptually related words that are organized by hierarchical and functional relations.

The theory identifies several foundational semantic relationships that structure lexical items into coherent fields. Each relation models a specific dimension of conceptual organization and thematic cohesion.

- Synonymy reflects semantic similarity, allowing clustering of lexically varied but conceptually equivalent terms—for example, *nisa'* and *imra'ah* as interchangeable terms for “woman.”
- Hyponymy structures lexicon hierarchically: specific terms (e.g., *ṣalāh*, ritual prayer) are subsumed under broader categories (e.g., *'ibādāt*, acts of worship).
- Meronymy captures part–whole relationships, as in *taslīm* (salutation) being a constituent of *ṣalāh*.
- Antonymy models binary oppositions like *ḥalāl* versus *ḥarām*, which are central to legal and ethical reasoning.
- Syntagmatic association maps co-occurrence patterns—such as the frequent pairing of *ṣawm* (fasting) and *ṣabr* (patience)—highlighting thematic resonance in exhortative contexts.

To illustrate these dynamics, Figure 1 presents an expanded semantic map of *'ibādāt* as represented in hadith discourse. This diagram organizes ritual lexemes not by form, but by their conceptual function—revealing field-based structures that enable nuanced annotation and facilitate more accurate thematic retrieval in digital systems.

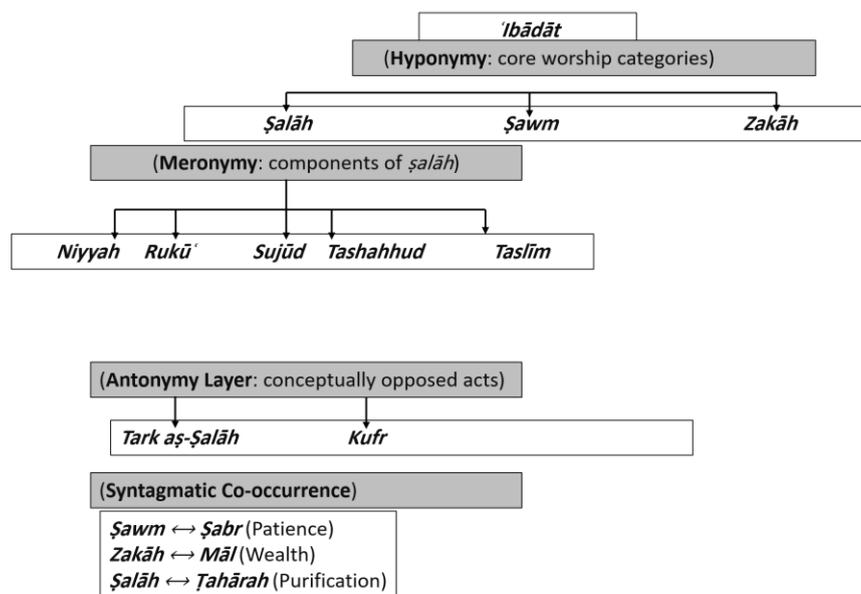


Figure 1: Expanded Semantic Field of 'Ibādāt in SFT Framework

This diagram visualizes the semantic relations—hyponymy, meronymy, antonymy, and syntagmatic co-occurrence—that structure the field of 'ibādāt within hadith texts. It shows how ritual terms like *ṣalāh*, *zakāh*, and *ṣawm* interrelate through hierarchical, oppositional, and co-occurring chains, supporting granular semantic tagging.

Importantly, semantic fields are not static. Lehrer (1974) emphasizes that their boundaries shift depending on genre, discourse context, and speaker intent. Cruse (1986) adds a distinction between paradigmatic relations (e.g., synonym sets) and syntagmatic ones (e.g., habitual co-occurrence). Both types of relations are especially pertinent in hadith, where repetition reinforces moral emphasis and interpretive inference.

Applied to Arabic texts, SFT aligns naturally with classical philological disciplines. *Ṣarf* (morphology) organizes lexemes around root families, echoing semantic field structures. *Naḥw* (syntax) and *balāghah* (rhetoric) mediate sentence-level cohesion and contextual nuance. Similarly, *munāsabah* (semantic interrelation)—central in *tafsīr* tradition—reflects SFT's foundational principle: that words derive meaning through embedded, relational constellations.

Indeed, Muslim scholars have long engaged in proto-SFT reasoning. Ahmad Ghafar et al. (2016) outline a four-phase chronology that traces SFT-like developments within the Islamic intellectual tradition:

1. Epistolary Phase: Early lexicographers such as al-Shaybānī, al-Aṣma'ī, and Qutrub compiled thematic treatises that grouped Arabic lexemes by cultural domains (e.g., *al-khayl* [horses], *al-ibil* [camels], *khalq al-insān* [human creation]).
2. Thematic Lexicons: Works such as *al-Gharīb al-Muṣannif* (al-Ḥarawī) and *Fiqh al-Lughah* (al-Tha'labī) advanced systematic lexical categorization, emphasizing synonymy, polysemy, and semantic overlap.

3. Historical Validation: From the early 2000s, scholars revisited classical sources like *al-Mukhaṣṣaṣ* and *Fiqh al-Lughah* to trace embedded field-based principles within Arabic linguistic thought.
4. Computational Application: In the current phase, researchers have begun applying SFT to Qur'anic and hadith corpora, clustering lexemes thematically in areas such as ethics (*akhlāq*), ritual (*'ibādah*), and social transactions (*mu'āmalāt*), particularly within *Ṣaḥīḥ al-Bukhārī*.

Taken together, these phases demonstrate that while SFT has its roots in Western structuralist linguistics, its conceptual foundations resonate deeply with Islamic scholarly traditions. As both a classificatory system and a computational scaffold, SFT enables semantic indexing of sacred texts in a way that preserves theological reverence while enhancing digital accessibility. It offers a productive convergence point—linking classical Arabic semantics with modern information modelling in the service of digital Islamic scholarship.

Computational Implementation: Ontologies and Hadith Data Structuring

The development of Arabic natural language processing (NLP) has made significant strides in recent years, enabling more structured, semantically aware engagement with classical texts. The SAFAR framework (Semantic Analysis Framework for Arabic Resources) exemplifies this trajectory. As a modular, ontology-based infrastructure, SAFAR supports semantic parsing, word sense disambiguation, and conceptual annotation of Arabic corpora. It integrates foundational lexical resources such as Arabic WordNet and Arabic VerbNet to enhance corpus-based information retrieval and language understanding (Nasri et al., 2018; Abouenour et al., 2014).

Arabic WordNet (AWN), initially modelled on Princeton WordNet, has undergone major revisions. A 2024 update by Freihat et al. improved over 58% of synsets, standardized glosses, and expanded semantic coverage—greatly increasing AWN's suitability for field-sensitive annotation in semantically dense domains such as hadith literature (Freihat et al., 2024).

Complementing this infrastructure is AraSAS (Arabic Semantic Annotation Scheme), a discourse-level model that applies hierarchical tagsets across syntactic, semantic, and pragmatic dimensions. AraSAS enables multi-level annotation and captures discourse relations, semantic roles, and context-sensitive patterns. It has proven effective in annotating narratively complex and ethically encoded genres, including sacred literature (El-Haj et al., 2022).

Together, SAFAR and AraSAS demonstrate the increasing maturity of Arabic NLP and its potential to support conceptually rich applications—such as the Semantic Field Theory–Hadith Extraction Model (SFT-HEM) proposed in this study.

Within the context of hadith, SFT brings several key affordances:

- Lexeme Clustering and Thematic Classification: SFT enables scholars to group semantically related narrations—even when lexically varied—into conceptual domains such as *'ibādāt* (acts of worship), *'uqūd* (contracts), and *akhlāq* (ethics) (Ahmad & Anwari, 2024).
- Hierarchical Structuring: Field mapping allows domain–subdomain layering; for example, *ṣalāh* is a subset of *'ibādāt*, and associated terms such as *rukū'*, *sujūd*, and *niyyah* can be semantically tagged for scalable annotation.

- Relational Semantics: Prior studies in Qur'anic hermeneutics and Arabic poetics show that SFT detects rhetorical and thematic coherence across lexically divergent texts (Faqeeh et al., 2024; Al-Mousa & Al-Rawashdeh, 2015).
- Ontology Development: Projects like SemanticHadith (Kamran et al., 2023) and other Islamic ontological initiatives (Fairouz & Nora, 2018) have begun incorporating SFT to generate RDF-compatible knowledge graphs for sacred corpora.

These affordances indicate that SFT is not merely a theoretical model but a computationally implementable framework—one that approximates the semantic intuition historically exercised by scholars of Arabic and hadith sciences.

More broadly, SFT offers a methodological bridge between classical hermeneutics and machine-readable environments. It supports fine-grained categorization while preserving lexical nuance, thereby increasing the semantic granularity of hadith corpora and enhancing the precision of digital annotation and retrieval systems.

Ṣaḥīḥ al-Bukhārī as Corpus for Thematic Structuring

Ṣaḥīḥ al-Bukhārī, compiled by *Imām Muḥammad ibn Ismā'īl al-Bukhārī* (d. 870 CE), holds canonical status in Sunni Islam and is widely regarded as the most authoritative hadith collection after the Qur'ān. Comprising over 7,000 narrations—including repetitions—*Ṣaḥīḥ al-Bukhārī* is meticulously organized into more than 90 *kitāb* (books) and over 3,000 *bāb* (chapters). Each entry typically includes both the *isnād* (chain of transmission) and the *matn* (narrative content), enabling layered legal, ethical, and theological engagement. *Al-Bukhārī's* methodology combined oral and written transmission pathways—*ijāzah*, *wajāda*, *samā'*, and *'arḍ wa-munāwalah*—and drew upon foundational sources such as *al-Muwaṭṭa'* and *al-Musannaf* (Obeid & Hussein, 2023).

The internal thematic structure and semantic cohesion of *Ṣaḥīḥ al-Bukhārī* render it particularly suitable for lexeme-driven thematic modelling using Semantic Field Theory (SFT). The present study extracted only the *matn* to ensure conceptual consistency and eliminate duplication associated with multiple *isnād* variants. This decision aligns with the goals of SFT, which relies on coherent lexical fields for clustering, mapping, and semantic proximity analysis. Furthermore, the selection of *Ṣaḥīḥ al-Bukhārī* is supported by recent digital hadith research that leverages its structured layout and rich linguistic features for classification, mining, and ontology development (Binbeshr et al., 2021; Saloot et al., 2016).

Nevertheless, the epistemological prestige of *Ṣaḥīḥ al-Bukhārī* has not shielded it from modern criticism. Contemporary voices such as Rashid Ailal have challenged the epistemic reliability of its transmission system and questioned the historiographical assumptions underlying its compilation. His monograph *Ṣaḥīḥ al-Bukhārī: Nihāyat Ustūrah* argues that the codification process was shaped by subjective filters and narrative approximations (Zainudin, 2022). Similarly, Western scholars such as Goldziher and Schacht have questioned the historicity of hadith transmission and the sanctity often attributed to canonical compilations (Husayn, 2022).

Despite these critiques, *Ṣaḥīḥ al-Bukhārī* continues to serve as a cornerstone in both traditional pedagogy and computational scholarship. In Southeast Asia, particularly Indonesia, it is taught extensively in *pesantren* (Islamic boarding schools) and higher Islamic education

using a range of methods—from memorization to thematic and critical approaches (Noorhidayati & Aziz, 2023). This pedagogical adaptability and its wide digital availability through platforms such as *Maktabah Shāmilah* strengthen its viability as a corpus for SFT-driven structuring. Thus, its canonical status, semantic depth, and educational influence justify its selection for this study’s semantic field extraction model.

The Methodological Logic of Repetition in Ṣaḥīḥ al-Bukhārī

One of the distinctive features of Imam *al-Bukhārī*’s methodology is his deliberate repetition of hadiths across different chapters (*abwāb*) within his *Ṣaḥīḥ*. These repetitions are not arbitrary but arise from his nuanced juristic reasoning (*istinbāṭ*) and precise thematic alignment. Rarely does *al-Bukhārī* reproduce the same narration with identical *isnād* and *lafz*; rather, he often introduces narrations through alternative transmission routes (*ṭuruq*) or subtle variations in wording, adapted to the specific legal or thematic focus of the chapter (Ibn Ḥajar, n.d., 1:10–12).

The objectives behind this repetition include: (1) elevating hadiths from the *gharīb* (solitary) category by supporting them with multiple chains of transmission (al-Kattānī, 1995); (2) resolving potential doubts about narrators who transmit both full and abridged versions of a narration (al-Sakhāwī, 1997); (3) highlighting semantic nuances in variant wordings from different narrators (al-Baghdādī, 2004); and (4) reconciling apparent contradictions between *mawṣūl* and *mursal* transmissions, or between *marfūʿ* and *mawqūf* attributions, in a way that preserves the epistemic integrity of the stronger chain (Ibn al-Ṣalāḥ, 2002).

Understanding this methodology is critical for studies such as the present one, which aims to filter hadiths based on thematic *matn* consistency. It also justifies the exclusion of *isnād* in our corpus extraction phase, as well as the removal of duplicate *matn* that reappear across structurally distinct chapters. Far from flattening the text, this approach acknowledges *al-Bukhārī*’s epistemic strategy while aligning it with semantic field theory’s emphasis on lexical clustering and thematic coherence.

Comparative Analysis: SFT versus Existing Hadith Extraction Methods

A comparative lens helps situate SFT alongside the two dominant paradigms in current hadith data extraction: keyword search and RDF-based ontologies. Each has merits but also limitations that SFT addresses more holistically.

Table 1

Comparison of Data Extraction Methods and the Semantic Field Theory (SFT) Advantage

Method	Strengths	Limitations	SFT Advantage
Keyword Search	Fast, scalable, and widely adopted in digital archives	Surface-level; misses synonymy, metaphor, and semantic drift	Captures lexical variation through field-based mapping and meaning prioritization
RDF/Ontologies	Enables structured queries; supports linked data	Rigid taxonomies; poor fit for fluid, context-sensitive semantics	Allows corpus-driven, linguistically informed construction of semantic fields

Keyword-based systems depend on literal matches. While efficient, they often fragment meaning—failing to retrieve narrations that differ lexically but share thematic or conceptual overlap. RDF models, by contrast, encode relational knowledge using subject–predicate–object triples. They offer structured annotation and interoperability but tend to mirror the rigidity of Western taxonomies. This can render them insufficiently flexible for the rhetorical subtleties and moral gradations characteristic of prophetic discourse.

SFT offers a fundamentally different approach. Rather than imposing external classifications, it derives structure from the corpus itself. Words are grouped by conceptual affinity—via synonymy, hyponymy, and meronymy—rather than by form alone. This yields semantically coherent clusters that better reflect traditional Arabic linguistic logic and hadith semantics. While this study recognizes the value of RDF tagging, it also acknowledges the diversity of Islamic ontological frameworks in development. Technologies such as OWL (Web Ontology Language) and SKOS (Simple Knowledge Organization System) provide powerful structuring capabilities, including multilingual glossing and cross-genre linkages between Qur’anic and hadith texts. However, these tools also introduce epistemological risks. Translating divine categories or jurisprudential norms into machine-readable code may inadvertently distort meaning. Ontology design in sacred domains thus demands careful balance—between technical expressiveness and theological integrity.

Methodology

We adopted a qualitative document analysis approach grounded in Semantic Field Theory (SFT) to systematically structure, extract, and classify narrations from *Ṣaḥīḥ al-Bukhārī* that address women’s familial roles—specifically as mothers, wives, and daughters. Rather than engaging in juridical rulings or theological interpretation, our focus was on semantically structuring textual data to support conceptual modelling and enable enriched digital retrieval. SFT, first proposed by Trier (1931) and later refined by Lehrer (1974) and Cruse (1986), emphasizes that lexical meaning arises through structured relationships such as synonymy, hyponymy, and meronymy. Accordingly, this methodology unfolds in three interrelated phases: (1) lexical identification, (2) corpus retrieval, and (3) thematic extraction based on role-specific semantics.

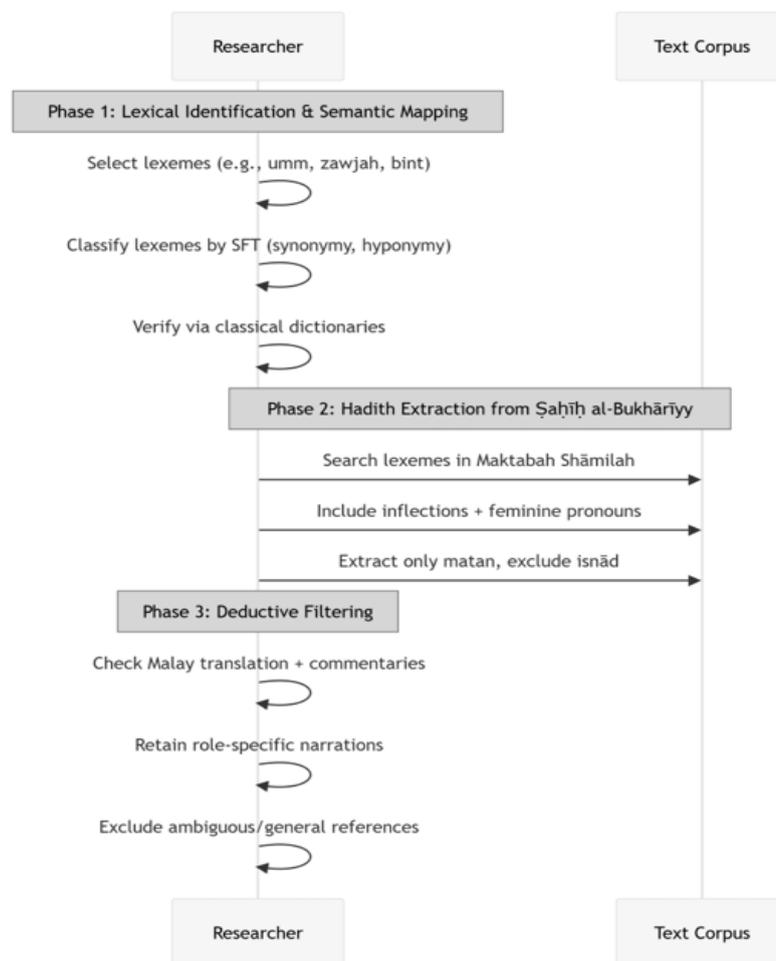


Figure 2: SFT-Hadith Extraction Model (SFT-HEM): Structuring Thematic Narratives from Ṣaḥīḥ al-Bukhārī

This workflow depicts researcher interactions with the Ṣaḥīḥ *al-Bukhārī* corpus through (1) Lexical Identification & Semantic Mapping, (2) Hadith Extraction, and (3) Thematic Extraction Based on Role-Specific Criteria.

Lexical Identification and Semantic Mapping

The first phase involved manually identifying Arabic lexemes associated with mar'ah ("woman"), guided by SFT's emphasis on semantic field structuring. Core terms—imra'ah, nisā', umm (mother), zawjah (wife), and bint (daughter)—were selected for their relevance to kinship roles and verified against classical dictionaries (Lisān al-'Arab, Tāj al-'Arūs, al-Mu'jam al-Wasīṭ, al-Muḥīṭ).

Lexeme selection followed a three-step process: (1) initial identification based on thematic salience in familial contexts, (2) lexical validation using classical dictionaries, and (3) cross-referencing with hadith concordances for frequency and contextual fit. Terms exhibiting

ambiguity or polysemy were excluded to maintain semantic coherence and ensure inter-coder reliability.

Each lexeme was then classified into one of four SFT-defined semantic relations

Table 2

Sample Lexemes by SFT Relation

Semantic Relation	Sample Lexemes
Synonymy	<i>imra'ah, nisā', zawjah, ḥarām</i>
Hyponymy	<i>bint, ṣāliḥah, mu'minah</i>
Meronymy	<i>umm, ukht, jaddah, 'ammah, khālah</i>
Antonymy	<i>rajul, dhakar, ṭifl, ṣabiyy</i>

Figure 4 visualizes the structured semantic field of mar'ah, showing how each term branches into a coherent conceptual cluster that guided subsequent thematic extraction.

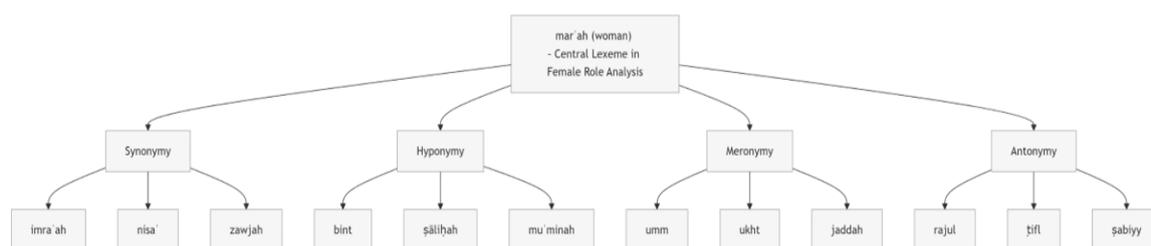


Figure 3: Figure 4. Structured semantic field of mar'ah showing thematic branches
Figure 4. Semantic Structuring of the Lexeme mar'ah ('Woman') Based on SFT Relations for Thematic Extraction

Hadith Extraction from Ṣaḥīḥ al-Bukhārī

In the second phase, we used the validated lexemes to extract narrations from *Ṣaḥīḥ al-Bukhārī* via *Maktabah Shāmilah v2.11*. The search process accounted for multiple inflectional variants, including singular and plural forms, definite and indefinite markers, and gendered expressions, as well as feminine *ḍamīr* (pronouns) that implicitly referred to women. We limited the focus to the *matn* (narrative body), deliberately excluding *isnād*, and prioritized thematic books such as *Kitāb al-Adab* and *Kitāb al-ʿIlm*.

Although this step was conducted manually, it is well-suited to enhancement through natural language processing (NLP) tools. Future studies could use platforms such as *HadithEnc* and *Qur'anBERT* to perform synonym-aware and context-sensitive searches. RDF schemas and conceptual graphs may also support ontology-ready tagging.

Deductive Filtering for Thematic Relevance

From the initial 1,463 narrations, 460 duplicates were removed (e.g., identical matn under different headings), yielding 1,003 unique texts. Thematic extraction was guided by strict inclusion criteria:

Table 3

Inclusion and Exclusion Criteria for Thematic Filtering of Ḥadīth Corpus

Inclusion Criteria	Exclusion Criteria
<ol style="list-style-type: none"> 1. Explicit mention of mother (<i>umm</i>), wife (<i>zawjah</i>), or daughter (<i>bint</i>)—either via direct lexeme or unambiguous gendered pronoun. 2. Thematic focus confirmed by Malay JAKIM translation and at least one classical commentary (<i>Fath al-Bārī</i>, <i>Irshād al-Sārī</i>, or <i>Umdah al-Qārī</i>). 	<ol style="list-style-type: none"> 1. Generic references to women (<i>nisa'</i>, <i>imra'ah</i>) without role specificity. 2. Incidental mentions of role-lexemes within lists of virtues where familial role is not central. 3. Lexical ambiguity unresolved by translation or commentary triangulation.

This structured approach yielded a refined set of 60 narrations (6% of the deduplicated corpus), each thematically grounded in one of the three familial roles. This alignment of thematic structure with semantic theory positions the model for integration into multilingual ontology frameworks and hadith-based digital knowledge graphs.

Validity and Reliability

To establish the reliability of our manual SFT-HEM application, we conducted an inter-rater agreement study on a random sample of 30 narrations from the initial 1,463-hadith corpus. Two independent coders—both trained on the SFT-HEM decision rules and lexeme lists—assigned each hadith to one of four categories: *mother*, *wife*, *daughter*, or *none*. We then computed Cohen's κ to quantify agreement beyond chance and report simple percent agreement for transparency.

The coders achieved $\kappa = 0.78$ (substantial agreement) and 86.7 % raw agreement. Disagreements were most frequent between the *daughter* and *no* categories, suggesting that our lexeme list for “daughter” may require further refinement. These results demonstrate that the SFT-HEM framework can be applied consistently by different researchers.

Theoretical Alignment and Interpretive Safeguards

Although SFT-HEM functions as a structural—not interpretive—model, reliance on the JAKIM Malay translation and classical commentaries unavoidably introduced interpretive layers. To mitigate this, we implemented four safeguards:

1. Data-Source Triangulation

Translations and commentaries were used only for disambiguation, not for defining thematic categories. Their role was to clarify ambiguous pronouns or lexemes.

2. Limitation of Interpretive Scope

Commentaries and translations helped confirm primary meaning within kinship context. No tafsīr-based reasoning (e.g., *asbāb al-nuzūl*, legal exegesis) was applied.

3. Cross-Validation Across Sources

Every translation-informed decision was checked against multiple commentaries. Where disagreement arose, the narration was flagged or excluded to ensure internal consistency.

4. Anchoring in SFT Semantic Relations

All classification followed SFT's core semantic logic—synonymy, hyponymy, and meronymy. Vernacular insights were supplementary and never overrode semantic field logic.

By treating translation and commentary as tools for semantic verification—rather than engines of interpretation—we preserved the structural integrity of SFT-HEM while adhering to the epistemic humility required when analysing sacred texts.

Results and Discussion

The SFT-HEM pipeline progressed through a three-phase workflow, beginning with a lexeme-driven query of *Ṣaḥīḥ al-Bukhārī*, which yielded 1,463 candidate narrations. In Phase 1, we removed 460 duplicate *matn*—narrative bodies appearing under different *isnād* or chapter headings—resulting in 1,003 unique *Ḥadīth* (a 31.4% reduction).

In Phase 3, the deductive filtering phase, each narration was evaluated for thematic relevance. To be retained, a hadith had to (1) reference a specific kinship role—*mother*, *wife*, or *daughter*—via explicit lexemes or unambiguous gendered pronouns, and (2) meet a semantic verification standard using the Malay JAKIM translation and at least one classical commentary. This process refined the corpus to 60 narrations (6% of the deduplicated set), each directly engaging with one of the three familial roles.

Table 4

Thematic Distribution of Narrations by Familial Role in Ṣaḥīḥ al-Bukhārī

Role	No of Hadiths	Percentage (%)
Mother	27	45
Wife	24	40
Daughter	9	15
Total	60	100

A semantic and ethical analysis of the final corpus revealed three thematic clusters, corresponding to the kinship roles. Each cluster is anchored in recurring rhetorical patterns and SFT-based semantic relations.

a. Mother (27; 45 %)

Mothers emerge as the most frequently invoked role, reflecting their profound spiritual and moral centrality in prophetic teachings. The narrations in this cluster foreground values such as care, responsibility, and moral guardianship, positioning mothers as ethical stewards of the domestic sphere.

“وَالْمَرْأَةُ رَاعِيَةٌ عَلَى أَهْلِ بَيْتِ رَوْحِهَا، وَوَلَدِهِ وَهِيَ مَسْئُولَةٌ عَنْهُمْ.”

“The woman is a shepherd or guardian over the household of her husband and his children, and she is accountable for them.”

(*Ṣaḥīḥ al-Bukhārī*, *Kitāb al-Nikāḥ*, *ḥadīth no. 4904*)

This authoritative narration encapsulates the spiritual significance of motherhood, framing it as a form of entrusted leadership with ethical weight and divine accountability. The semantic consistency across these hadiths—centred on concepts of stewardship (*ri'āyah*), trust

(*amānah*), and responsibility (*mas'ūliyyah*)—reinforces their classification under the “mother” field.

b. Wife (24; 40 %)

Wives are portrayed in the Prophetic narrations as partners embodying care, ethical companionship, and emotional resilience within the household. These reports not only stress fairness and kindness in conjugal relations but also spotlight the moral strength of women in navigating hardship with patience and wisdom.

كَانَ ابْنُ لِأَيِّ ظَلْحَةَ يَسْتَكِي ... فَقَالَ رَسُولُ اللَّهِ ﷺ: أَعْرَسْتُمْ اللَّيْلَةَ؟ قَالَ: نَعَمْ، قَالَ: اللَّهُمَّ بَارِكْ لَهُمَا، فَوَلَدَتْ غُلَامًا ... فَسَمَّاهُ النَّبِيُّ ﷺ عَبْدَ اللَّهِ

“Anas bin Mālik reported: A son of *Abū Ṭalḥah* fell ill and died in his absence. When he returned and asked about the child, Umm Sulaym replied: ‘He is now at peace.’ She then served him supper, and they were together that night. In the morning, *Abū Ṭalḥah* informed the Prophet ﷺ, who asked: ‘Did you come together last night?’ He replied yes. The Prophet ﷺ prayed: ‘O Allah, bless them both.’ Later, Umm Sulaym gave birth to a son, and the Prophet ﷺ performed *tahnīk* for him and named him ‘*Abd Allāh*.’”

(*Ṣaḥīḥ al-Bukhārī, Kitāb al-‘Aqīqah, ḥadīth no. 5470*)

This narration encapsulates a wife’s role as a source of emotional stability and wisdom in moments of crisis. Umm Sulaym’s composure illustrates a profound ethic of *ṣabr* (patience) and *hikmah* (prudence), reinforcing the image of the wife not only as a partner in affection but also as a moral anchor in adversity. The semantic field in this cluster spans terms of companionship (*zawjiyyah*), kindness (*iḥsān*), and resilience (*ṣabr*), anchoring the role of wives within both relational and spiritual stewardship.

c. Daughter (9; 15 %)

Although limited in number, narrations concerning daughters demonstrate marked semantic density and moral depth. Their thematic rarity is counterbalanced by concentrated references to patience, filial dignity, and eschatological reward—framed in direct contrast to pre-Islamic norms that devalued female offspring. These *ḥadīths* reframe daughterhood as a locus of divine favour, spiritual refinement, and protective intercession in the Hereafter.

خَرَجَ عَلَيْنَا النَّبِيُّ ﷺ وَأَمَامَهُ بِنْتُ أَبِي الْعَاصِ عَلَى عَاتِقِهِ فَصَلَّى، فَإِذَا رَكَعَ وَضَعَهَا، وَإِذَا رَفَعَ رَفَعَهَا

“*Abū Qatādah* reported: The Prophet ﷺ came out to us with *Umāmah*, the daughter of *Abū al-‘Āṣ*, on his shoulder. He prayed, and whenever he bowed, he put her down, and when he rose he lifted her up again.”

(*Ṣaḥīḥ al-Bukhārī, Kitāb al-Adab, ḥadīth no. 5996*)

This narration embodies a vivid ethic of tenderness and dignity toward daughters, situating them within the rhythms of worship itself. The Prophet’s act of carrying *Umāmah* during prayer underscores the integration of familial affection into spiritual life, redefining daughterhood as a site of mercy and blessing rather than burden. Lexically, these narrations cluster around meronymic and hyponymic relations, linking daughters to broader familial domains (*ahl*, *‘iyāl*) and to divine reward structures (*‘afw*, *nārah*). Their numerical scarcity does not undermine their semantic coherence or ethical resonance; rather, it highlights the rhetorical intensity and moral intentionality embedded within each narration.

Collectively, these findings demonstrate that SFT-HEM not only isolates role-specific hadiths with high semantic precision but also preserves their moral architecture and rhetorical subtlety. Each cluster is grounded in well-defined SFT relations—allowing for more faithful representation than keyword or RDF-based methods can typically achieve.

By grouping narrations based on conceptual affinity rather than surface lexemes, SFT-HEM transcends the limitations of literalism. It offers a replicable model for structuring hadith discourse that is both computationally rigorous and theologically respectful.

This layered approach—rooted in classical Arabic semantics and supported by digital annotation tools—suggests a viable pathway for future hadith corpus design, ontology development, and pedagogy in digital Islamic scholarship.

Conclusion

This study has reaffirmed the centrality of women’s familial roles—as mothers, wives, and daughters—in Islamic ethical and social thought. Prophetic narrations situate these roles as pillars of responsibility, moral education, and communal continuity. Their importance is consistently affirmed in contemporary scholarship, yet within *Ṣaḥīḥ al-Bukhārī* these narrations appear dispersed across books and chapters, weakening thematic integration and obscuring continuity in women’s religious agency.

Although prior scholarship recognizes the moral and religious significance of these narrations, it also highlights the absence of systematic methods for organizing them in linguistically coherent and conceptually rigorous ways. The lack of structured tools limits both pedagogical application and scholarly interpretation, leaving women’s familial roles underrepresented in curricular, digital, and theological contexts where accessibility and clarity are most needed.

To address this gap, the study introduced the Semantic Field Theory–Hadith Extraction Model (SFT-HEM), a hybrid framework that clusters narrations through lexical relations such as synonymy, hyponymy, and meronymy. Unlike keyword-based or RDF-driven systems, SFT-HEM enables thematically structured retrieval without compromising theological integrity or rhetorical nuance. A pilot application refined 60 role-specific narrations from a corpus of 1,463 candidates, demonstrating the model’s ability to reveal latent thematic patterns and illuminate the textual architecture of women’s roles within *Ṣaḥīḥ al-Bukhārī*.

The broader implications of SFT-HEM extend across education, digital humanities, and community life. For educators and students, it provides thematic access to narrations that support gender-aware moral instruction. For digital humanists, it establishes a foundation for nuanced ontologies, Arabic NLP pipelines, and discourse-sensitive retrieval systems. For the wider Muslim community, it reaffirms the ethical and spiritual significance of women’s familial contributions within Islamic tradition. By integrating classical semantic theory with computational tools, SFT-HEM advances a rigorous, digitally grounded, and theologically respectful approach to hadith scholarship, while opening future directions in automation, Qur’an–hadith intertextuality, and multilingual adaptation.

References

- Abed Alhakim, F., Khalilia, H. M., Bella, G., & Giunchiglia, F. (2024). Advancing the Arabic WordNet: Elevating content quality. Proceedings of the 6th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT) with Shared Tasks on Arabic LLMs Hallucination and Dialect to MSA Machine Translation @ LREC-COLING 2024, 74–83. ELRA & ICCL.
- Abouenour, L., Nasri, M., Bouzoubaa, K., & Rosso, P. (2014). Construction of an ontology for intelligent Arabic QA systems leveraging the Conceptual Graphs representation. *Journal of Intelligent and Fuzzy Systems*, 26(2), 861–874.
- Ahmad, H., Ghafar, N. A., & Bahjat, M. M. (2017). Teks prosa Zarzuriyyat Ibn Fakhaar: Analisa korpus bertema berdasarkan teori medan. *Journal of Islamic, Social, Economics and Development*, 2(2), 21–31.
- Ahmad, M. I., & Anwari, M. K. (2024). Semantic field extraction in Arabic online news using NLP and machine learning models. *Studies in English Language and Education*, 11(1), 32–51.
- Al-Baghdādī, H. (2004). *Manhaj al-Imām al-Bukhārī fī al-jam‘ bayn al-aḥādīth al-muta‘arīḍah fī ṣaḥīḥih*. Riyadh: Maktabah al-Rushd.
- Al-Bukhari, M. I. (1997). *Sahih al-Bukhari* (M. M. Khan, Trans.). Darussalam.
- Al-Kattānī, M. ibn Ja‘far. (1995). *Al-Risālah al-mustaṭrafah li-bayān mashhūr kutub al-sunnah al-muṣannafah*. Beirut: Dār al-Bashā‘ir al-Islāmiyyah.
- Al-Mousa, Y., & Al-Rawashdeh, B. (2015). Semantic relations in Al-Ibil by Al-Asma‘ī. *Dirasat: Human and Social Sciences*, 42(2), 523–540.
- Al-Sakhāwī, M. ‘A. (1997). *Faṭḥ al-mughīth fī sharḥ Alfiyyat al-ḥadīth* (Vol. 1). Beirut: Dār al-Kutub al-‘Ilmiyyah.
- Bendjamaa, F., & Taleb, N. (2024). OntoDin: An Islamic ontology of Quran and Hadith. *International Arab Journal of Information*.
- Binbeshr, F., Kamsin, A., & Mohammed, M. (2021). A systematic review on hadith authentication and classification methods. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 20(2), Article 34. <https://doi.org/10.1145/3434236>
- Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.
- El-Haj, M., Rayson, P., & Habash, N. (2022). AraSAS: The open-source Arabic semantic analysis system. In *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools at LREC 2022* (pp. 56–66).
- Erwani, I., & Siregar, A. S. (2025). The role of women in Islamic sacred texts: A critical study of women’s narratives and authority in Islamic tradition. *Pharos Journal of Theology*, 106(1), 1–14. <https://doi.org/10.46222/PHAROSJOT.106.6>
- Fairouz, B., & Nora, T. (2018). Computational ontologies for a semantic representation of the Islamic knowledge. In *IFIP Advances in Information and Communication Technology* (Vol. 539, pp. 322–330). Springer.
- Faqeeh, M. H. A., Alhourani, M. I., & Obeidat, A. M. B. (2024). “Words of war” as found in the writings by poet Abu Firas Al-Hamdani based on semantic fields theory. *Dirasat: Human and Social Sciences*, 51(1), 87–102.
- Ghafar, N. A., Ahmad, H., Mohd Ghazali, N., Mohd, R. A., & Mustapha, N. F. (2016). Semantic field theory studies among Muslim scholars. *Al-Qanatir: International Journal of Islamic Studies*, 4(2), 24–35.

- Husayn, N. (2022). Modern Muslim objections to Ṣaḥīḥ al-Bukhārī. *Studia Islamica*, 117(1), 127–170. <https://doi.org/10.1163/19585705-12341459>
- Ibn al-Ṣalāḥ. (2002). *ʿUlūm al-ḥadīth*. Cairo: Dār al-Ḥadīth.
- Ibn Ḥajar al-ʿAsqalānī. (n.d.). *Faṭḥ al-bārī bi sharḥ Ṣaḥīḥ al-Bukhārī* (Vol. 1). Cairo: Dār al-Maʿārif.
- Junaedi, S. I., Syafi'i, I., Hasyim, F., & Jawid Marwat, M. (2025). Transformational curriculum model through interpretation epistemology: Islamic boarding school curriculum management perspective. *Munaddhomah*, 6(3), 387–403. <https://doi.org/10.31538/munaddhomah.v6i3.1802>
- Kamran, A. B., Abro, B., & Basharat, A. (2023). SemanticHadith: An ontology-driven knowledge graph for the hadith corpus. *Journal of Web Semantics*, 79, 100740. <https://doi.org/10.1016/j.websem.2023.100740>
- Lehrer, A. (1974). *Semantic fields and lexical structure*. North-Holland.
- Mehfooz, M. (2021). Women and hadith transmission: Prolific role of Aisha in validation and impugment of prophetic traditions. *Al-Bayan*, 19(2), 200–227. <https://doi.org/10.1163/22321969-12340099>
- Nasri, M., Jaafar, Y., & Bouzoubaa, K. (2018). Semantic analysis of Arabic texts within SAFAR framework. In *Proceedings of the 6th Colloquium in Information Science and Technology (CIST)* (pp. 137–144). IEEE. <https://doi.org/10.1109/CIST.2018.8596637>
- Noorhidayati, S., & Aziz, T. (2023). Hadith studies in Indonesia: Vernacularization and teaching methods of Ṣaḥīḥ al-Bukhārī in traditional and contemporary Islamic educational institutions. *European Journal for Philosophy of Religion*, 15(3), 60–80. <https://doi.org/10.24204/EJPR.2023.4182>
- Obeid, S. M. S., & Hussein, F. S. (2023). Al-Bukhari's sources. *Jordan Journal for History and Archaeology*, 17(3), 44–62. <https://doi.org/10.35516/jjha.v17i3.434>
- Saloot, M. A., Idris, N., Mahmud, R., Ja'afar, S., Thorleuchter, D., & Gani, A. (2016). Hadith data mining and classification: A comparative analysis. *Artificial Intelligence Review*, 46(1), 113–128. <https://doi.org/10.1007/s10462-016-9458-x>
- Selsuk, M. (2021). Religious education, pedagogy and curriculum in plural context: A Muslim perspective in Turkey. In *New Directions in Religious and Values Education: International Perspectives* (pp. 235–250). Peter Lang.
- Stowasser, B. F. (2011). *Women in the Qur'an, traditions, and interpretation*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195111484.001.0001>
- Trier, J. (1931). *Der deutsche Wortschatz im Sinnbezirk des Verstandes: Die Geschichte eines sprachlichen Feldes*. Carl Winter.
- Zainudin. (2022). Ailal Rashid's critical contribution to Ṣaḥīḥ al-Bukhārī in the book Ṣaḥīḥ al-Bukhārī: Nihāyat Ustūrah. *Jurnal Studi Ilmu-ilmu al-Qur'an dan Hadis*, 23(1), 1–18. <https://doi.org/10.14421/qh.2022.2301-01>