https://jpll.ui.ac.ir/?lang=en
Research on Mystical Literature
E-ISSN: 2476-3292
Document Type: Research Paper
Vol. 19, Issue 1, No.54, 2025, pp. 187-207
Received: 12/10/2025 Accepted: 01/11/2025

In Search of Meaning: Knowledge Graph-Based Conceptual Retrieval in Persian Mystical Texts - A Case Study of the *Masnavi Ma'navi*

Mohammad Amini
Post-doctoral Researcher, Faculties of Literature and Humanities and Computer Engineering, University of Isfahan, Isfahan, Iran
m.amini@ltr.ui.ac.ir

Ehsan Reisi *
Associate Professor of Persian Language and Literature, Department of Persian Language and Literature, Faculty of Literature and Humanities, University of Isfahan, Isfahan, Iran
e.reisi@ltr.ui.ac.ir

Mohammad Ali Nematbakhsh
Professor, Faculty of Computer Engineering, University of Isfahan, Isfahan, Iran
nematbakhsh@eng.ui.ac.ir

Abstract
Effective search within Persian literary works plays a crucial role in extracting new knowledge and insights from the intellectual heritage of Iran's literary masters. However, classical Persian literature, particularly mystical works like Rumi's *Masnavi Ma'navi*, is replete with metaphorical layers and polysemous concepts. This characteristic renders conventional text search methods inadequate for the accurate and meaningful retrieval of such texts, as these algorithms are fundamentally reliant on word frequency and overlook the latent semantic relationships between concepts. In this research, we propose a novel approach for semantic search in Persian mystical texts, grounded in the integration of classical statistical methods and conceptual knowledge graphs. The methodology begins with an initial keyword-based search, followed by the extraction of key concepts from the retrieved verses. These concepts are then expanded by leveraging lexical semantic networks, textual co-occurrence, synonyms, and hypernyms, and are represented as a conceptual graph. The final search is conducted by integrating the findings of the initial search with the expanded concepts. Experimental results demonstrate that this method not only enhances search precision but also offers greater explainability and transparency for the user, as the conceptual relationships are explicitly visualized within the graph. The applications of this research extend beyond text retrieval and can be leveraged in domains such as Digital Humanities, literary research, and mystical studies.

Keywords: Conceptual Search, Knowledge Graph, Semantic Similarity, *Masnavi Ma'navi*.

* Corresponding author
Amini, M. , Reisi, E. and Nematbakhsh, M. A. (2025). In Search of Meaning: Knowledge Graph-Based Conceptual Retrieval in Persian Mystical Texts - A Case Study of the Masnavi Ma'navi. *Research on Mystical Literature*, 19(1), 187-207.

2476-3292 © The Author(s).

This is an open access article under the CC BY-NC 4.0 License (https://creativecommons.org/licenses/by-nc/4.0

10.22108/jpll.2025.147019.1959

Introduction

Persian mystical literature, particularly *Rumi's Masnavi Maʿnavi*, abounds with layers of metaphor, symbolism, and conceptual depth. Scholars seeking to retrieve verses that express particular spiritual or philosophical ideas face persistent challenges. Traditional keyword-based retrieval methods—effective in factual or scientific texts—are less suitable for the figurative and polysemous nature of mystical poetry, where a single term may convey multiple, context-dependent meanings. These limitations highlight the need for semantic approaches that move beyond surface words to uncover deeper conceptual linkages among verses.

A central problem in text retrieval today is the semantic gap—the disparity between surface-level word matching and true meaning comprehension. Frequency-based models treat words as isolated units, ignoring synonymy, metaphor, and inter-conceptual relations. Semantic knowledge graphs offer a solution by structuring knowledge as interconnected nodes (concepts) and edges (relations), capturing both lexical and semantic dimensions of meaning. As an advanced framework in artificial intelligence and natural language processing, this approach enhances information retrieval by modeling semantic relations between entities, improving disambiguation, and supporting dynamic query expansion (Haslhofer et al., 2018).

The present study proposes a hybrid method that integrates classical statistical models with a semantic graph-based representation to enable conceptual search in Persian mystical texts. An initial lexical search using the BM25 model (Robertson & Zaragoza, 2009) retrieves base verses. Then, by employing the Persian semantic network *FarsNet* (Mousavi & Faili, 2017), along with co-occurrence, synonymy, and hypernym relations, a semantic knowledge graph is dynamically built to identify conceptually related verses. This approach allows user queries to move from word-level to meaning-level retrieval, returning verses that reflect the intended concepts rather than mere keyword matches. Ultimately, the model aims to enhance retrieval accuracy, interpretability, and depth of analysis, contributing to the broader goals of digital humanities and the computational study of Persian mystical heritage.

Literature Review

Semantic knowledge graphs have become a central framework in modern information retrieval, enabling systems to move from surface-level keyword matching to meaning-based understanding. By modeling concepts as nodes and their semantic relations as edges, such graphs improve the accuracy and contextual relevance of search results. Since Google's introduction of the *Knowledge Graph* (Peng et al., 2023), this approach has been widely adopted across domains—from biomedical retrieval systems that connect research data and treatments (Waldrop et al., 2022) to e-commerce platforms that model relationships among products and users (Kiff, n.d.). Recent advances, such as causal reasoning graphs (Bektemyssova & Sabdenov, 2024) and large-scale resources like *CS-KG 2.0* (Dessí et al., 2025), demonstrate the scalability and analytical power of this paradigm.

In the Persian context, progress in lexical and semantic resources has laid important groundwork for conceptual modeling. *FarsNet* (Shamsfard et al., 2010) serves as the Persian equivalent of WordNet and has been expanded through unsupervised distributional methods (Montazery & Faili, 2011) and supervised bilingual mapping techniques (Mousavi & Faili, 2017). Traditional pattern-based methods, such as Hearst patterns (Hearst, 1992), have been complemented by embedding-based models like Word2Vec, fastText, and ParsBERT, which detect semantically related words beyond synonymy (Zhong et al., 2023). These developments underscore a broader methodological trend toward hybrid, multi-source approaches for low-resource languages (Zhao et al., 2023).

Semantic search systems extend these ideas by integrating structured knowledge and statistical models to retrieve conceptually related information. Prior studies have applied concept mapping, explicit semantic analysis (Egozi et al., 2011), and ontology-based retrieval (Wen et al., 2006) to enhance meaning-level understanding. In Persian NLP, resources such as FNLP-ONT (Hosseini Pozveh et al., 2018), *PersPred* (Samvelian & Faghiri, 2013), and *Perspell* (Dastgheib et al., 2017) have improved language understanding and relation extraction. However, semantic retrieval in Persian classical literature remains limited. The current study addresses this gap by proposing a dynamic graph-based framework that combines BM25 with FarsNet-driven query expansion to enable conceptual, interpretable search within Rumi's *Masnavi*.

Methodology

The proposed retrieval framework consists of two main components: a search core and a semantic layer. The search core employs the classical BM25 model implemented through the Lucene engine to retrieve the initial set of candidate verses. BM25, a probabilistic ranking function based on term frequency and document length normalization (Robertson & Zaragoza, 2009), estimates the relevance of each document to a query by weighting words according to their statistical significance. While highly effective for contemporary and structured texts, BM25 alone cannot capture the figurative, polysemous, and metaphorical nature of classical Persian poetry. To address this limitation, the system integrates a semantic layer that expands user queries through a concept-based knowledge graph built from *Masnavi Maʿnavi* and the Persian WordNet (FarsNet). This layer enriches the

retrieval process by identifying conceptual relations such as synonymy, hypernymy, and co-occurrence, thereby transforming lexical search into conceptual search.

1) Semantic Knowledge Graph Construction

The semantic graph models the network of interrelated mystical concepts in Rumi's Masnavi. It is a multi-layer graph, where each layer encodes a distinct type of relation: (1) co-occurrence extracted from verses, (2) lexical relations from FarsNet, and (3) distributional similarity based on word embeddings (fastText). Each node represents a concept or keyword, and each edge denotes a weighted semantic relation. Two foundational mappings were created to support graph construction:

(a) a verse-to-concept map linking every verse (25,636 in total) to its associated concepts, extracted via unigram/bigram statistics, pointwise mutual information (PMI), and expert-curated lists of mystical terms; and

(b) a concept-to-verse index that enables reverse lookup and efficient retrieval.

In addition, FarsNet provided synonym and hypernym relations, forming two auxiliary datasets that expand conceptual coverage. These resources collectively allow the system to construct a dynamic knowledge graph at query time, representing both explicit and implicit semantic connections among the retrieved verses.

2) Enhanced Semantic Retrieval and Re-ranking

Once a user submits a query, an initial BM25 search retrieves the most relevant verses. Concepts associated with these verses are then expanded using the semantic graph—incorporating synonyms, hypernyms, and distributionally similar terms. Each concept becomes a node, and edges between nodes are assigned composite weights based on co-occurrence frequency, lexical relation strength, and embedding similarity. Low-weight edges are pruned to maintain graph efficiency and focus on semantically meaningful connections. The expanded graph is then traversed to identify additional verses linked through conceptual similarity, even if they do not contain the original query terms.

Finally, the retrieved results are re-ranked using a hybrid scoring model that combines the statistical BM25 score with a semantic relevance score derived from the graph structure. The final rank of each verse (d) is computed as a weighted combination of the two scores, with the weighting coefficient optimized empirically. This integration ensures that verses conceptually aligned with the user's intent—but lacking direct lexical overlap—are ranked more highly. The resulting system thus bridges the gap between statistical and semantic retrieval, providing interpretable, conceptually coherent search results suitable for the intricate symbolic structure of Persian mystical literature.

Results

To evaluate the effectiveness of the proposed semantic knowledge graph–based retrieval method, two systems were compared: (1) a **baseline BM25** model (Robertson & Zaragoza, 2009), and (2) the proposed **hybrid KG-BM25** model, which integrates BM25 with a semantic knowledge graph constructed from *Masnavi Maʿnavi* and *FarsNet.*

Eighteen conceptual queries were designed, representing key mystical notions such as *love (ʿeshq)*, *unity (tawḥid)*, *reason (ʿaql)*, *self (nafs)*, and *prophethood*. Each query retrieved the top ten verses per method. The results were independently assessed by three domain experts in Persian mysticism and classical literature, who rated each verse on a 1–5 scale for semantic relevance. The final score for each method was obtained by averaging expert ratings across all queries.

The comparative analysis demonstrated that the **KG-BM25 method consistently outperformed the baseline BM25** in terms of conceptual retrieval quality. On average, KG-BM25 improved expert-rated relevance scores by approximately **9%** across all queries, with the improvement reaching **up to 40%** for deeply metaphorical queries such as *"union of the apparent and the manifest."* Unlike BM25—which primarily retrieved verses containing the literal query term—KG-BM25 successfully identified semantically related verses that expressed equivalent or complementary concepts (e.g., *appearance*, *manifestation*, and *unity*). These findings confirm that the semantic graph effectively captured latent conceptual and metaphorical connections embedded within Rumi's text.

In a few cases, the hybrid model performed slightly below the baseline, primarily due to **semantic mismatches** between mystical meanings and general lexical resources. For example, words like *wine (sharāb)* metaphorically denote *divine love* in Sufi discourse but are associated with their literal meanings in *FarsNet* or embedding models. Such gaps highlight the limitations of current lexical coverage and the need for domain-specific enrichment of Persian semantic resources. Nonetheless, expert feedback emphasized that the **KG-BM25 results were more interpretable**, as the semantic relationships between query and verse could be traced through the graph structure.

Overall, the findings indicate that integrating semantic knowledge graphs into classical text retrieval significantly enhances conceptual accuracy and interpretability. The KG-BM25 model thus provides a robust foundation for **conceptual search and digital humanities applications**, enabling deeper semantic exploration of Persian mystical literature beyond surface-level lexical matching.

نشریهٔ علمی پژوهش‌های ادب عرفانی
سال نوزدهم، شمارهٔ اول، پیاپی ٥٤، ١٤٠٤، صص. ١٨٧-٢٠٧
تاریخ وصول: ١٤٠٤/٧/٢٠ تاریخ پذیرش: ١٤٠٤/٨/١٠

مقاله پژوهشی

در جست‌وجوی معنا: بازیابی مفهومی مبتنی‌بر گراف دانش در متون عرفانی فارسی (مطالعهٔ موردی مثنوی معنوی)

محمد امینی، پژوهشگر پسادکتری، دانشکدهٔ ادبیات و علوم انسانی و دانشکدهٔ مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان، ایران
m.amini@ltr.ui.ac.ir
احسان رئیسی*، دانشیار زبان و ادبیات فارسی، دانشکدهٔ ادبیات و علوم انسانی، دانشگاه اصفهان، اصفهان، ایران
e.reisi@ltr.ui.ac.ir
محمدعلی نعمت‌بخش، استاد دانشکدهٔ مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان، ایران
nematbakhsh@eng.ui.ac.ir

چکیده

جست‌وجوی مؤثر در آثار ادبی فارسی نقش مهمی در استخراج دانش‌ها و بینش‌های جدید از اندیشهٔ مفاخر ادبی ایران ایفا می‌کند؛ اما ادبیات کلاسیک فارسی، به‌ویژه آثار عرفانی همچون *مثنوی معنوی* مولانا، سرشار از لایه‌های استعاری و مفاهیم چندمعنایی است. این ویژگی موجب می‌شود که روش‌های متداول جست‌وجوی متنی در بازیابی دقیق و معنادار این متون ناتوان باشند؛ زیرا چنین الگوریتم‌هایی اساساً بر بسامد واژگان متکی‌اند و روابط پنهان و معنایی میان مفاهیم را در نظر نمی‌گیرند. در این پژوهش، رویکردی نوین برای جست‌وجوی معنایی در متون عرفانی فارسی ارائه شده که بر ترکیب روش‌های آماری کلاسیک و گراف‌های دانش مفهومی استوار است. بدین‌سان، ابتدا جست‌وجوی پایه مبتنی‌بر کلیدواژه صورت می‌گیرد و سپس مفاهیم کلیدی از ابیات بازیابی‌شده استخراج می‌شوند. این مفاهیم با بهره‌گیری از شبکهٔ معنایی واژگان ، هم‌رخدادی در متن، مترادف‌ها و ابرده‌ها بسط داده شدند و در قالب یک گراف مفهومی بازنمایی می‌شوند. جست‌وجوی نهایی با ترکیب یافته‌های جست‌وجوی پایه و مفاهیم بسط داده‌شده انجام می‌شود. آزمایش‌های انجام‌شده نشان می‌دهد که این روش نه‌تنها دقت جست‌وجو را افزایش می‌دهد، امکان تبیین‌پذیری و شفافیت بیشتری نیز برای کاربر فراهم می‌سازد؛ زیرا روابط مفهومی در قالب گراف به‌روشنی آشکار می‌شوند. کاربردهای این پژوهش فراتر از جست‌وجوی متنی بوده و می‌تواند در حوزه‌های علوم انسانی دیجیتال، پژوهش‌های ادبی و مطالعات عرفانی به کار گرفته شوند.

واژه‌های کلیدی: جست‌وجوی مفهومی، گراف دانش، مشابهت معنایی، مثنوی مولوی.

---

* مسؤول مکاتبات

امینی، محمد، رئیسی، احسان و نعمت‌بخش، محمدعلی . (١٤٠٤). در جستجوی معنا: بازیابی مفهومی مبتنی بر گراف دانش در متون عرفانی فارسی- مطالعه موردی مثنوی معنوی، *پژوهش‌های ادب عرفانی*، ١٩(١):١٨٧-٢٠٧.

2476-3292 © The Author(s).
This is an open access article under the CC BY-NC 4.0 License (https://creativecommons.org/licenses/by-nc/4.0

10.22108/jpll.2025.147019.1959

۱. مقدمه

متون عرفانی فارسی، به‌ویژه شاهکار سترگ مولانا در *مثنوی معنوی*، آکنده از صور خیال، تمثیل‌ها و لایه‌های معنایی پیچیده‌اند. پژوهشگرانی که به مطالعهٔ این متون می‌پردازند، همواره با چالش بازیابی و یافتن ابیاتی مواجه بوده‌اند که مفاهیم خاصی را بازتاب می‌دهند. روش‌های جست‌وجوی متنی رایج، مانند جست‌وجوی مبتنی‌بر کلیدواژه که بر بسامد واژگان و مدل‌های احتمالاتی استوارند، گرچه در متون خبری یا علمی کارآمد می‌نمایند؛ در برابر پیچیدگی‌های زبانی و معنایی ادبیات عرفانی فارسی کارایی اندکی دارند. در این متون، یک واژه ممکن است در معانی گوناگون (اعم از استعاری، تمثیلی و یا رمزی) به کار رود و ارتباطات مفهومی فراتر از سطح واژگانی پدید آیند. بدین‌ترتیب، نیاز به روش‌هایی که از سطح ظاهر کلمات فراتر روند و پیوندهای مفهومی میان ابیات را آشکار سازند، به‌شدت احساس می‌شود.

در عصر انفجار اطلاعات، چالش بنیادین سیستم‌های بازیابی اطلاعات، عبور از تطبیق سطحی واژگان و رسیدن به درک عمیق معناست؛ چالشی که به «شکاف معنایی»[1] معروف است. روش‌های مبتنی‌بر بسامد کلمات، واژگان را به‌مثابهٔ واحدهایی منفهم و مجزا در نظر می‌گیرند و از جهان پیچیدهٔ روابط معنایی، هم‌معنایی و وابستگی مفهومی غافل می‌مانند. گراف دانش معنایی، ساختاری قدرتمند است که این چالش را با مدل‌سازی دانش به‌صورت شبکه‌ای از مفاهیم (گره‌ها) و روابط میان آن‌ها (یال‌ها) پاسخ می‌دهد.

گراف دانش معنایی[2] به‌عنوان یک چهارچوب پیشرفته در حوزهٔ هوش مصنوعی و پردازش زبان طبیعی، نمایانگر تحولی اساسی در نحوهٔ سازماندهی و بازیابی دانش از پیکره‌های متنی عظیم است (Haslhofer et al., 2018). این ساختار گراف‌محور با سازماندهی اطلاعات در قالب موجودیت‌ها و روابط معنادار بین آن‌ها، نقش بسزایی در غلبه بر چالش‌های ابهام معنایی و بهبود دقت بازیابی اطلاعات در پیکره‌های متنی ایفا می‌کند. گراف دانش معنایی روابط معنایی بین موجودیت‌ها (مانند کلمات، عبارات و مفاهیم) را از پیکره‌های متنی استخراج و مدل‌سازی می‌کند و می‌تواند دقت جست‌وجو را ازطریق گسترش پویای پرس‌وجو و رفع ابهام معنایی بهبود بخشد. از این منظر، گراف دانش نه‌تنها به موتورهای جست‌وجو توانایی تشخیص و تمایز بین مفاهیم هم‌نویسه (مانند «تاج‌محل» به‌عنوان یک بنا یا یک موسیقی‌دان) را می‌دهد، با ایجاد ساختاری یکپارچه، امکان کاوش اطلاعاتی عمیق‌تر و اکتشاف روابط غیرمنتظره بین موجودیت‌ها را برای پژوهشگران فراهم می‌سازد.

پژوهش حاضر تلاشی است در مسیر پیوندزدن رویکردهای آماری کلاسیک با روش‌های معناشناسانهٔ نوین. رویکرد پیشنهادی، نقشه‌ای پویا از دانش دامنهٔ مد نظر ایجاد می‌کند و به سیستم جست‌وجو اجازه می‌دهد تا پرس‌وجوی کاربر را از سطح واژگان به عمق مفاهیم ارتقا دهد. به این منظور، برای تقویت روش جست‌وجو از گراف مفهومی استفاده می‌شود. این گراف که بی‌درنگ و پس‌از پرس‌وجوی کاربر ایجاد می‌شود، با استخراج مفاهیم مرتبط با یک بیت و ایجاد پیوند براساس قرابت معنایی با دیگر مفاهیم در ابیات دیگر می‌تواند ارتباط معنایی بین ابیات را ایجاد کند؛ درنتیجه، جست‌وجو صرفاً براساس کلیدواژه و ظاهر آنها صورت نمی‌گیرد، بلکه ابیاتی در پاسخ به پرس‌وجوی کاربر بازگردانده می‌شوند که دارای مفاهیم موجود در پرس‌وجو هستند و معنایی درخواست‌شده از کاربر را منعکس می‌کنند. با بسط‌دادن حوزهٔ جست‌وجو به شمول مترادف‌ها، ابرده‌ها و مفاهیم مرتبطی که در گراف به هم متصل‌اند، روش حاضر، گذار از جست‌وجوی واژگانی[3] به

---

1. Semantic Gap
2. Semantic Knowledge Graph
3. Lexical Search

جست‌وجوی مفهومی[1] را ممکن می‌سازد و درنتیجه، دقت و مرتبط‌بودن نتایج را به‌شکل چشمگیری بهبود می‌بخشد.

در روش پیشنهادی، نخست یک جست‌وجوی پایه با روش مرسوم مبتنی‌بر بسامد واژگان (BM25) (Robertson & Zaragoza, 2009) انجام می‌گیرد تا ابیات اولیهٔ مرتبط با پرسش کاربر به دست آید. سپس، با بهره‌گیری از ابزارهایی همچون شبکهٔ معنایی فارسی (فارس نت) (Mousavi & Faili, 2017)، استخراج روابط هم‌رخدادی، مترادف‌ها و ابرردهها، مجموعه‌ای از مفاهیم مرتبط با جست‌وجو گردآوری می‌شود. این مفاهیم در قالب یک گراف دانش سازمان می‌یابند و روابط میان آنها به‌صورت بصری و تحلیلی آشکار می‌شود. سپس، با پایش گراف دانش معنایی هرکدام از مفاهیم نیز در پیکرهٔ اشعار مثنوی جست‌وجو می‌شوند و به‌عنوان ابیات مرتبط بازگردانده می‌شوند. نتیجهٔ نهایی جست‌وجو، ترکیبی از نتایج جست‌وجوی مبنا و جست‌وجوی مفاهیم است که با استفاده از یک الگوریتم رتبه‌بندی ایجاد می‌شوند.

هدف اصلی این مقاله، ارائهٔ روشی نوین برای جست‌وجوی معنایی در متون عرفانی فارسی است که هم دقت بازیابی را افزایش می‌دهد و هم فرایند جست‌وجو را برای پژوهشگران علوم انسانی تبیین‌پذیر و معنادار می‌سازد. افزون‌بر این، چنین رویکردی می‌تواند در گسترش حوزهٔ علوم انسانی دیجیتال و درک ژرف‌تر از میراث عرفانی ایران اسلامی نقشی بنیادین ایفا کند. در ادامه، ابتدا مروری بر تحقیقات مرتبط با موضوع مقاله ارائه می‌شود و در بخش سوم، مدل پیشنهادی جست‌وجو تبیین خواهد شد. در بخش چهارم نیز ارزیابی روش جست‌وجو و تحلیل آنها ارائه می‌شود؛ درنهایت، مقاله با جمع‌بندی و نتیجه‌گیری به پایان می‌رسد.

۲. پیشینهٔ پژوهش

در این بخش، تحقیقات پیشین درزمینهٔ پژوهش حاضر ارائه می‌شود. این تحقیقات با درنظرگرفتن دو موضوع کلیدی بررسی می‌شوند: الف) گراف دانش معنایی؛ ب) جست‌وجوی معنایی در آثار فارسی.

۲ـ۱. گراف دانش معنایی

گراف‌های دانش معنایی در جست‌وجوی معنایی، پایه‌ای برای درک ارتباطات بین داده‌ها و بهبود بازیابی اطلاعات فراهم می‌کنند. این گراف‌ها با مدل‌سازی مفاهیم و روابط بین آن‌ها، جست‌وجو را از سطح کلیدواژه به سطح معنا ارتقا می‌دهند و در حوزه‌های مختلف کاربردهای موفقی داشته‌اند. شرکت گوگل در سال ۲۰۱۲م. گراف دانش خود را معرفی کرد تا نتایج جست‌وجو را فراتر از تطابق کلیدواژه‌ها برده و ارتباط بین اشیا، اشخاص و مکان‌ها را درک کند. استفاده از گراف باعث شد پاسخ‌های مستقیم، پیشنهادهای زمینه‌ای و خلاصهٔ اطلاعات مرتبط به کاربر نمایش داده شود (Peng et al., 2023).

در ادبیات جهانیِ ساخت گراف دانش و استخراج روابط معنایی، طی سال‌های اخیر چند گرایشِ قوی شکل گرفته است. مرور ادبیات پژوهش‌ها دربارهٔ گراف‌های دانش معنایی نشان می‌دهد که این ساختارها به‌منزلهٔ زیرساخت اصلی برای بهبود دقت و هوشمندی سیستم‌های جست‌وجوی معنایی شناخته می‌شوند. یافته‌های اخیر از مقالات مختلف چنین روندی را تأیید می‌کنند: سامانه‌های جست‌وجوی زیست‌پزشکی مانند ابزار Dug از گراف‌های دانش استفاده می‌کنند تا باتکیه‌بر هستان‌شناسی‌ها، بین داده‌های پژوهشی، بیماران و درمان‌ها پیوند برقرار کنند (Waldrop et al., 2022). این رویکرد جست‌وجوی معنایی را برای داده‌های پیچیده علمی ممکن می‌سازد. پلتفرم‌های تجارت الکترونیکی از گراف‌های دانش برای

[1]. Conceptual Search

تحلیل روابط بین محصولات، برندها و ترجیحات کاربران بهره می‌برند (Kiff, n.d.). این گراف‌ها به بهبود جست‌وجو، رتبه‌بندی نتایج و پیشنهادات هوشمند کمک می‌کنند. در یکی از تحقیقات اخیر، ساخت گراف دانش معنایی متناسب با پاسخ‌های علّی، با تمرکز بر ویژگی‌های جست‌وجوی معنایی عمیق عصبی بررسی شده است (Bektemyssova & Sabdenov, 2024). در این مقاله، روش‌های تحلیل سیستم برای ساخت گراف‌های دانش، ضمن تحلیل جنبه‌های جست‌وجوی معنایی عمیق عصبی بررسی شده است. نتایج این تحقیق بینش‌هایی را دربارهٔ ساخت یک مدل جست‌وجوی گراف دانش معنایی برای پاسخ‌های علّی ارائه می‌دهد و به چالش‌ها و روش‌های مربوط به ساخت چنین مدلی می‌پردازد. این امر بر اهمیت گراف‌های دانش در سیستم‌های اطلاعاتی مدرن و کاربردهای بالقوه آنها در حوزه‌های مختلف تأکید می‌کند. استفاده از گراف دانش همچنین در دسترسی آزاد به منابع علمی و تحقیقاتی نقش مهمی ایفا کرده است. آخرین نسخه از گراف دانش علوم رایانه (CS-KG 2.0)، یک پایگاه دانش گستردهٔ تولیدشده از ۱۵ میلیون مقالهٔ تحقیقاتی را ارائه می‌دهد (Dessí et al., 2025). این گراف، ۲۵ میلیون موجودیت مرتبط با ۶۷ میلیون رابطه را توصیف می‌کند و نمایش ظریفی از دانش علمی در حوزهٔ علوم کامپیوتر را عرضه می‌کند.

مطالعات در منابع ایرانی نشان می‌دهند که ادغام دانش واژگانی در فرایند گسترش کوئری یا بازنویسی پرس‌وجو می‌تواند اثر محسوسی بر کیفیت بازیابی داشته باشد، به‌ویژه در متونی که اصطلاحات یا وارونگی‌های معنایی دارند. پژوهش‌های خطی بر متون خبری/ علمی تفاوت‌هایی با متون ادبی دارند؛ فضای استعاری و تمثیلی متون عرفانی، نیازمند تلفیق دقیق‌تر بین سطح و معناست. استفاده از فارس‌نت به‌عنوان یک منبع ساخت‌یافته، هنگامی که با روش‌های استخراج از متن نیز ترکیب شود، امکان پیداکردن روابط ضمنی را افزایش می‌دهد. توسعهٔ واژه‌نامک‌ها و WordNet فارسی از سال‌ها پیش در دستور کار پژوهشگران ایرانی قرار گرفته است. پروژهٔ فارس‌نت، که هدف آن ایجاد یک WordNet فارسی و نگاشت آن به پرینستون وردنت است، از شبکه‌های نیمه‌خودکار و انسانی بهره برده و در نگارش‌ها و گزارش‌های فنی آن جزئیات دربارهٔ ساختار مجموعهٔ مترادف‌ها و نگاشت‌ها ارائه شده است(Shamsfard et al., 2010)؛ این منبع به‌عنوان ستون فقرات بسیاری از پژوهش‌های بعدی در زبان فارسی عمل کرده است.

دو رستهٔ روشِ برنامه‌محور برای ساخت Persian WordNet (خودکار/نیمه‌خودکار) در ادبیات فارسی بارها پیشنهاد شده‌اند: روش‌های بدون ناظر [1] که بر شباهت توزیعی یا معیارهای هم‌آیندی مبتنی‌اند (Montazery & Faili, 2011) و روش‌های نظارت‌شده [2] که از نگاشت‌های دوزبانه و مدل‌های طبقه‌بندی برای تشخیص نگاشت‌ها استفاده می‌کنند؛ نمونهٔ اخیرِ یک روش نظارت‌شده، که نتایج با دقت بالا تولید کرده، در پژوهش موسوی و فیلی (۲۰۱۷) مشاهده شده است (Mousavi & Faili, 2017).

الگوهای واژگانی ـ نحویِ Hearst (مانند «مانند»، «از قبیل»، «نظیر») از قدیمی‌ترین و معتبرترین روش‌ها برای کشف رابطهٔ ترادف/ابردگی [3] هستند؛ اگرچه دقت اولیهٔ آن‌ها بالا است، فراخوانی کامل روابط [4] را به‌تنهایی تأمین نمی‌کنند و معمولاً نیاز به توسعهٔ الگوها یا استفاده از ساختارهای وابستگی نحوی دارند (Hearst, 1992; Roller et al., 2018). در مقابل،

---

1. Unsupervised
2. Supervised
3. hyponymy/hypernymy
4. Recall

روش‌های توزیعی مبتنی‌بر بردارهای واژه یا عبارات مانند Word2Vec, fastText و مدل‌های مبتنی‌بر زمینه[1] مانند BERT و مدل تنظیم‌شدهٔ فارسی آن ParsBERT توانایی شناسایی واژگان «هم‌زمینه» یا «نزدیک معنایی» را دارند و می‌توانند به‌منزلهٔ منبع تولید کاندیدای مترادف یا هم‌زمینه به کار روند؛ بااین‌حال، تمایز میان مترادف واقعی و تنها «مرتبط» نیازمند پالایش‌های اضافی است (Zhong et al., 2023).

به‌طورکلی، ساختن گراف دانش معنایی در سه مرحلهٔ کلی انجام می‌پذیرد: (۱) کسب دانش (استخراج موجودیت‌ها و روابط)، (۲) پالایش و یکپارچه‌سازی[2] و (۳) تکمیل و تحول گراف.

در سطح روش‌شناختی، رویکرد مؤثر برای زبان‌های کم‌منبع، ترکیب چندمنبعی است: الگوهای دست[3]، استخراج توزیعی[4] و یادگیری نظارت‌شده یا دور ـ نظارتی[5] که از نگاشت‌های موجود به‌مثابه برچسب‌گذار خودکار بهره می‌برند. مروری جامع که بیش از ۳۰۰ روش را دسته‌بندی کرده و تفاوت‌های معماریِ موجود را توضیح می‌دهد، در تحقیق ژائو و همکاران ارائه شده که می‌تواند راهنمای طراحی سیستم‌های جدید باشد (Zhao et al., 2023).

۲ـ۲. جست‌وجوی معنایی در آثار فارسی

جست‌وجوی مفهومی و بازیابی اطلاعات معنایی، پارادایم‌های مکملی را ترکیب می‌کنند که در آن، پوشش واژگانی فدای دقت معنایی می‌شود. سیستم‌ها یا متن را با استفاده از منابع دانش ساختاریافته به فضاهای مفهومی ارتقا می‌دهند یا گراف‌های مفهومی ضمنی را از الگوهای آماری هم‌رخدادی می‌سازند (Egozi et al., 2011; Giunchiglia et al., 2009). این روش‌ها اصولاً از ۴ رویکرد مشخص استفاده کرده‌اند:

- نگاشت مفهومی[6]: تبدیل واژگان به مفاهیم با استفاده از منابع واژگانی یا ابهام‌زدایی از معنای واژگان، پیش از نمایه‌سازی یا تطبیق، که به کاهش خطاهای مترادفی و چندمعنایی منجر می‌شود (Manjula et al., 2003).
- ویژگی‌های دانش صریح[7]: نمایش اسناد و پرس‌وجوها به‌صورت بردارهایی برحسب مفاهیم دانش بشری (مانند مفاهیم ویکی‌پدیا و مجموعه‌های مترادف واژگان) به‌جای واژگان خام، همان‌طور که در تحلیل معنایی صریح (ESA) پیاده‌سازی شده است (Egozi et al., 2011; Manjula et al., 2003).
- رویکردهای مبتنی‌بر گراف: ساخت گراف‌های مفهومی از روی الگوهای هم‌رخدادی یا ساختارهای پیوندی ویکی‌پدیا و استفاده از فاصله‌های گرافی برای تطبیق پرس‌وجو با سند (Chen & Schatz, 1994).
- تطبیق مبتنی‌بر هستان‌شناسی[8]: استفاده از هستان‌شناسی حوزه، سه‌تایی‌های RDF و استنتاج برای درک بافتار مفهومی و روابط دامنه به‌منظور دستیابی به بازیابی با محوریت دقت (Wen et al., 2006).

به‌منظور پیاده‌سازی جست‌وجوی معنایی در ادبیات این حوزه، از روش‌ها و الگوریتم‌های خاصی استفاده شده است که مهم‌ترین آنها به شرح زیر هستند:

---

1. Contextual
2. Entity linking, canonicalization
3. Pattern-based
4. Embedding-based
5. Distant supervision
6. Concept Mapping
7. Explicit Knowledge Features
8. Ontology-Based Matching

- تحلیل معنایی صریح[1]: این روش با تبدیل متن به بردارهای مفهومی با ابعاد زیاد در ویکی‌پدیا، شباهت را در فضای مفهومی محاسبه می‌کند. بهره‌گیری از گزینش ویژگی و گزینش نظارت‌شده، به افزایش تمرکز و بهبود عملکرد رتبه‌بندی در وظایف مشخص‌شده در کنفرانس بازیابی اطلاعات متنی[2] منجر می‌شود و پیشرفت‌های پایداری نسبت به روش‌های مبنا ایجاد می‌کند (Egozi et al., 2011).
- مدل‌های بازتابی مبتنی‌بر مفهوم: این روش‌ها توزیع‌های تناسب را در فضای مفهومی تخمین می‌زنند، هم‌زمان وابستگی‌های واژگانی درون‌مفهومی را حفظ کرده و گونه‌های سطحی[3] را یکپارچه می‌سازند. کاربردهای این روش‌ها در بازیابی اطلاعات پزشکی، بهبودهای ۹ تا ۲۰ درصدی را نسبت به مدل‌های مبتنی‌بر واژگان نشان می‌دهند (Wang & Akella, 2015).
- توسعهٔ پرس‌وجو بر پایهٔ گراف: در این روش با به‌کارگیری فاصله‌های تناسب یا مرتبط‌بودن معنایی، پرس‌وجوها را روی گراف‌های مفهومی غیرجهت‌دار توسعه می‌دهد تا نرخ فراخوانی (Recall) افزایش یافته و رانش معنایی[4] نیز کنترل شود(Boubacar & Niu, 2013) .
- ابهام‌زدایی معنای واژگان و نمایه‌سازی مفهومی: ابتدا با ابهام‌زدایی از معنای واژگان[5]، واژگان به شبکه‌ای از معانی[6] نگاشت می‌شوند. سپس، این شبکه‌های مرتبط در قالب مفاهیم، گروه‌بندی و نمایه‌سازی می‌شوند. در این فرایند، از وزن‌دهی مفهومی به‌جای فراوانی اولیهٔ واژگان استفاده می‌شود (Boubacar & Niu, 2013; Manjula et al., 2003).

مسیر تکامل پردازش زبان طبیعی فارسی، از ابزارهای جزیره‌ای و قانون‌محور به‌سوی جعبه‌ابزارهای یکپارچهٔ عصبی، پیکره‌های عظیم و منتخب و مدل‌های زبان بزرگ بومی، که زمینه‌ساز کاربردهای معنایی و سیستم‌های مکمل‌یافته با بازیابی هستند، طی شده است. این پیشرفت، چالش‌های ویژهٔ پردازش زبان فارسی را مد نظر قرار می‌دهد؛ چالش‌هایی نظیر ساختار ریخت‌شناسی پیچیده، نگارش راست‌به‌چپ، و محدودیت منابع داده‌ای حاشیه‌نویسی‌شده در مقایسه با زبان‌های دارای منابع غنی. در حوزهٔ پردازش زبان طبیعی و جست‌وجوی معنایی در زبان فارسی کارهای مختلفی برای پردازش مفهومی و هستان‌شناسی انجام شده است:

- هستان‌شناسی‌های فارسی: هستان‌شناسی FNLP-ONT ، که به‌طور خاص برای وظایف پردازش زبان طبیعی فارسی طراحی شده است، ویژگی‌های معنایی را در بر می‌گیرد که هنگام ترکیب با مدل‌های یادگیری ماشین روی پیکره‌های فارسی، دقت را به میزان ۴٫۵٪ تا ۲۳٪ بهبود می‌بخشد؛ این بهبود، به‌ویژه برای وظایفی نظیر ابهام‌زدایی از معنای واژگان و شناسایی فعل مرکب، چشمگیر است (Hosseini Pozveh et al., 2018).
- استخراج روابط: سیستم‌هایی که روش‌های مبتنی‌بر الگو، ساختاری و آماری را ترکیب می‌کنند، روابط مفهومی طبقه‌بندی‌شده[7] و غیرطبقه‌بندی‌شده[8] را از متن فارسی و ساختارهای ویکی‌پدیا استخراج می‌کنند و تکمیل خودکار گراف

---

1. Explicit Semantic Analysis
2. Text REtrieval Conference (TREC)
3. Surface Variant
4. Semantic Drift
5. Word Sense Disambiguation
6. Synsets
7. (Taxonomic)
8. Non-Taxonomic

دانش از منابع فارسی را ممکن می‌سازند (Fadaei & Shamsfard, 2010).

- پایگاه‌های دادهٔ واژگانی PersPred: با فراهم‌آوردن یک پایگاه دادهٔ نحوی و معنایی دستی پالایش‌شده برای گزاره‌های پیچیدهٔ فارسی (حدود ۷۰۰ مدخل)، حوزه‌های واژگانی، نحوی و معنایی را کدگذاری می‌کند. این پایگاه برای درک روابط اصطلاحی در پایگاه‌های دانش و بهبود جست‌وجوی مفهومی برای گزاره‌های چندواژی حیاتی است (Samvelian & Faghiri, 2013).
- پیش‌پردازش معنایی Perspell: یک سیستم تصحیح املای آگاه به معنا، به میزان تقریبی ۹۵٪ در تشخیص خطای واژگان واقعی[1] موفق عمل می‌کند و نویز را در خط لوله‌های پایین‌دستی گراف دانش و جست‌وجوی معنایی کاهش می‌دهد (Dastgheib et al., 2017).

تاکنون پژوهش‌های اندکی درزمینهٔ جست‌وجوی معنایی در آثار ادبی کلاسیک فارسی انجام شده است. سیستم‌های جست‌وجو در شعر کلاسیک در دنیا از استراتژی‌های بازیابی متنوعی بهره می‌برند که به نیازهای پژوهشی گوناگون و محدودیت‌های فنی پاسخ می‌دهند. ازجملهٔ این استراتژی‌ها می‌توان به موارد زیر اشاره کرد:

الف) جست‌وجوی متن کامل و زبان‌شناختی: جست‌وجوی استاندارد روی پیکره‌های دیجیتالی‌شده (مانند کتابخانهٔ دیجیتال پرسئوس) با بهره‌گیری از پیش‌پردازش زبان‌شناختی برای زبان‌های کلاسیک صرفی، امکان جست‌وجوی ریشه (Lemma) و گونه‌های واژگانی و همچنین، ارجاعات استاندارد را فراهم می‌آورد (Smith et al., 2000). این سیستم‌ها پایه و اساس بیشتر سیستم‌های بازیابی متن کلاسیک را تشکیل می‌دهند.

ب) روش‌های بازیابی خاص: این روشها باتوجه‌به شرایط و ایده‌های مرتبط با زبان و متن انجام می‌شوند. جست‌وجوی مبتنی‌بر بیت[2] سیستم‌هایی هستند که اولین بیت‌های ناقص را می‌پذیرند، موارد وقوع و متون ثانویه را ازطریق نمایه‌سازی مفهومی به‌جای تطبیق تحت‌اللفظی بازمی‌گردانند (Sanyal, 2018). در روش تشخیص بازاستفادهٔ بینامتنی[3]، موتورهای مقایسه‌ای دوتایی مانندTesserae ، با شناسایی واژگان مشترک، به کشف الگوهای تل میح و بازاستفاده در میان متون می‌پردازند(Bernstein & Gervais, 2015) . روش جست‌وجوی الگوی وزنی پیکره‌های مبتنی‌بر TEI که ازنظر آوای حاشیه‌نویسی[4] شده‌اند، امکان پرس‌وجوی الگوهای آهنگی و ساختارهای سونت را فراهم می‌آورند (Candela et al., 2017).

۲ـ۳. خلأها و انگیزهٔ پژوهش حاضر

بررسی پژوهش‌های پیشین نشان می‌دهد که تحقیقات مرتبط با جست‌وجوی مفهومی در حوزهٔ زبان و ادب فارسی بسیار محدود بوده و تاکنون پژوهش درخور توجهی درزمینهٔ جست‌وجوی مفهومی با به‌کارگیری روش‌های نوین درزمینهٔ آثار شعر فارسی انجام نشده است؛ گرچه منابعی همچون FarsNet بسیاری از واژگان فارسی را پوشش داده‌اند، به‌کارگیری آنها در متون ادبیِ کلاسیک نیازمند انطباق‌ها و گسترش‌هایی است:

الف) مفاهیم استعاری که در متون عرفانی پدیدار می‌شوند لزوماً در منابع واژگانی عمومی نیامده‌اند،

ب) ساخت گراف‌های مفهومی پویا که مختص پرسش کاربر تولید شوند (و نه صرفاً یک گراف ایستا از کلِ مجموعه)

---

1. Real-Word Error
2. Line-oriented
3. Intertextual Reuse
4. Prosody

هنوز به‌طور کامل بررسی نشده‌اند.

پژوهش حاضر می‌کوشد این خلأ را با ارائهٔ چهارچوبی ترکیبی ازطریق تلفیق روش جست‌وجوی مبنای کلیدواژهٔ BM25 با گراف‌های مفهومی پویا و منابع واژگانی (FarsNet) پُر کند؛ درواقع، روش پیشنهادی این پژوهش مبتنی‌بر توسعهٔ پرس‌وجو بر پایهٔ گراف است که پرس وجوی کاربر را با استفاده از ایجاد گراف مفاهیم مرتبط در هنگام اجرای جست‌وجو گسترش داده و درنتیجه با افزایش میزان تناسب معنایی نتایج، پوشش حداکثری آنها را موجب می‌شود. افزون‌بر این، تأکید ویژه‌ای بر تبیین‌پذیری[1] نتایج جست‌وجو و ارائهٔ نمای گرافی از چرایی بازگشت یک بیت گنجانده شده است.

۳. روش پیشنهادی

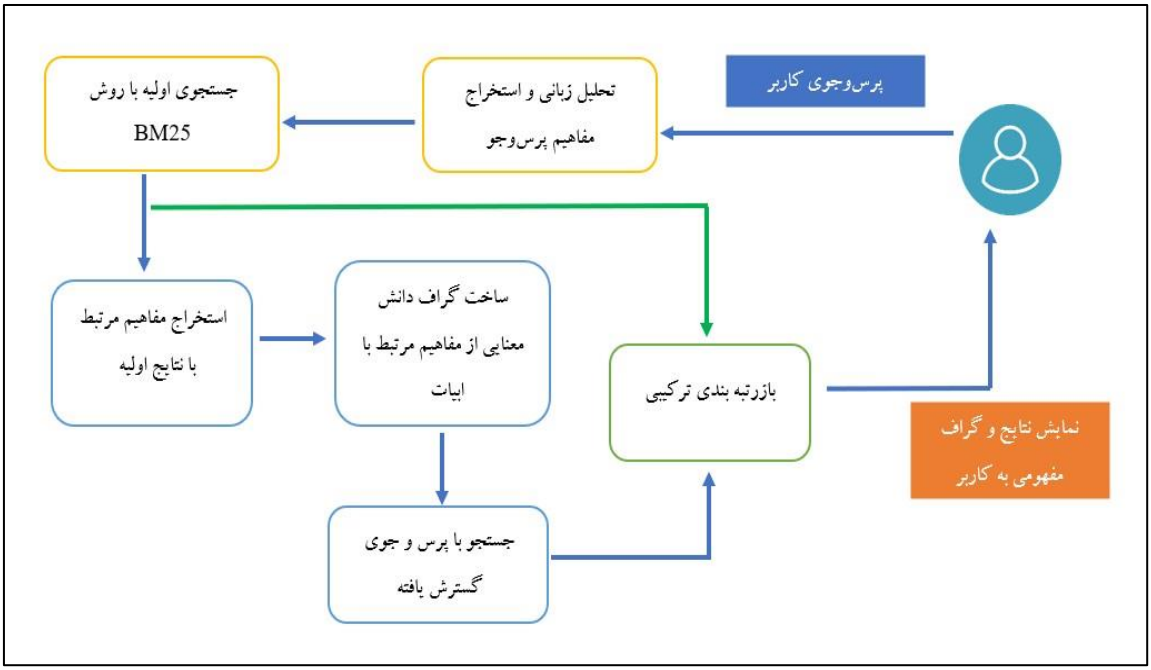
۳ـ۱. فرایند کلی جست‌وجو

روش پیشنهادی جست‌وجو از دو مؤلفهٔ اصلی تشکیل شده است:

۱. هستهٔ جست‌وجو[2]: این بخش یک روش جست‌وجوی مبنا را انجام می دهد که بر پایهٔ Lucene/BM25 عمل می‌کند و مسئول بازیابی اسناد اولیه است.

۲. لایهٔ معنایی[3]: این بخش از گراف دانش مفهومی ساخته‌شده از مثنوی معنوی و شبکهٔ واژگان فارس‌نت برای گسترش مفاهیم، بسط پرس‌وجوی کاربر و بازرتبه‌بندی نتایج بهره می‌گیرد.

فرایند کلی روش جست‌وجو در قالب نمودار شکل ۱ توصیف می‌شود:



شکل ۱. نمودار روش جست‌وجوی مفهومی در اثر مثنوی مولانا

[1]. Explainability
[2]. Retrieval Core
[3]. Semantic Layer

روش BM25 یکی از مدل‌های کلاسیک بازیابی اطلاعات مبتنی‌بر احتمال و فراوانی واژگان است که میزان ارتباط یک سند با پرس‌وجوی کاربر را براساس بسامد واژه‌ها و میزان پراکندگی آن‌ها در مجموعه اسناد محاسبه می‌کند. در این مدل، وزن هر واژه با درنظرگرفتن بسامد در سند (TF)، فراوانی معکوس در کل اسناد (IDF) و طول نرمال‌شدهٔ سند تعیین می‌شود تا از تأثیر اسناد بسیار بلند یا کوتاه جلوگیری شود. به بیان دیگر، BM25 به‌جای صرف تطابق سطحی واژه‌ها، با وزن‌دهی آماری، احتمال مرتبط‌بودن هر سند با پرس‌وجوی کاربر را برآورد می‌کند و بدین‌ترتیب مبنایی کارآمد برای رتبه‌بندی نتایج جست‌وجو فراهم می‌سازد.

روش BM25 باوجود کارایی بالا در بازیابی متون معاصر و اطلاعات ساخت‌یافته، در مواجهه با متون کلاسیک و استعاری، به‌ویژه متون عرفانی فارسی مانند *مثنوی معنوی*، با محدودیت‌های جدی روبه‌رو است. این مدل صرفاً بر پایهٔ تطابق آماری واژه‌ها عمل می‌کند و از درک مفاهیم پنهان، روابط استعاری و چندمعنایی واژگان ناتوان است. به همین دلیل، در پژوهش حاضر، لایه‌ای معنایی مبتنی‌بر گراف دانش مفهومی بر مدل آماری BM25 افزوده شده است تا با بهره‌گیری از روابطی همچون ترادف، ابرده‌گی، و هم‌رخدادی مفاهیم، جست‌وجویی ژرف‌تر و معنادارتر ممکن شود؛ بدین‌ترتیب، جست‌وجو نه بر سطح واژه، بلکه بر شبکهٔ مفاهیم و پیوندهای معنایی میان آن‌ها استوار می‌شود.

۳ـ۲. روش ساخت گراف دانش معنایی

در این بخش، جزئیات ساخت، سازمان‌دهی و وزن‌دهی به گراف دانش مفهومی برای مثنوی ارائه می‌شود. هدف این بخش ایجاد مدلی ساختاری است که بتواند روابط معنایی عمیق بین مفاهیم عرفانی در *مثنوی معنوی* را بازنمایی کرده و از آن برای جست‌وجوی معنایی و استنتاج مفهومی بهره گیرد.

۳ـ۲ـ۱. هدف و منطق نظری

متون عرفانی فارسی، ازجمله *مثنوی معنوی*، دارای شبکه‌ای غنی از مفاهیم هم‌بسته‌اند که اغلب در قالب نماد، تمثیل و ارجاع متقابل بیان می‌شوند. برای درک این شبکه، لازم است بین مفاهیم نه‌تنها روابط واژگانی (مانند مترادفی)، روابط مفهومی و استعاری نیز شناسایی شود. گراف دانش به‌عنوان یک مدل شبکه‌ای از دانش، این امکان را فراهم می‌کند تا:

- هم‌وابستگی مفاهیم براساس کاربرد در متن شناسایی شود.
- پیوندهای واژگانی از FarsNet یا منابع مشابه افزوده شود.
- با افزودن اطلاعات توزیعی از مدل‌های زبانی مدرن، پیوندهای معنایی ضمنی نیز آشکار شود.

۳ـ۲ـ۲. ساختار کلی گراف

گراف مفهومی پیشنهادی یک گراف چندلایه (multi-layer graph) است که در آن هر نوع رابطه در لایه‌ای مجزا مدل می‌شود. در این گراف سه لایه وجود دارد: لایهٔ هم‌رخدادی، لایهٔ واژگانی و لایهٔ توزیعی. جدول (۱) لایه‌های مختلف این گراف و عملکرد هرکدام را نشان می‌دهد. هر گره (Node) نشان‌دهندهٔ یک مفهوم یا واژهٔ کلیدی است و هر یال (Edge) حامل نوع و وزن رابطه میان آن‌هاست.

جدول ۱. ساختار گراف دانش مفهومی برای بسط جست‌وجو

لایه	نوع رابطه	منبع داده	نمونه رابطه در متن مثنوی
لایهٔ هم‌رخدادی(Co-occurrence Layer)	هم‌ظهوری مفاهیم در یک بیت یا مصرع	تحلیل درون‌متنی	«نور» ↔ «دل» در بیت‌های مربوط به تجلی
لایهٔ واژگانی(Lexical Layer)	مترادف، ابرراده، جزء-کل	FarsNet، WordNet فارسی	«عقل» ↔ «خرد»، «خدا» ↔ «معبود»
لایهٔ توزیعی(Distributional Layer)	شباهت معنایی در فضای embedding	ParsBERT، fastText	«عشق» ↔ «محبت»، «نور» ↔ «روشنایی»

برای ساخت گراف معنایی از هر پرس‌وجوی کاربر، دو ساختار داده‌ای مبنا از قبل تشکیل می‌شود تا هنگام پرس‌وجو گراف پویا با استفاده از آنها ساخته شود:

الف) مجموعه‌مفاهیم مرتبط با هر بیت از مثنوی (نگاشت بیت به مفهوم): هر بیت و مفاهیم مرتبط با آن در یک رکورد ذخیره می‌شوند. بنابراین، کل این مجموعه دارای ۲۵،۶۳۶ رکورد است که برابر با تعداد ابیات مثنوی مولوی است. برای ساخت این مجموعه، ابتدا کلمات و واژه‌های پرتکرار به‌صورت unigram و bigram استخراج شدند و سپس اسم‌های مفهومی با PMI و فراوانی بالا گزینش شدند. درنهایت، این مفاهیم استخراج‌شده با مفاهیم حاصل از یک لیست ازپیش‌تعیین شده توسط قطب علمی تحقیق در متون حکمی و عرفانی دانشگاه اصفهان ادغام شد. این لیست مفاهیم برای دفتر اول و براساس نظر خبرگان حوزهٔ ادبیات عرفانی تدوین شده بود.

ب) مجموعه ابیاتی که دارندهٔ هر مفهوم هستند (نگاشت مفهوم به بیت): هر مفهوم و ابیاتی که دارای آن مفهوم هستند به روش ایندکس معکوس ذخیره می‌شوند تا هم برای پردازش الگوریتمی و هم برای نمایش بصری استفاده شوند. برای این کار، ساختار نگاشت (الف) را پردازش کرده و از آن یک ایندکس معکوس ایجاد می‌شود. تعداد رکوردهای این مجموعه با تعداد مفاهیم استخراج‌شده از ابیات متناسب است. برای پژوهش فعلی، باتوجه‌به روش‌های استخراج مفاهیم، ۱۵۴۴ مفهوم استخراج شد و درنتیجه، این ساختار نگاشت دارای ۱۵۴۴ رکورد است.

جدول (۲) یک نمونه از ساختار نگاشت بخش (الف) را نشان می دهد که در آن مفاهیم مرتبط با هر بیت ذخیره می‌شوند. در جدول (۳) نیز ساختار نگاشت مفهوم به بیت را برای نمایش ابیاتی که هر مفهوم را در خود دارند، نشان می‌دهد.

جدول ۲. نگاشت بیت به مفهوم

بیت		مفاهیم
بشنو از نی چون حکایت می‌کند	از جدایی‌ها شکایت می‌کند	حکایت، شکایت، جدایی، نی
بهر آنست امتحان نیک و بد	تا بجوشد بر سر آرد زر زبد	امتحان، زر، بدی، نیکی
هم ز بیم معجزات انبیا	سر کشیده منکران زیر گیا	نبوت، بیم، معجزه
شه که او از حرص قصد هر حرام	می‌کند او را گدا گوید همام	شاه، حرص
روح را توحید الله خوشترست	غیر ظاهر دست و پای دیگرست	الله، توحید، دست

افزون‌بر دو ساختار نگاشتی معرفی‌شده، با استفاده از شبکه معنایی فارس‌نت، دو ساختار داده‌ای برای رابطهٔ ترادف (هم معنی‌ها)[1] و نیز ابرردگی[2] از میان ابیات مثنوی ساخته شدند. درواقع، یک بخش از شبکهٔ فارس‌نت که دارای مجموعه‌کلمات هم‌معنی است و بخش دیگر که شامل روابط ابرردگی است، برای ایجاد مفاهیم مرتبط از مفاهیم اولیهٔ مثنوی استفاده می‌شوند. به‌این‌ترتیب، دو فایل مبنا ایجاد می‌شود که در آن مجموعه ترادف و ابرردگی برای مفاهیم استخراج‌شده از *مثنوی* به دست می‌آید و این کار گسترش معنایی مفاهیم را در پی دارد. مفاهیم مستخرج از این دو مجموعه برای گسترش پرس‌وجو به کار می‌روند.

جدول ۳. نگاشت مفهوم به بیت

مفهوم	ابیات مرتبط
حکایت	بشنو از نی چون حکایت می‌کند از جدایی‌ها شکایت می‌کند این سخن پایان ندارد ای غلام روز بیگه شد حکایت کن تمام در حکایت گفته‌ایم احسان شاه در حق آن بی‌نوای بی‌پناه این حکایت گفته شد زیر و زبر / همچو فکر عاشقان بی پا و سر ...
زمان	این جهان ویران شدی اندر زمان حرصها بیرون شدی از مردمان غیر آن قطب زمان دیده‌ور کز ثباتش کوه گردد خیره‌سر نه که قلب و قالبم در حکم اوست لحظه‌ای مغزم کند یک‌لحظه پوست ...
هستی	اندرونی کاندرونها مست از اوست نیستی کین هستهامان هست از اوست تو تبرا کن که هستت دستگاه ای تبرای ترا جان عذرخواه زانک آلت دعوی است و هستی است کار در بی‌آلتی و پستی است ....

با استفاده از مجموعه‌داده‌های مفهومی که پیش‌تر نیز ذکر شد، پس‌از دریافت پرس‌وجوی کاربر، یک پرس‌وجوی اولیه با روش مبتنی‌بر کلیدواژه یا BM25 اجرا می‌شود و ابیات استخراج‌شده برای ساخت گراف دانش پویا به‌صورت زیر استفاده می‌شوند:

۱- همهٔ مفاهیم مرتبط با ابیات استخراج‌شده، ابتدا ازطریق نگاشت «بیت به مفهوم» استخراج می‌شوند و در یک مخزن قرار می‌گیرند. افزون‌بر این، مفاهیم مرتبط با هر مفهوم نیز از دو ساختار ترداف و ابرردگی که از *مثنوی* و با استفاده از فارس‌نت استخراج شده‌اند، به دست می‌آیند و به این مخزن افزوده می‌شوند. همچنین، با استفاده از شباهت توزیعی مبتنی‌بر مدل‌های جانمایی fasttext، کلماتی که بیشترین شباهت را با مفاهیم مستخرج دارند، از مثنوی استخراج شده و در همین مخزن ذخیره می‌شوند.

۲- در گراف دانش معنایی برای هر مفهوم یک گره اختصاص می‌یابد. مفاهیم مرتبط با استفاده از یال به یکدیگر متصل می‌شوند. به هر یال یک وزن اختصاص می‌یابد که نشان‌دهندهٔ میزان قوت آن ارتباط است. وزن یال ترکیبی است از رابطهٔ هم‌رخدادی، ترادف و ابرردگی و نیز شباهت توزیعی.

---

[1]. sysnonyms
[2]. hypernyms

بنابراین، اگر وزن بین دو مفهوم i و j باشد؛ آنگاه این وزن ازطریق زیر محاسبه می‌شود:

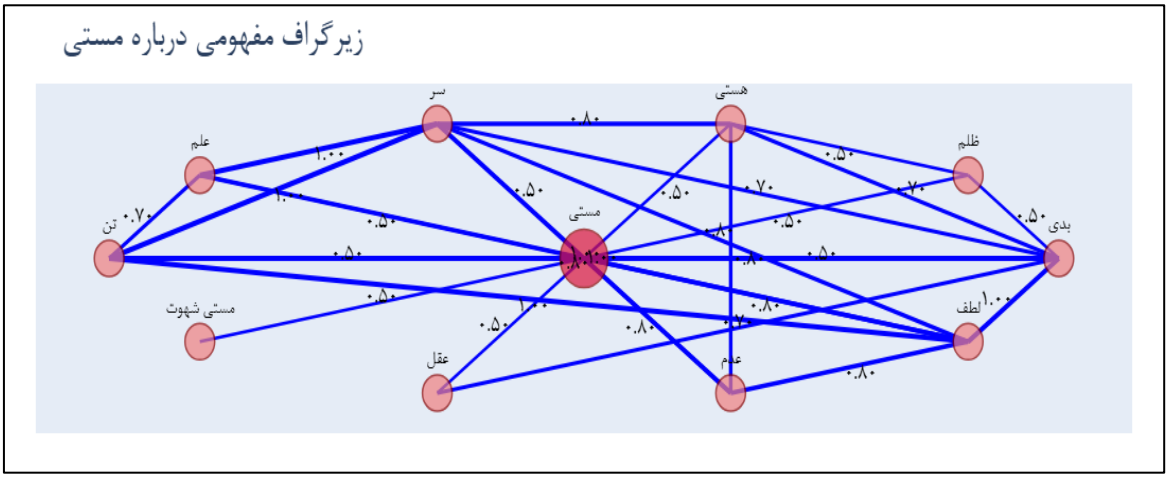
$$W_{ij} = \alpha.C_{ij} + \beta.S_{ij} + \gamma.H_{ij} + \delta.D_{ij} \quad (۱)$$

که در آن :

- C_{ij} وزن نرمال‌شدهٔ هم‌رخدادی در ابیات مشترک
- S_{ij} وجود رابطهٔ ترادف (۰ یا ۱)
- H_{ij} وجود رابطهٔ ابرده (۰ یا ۱)
- D_{ij} فاصلهٔ توزیعی بین دو مفهوم
- α، β، γ، δ ضرایب وزنی که می‌توان تنظیم کرد (در این پژوهش به‌ترتیب مقادیر ۰٫۴، ۰٫۲ ، ۰٫۱ ، ۰٫۳)

۳- گراف تشکیل‌شده پالایش می‌شود تا گره‌هایی که یال‌هایی با وزن کمتر از یک آستانه خاص دارند حذف شوند. به این طریق، هم گراف کوچک‌تر می‌شود و پردازش آن سبک‌تر خواهد بود و هم دامنهٔ مفاهیم مرتبط غنی‌تر می‌شود و به جست‌وجوی معنایی بهتری منجر خواهد شد.

با استفاده از گام‌های بیان‌شده، یک گراف دانش معنایی از مفاهیم مرتبط با پرس‌وجوی کاربر به‌صورت پویا و بی‌درنگ در زمان جست‌وجو ایجاد می‌شود که در مرحلهٔ بعد برای بهبود جست‌وجو به کار خواهد رفت. شکل (۲) نمونهٔ یک گراف دانش مفهومی را برای یک مفهوم پرسش‌شده، یعنی «مستی» نشان می‌دهد. در این شکل، هر یال نشانگر یک ارتباط مفهومی بین یک مفهوم با مفهوم مرکزی «مستی» است. همچنین، وزن هر یال بیانگر قوت آن ارتباط است که از رابطهٔ (۱) محاسبه شده است. هرچه پیوند مفهومی حاصل قوی‌تر باشد و از عناصر بیشتری طبق رابطهٔ (۱) بهره گیرد، آن پیوند وزن بیشتری خواهد داشت و ضخامت یال آن نیز بیشتر است.



شکل ۲. نمونهٔ یک گراف دانش مفهومی برای واژهٔ «مستی» در مثنوی مولوی

۳ـ۳. جست‌وجوی بهبودیافته

پس‌از استخراج گراف دانش معنایی از مفاهیم مرتبط با پرس‌وجو، پژوهشگران می‌توانند این مفاهیم را برای گسترش پرس‌وجو به کار گیرند و جست‌وجوی جدیدی را انجام دهند. به این منظور، جست‌وجوی جدید برای هر مفهوم در گراف با پیمایش آن صورت می‌گیرد؛ درنتیجه، ابیات مرتبط با هر مفهوم پیمایش‌شده از گراف دانش استخراج می‌شود. بنابراین،

اکنون فهرستی گسترش‌یافته از ابیاتی که با مفاهیم مستخرج از پرس‌وجوی کاربر مرتبط هستند، به دست می‌آید. بسیاری از این ابیات ممکن است کلمات موجود در پرس‌وجوی اولیهٔ کاربر را در خود نداشته باشند؛ اما به‌سبب استفاده از گراف دانش معنایی، دارای معنای مرتبط هستند.

پس‌از به‌دست‌آمدن فهرست نهایی نتایج از فهرست گسترش‌یافتهٔ ابیات، مرحلهٔ بازرتبه‌بندی انجام می‌شود. برای رتبه‌بندی نهایی نتایج، از یک مدل ترکیبی آماری ـ مفهومی استفاده می‌شود. بدین‌صورت که امتیاز نهایی هر بیت (d) براساس ترکیبی از امتیاز آماری (مانند BM25) و امتیاز مفهومی حاصل از گراف محاسبه می‌شود. این ترکیب سبب می‌شود که نتایج نه‌تنها ازنظر ظاهری، از منظر معنایی نیز به پرس‌وجوی کاربر نزدیک‌تر شوند. در ادامه، شیوهٔ محاسبهٔ امتیاز مفهومی و امتیاز نهایی نشان داده خواهد شد.

۳ـ۳ـ۱. محاسبهٔ امتیاز مفهومی[1]

این امتیاز برای هر بیت (d) و پرس‌وجوی گسترش‌یافته (q') که مشتمل بر مفاهیم اصلی و مفاهیم استخراج‌شده از گراف است، محاسبه می‌شود.

$$CS(d,q') = \frac{\left[\Sigma_{\{c \in Concepts(q')\}} w_c \times \mathbb{1}\{c \in Concepts(d)\}\right]}{\left[\Sigma_{\{c \in Concepts(q')\}} w_c\right]} \qquad (۲)$$

که در آن:

- w_c وزن مفهوم c در گراف (که از مجموع وزن یال‌های متصل به آن محاسبه می‌شود).
- $\mathbb{1}\{c \in Concepts(d)\}$ تابع شاخصی که درصورت حضور مفهوم c در بیت d مقدار ۱ می‌گیرد.

۳ـ۳ـ۲. محاسبهٔ امتیاز بازرتبه‌بندی ترکیبی

این امتیاز به‌صورت زیر محاسبه می‌شود:

$$FS(d,q) = \gamma\ \hat{Z}_{BM25}(d,q) + (1 - \gamma)\ CS(d,q') \qquad (۳)$$

که در آن:

- $\hat{Z}_{BM25}(d,q)$ امتیاز نرمال‌شده BM25 در بازهٔ [0,1] است.
- $CS(d, q')$: امتیاز مفهومی حاصل از گراف است
- $\gamma \in [0,1]$ ضریب وزن‌دهی برای تعادل میان دو مؤلفه (پیشنهاد می‌شود $\gamma = 0.7$ در نظر گرفته شود).

بدین‌ترتیب، نتیجهٔ نهایی نه‌تنها بر پایهٔ تطبیق واژگان ظاهری، براساس شبکهٔ ارتباطات معنایی میان مفاهیم عرفانی در متن به دست می‌آید؛ درنتیجه، ابیاتی که ازنظر معنایی با پرس‌وجو هم‌خوانی دارند ولی فاقد واژه‌های سطحی مشابه‌اند نیز در رتبه‌های بالاتر ظاهر می‌شوند. نکتهٔ درخور توجه این است که انتخاب ضریب وزن‌دهی مناسب برای تعادل میان دو مؤلفه بسیار حیاتی است؛ زیرا امتیاز مفهومی نهایی به این ضریب حساسیت دارد. ازاین‌رو، این ضریب باید با فرایند تنظیم پارامتر و طی آزمایش‌های متعدد تعیین شود و مقدار صحیح آن وابسته به دامنه و مسئلهٔ جست‌وجوی مد نظر انتخاب می‌شود.

---

[1]. Conceptual Score

٤. ارزیابی و تحلیل

برای ارزیابی اثربخشی روش پیشنهادی مبتنی‌بر گراف دانش معنایی، دو رویکرد جست‌وجو با هم مقایسه شدند:

۱. روش مبنا (Baseline) که براساس الگوریتم آماری BM25 عمل می‌کند (Robertson & Zaragoza, 2009).

۲. روش پیشنهادی ترکیبی که در این پژوهش با عنوان KG-BM25 شناخته می‌شود و تلفیقی از مدل آماری BM25 و گراف دانش معنایی (Knowledge Graph) است.

هدف از این مقایسه، سنجش تأثیر گراف دانش معنایی با روش پیشنهادی در بهبود بازیابی مفهومی از متون عرفانی، به‌ویژه در فضای استعاری و چندلایهٔ مثنوی معنوی است.

٤ـ۱. مجموعه دادهٔ ارزیابی

برای سنجش عملکرد دو روش، مجموعه‌ای از ۱۸ پرس‌وجو (Query) طراحی شد که هریک بازتاب‌دهندهٔ مفاهیم کلیدی در اندیشهٔ مولوی است؛ مفاهیمی همچون «عشق»، «توحید»، «عقل»، «نفس»، «فنا»، و «نبوت و رسالت».

در انتخاب پرس‌وجوها، تلاش شد تا طیف متنوعی از ساختارهای زبانی و معنایی پوشش داده شود، به‌گونه‌ای که برخی از آنها شامل ترکیبات کنایی یا استعاری و برخی دیگر مفاهیم مجرد و فلسفی باشند. برای هر پرس‌وجو، ۱۰ بیت برتر بازگردانده‌شده توسط هر روش ذخیره شدند و نتایج دو روش به‌صورت موازی در اختیار ارزیابان قرار گرفت.

٤ـ۲. شیوهٔ ارزیابی

باتوجه‌به ماهیت متون عرفانی و گسترهٔ معناشناختی در آثار مولوی، ایجاد یک مجموعه‌دادهٔ استاندارد طلایی[1] برای قضاوت ماشینی عملاً امکان‌پذیر نیست؛ زیرا ارزیابی شباهت مفهومی در این حوزه، وابسته به درک فلسفی و عرفانی از معناست. ازاین‌رو، در این پژوهش از روش امتیازدهی خبره[2] (امتیاز دهی قضاوت) استفاده شد. سه تن از خبرگان حوزهٔ مولوی‌پژوهی و ادبیات عرفانی، با تجربهٔ تدریس و پژوهش در حوزهٔ متون کلاسیک فارسی، خروجی‌های هر دو سیستم را به‌صورت مستقل قضاوت کردند. برای هر بیت بازگردانده‌شده، امتیازی در بازهٔ ۱ تا ۵ براساس معیارهای جدول (۴) تخصیص یافت. میانگین امتیازات سه خبره به‌عنوان نمرهٔ نهایی هر بیت محاسبه شد. سپس برای هر پرس‌وجو، میانگین امتیاز ده بیت نخست تعیین شد. درنهایت، میانگین کلی امتیازات ۱۸ پرس‌وجو برای هر روش محاسبه و با یکدیگر مقایسه شد.

جدول ٤. روش امتیازدهی براساس میزان تناسب نتایج هر پرس‌وجو

امتیاز	معیار ارزیابی تناسب
۵	بیت کاملاً مرتبط با معنای پرس‌وجو (ارتباط مستقیم و روشن)
۴	بیت مرتبط ازنظر مضمون یا مفهوم عرفانی مشابه
۳	ارتباط ضعیف یا ضمنی با مفهوم پرس‌وجو
۲	بیت دارای شباهت زبانی اما نه مفهومی
۱	بیت نامرتبط یا دارای معنای غیرقابل اتصال به پرس‌وجو

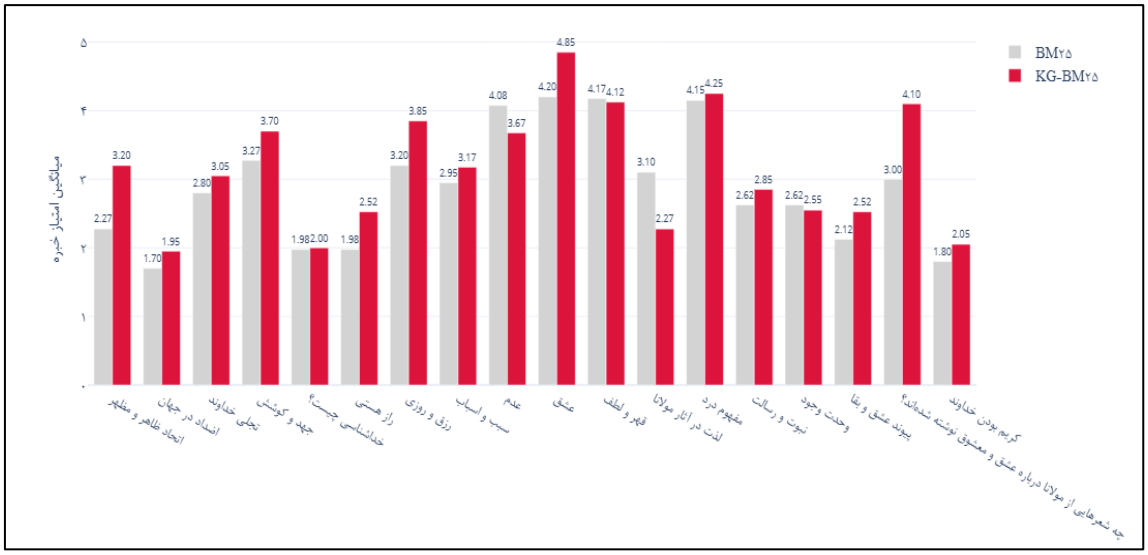
[1]. Gold standard
[2]. Expert Judgment Scoring

٤ـ۳. نتایج ارزیابی

جدول (۵) میانگین امتیازات خبرگان را برای ۱۰ بیت نتیجهٔ دریافت‌شده از هر پرس‌وجو نشان می‌دهد. همچنین، در ستون بهبود، میزان ارتقای نتایج توسط روش KG-BM25 در مقایسه با روش مبنای BM25 مشاهده می‌شود. از نتایج مقایسه‌ای چنین برمی‌آید که گراف دانش معنایی پیشنهادی توانسته است در بیشتر پرس‌وجوها، به استثنای سه پرس‌وجو، میزان تناسب نتایج برگشتی را بهبود بخشد. همچنین، در شکل (۳) نمودار میله‌ای برای مقایسهٔ عملکرد روش‌های جست‌وجو برای هر پرس‌وجو نمایش داده شده است.

جدول ۵. میانگین امتیازات خبرگان برای نتایج هر پرس‌وجو

پرس‌وجو	BM25	KG-BM25	بهبود (٪)
اتحاد ظاهر و مظهر	2.275	3.2	40.65934
اضداد در جهان	1.7	1.95	14.70588
تجلی خداوند	2.8	3.05	8.928571
جهد و کوشش	3.275	3.7	12.9771
خداشناسی چیست؟	1.975	2	1.265823
راز هستی	1.975	2.525	27.8481
رزق و روزی	3.2	3.85	20.3125
سبب و اسباب	2.95	3.175	7.627119
عدم	4.075	3.675	-9.81595
عشق	4.2	4.85	15.47619
قهر و لطف	4.175	4.125	-1.1976
لذت در آثار مولانا	3.1	2.275	-26.6129
مفهوم درد	4.15	4.25	2.409639
نبوت و رسالت	2.625	2.85	8.571429
وحدت وجود	2.625	2.55	-2.85714
پیوند عشق و بقا	2.125	2.525	18.82353
چه شعرهایی از مولانا دربارهٔ عشق و معشوق نوشته شده‌اند؟	3	4.1	36.66667
کریم‌بودن خداوند	1.8	2.05	13.88889



شکل ۳. نمودار نتایج مقایسه‌ای دو روش جست‌وجو براساس میانگین امتیازات خبرگان

جدول (۶) میانگین کل امتیازات تناسب ارزیابی‌شده توسط خبرگان را برای کل ۱۸۰ بیت برگشتی نشان می‌دهد. به این صورت، امکان یک مقایسهٔ جامع دربارهٔ عملکرد مدل‌های جست‌وجو فراهم می‌شود. باتوجه‌به این جدول، می‌توان نتیجه گرفت که روش جست‌وجوی مفهومی مبتنی‌بر گراف دانش معنایی استخراج‌شده قادر است تا حدود ۹ درصد میزان اثربخشی نتایج را بهبود بخشد؛ البته این بهبود درصورتی‌که جست‌وجو بیشتر مرتبط با استخراج معنا باشد، چشمگیرتر خواهد بود.

جدول ٦. میانگین کل امتیازات نتایج برای دو روش جست‌وجو

روش	BM25	KG-BM25	بهبود (%)
میانگین کل امتیازات تناسب	2.89	3.15	8.99

٤ـ٤. تحلیل نتایج

براساس نتایج به‌دست‌آمده، روش KG-BM25 در اغلب پرس‌وجوها توانست امتیاز بالاتری نسبت به روش کلاسیک BM25 کسب کند. در پرس‌وجوهایی که بار معنایی عمیق‌تری داشتند (مانند «عشق و معشوق»، «رزق و روزی»، یا «اتحاد ظاهر و مظهر»)، اختلاف امتیاز میان دو روش چشمگیرتر بود. این امر نشان می‌دهد که گراف دانش توانسته است بخشی از پیوندهای استعاری و مفهومی موجود در لایه‌های زبانی مثنوی را آشکار سازد و ابیاتی را بازیابی کند که دسترسی به آنها در روش‌های آماری سنتی امکان‌پذیر نبود. برای نمونه، در پرس‌وجوی «اتحاد ظاهر و مظهر»، روش KG-BM25عمدتاً ابیاتی را بازمی‌گرداند که واژهٔ «اتحاد» یا ترکیبات آن در آنها تکرار شده بود؛ درحالی‌که روش BM25ابیاتی را نیز یافت که مفاهیمی همچون «ظاهر»، «مظهر»، و «اتحاد بین آن دو» در آنها به کار رفته بود؛ مفاهیمی که بیشتر با مفهوم مستتر در پرس‌وجو ارتباط معنایی داشتند؛ درنتیجه، این روش بهبودی به‌اندازهٔ ۴۰ درصد در مقایسه با روش مبنای جست‌وجو ارائه می‌دهد. در جدول (۷) نیز نمونهٔ ابیات برگشتی از دو روش جست‌وجو برای این پرس‌وجو و نیز امتیازی که خبرگان برای هر بیت منظور کرده‌اند، مشاهده می‌شود.

در مواردی معدود که روش ترکیبی KG-BM25 نسبت به روش پایهٔ BM25 عملکرد ضعیف‌تری نشان داده است، باید این نکته را در نظر گرفت که انتظار نمی‌رود جست‌وجوی معنایی در همهٔ موارد و پرس‌وجوها نتیجه بهتر ارائه دهد؛ بلکه عملکرد کلی روش باید بهتر باشد. بااین‌حال، بررسی دقیق‌تر نشان داد که علت اصلی، به ماهیت خاص زبان عرفانی و استعاری مثنوی بازمی‌گردد. در برخی ابیات، واژگان یا ترکیبات زبانی دارای چندلایگی معنایی هستند که در شبکهٔ مفهومی، به‌ویژه در فارس‌نت یا مدل‌های embedding، با معنای متداول آن‌ها هم‌پوشانی ندارد. برای نمونه، واژه‌ای مانند «شراب» در معنای عرفانی به «عشق الهی» یا «معرفت» دلالت دارد؛ اما در منابع زبانی عمومی با معنای مادی آن مرتبط است. در چنین مواردی، گسترش معنایی مبتنی‌بر گراف ممکن است مسیرهای معنایی نادرست را نیز فعال کند و درنتیجه دقت بازیابی کاهش یابد. عامل دیگر، ناقص‌بودن پوشش واژگانی فارس‌نت برای حوزهٔ عرفان و ادبیات کلاسیک است؛ برخی از مفاهیم خاص مثنوی هنوز در این شبکه تعریف نشده یا روابط معنایی کافی میان آن‌ها وجود ندارد. افزون‌بر این، وزن‌دهی یال‌ها در گراف براساس ترکیب خطی روابط (هم‌رخدادی، ترادف و ابرده‌گی) صورت گرفته است و ممکن است در برخی پرس‌وجوها، وزن‌های بهینه برای هر نوع رابطه متفاوت باشد.

همچنین، بررسی بازخوردهای خبرگان نشان داد که ابیات بازگردانده‌شده توسط KG-BM25 ازنظر توضیح‌پذیری[1] نیز وضعیت بهتری دارند؛ زیرا ارتباط مفهومی میان پرس‌وجو و بیت ازطریق گراف ردیابی می‌شود (برای مثال، از مسیر ابرده یا ترادف).

به‌طور کلی، نتایج نشان دادند که استفاده از گراف دانش معنایی می‌تواند دقت بازیابی و میزان تناسب معنایی را به‌طور معناداری افزایش دهد. به‌عبارت‌دیگر، ترکیب مدل آماری با ساختار معنایی باعث غنای محتوایی در نتایج جست‌وجو و درک بهتر روابط بینامتنی در متون عرفانی شد. اما نکته مهم آن است که این روش بیشتر در جست‌وجوی مفهومی یا معنایی کاربرد دارد و درصورتی‌که کاربر قصد جست‌وجوی کلیدواژه یا لفظی داشته باشد بهبود چشمگیری را در مقایسه با روش جست‌وجوی مبنا ارائه نمی‌دهد.

در مجموع، نتایج ارزیابی نشان داد که مدل پیشنهادی KG-BM25 با بهره‌گیری از ساختار گراف دانش معنایی توانسته است فراتر از سطح واژگانی، روابط مفهومی را در جست‌وجوی متون کلاسیک فارسی مدل‌سازی نماید. این ویژگی، مسیر تازه‌ای را برای پژوهش‌های مرتبط با علوم انسانی دیجیتال، تحلیل متون عرفانی و بازیابی معنایی در ادبیات فارسی فراهم می‌سازد.

---

[1]. Interpretability

جدول ۷. نتایج برگشتی برای پرس‌وجوی «اتحاد ظاهر و مظهر» و ارزیابی خبرگان از میزان تناسب و ارتباط معنایی نتایج

KG-BM25	امتیاز خبره	BM25	امتیاز خبره
جان اول مظهر درگاه شد / جان جان خود مظهر الله شد	5	جان اول مظهر درگاه شد / جان جان خود مظهر الله شد	5
مظهر لطف خدایی گشته‌ام / لوح شرح کبریایی گشته‌ام	4	جان حیوانی ندارد اتحاد / تو مجو این اتحاد از روح باد	4
بهر دیده‌روشنان یزدان فرد / شش جهت را مظهر آیات کرد	3	بودشان تمییز کان مظهر کند / لیک حرص و آز کور و کر کند	2
آدم اصطرلاب اوصاف علوست / وصف آدم مظهر آیات اوست	4	در جهان خاک ابر و آب بود / مظهر بخشایش وهاب بود	3
ظاهر قرآن چو شخص آدمیست / که نقوشش ظاهر و جانش خفیست	3	چونک جامه چست و دوزیده بود / مظهر فرهنگ درزی چون شود	4
دست و پای ما می آن واحدست / دست ظاهر سایه است و کاسدست	3.5	مظهر عزست و محبوب به حق / از همه کروبیان برده سبق	3
چونک جامه چست و دوزیده بود / مظهر فرهنگ درزی چون شود	4	چون شد آدم مظهر وحی و وداد / ناطقهٔ او علم الاسما گشاد	4.5
اتحاد یار با یاران خوشست / پای معنی‌گیرْ، صورت سرکشست	1	می‌فتد این عقل‌ها در افتقاد / در مغاکی حلول و اتحاد	2
پادشاهان مظهر شاهی حق / فاضلان مرآت آگاهی حق	4	ظاهر قرآن چو شخص آدمیست / که نقوشش ظاهر و جانش خفیست	3
گر بنالیدی و مستغفر شدی / نورِ رفته از کرم ظاهر شدی	2	آن هلال و بدر دارند اتحاد / از دوی دورند و از نقص و فساد	1
جان حیوانی ندارد اتحاد / تو مجو این اتحاد از روح باد	4	که اتحاد جسمهای آب و طین / هست ناقص جان نمی‌ماند بدین	2
بودشان تمییز کان مظهر کند / لیک حرص و آز کور و کر کند	2	تا میان قهر و لطف آن خفیه‌ها / ظاهر آید ز آتش خوف و رجا	1
روح را توحید الله خوش‌ترست / غیر ظاهر دست و پای دیگرست	3	پادشاهان مظهر شاهی حق / فاضلان مرآت آگاهی حق	4
در جهان خاک ابر و آب بود / مظهر بخشایش وهاب بود	3	آدم اصطرلاب اوصاف علوست / وصف آدم مظهر آیات اوست	4
میل اندر مرد و زن حق زان نهاد / تا بقا یابد جهان زین اتحاد	3	بهر دیده‌روشنان یزدان فرد / شش جهت را مظهر آیات کرد	3
چون شد آدم مظهر وحی و وداد / ناطقهٔ او علم الاسما گشاد	4.5	هست ازل را و ابد را اتحاد / عقل را ره نیست آن سو ز افتقاد	2
که اتحاد جسم‌های آب و طین / هست ناقص جان نمی‌ماند بدین	2	میل اندر مرد و زن حق زان نهاد / تا بقا یابد جهان زین اتحاد	3
مظهر عزست و محبوب به حق / از همه کروبیان برده سبق	3	آنک تن را مظهر هر روح کرد / وآنک کشتی را براق نوح کرد	3
مظهر لطف خدایی گشته‌ام / لوح شرح کبریایی گشته‌ام	4	اتحاد یار با یاران خوشست / پای معنی‌گیر، صورت سرکشست	1
روز و شب ظاهر دو ضد و دشمنند / لیک هر دو یک حقیقت می‌تنند	4	انبیا بودند ایشان اهل ود / اتحاد انبیاام فهم شد	1

۵. نتیجه‌گیری

در این پژوهش، تلاش شد تا از رهگذر پیوند فناوری‌های نوین معناشناختی با متون عرفانی کلاسیک فارسی، گامی در مسیر فهم و بازیابی دقیق‌تر مفاهیم ژرف مثنوی معنوی برداشته شود. مسئلهٔ اصلی در جست‌وجوهای متنی در آثار ادبی و عرفانی، محدودیت روش‌های آماری صرف ــ نظیر BM25 ــ در درک لایه‌های استعاری و مفهومی زبان است؛ زیرا زبان عرفانی، نه‌تنها بر واژگان، که بر شبکه‌ای از اشارات، تمثیل‌ها و مفاهیم درونی بنا شده است. ازاین‌رو، در این تحقیق، الگویی ترکیبی با عنوان KG-BM25 ارائه شد که ضمن بهره‌گیری از توان روش‌های آماری در بازیابی اولیه، از گراف دانش معنایی برای گسترش و بازرتبه‌بندی نتایج استفاده می‌کند.

در این چهارچوب، گراف دانش از روابط سه‌گانهٔ هم‌رخدادی، ترادف و ابررده‌گی میان مفاهیم استخراج‌شده از ابیات مثنوی ساخته شد. با اتکا به این گراف، مفاهیم مرتبط با هر پرس‌وجو به‌صورت پویا گسترش یافتند و بدین‌ترتیب، جست‌وجو قادر شد ابیاتی را نیز بازیابی کند که گرچه فاقد واژگان ظاهری پرس‌وجو بودند، در سطح معنا با آن پیوند داشتند. نتایج ارزیابی با مشارکت خبرگان حوزهٔ مولوی‌پژوهی نشان داد که روش KG-BM25 در مقایسه با روش پایهٔ BM25، ازنظر امتیاز تناسب معنایی به‌طور میانگین افزایشی معنادار در دقت و غنای مفهومی نتایج دارد. همچنین، تحلیل‌های کیفی خبرگان آشکار ساخت که این روش نه‌تنها ابیات مرتبط‌تری را ارائه می‌دهد، امکان توضیح‌پذیری و تبیین روابط معنایی میان ابیات را نیز فراهم می‌آورد. از منظر کاربردی، روش پیشنهادی می‌تواند بستری برای توسعهٔ سامانه‌های هوشمند جست‌وجو و تحلیل در متون عرفانی فارسی باشد؛ سامانه‌هایی که پژوهشگر را نه صرفاً در یافتن واژه‌ها، بلکه در کشف پیوندهای مفهومی میان مفاهیم «عشق»، «قرب»، «فنا»، «عقل» و دیگر بن‌مایه‌های عرفان یاری رسانند.

یکی از چالش‌های اساسی در این پژوهش، محدودیت منابع زبانی موجود برای زبان فارسی در حوزهٔ متون عرفانی کلاسیک است. شبکه‌های واژگانی و مدل‌های توزیعی نظیر FarsNet و FastText فارسی اگرچه ابزارهای قدرتمندی برای بازنمایی معنایی واژگان‌اند، پوشش آن‌ها نسبت به واژگان تخصصی عرفانی، استعاره‌های عرفانی و ترکیبات اصطلاحی خاص متون مولوی ناکافی است. بسیاری از مفاهیم محوری در مثنوی ــ مانند فنا فی‌اللّه، سِرّ وحدت، نفس رحمانی، یا جذبهٔ الهی ــ یا در این منابع وجود ندارند یا در قالب‌هایی ثبت شده‌اند که با ساختار معنایی متون عرفانی سازگار نیستند. محدودیت دیگر، وابستگی نسبی مدل به کیفیت داده‌های استخراج‌شده از متن است. در متون کلاسیک فارسی، وجود پدیده‌هایی همچون املای متغیر، آرایه‌های ادبی و حذف یا تغییر ساخت‌های نحوی، فرایندهای استخراج مفاهیم، تشخیص هم‌رخدادی‌ها و ساخت نگاشت‌های دقیق را دشوار می‌کند.

برای غلبه بر این محدودیت‌ها، مسیرهای پژوهشی زیر پیشنهاد می‌شود:

۱. ایجاد WordNet عرفانی فارسی با محوریت واژگان استخراج‌شده از متون بنیادین عرفانی همچون *مثنوی معنوی*، *گلشن راز*، و *فصوص‌الحکم* و پیوند آن با FarsNet عمومی.

۲. آموزش مدل‌های embedding خاص متون عرفانی مانند *SufiBERT* یا *Mystic2Vec* که با داده‌های حوزهٔ عرفان فارسی آموزش یابند و معنای استعاری و تمثیلی را بهتر بازنمایی کنند.

۳. به‌کارگیری تحلیل صرفی و نحوی عمیق‌تر برای تفکیک معانی استعاری از معانی حقیقی در متون شعری.

۴. طراحی چهارچوب ارزیابی انسانی ترکیبی که افزون‌بر امتیازدهی خبره، از تحلیل هم‌پوشانی مفهومی با سایر متون عرفانی بهره گیرد تا قضاوت‌ها نظام‌مندتر شود.

۵. گسترش گراف دانش به متون عرفانی دیگر به‌منظور مقایسهٔ میان‌متنی مفاهیم و بررسی شبکه‌های معنایی مشترک در

آثار مولوی، عطار و ابن‌عربی.

به باور نگارندگان، ترکیب این مسیرها می‌تواند گام‌های مؤثری در مسیر ایجاد یک زیرساخت دانش‌محور برای مطالعات عرفانی دیجیتال در زبان فارسی فراهم آورد. چنین رویکردی می‌تواند در افق آینده، راه را برای پدیدآمدن علوم انسانی دیجیتالِ فارسی‌محور و بازخوانی داده‌محور میراث معنوی ایران هموار سازد.

اعلام استفاده از هوش مصنوعی مولد:

در تهیه این مقاله از هوش مصنوعی مولد صرفاً برای سازمان‌دهی و انسجام بخشی متن بخش پیشینه پژوهش (و نه تولید خودکار متن) و نیز برای ترجمه و سازمان‌دهی چکیده گسترده انگلیسی استفاده شده است.

References

Candela, G., Escobar, P., & Navarro-Colorado, B. (2017). In search of poetic rhythm: Poetry retrieval through text and metre. In *Proceedings of the 2nd International Conference on Digital Access to Textual Cultural Heritage* (pp. 53–57). https://doi.org/10.1145/3078081.3078085

Chen, H., & Schatz, B. R. (1994). *Semantic retrieval for the NCSA Mosaic*. https://repository.arizona.edu/handle/10150/105613

Dastgheib, M. B., Fakhrahmad, S. M., & Jahromi, M. Z. (2017). Perspell: A new Persian semantic-based spelling correction system. *Digital Scholarship in the Humanities*, *32*(3), 543–553. https://doi.org/10.1093/llc/fqw015

Dessí, D., Osborne, F., Buscaldi, D., Reforgiato Recupero, D., & Motta, E. (2025). CS-KG 2.0: A Large-scale Knowledge Graph of Computer Science. *Scientific Data*, *12*(1), 964. https://doi.org/10.1038/s41597-025-05200-8

Egozi, O., Markovitch, S., & Gabrilovich, E. (2011). Concept-based information retrieval using explicit semantic analysis. *ACM Transactions on Information Systems (TOIS)*, *29*(2), 1–34. https://doi.org/10.1145/1961209.1961211

Fadaei, H., & Shamsfard, M. (2010). Extracting conceptual relations from Persian resources. *2010 Seventh International Conference on Information Technology: New Generations* (PP.244–248). https://doi.org/10.1109/ITNG.2010.191

Giunchiglia, F., Kharkevich, U., & Zaihrayeu, I. (2009). Concept Search. In L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. Hyvönen, R. Mizoguchi, E. Oren, M. Sabou, & E. Simperl (Eds.), *The Semantic Web: Research and Applications* (pp. 429–444). Springer. https://doi.org/10.1007/978-3-642-02121-3_33

Haslhofer, B., Isaac, A., & Simon, R. (2018). Knowledge graphs in the libraries and digital humanities domain. In *Encyclopedia of big data technologies* (pp. 1080-1087). Springer. https://arxiv.org/pdf/1803.03198

Hearst, M. A. (1992). Automatic Acquisition of Hyponyms from Large Text Corpora. *COLING 1992 Volume 2: The 14th International Conference on Computational Linguistics*. https://aclanthology.org/C92-2082/

Hosseini Pozveh, Z., Monadjemi, A., & Ahmadi, A. (2018). FNLP-ONT: A feasible ontology for improving NLP tasks in Persian. *Expert Systems*, *35*(4), e12282. https://doi.org/10.1111/exsy.12282

Kiff, L. (n.d.). *Knowledge Graph Use Cases | Tom Sawyer Software*. Retrieved October 20, 2025, from https://blog.tomsawyer.com/knowledge-graph-use-cases

Manjula, D., Kulandaiyan, S., Sudarshan, S., Francis, A., & Geetha, T. V. (2003). Semantics based information retrieval using conceptual indexing of documents. *International Conference on Intelligent Data Engineering and Automated Learning* (PP. 685–692). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-45080-1_92

Montazery, M., & Faili, H. (2011). Unsupervised Learning for Persian WordNet Construction. In R. Mitkov & G. Angelova (Eds.), *Proceedings of the International Conference Recent Advances in*

Natural Language Processing 2011 (pp. 302–308). Association for Computational Linguistics. https://aclanthology.org/R11-1042/

Mousavi, Z., & Faili, H. (2017). *Persian Wordnet Construction using Supervised Learning* (No. arXiv:1704.03223). arXiv. https://doi.org/10.48550/arXiv.1704.03223

Peng, C., Xia, F., Naseriparsa, M., & Osborne, F. (2023). Knowledge Graphs: Opportunities and Challenges. *Artificial Intelligence Review*, *56*(11), 13071–13102. https://doi.org/10.1007/s10462-023-10465-9

Robertson, S., & Zaragoza, H. (2009). The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends® in Information Retrieval*, *3*(4), 333–389. http://dx.doi.org/10.1561/1500000019

Roller, S., Kiela, D., & Nickel, M. (2018). Hearst Patterns Revisited: Automatic Hypernym Detection from Large Text Corpora. In I. Gurevych & Y. Miyao (Eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 358–363). Association for Computational Linguistics. https://doi.org/10.18653/v1/P18-2057

Samvelian, P., & Faghiri, P. (2013). Introducing PersPred, a syntactic and semantic database for Persian complex predicates. *The 9th Workshop on Multiword Expressions* (PP. 11–20). https://shs.hal.science/halshs-00974259/

Sanyal, D. K. (2018). Semantically Enriched Line Search in a Humanities Digital Library. *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (PP. 2169–2174). IEEE. https://doi.org/10.1109/ICACCI.2018.8554885

Shamsfard, M., Fadaei, H., & Fekri, E. (2010). Extracting Lexico-conceptual Knowledge for Developing Persian WordNet. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA). https://aclanthology.org/L10-1540/

Smith, D. A., Rydberg-Cox, J. A., & Crane, G. R. (2000). The Perseus Project: A digital library for the humanities. *Literary and Linguistic Computing*, *15*(1), 15–25. https://doi.org/10.1093/llc/15.1.15

Waldrop, A. M., Cheadle, J. B., Bradford, K., Preiss, A., Chew, R., Holt, J. R., Kebede, Y., Braswell, N., Watson, M., & Hench, V. (2022). Dug: A semantic search engine leveraging peer-reviewed knowledge to query biomedical data repositories. *Bioinformatics*, *38*(12), 3252–3258. https://doi.org/10.1093/bioinformatics/btac284

Wang, C., & Akella, R. (2015). Concept-based relevance models for medical and semantic information retrieval. *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management* (PP.173–182). https://doi.org/10.1145/2806416.2806497

Wen, K., Lu, Z., Li, R., Sun, X., & Wang, Z. (2006). A semantic search conceptual model and application in security access control. *Asian Semantic Web Conference* (pp. 366-376). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/11836025_37

Zhao, Z., Luo, X., Chen, M., & Ma, L. (2023). A Survey of Knowledge Graph Construction Using Machine Learning. *Computer Modeling in Engineering & Sciences*, *139*(1), 225–257. https://doi.org/10.32604/cmes.2023.031513

Zhong, L., Wu, J., Li, Q., Peng, H., & Wu, X. (2023). *A Comprehensive Survey on Automatic Knowledge Graph Construction* (No. arXiv:2302.05019). arXiv. https://doi.org/10.48550/arXiv.2302.05019