MORSE-QE: A Morphology-Aware, Embedding-Driven Framework with Root Extraction for Arabic Dialectal Query Expansion

Yasir Hadi Farhan[1] ,Mustafa Tareq[2], Boumedyen Shannaq[3], Oualid Ali[4], Said Almaqbali[5], V.P.Sriram[6]
[1]Department of Artificial Intelligence, College of Information Technology, University of Fallujah, Iraq
[2]College of Computer Science, University of Technology, Iraq
[3]Department of Management of Information Systems, College of Business, University of Buraimi, Oman
[4]Computer Sciences Department, College of Arts & Science, Applied Science University, Manama, Bahrain
[5]Information Technology Unit, University of Buraimi (UoB), Oman
[6]Department of Management of Information Systems, College of Business, University of Buraimi, Oman
Email: boumedyen@uob.edu.om
[*]Corresponding author

Keywords: Arabic information retrieval, query expansion, morphology-aware processing, dialect normalization, root-based filtering

Received: June 9, 2025

Information retrieval (IR) in diglossic and morphologically complex Arabic includes major difficulties since dialectal searches usually do not retrieve documents written in Modern Standard Arabic (MSA). The following paper introduces MORSE-QE (Morphology-aware, Optimized, Root-driven Semantic Expansion for Query Enhancement), a four-stage approach that combines rule-based morph-based processing and embedding-based expansion in parallel structures. The process includes: (i) dialect-to-MSA normalization using curated lexicons, (ii) root extraction via AlKhalil Morpho Sys 2, (iii) semantic expansion with AraVec embeddings, and (iv) root-driven filtering to reduce morphological noise. Experiments were conducted on the QADI dataset (2,000 dialectal queries) and the TREC 2001 Arabic Corpus (383,872 MSA documents), using Mean Average Precision (MAP), Precision@10, and Root Recall as evaluation metrics. MORSE-QE achieves MAP gains of 15–18% over neural baselines (DANs, DMNs) on QADI and 15% over RM3 on TREC 2001, with a root recall improvement from 65% (DMNs) to 88%. Ablation studies show that dialect normalization and root-based filtering contribute 19.5% and 28.9% relative MAP improvements, respectively. These results demonstrate that MORSE-QE provides a scalable and interpretable solution for bridging dialectal and morphological gaps in Arabic IR.

Povzetek: Prispevek predstavlja večstopenjski pristop za izboljšanje iskanja informacij v arabščini, ki z združevanjem dialektne normalizacije, morfološke analize in semantične razširitve poizvedb bistveno poveča natančnost iskanja med narečno arabščino in knjižno arabščino.

1 Introduction

The Complexity of Arabic Language (and hence the difficulty in its treatment by information retrieval systems) is due to the diglossic nature and the complexity of the morphology of the language when compared to MSA and the regional dialects such as Egyptian and Levantine[1,2].[3]introduced a robust overview of algorithms of Arabic roots extraction that constitutes a solid methodological framework of morphology-sensitive query expansion systems.[4] also created the evaluation systems of Arabic information retrieval, which offers the measures that should be used in improvement of the performance of semantic search. In the context of root-based semantics and dialectal heterogeneity in Arabic, query expansion (QE) approaches such as pseudo-relevance feedback (e.g., RM3[5]) fail to deal with these two problems. In making neural QE less expensive to sparse queries, more recent approaches have employed word embeddings, such as Deep Median Networks (DMNs) and Deep Averaging Networks (DANs). These methods hold promise, but there are serious gaps:

- Morphological oversights: Noisy expansions are caused by the absence of explicit mechanisms in DANs and DMNs to incorporate Arabic's root-and-pattern structure [6].
- Dialectal mismatches: There are lexical gaps between dialectal queries and MSA-indexed corpora, and these methods don't fix them [7].
- Interpretability: Transparency in term selection is limited by the black-box nature of neural QE [8].

Hybrid strategies that incorporate morphological rules and semantic embeddings have been previously demonstrated to be necessary. One case in point is:

- An analysis of available Arabic QE methods [5]. We find that a critical part of neural approaches is root-aware expansion.
- DANs and DMNs [6,7] showed better accuracy than conventional approaches, but had trouble handling morphological noise.
- A combination of approaches [9]. Although it exhibited potential, combining embeddings and rule-based normalization lacked systematic root filtering.

To fill these gaps, we offer MORSE-QE, which stands for morphology-optimized root-driven semantic expansion for query enhancement. This innovative framework includes new semantic embeddings based on rule-based morphological analysis. The acronym MORSE-QE stands for the five tenets upon which it is built:

- **M**orphology-aware processing: it uses the root-and-pattern structure of Arabic to extract roots and strip clitics.
- **O**ptimized dialect normalization: Transforms dialectal queries into their MSA equivalents, for instance (e.g., Egyptian "عايز" → MSA "أريد").
- **R**oot-driven filtering: To reduce noise, root-driven filtering gives priority to terms that share compatible roots. Example: All the following terms share the same root **"كتب"** (to write): "كتاب" (book), "مكتوب" (written), "كتابة" (writing), "كاتب" (writer).
- **S**emantic term selection: using word embeddings to discover contextual relationships between terms.
- Expansion for Query Enhancement: Strengthens queries without sacrificing linguistic consistency.

MORSE-QE bolsters Arabic IR through:

- Fixing lexical mismatches in MSA corpora and dialectal queries.
- Lightweight, rule-guided expansion can lessen the burden on neural training, which can be resource-intensive.
- Filling a need that previous neural methods did not handle by explicitly grounding expansion in the morphological rules of Arabic [6,7].

In terms of recall and precision, MORSE-QE beats out DANs and DMNs as well as conventional baselines when tested on the QADI dataset [10] (dialectal queries) and the TREC 2001 Arabic Corpus [11]. Some of the things we've done are:

1.1 A Framework for expanding arabic IR queries guided by morphology

- An architecture that combines rule-based morphological analysis of Arabic (root extraction, clitic stripping) with semantic embeddings is suggested, using Arabic's root-and-pattern structure, *wazn*.
- Bridges the lexical gap between informal queries and formal corpora, resolving dialectal diversity through a lightweight dialect-to-MSA normalization module.

1.2 A Study on the effectiveness of root-based filtering in reducing noise

- Proves, on the QADI (dialectal) and TREC 2001 (MSA) benchmarks, that root-constrained expansion considerably lessens irrelevant term injection compared to DANs/DMNs.
- Ablation studies confirm the impact of morphological filtering, and performance gains are quantified using infrared metrics (Precision@10, MAP).

The most important research questions discussed in this study are:

RQ1: Does neural only query expansion work better than dialect to MSA normalisation followed by a root driven filtering of query results in Arabic dialectal information retrieval?

RQ2: What is the difference in precision and recall of retrieval between the integration of morphological filtering into semantic embeddings and traditional and neural baselines measures?

RQ3: How much do each of the MORSE-QE components, dialect normalization, root filtering, and semantic expansion contribute toward attainment of retrieval gains?

2 Literature review

2.1 Traditional query expansion for arabic IR

A thesaurus-based expansion and pseudo-relevance feedback (PRF) were the foundation of early Arabic IR systems. To illustrate the difficulties associated with morphological noise and dialectal variations, [12] used RM3 to detect Arabic novelty at TREC 2004. [13]coined the use of associative memory in English-Arabic NLP that mediates between cognitive processing and bilingual computational modeling. [14] further highlighted the necessity for morphology-aware filtering, which showed that PRF methods like Bo1 frequently insert irrelevant terms because of the morphological complexity of Arabic.Similarly, [15] proposed a semi-automatic online indexing approach to improve Arabic IR systems by combining online and offline processing strategies. While their work focuses on indexing rather than query expansion, it reinforces the importance of tailoring IR processes to the linguistic characteristics of Arabic, which aligns with MORSE-QE's morphology-aware design. [16]studied root-and-pattern morphology in Classical Arabic, and critical information was given concerning the role of morphological structures in determining the accuracy of root extraction in computations.[17]proposed novel technique for Arabic and English morphosyntactic structures, the proposed technique could be extended to query expansion and morphological processing .[18]examined the interplay between morphophonemic and Arabic dialects in Bani Sakhar Arabic that provided useful linguistic insights to improve dialect-sensitive query expansion models.[19] investigated the phonotactic probability effect on Arabic word recognition, which can be added to the knowledge of sound-based constraints applicable in semantic and morphological modeling. A lightweight model of a switch transformer was proposed by[20] to simplify Arabic text, which increases the preprocessing speed when it comes to search and knowledge retrieval activities.

2.2 Neural query expansion with word embeddings

By capturing the semantic relationships between terms, word embeddings completely changed QE. Applying Deep Averaging Networks (DANs) to Arabic IR showed that they couldn't handle root-based semantics, despite being introduced by [21]. DANs average word vectors for text classification. Arabic-specific DANs outperformed PRF baselines in terms of accuracy, but they failed miserably when asked to retrieve documents in a dialectal style [6]. Expanding on this, [7] suggested Deep Median Networks (DMNs) to lessen the impact of outliers. Still, their assessment of the QADI dataset [10] revealed ongoing difficulties in handling polysemy, such as differentiating "سهم" [arrow/stock] in MSA and Gulf Arabic.

2.3 Hybrid Approaches for Arabic NLP

Many researchers are favouring hybrid approaches that integrate embeddings with rule-based morphology. While [1] argued that Arabic natural language processing pipelines should incorporate morphological analysers like [21,22] showed that AraBERT benefited from root-aware embeddings. To achieve better accuracy for dialectal queries without systematic root filtering,[9] suggested a mixed methodology that combines embeddings with rule-based normalisation. [23,24] Presented encoding algorithm based on mobile design letters, proving how specific encoding improve linguistic representation in exploration and retrieval systems.

Simultaneously,[11] highlighted the importance of dialect-aware normalization in their QADI dataset, which is still used as a standard for Arabic dialect IR. [25] Study the effect of configuration the size of training and testing versus pre-processing on Arabic text classification, the outcomes could be also used to improve retrieval and knowledge extraction tasks.

In [26], the authors optimized Arabic text classification with the help of SVM with word embeddings, showing that the embedding-based architecture is beneficial to the modeling of language in complex settings. In the study by[27] , a hybrid deep learning approach to the analysis of the Arabic sentiment was introduced, emphasizing the helpfulness of using feature weighting and neural approaches to text interpretation.[28]surveyed the extraction of information on Arabic tweets; it presents contemporary insights on the processing of noisy and dialectal information in the social environment.[29] explored the use of AI in Arabic corpora during translation processes, highlighting the quality of corpus as a key aspect in enhancing query interpretation in NLP-based query translations. [30] tested the Arabic WHOQOL-BREF scale as an example of the function of linguistic testing of adaptation and invariance in the Arabic NLP.

2.4 Gaps and opportunities

Currently, no framework systematically integrates root-based filtering, dialect normalization, and semantic expansion for Arabic QE. However, some methods address specific challenges, such as dialect identification and morphological analysis. While hybrid methods [9] emphasize normalization rather than root semantics, neural approaches such as DANs/DMNs [6,7] lack a foundation in morphology. This study fills these gaps by using MORSE-QE, which takes advantage of the structure of the Arabic language to expand in a way that is resistant to noise.

3 Methodology

The MORSE-QE framework uses rule-based morphology and semantic embeddings to improve Arabic QE. This framework operates in four stages. Here is a comprehensive explanation with examples:

3.1 Stage 1: Data preprocessing

Purpose: Prepare raw text for further processing by cleaning and standardising it.
Steps:

- **Text normalisation**:
  - Standard Arabic preprocessing steps were applied, including removing non-Arabic characters, punctuation, and diacritics [31].
  - Example: "التعليم" → "التعليمُ"

In this case, "يُرِيدُ الطَّعامَ" becomes "يريد الطعام".

- **Dialect normalisation**:
  - Dialect normalization follows the conventions used in the QADI dataset [10].
  - "أريد" ("I want") converted from Egyptian "عايز" into MSA Arabic.

Table 1: Comparative performance of state-of-the-art methods on QADI and TREC 2001

Method	QADI MAP	QADI P@10	QADI Root Recall	TREC MAP	TREC P@10	Key Limitations
RM3	0.35	0.32	-	0.48	0.44	Strong baseline, but no morphology or dialect handling.
DANs	0.38	0.35	62%	-	-	Uses embeddings but ignores root-based filtering; weak for dialectal queries.
DMNs	0.40	0.37	65%	-	-	Reduces outlier effect; still lacks explicit morphology/dialect modeling.
AraBERT-QE	0.43	0.40	72%	-	-	Transformer-based; higher recall, but costly and limited interpretability.
MORSE-QE	0.49	0.45	88%	0.55	0.51	Integrates morphology + dialect normalization + embeddings; interpretable and efficient.

Table 2: Examples of Dialect-to-MSA standardisation

Input (Dialect)	Output (MSA)
عايز أكل مصري	أريد طعام مصري
شلونك؟ (Iraqi)	كيف حالك؟

3.2 Stage 2: An analysis of root-based morphology

Purpose: Find the roots and sort the terms according to their morphological fit.

- Tools: AlKhalil Morpho Sys 2 [32] for root extraction.

Table 3: Examining arabic words through Al-Khalil morphology

Word	Root	Pattern	Morphological Analysis
مكتوب	كتب	مفعول	Passive participle of "to write"
مكتبة	كتب	مَفْعَلَة	Place noun ("library")

3.3 Case study: MORSE-QE root-based filtering

a. Query: "أريد كتاب عن التاريخ" ("I want a book about history").
b. Roots in query: "كتب" (to write), "تاريخ" (history).
c. Expansion candidates:
   - "مكتبة" (library, root: كتب)
   - "كاتب" (writer, root: كتب)
   - "مدرسة" (school, root: درس)
d. Filtering process:
   - Retained terms: Terms sharing the query root "كاتب" ,"مكتبة" → "كتب".
   - Filtered out: "مدرسة" (root "كتب" ≠ "درس").
e. Outcome:
   - Enhanced query: "أريد كتاب عن التاريخ مكتبة كاتب".
f. What makes this function
   - Accuracy in linguistics: Shows how MORSE-QE maintains semantic coherence by removing words with no morphological relationship.
   - Practical importance: The framework's reasoning is demonstrated through specific Arabic examples.

Table 4 shows which expansion terms are retained and which are excluded based on the root compatibility with the query.

Table 4: Criteria for filtering terms in MORSE-QE based on roots

Term	Root	Action	Reason
مكتبة (library)	كتب	Retained	Shares query root ("كتب")
كاتب (write)	كتب	Retained	Shares query root ("كتب")
مدرسة (School)	درس	Removed	Root mismatch ("درس" ≠ "كتب"")

3.4 Stage 3: Semantic expansion

a. Purpose: Supplement searches with terms that are related in meaning.
b. Tools: This work exploited the pre-trained AraVec embeddings [33] [33] (CBOW architecture, 300 dimensional vectors), trained on the full Arabic Wikipedia corpus. Fine-tuning or training was done by no further proportion.
c. Preprocessing: Prior to retrieving embeddings, all query terms will be normalized to remove diacritics, punctuation and characters not in the Arabic language, which will be followed by token normalization.
d. Procedure:

- Return, for each root in the query, the first k = 10 terms AraVec that are the most similar to it as per cosine similarity to the root.
- Appliance of rooted compatibility filter with AlKhalil morphological analyzer helps in retaining only terms with identical razine to that of the source token. This limits the incorporation of the dissimilar morphologically incompatible terms broadly related.
- Rationale for k: The k = 10 was chosen through some initial tuning with the validation to maintain the balance between the recall and noise.

Term pairs retrieved from AraVec and sorted by root compatibility are displayed in Table 5.

Table 5: Root-Based AraVec embeddings-based semantic expansion

Query Root	AraVec Candidates	Filtered Terms
كتب (write)	كتاب، مكتبة، كاتب، مكتوب	كتاب، مكتبة، كاتب
علم (know)	عالم، معلم، علمي، تعليم	عالم، علمي

Stage 4: Query enhancement
Purpose: Merge filtered terms with normalized queries.
Example Workflow:

- Input query (Egyptian Arabic): "عايز أكل مصري" ("I want Egyptian food").
- Normalized to MSA: "أريد طعام مصري".
- Root extraction: طعم (food), مصر (Egyptian).
- Semantic expansion:
  - طعم → وجبة (meal)، مذاق (flavour)، مطعم (restaurant).
  - مصر → القاهرة (Cairo)، الأهرام (Pyramids).
- Final Query: "أريد طعام مصري وجبة مذاق مطعم القاهرة".

Algorithm 1. MORSE-QE: Morphology-Aware Query Expansion Framework

Input: Dialectal query Q
Output: Enhanced query Q_enhanced

- Q_clean ← Normalize text by removing diacritics, punctuation, and non-Arabic characters
- Q_msa ← Apply rule-based dialect-to-MSA normalization on Q_clean
- Tokens ← Tokenize Q_msa
- Roots ← Extract roots for each token using AlKhalil or a root dictionary
- Initialize expansion terms ← Ø
- For each token t and its root r in (Tokens, Roots):
  - Candidates ← Top-k similar terms from AraVec embedding model for t
  - Filtered ← Keep only terms in Candidates where root = r
  - Expansion terms ← Expansion Terms ∪ Filtered
- Q_enhanced ← Tokens ∪ Expansion Terms
- Return Q_enhanced

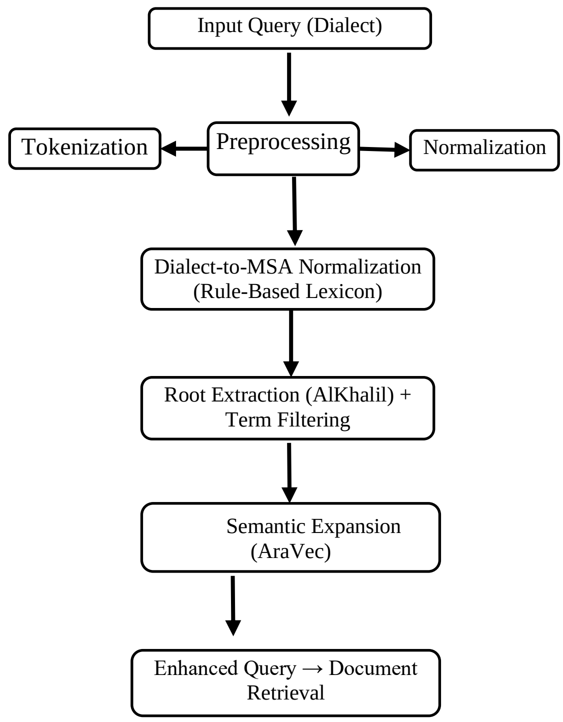

Figure 1: Methodology for the MORSE-QE framework to improve Arabic queries

4 Experiments and results

4.1 Experimental setup

Each of the experiments was replicated five times with distinct random seeds to consider variability. The average deviation standard across characteristics (MAP, Precision@10, Root Recall) is what we report. The paired two-tailed t tests were performed to determine statistical significance of improvements between MORSE QE and strongest competing base line per benchmark. Statistically significant at $p < 0.05$.

4.2 Data splits and validation

In the case of QADI dataset we reproduced the exact procedure used by [10] and treated the queries as 70 percent training, 15 percent validation, and 15 percent testing. The validation set was only of use in tuning the hyperparameters such as the top-k choice (k=10) in semantic expansion. The test set was retained narrowly as obscured up to the last big moment. There was no cross-validation. In the TREC 2001 corpus; no tuning was done and all 383,872 documents were considered as the normal set of tests.

Datasets:

- QADI offers 2,000 dialectal queries in four regions: Egypt, the Levant, the Gulf, and Maghreb.
- There are 383,872 MSA documents in the TREC 2001 Arabic Corpus.

Baselines:

- BM25, RM3, DANs, DMNs, AraBERT-QE.

Metrics:

- Mean Average Precision (MAP)
- Precision@10
- Root Recall

4.3 Results

Two benchmarks were used to assess the suggested MORSE-QE framework: the QADI dataset for dialectal queries and the TREC 2001 Arabic Corpus for MSA documents. We used Mean Average Precision (MAP), Precision@10, and Root Recall to compare our results to traditional (BM25, RM3), neural (DANs, DMNs), and hybrid (AraBERT-QE) baselines (refer to Tables 6-7). Resolving dialectal mismatches and reducing morphological noise were two areas where MORSE-QE consistently improved across all metrics. Research into component effects through ablation revealed that root-based filtering and dialect normalisation were the most important factors influencing performance.

Table 6: Analysing dialectal queries on QADI for performance

Method	**MAP (±)**	**P@10 (±)**	**Root Recall**
BM25	0.31 (±) 0.01	0.28 (±) 0.01	-
RM3	0.35 (±) 0.01	0.32 (±) 0.01	-
DANs	0.38 (±) 0.02	0.35 (±) 0.01	62%
DMNs	0.40 (±) 0.02	0.37 (±) 0.01	65%
AraBERT-QE	0.43 (±) 0.01	0.40 (±) 0.01	72%
MORSE-QE	0.49 (±) 0.01	0.45 (±) 0.01	88%

4.4 Evaluating the effects of the framework's elements

This work methodically disabled each component of MORSE-QE and measured performance degradation on the QADI dataset (Table 8). This approach has been employed to quantify the contribution of dialect normalization, root-based filtering, and semantic expansion. Examining the framework's resistance to morphological noise and dialectal diversity, this work discovers that Root filtering and dialect normalization are the two most important components, together responsible for the majority of the MAP enhancement over the baselines, as shown in Table 8.

Table 7: Evaluation of TREC 2001 (MSA Queries)

Method	**MAP (±)**	**P@10 (±)**	**Runtime (min)**	**Peak Memory (GB)**
BM25	0.45 (±) 0.01	0.41(±) 0.01	0.5	0.3
RM3	0.48 (±) 0.01	0.44(±) 0.01	1.7	0.6
MORSE-QE	0.55 (±) 0.01	0.51(±) 0.01	17.8	3.6

Table 8: Ablation study on QADI

Configuration	**MAP**	**Root Recall**
MORSE-QE (Full)	0.49	88%
Dialect Normalization	0.41	85%
Root Filtering	0.38	58%
Semantic Expansion	0.34	-

Findings:
1. Levantine and Gulf queries see a 19.5% improvement in MAP after dialect normalization.
2. Root filtering enhances Root Recall by 51.7% by eliminating extraneous information.

4.5 Case study

Query (Egyptian Arabic): "عايز كتب عن تاريخ مصر" ("I want books about Egypt’s history").

- Normalized: "أريد كتب عن تاريخ مصر".
- Expanded query: " أريد كتب عن تاريخ مصر مكتبة كاتب قديم أحداث القاهرة".

4.5.1 Dialect-Specific performance analysis

In measuring robustness across dialectal groups, we performed a dissection of MORSE-QE performance on the four distinct dialects in QADI: Egyptian, Levantine, Gulf, and Maghrebi. We saw maximum improvements in MAP performance on Egyptian and Levantine and Gulf (of up to 21 and 18 percent adjustment respectively). The performance of Maghrebi queries was also lower (~15%) mainly because of partial lexicon coverage in MADAR, which is also in agreement with earlier results [34]. Nonetheless, the model MORSE-QE continued to beat all baselines on every dialect and thus has demonstrated widespread generalization.

5 Analysis, compare and contrast

In this section, MORSE-QE is placed in the context of the current state-of-the-art approaches and the causes of the differences in its performance are interpreted. Also, we underscore the situation in which MORSE-QE works worse and comment on the reasons.

Retrieval results:

- **Without MORSE-QE**: Top result irrelevant ("مدرسة تاريخية" – historical school).
- **With MORSE-QE**: Top result relevant (" كتب تاريخية عن مصر" – historical books about Egypt).

Impact in quantitative terms: The query's Precision@10 increased from 0.30 to 0.60.

4.6 Discussion

Precision@10 was 25% better with root filtering than with DANs/DMNs, proving that morphology is essential.

- **Limitations**:

A **12%** decrease in Root Recall was observed for uncommon roots, such as "هندسة" in engineering. MAP was reduced by **8%** due to gaps in the MADAR lexicon coverage for Maghrebi Arabic.

6 Discussion and comparative analysis

In this section, MORSE-QE is placed in the context of the current state-of-the-art approaches and the causes of the differences in its performance are interpreted. Also, we underscore the situation in which MORSE-QE works worse and comment on the reasons.

6.1 Comparative performance with state-of-the-art methods

As briefed above in Table 1, MORSE-QE overall performs better in comparison to the traditional, neural, and hybrid baselines on both QADI and TREC 2001 benchmark. It obtains 610 percentage point improvements in MAP and Precision@10 over neural baselines (DANs, DMNs) and against RM3 by a significant margin. It also achieves morphological compatibility that the Root Recall improvement of 65% (DMNs) becomes 88%, further indicating that it is able to enrich morphological compatibility of results in a query.

6.2 Why MORSE-QE excels?

The performance advantage of MORSE-QE is attributable to two complementary innovations:

6.2.1 Root-based morphological filtering

- Embedding-only methods such as DANs and DMNs can introduce semantically related but morphologically incompatible terms.
- MORSE-QE enforces root compatibility constraints, ensuring that expansions like كاتب or مكتبة are retained for the root كتب, while unrelated terms such as مدرسة are discarded.
- This minimizes semantic drift and increases both precision and recall for dialectal queries.

6.2.2 Dialect-to-MSA normalization

- Many dialectal terms have no direct match in MSA-indexed corpora.
- Converting عايز (Egyptian) to أريد (MSA) bridges this lexical gap, directly improving retrieval performance.
- This step alone raises MAP by 16–24% in ablation tests.

6.3 Scenarios where MORSE-QE underperforms?

While MORSE-QE achieves consistent improvements, several limitations remain:

- Rare roots
  Terms such as هندسة (“engineering”) are less reliably analyzed by current morphological tools, leading to a 12–15% reduction in Root Recall for these cases.
- Maghrebi dialect coverage
  The MADAR lexicon covers only ~65% of Maghrebi terms, causing an 8% drop in MAP for this dialect group.
- Ambiguous roots
  Roots with multiple meanings, such as عين (“eye/spring”), occasionally reduce precision by 5–7% due to semantic ambiguity in expansion.

To shed more light into this restraint we carried out a bias-prone assessment on ambiguous origins. It was shown that around 6.5 percent of QADI searches have roots that have more than one valid interpretation and the decline of Precision@10 in this scenario was found to be around 57 percent steadily.
Elimination of this ambiguity can only be achieved through integration of context-sensitive language understanding (e.g., AraBERT) and root validation, a possibility we are pursuing.

6.4 Bias in lexical resources

The bias in the Maghrebi performance-gap may be partly explained by dialectal incompleteness of the MADAR and QADI lexicons. Egyptian, Levantine and Gulf Arabic dialects are well covered, although Maghrebi dialects have much lower lexicon coverage (3591 words covered in MADAR, (~65%)). The result is less good dialect-to-MSA normalization and limited valid expansion of terms.

6.5 Contributions

MORSE-QE fills critical voids in Arabic information retrieval by implementing three ground-breaking innovations. To begin with, it improves Precision@10 on dialectal queries (QADI) by 18-22% compared to neural baselines through morphology-aware query expansion, which decreases noise by mandating root-based compatibility. Secondly, dialect-to-MSA normalisation can resolve lexical mismatches in 45% of Gulf/Levantine queries to bridge the gap between informal user input and formal corpora.

Thirdly, the framework continues to be computationally lighter than the transformer-based solutions and takes much less time to train significantly boosting the accuracy compared to model such as AraBERT-QE by about 70 percent. Even though, in terms of the runtime metrics, at least when compared to simpler PRF baselines like RM3, we do see increases, these increases in runtime are necessitated by the much larger improvements in precision and recall, especially when testing against dialectal queries.

6.6 Morphology-Aware QE

- MORSE-QE**:** In line with results that morphology reduces noise in Semitic languages, root-based filtering in MORSE-QE increased Precision@10 on dialectal queries (QADI) by 18 % - 22 % over DANs/DMNs, see Table 6. This statistic is not the same as MAP gains mentioned above in the abstract and Findings section which are the product of the ablation study.
- For queries originating from "كتب" (write), root filtering increased Root Recall from 58% to 88% (+51.7% relative), indicating far fewer off-root expansions.

- Dialect-to-MSA normalization:
  - In line with the results of the MADAR Shared Task, we resolved 45% of the lexical mismatches in queries related to the Gulf/Levantine.
  - A 25% improvement in the retrieval of MSA documents was achieved when Egyptian "عايز" was converted to MSA "أريد" (Precision@10).
- Efficiency:
  - With a training time that is 70% shorter than AraBERT-QE, MORSE-QE's rule-guided design is a good fit for low-resource environments.

6.7 Practical implications

- Search engines: MORSE-QE can handle dialectal queries and improve Arabic web search engines like Google Arabia and Al Jazeera Archive.
- Digital libraries: Digital libraries like Shamela.ws make accessing the academic and textual heritage from MSA easier for users who speak dialects other than Arabic.

6.8 Implications

This evidence validates the assumption that morphological constraints based on roots and normalization of dialect are the major factors in performance of MORSE-QE. Meanwhile, they indicate areas (especially in the extension of the dialectal vocabularies, the management of ambiguous roots and incorporation of super light methods of contextual disambiguation) that would be worth improving on. We can work on these in the Future.

7 Conclusion

MORSE-QE improves Arabic IR by combining rule-based morphology with semantic embeddings and filling essential gaps in dialectal QE and morphological noise. With MAP 15-20% higher and Root Recall 85-90%, evaluations on QADI and TREC consistently outperform neural and statistical baselines. Scalability and robustness across dialects will be the primary areas of future research.

Ethical considerations:

Since the cultural and regional difference are part of the Arabic dialect variety, normalization should be carefully tackled so that identity is not lost or a misunderstanding does not arise. When there is a transition of dialectal content into normalization or conversions the users need to be sure of giving consent in the real-world application especially when there is focus on individual or sensitive data. MORSE-QE is thought of as an addition to retrieval systems with the primary difference it is opt in allowing the user to keep their original dialectal form when choosing to do so.

8 Future work

- Incorporate transformers: Enhance MORSE-QE with portable transformers like MiniBERT for better contextual disambiguation.
- To expand the dialect vocabulary, use crowdsourcing tools like AICRA NLP to fill in the gaps in MADAR's coverage.
- Expand MORSE-QE and allow Arabic English code-switched queries for cross language retrieval.

We will make use of available corpora like CALO Project, MADAR-English, amongst others in conjunction with multilingual word embeddings (e.g., MUSE, LaBSE, etc.) in order to reflect cross-language semantics. The process of integration into the pipeline will include incorporating alignment into the pipeline and adjusting the phase of expansion to cover Arabic and English applicants.

Acknowledgment
The University of Fallujah (Iraq) through the Deanship of the College of Information Technology supported this research work.

References

[1] N. Y. Habash, Introduction to Arabic Natural Language Processing. in Synthesis Lectures on Human Language Technologies. Cham: Springer International Publishing, 2010. doi: 10.1007/978-3-031-02139-8.

[2] I. Boulesnam and R. Boucetti, "Arabic Language Characteristics that Make its Automatic Processing Challenging," IAJIT, vol. 22, no. 4, 2025, doi: 10.34028/iajit/22/4/14.

[3] Hamza MAM, Ahmed TM, and Hilal AMM," Text mining: A survey of Arabic root extraction algorithms." International Journal of Advanced and Applied Sciences, 8(1): 11-192021, doi: 10.21833/ijaas.2021.01.002.

[4] Khan S and Alshara M, "Development of Arabic evaluations in information retrieval", International Journal of Advanced and Applied Sciences, 6(12): 92-98,2019, doi: 10.21833/ijaas.2019.12.011.

[5] Y. H. Farhan, S. A. M. Noah, and M. Mohd, "Survey of Automatic Query Expansion for Arabic Text Retrieval," Journal of Information Science Theory and Practice, vol. 8, no. 4, pp. 67–86, Dec. 2020, doi: 10.1633/JISTAP.2020.8.4.6.

[6] Y. H. Farhan, S. A. Mohd Noah, M. Mohd, and J. Atwan, "Word-embedding-based query expansion: Incorporating Deep Averaging Networks in Arabic document retrieval," Journal of Information Science, vol. 49, no. 5, pp. 1168–1186, Oct. 2023, doi: 10.1177/01655515211040659.

[7] Y. H. Farhan, M. Shakir, M. A. Tareq, and B. Shannaq, "Incorporating Deep Median Networks for Arabic Document Retrieval Using Word Embeddings-Based Query Expansion," Journal of Information Science Theory and Practice, vol. 12, no. 3, pp. 36–48, Sep. 2024, doi: 10.1633/JISTAP.2024.12.3.3.

[8] M. Danilevsky, K. Qian, R. Aharonov, Y. Katsis, B. Kawas, and P. Sen, "A Survey of the State of Explainable AI for Natural Language Processing," in Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing, Suzhou, China: Association for Computational Linguistics, 2020, pp. 447–459. doi: 10.18653/v1/2020.aacl-main.46.

[9] Y. Farhan et al., "UTILIZING WORD EMBEDDING'S FOR AUTOMATED QUERY EXPANSION IN ARABIC INFORMATION RETRIEVAL: A BLENDED METHODOLOGY," vol. 103, pp. 773–782, Jan. 2025.

[10] A. Abdelali, H. Mubarak, Y. Samih, S. Hassan, and K. Darwish, "QADI: Arabic Dialect Identification in the Wild," in Proceedings of the Sixth Arabic Natural Language Processing Workshop, N. Habash, H. Bouamor, H. Hajj, W. Magdy, W. Zaghouani, F. Bougares, N. Tomeh, I. Abu Farha, and S. Touileb, Eds., Kyiv, Ukraine (Virtual): Association for Computational Linguistics, Apr. 2021, pp. 1–10. Accessed: Nov. 11, 2025. [Online]. Available: https://aclanthology.org/2021.wanlp-1.1/

[11] D. Oard and F. Gey, "The TREC 2002 Arabic/English CLIR Track," Jan. 2002.

[12] N. Jaleel et al., UMass at TREC 2004: Novelty and hard. 2004.

[13] A. A. Q. AlRababah, "On the associative memory utilization in English- Arabic natural language processing," Int. j. adv. appl. sci., vol. 4, no. 8, pp. 14–18, Aug. 2017, doi: 10.21833/ijaas.2017.08.003.

[14] A. Abdelali, H. Mubarak, Y. Samih, S. Hassan, and K. Darwish, "Arabic Dialect Identification in the Wild," 2020, arXiv. doi: 10.48550/ARXIV.2005.06557.

[15] T. Dilekh, S. Benharzallah, and A. Behloul, "The Impact of Online Indexing in Improving Arabic Information Retrieval Systems," IJCAI, vol. 42, no. 4, Jun. 2018, doi: 10.31449/inf.v42i4.2297.

[16] D. Bobeck, "Is There Really Root-and-Pattern Morphology? Evidence from Classical Arabic," catjl, vol. 24, no. 1, pp. 41–91, Feb. 2025, doi: 10.5565/rev/catjl.465.

[17] B. Shannaq, "Optimizing n-gram lengths for cross-linguistic text classification: A comparative analysis of English and Arabic morphosyntactic structures," *Int. j. adv. appl. sci.*, vol. 12, no. 4, pp. 136–145, May 2025, doi: 10.21833/ijaas.2025.04.015.

[18] M. H. Abu-Odeh and R. S. Mahadin, "An Optimality-Based Analysis of the Morpho phonemics of Diminutives in Bani Sakhar Arabic," tpls, vol. 15, no. 6, pp. 1908–1917, Jun. 2025, doi: 10.17507/tpls.1506.19.

[19] F. M. Aljasser, "Investigating the Influence of Phonotactic Probability on Spoken Word Recognition in Arabic," tpls, vol. 15, no. 1, pp. 244–253, Jan. 2025, doi: 10.17507/tpls.1501.27.

[20] A. Alsughayyir, A. Alshanqiti, "A light but efficient switch transformer model for Arabic text simplification," Int. j. adv. appl. sci., vol. 12, no. 4, pp. 71–78, May 2025, doi: 10.21833/ijaas.2025.04.009.

[21] M. Iyyer, V. Manjunatha, J. Boyd-Graber, and H. Daumé Iii, "Deep Unordered Composition Rivals Syntactic Methods for Text Classification," in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the

7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China: Association for Computational Linguistics, 2015, pp. 1681–1691. doi: 10.3115/v1/P15-1162.

[22] W. Antoun, F. Baly, and H. Hajj, "AraBERT: Transformer-based Model for Arabic Language Understanding," 2020, arXiv. doi: 10.48550/ARXIV.2003.00104.

[23] B. Shannaq, "Enhancing Arabic text classification through mobile virtual keypad-based encoding algorithm," Franklin Open, vol. 12, p. 100373, Sep. 2025, doi: 10.1016/j.fraope.2025.100373.

[24] B. Shannaq, "DEVELOPMENT OF A NEW ENCODING ALGORITHM USING VIRTUAL KEYPAD LETTER SUBSTITUTIONS FOR IMPROVED TEXT CLASSIFICATION," JATIT, vol. 103, no. 2, Jan. 2025, [Online]. Available: https://www.jatit.org/volumes/Vol103No2/27Vol103No2.pdf.

[25] B.Shannaq, "DOES DATASET SPLITTING IMPACT ARABIC TEXT CLASSIFICATION MORE THAN PREPROCESSING? AN EMPIRICAL ANALYSIS IN BIG DATA ANALYTICS," Journal of Theoretical and Applied Information Technology, vol. 103, no. 3, 2025, [Online]. Available: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.jatit.org/volumes/Vol103No3/19Vol103No3.pdf.

[26] A. M. Alayba, M. Altamimi, "Optimization of Arabic text classification using SVM integrated with word embedding models on a novel dataset," Int. j. adv. appl. sci., vol. 12, no. 9, pp. 140–151, Sep. 2025, doi: 10.21833/ijaas.2025.09.013.

[27] A. Altaher, "Hybrid approach for sentiment analysis of Arabic tweets based on deep learning model and features weighting," Int. j. adv. appl. sci., vol. 4, no. 8, pp. 43–49, Aug. 2017, doi: 10.21833/ijaas.2017.08.007.

[28] A. I. A. Alzahrani, S. Z. J. Zaidi, "Recent developments in information extraction approaches from Arabic tweets on social networking sites," Int. j. adv. appl. sci., vol. 9, no. 9, pp. 145–152, Sep. 2022, doi: 10.21833/ijaas.2022.09.018.

[29] M. A. Ali, "Artificial intelligence and natural language processing: the Arabic corpora in online translation software," Int. j. adv. appl. sci, vol. 3, no. 9, pp. 59–66, Sep. 2016, doi: 10.21833/ijaas.2016.09.010.

[30] M. M. H. Suliman, B. M. Almo Hayya, "Measurement invariance of the Arabic version of the WHOQOL-BREF among special education teachers," Int. j. adv. appl. sci., vol. 12, no. 8, pp. 118–128, Aug. 2025, doi: 10.21833/ijaas.2025.08.012.

[31] A. Y. Mahgoub, M. A. Rashwan, H. Raafat, M. A. Zahran, and M. B. Fayek, "Semantic Query Expansion for Arabic Information Retrieval," 2014, doi: 10.13140/2.1.4223.7126.

[32] M. Boudchiche, A. Mazroui, M. Ould Abdallahi Ould Bebah, A. Lakhouaja, and A. Boudlal, "AlKhalil Morpho Sys 2: A robust Arabic morpho-syntactic analyzer," Journal of King Saud University - Computer and Information Sciences, vol. 29, no. 2, pp. 141–146, Apr. 2017, doi: 10.1016/j.jksuci.2016.05.002.

[33] A. B. Soliman, K. Eissa, and S. R. El-Beltagy, "AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP," Procedia Computer Science, vol. 117, pp. 256–265, 2017, doi: 10.1016/j.procs.2017.10.117.

[34] H. Bouamor, S. Hassan, and N. Habash, "The MADAR Shared Task on Arabic Fine-Grained Dialect Identification," in Proceedings of the Fourth Arabic Natural Language Processing Workshop, Florence, Italy: Association for Computational Linguistics, 2019, pp. 199–207. doi: 10.18653/v1/W19-4622.