International Journal of Leading Research Publication (IJLRP)
E-ISSN: 2582-8010 ● Website: www.ijlrp.com ● Email: editor@ijlrp.com

Enhancing Arabic Search Functionality for E-Commerce Websites: Challenges and Solutions

Devender Yadav

Abstract

The expanding e-commerce sector in the Arab world offers various opportunities and challenges. Online shopping is increasingly popular; however, the complexities of the Arabic language can hinder the user experience in product searches. This study examines the linguistic challenges that impede effective Arabic search on e-commerce platforms. This study examines issues related to morphological complexity, diacritical marks, dialectal variations, and the lack of standardized transliteration. This study proposes practical and innovative solutions utilizing natural language processing, machine learning, and an enhanced understanding of user search behavior within the Arabic-speaking context. The potential of utilizing phonetic algorithms and integrating voice search to enhance the online shopping experience in Arabic is also discussed. This study seeks to offer a framework for e-commerce enterprises to improve their search functionalities, resulting in increased user satisfaction and improved business performance in the Arabic-speaking market.

Keywords: Arabic language processing, e-commerce, search optimization, natural language processing (NLP), morphology, dialectal variation, user experience, Middle East and North Africa (MENA), online shopping, transliteration, voice search, stemming, diacritics

Introduction

The digital revolution has impacted the globe, including the Middle East and North Africa (MENA) region. E-commerce is undergoing significant growth, driven by rising internet access, smartphone usage, and a youthful, technologically adept demographic inclined towards online shopping. The distinct features of the Arabic language present particular challenges in facilitating a seamless and effective online shopping experience. The complexity of search functionality is central to this issue. Arabic's complex morphology, the application of diacritics, and the variety of dialects present significant challenges for e-commerce platforms aiming to link Arabic-speaking consumers with their desired products.

Consider it from the viewpoint of the user. An Arabic speaker is searching for a specific type of coffee pot online. One may input a term that possesses various potential spellings, contingent upon one's dialect or degree of formality. A commonly used term in your region may be well understood locally, yet it may not align with the official terminology employed in product listings. An inadequately sophisticated search engine on an e-commerce website may yield irrelevant results or, in the worst case, a "no results found" message. This results in frustration, inefficiencies, and possible revenue loss for the business.

The challenge is not solely technical; it is fundamentally cultural and linguistic. This involves analyzing the cognitive processes, search behaviors, and interactions of Arabic speakers with online platforms. It necessitates progressing beyond basic keyword matching to explore the semantic subtleties of the language. Despite advancements in Arabic natural language processing (NLP), its implementation in the e-commerce sector remains underdeveloped. This paper seeks to address the existing gap by providing a comprehensive analysis of the challenges and suggesting practical solutions to improve Arabic search functionality, thereby enhancing the online shopping experience for millions of users in the MENA region.

Problem Statement

The primary issue is that existing search functionalities on numerous e-commerce platforms inadequately address the complexities of the Arabic language, leading to suboptimal search results, user dissatisfaction, and missed business opportunities. The Arabic language, characterized by its complex morphology and varied dialects, poses distinct challenges:

1. **Morphological Complexity:**Arabic words exhibit morphological complexity, typically originating from a three-consonant root. Various prefixes, suffixes, and infixes are employed to generate diverse forms and meanings [1]. A single word can exhibit numerous variations, complicating the ability of a basic keyword-based search engine to identify all pertinent forms. A search for "book" (كتاب) may overlook results that include related terms such as "books" (كتب), "writer" (كاتب), or "library" (مكتبة).
2. **Diacritical Marks (Tashkeel):**Diacritical marks, known as Tashkeel, are frequently excluded in informal writing. They denote vowel sounds and can substantially change the meaning of a word. Their inconsistent application in product descriptions and user search queries generates ambiguity and obstructs precise matching. The term "مُعَلِّم" (teacher, with diacritics) is frequently rendered as "معلم" (without diacritics), potentially leading to confusion with "مَعْلَم" (landmark) [2].
3. **Dialectal Variations:**Arabic exhibits considerable regional dialectal variations, characterized by distinct differences in vocabulary, pronunciation, and grammatical structures. A user utilizing a specific dialect may not locate products that are described in Modern Standard Arabic (MSA) or in an alternative dialect. A user in Egypt may search for a specific type of cooking pot using a colloquial term that is not widely recognized in Saudi Arabia [3].
4. **Lack of Standardized Transliteration:**The absence of a universally accepted standard for the transliteration of Arabic words into Latin script is evident. This poses challenges when users input Arabic using Roman characters, as the search engine may fail to identify the intended Arabic term. Users may express the term for "phone" as "telefon," "telephone," or employ a mix of numbers and letters, such as "h4t3f," to approximate the phonetics of the Arabic word for phone [4].
5. **Contextual Ambiguity:**Arabic is a language rich in context, where the meaning of a word can change based on its surrounding words. Search engines lacking comprehension of these contextual nuances may yield irrelevant results.
6. **Right-to-Left Script:**Arabic, in contrast to Latin-based languages, is written from right to left. This may result in complications for search interfaces primarily designed for left-to-right languages, thereby causing usability challenges.
7. **Lack of Data:**In contrast to English, there exists a notable deficiency of high-quality, labeled Arabic text data for the training of advanced NLP models. This restricts the development of

advanced search algorithms designed to address the specific nuances of Arabic e-commerce.

8. **Search Intent:**The search functionality faces challenges in determining whether the user is seeking a specific product, a category of products, or general information.
9. **Word Order:**In contrast to English, Arabic exhibits significant variability in word order without altering the meaning. This may influence the manner in which search engines handle queries.

The linguistic challenges, along with the limitations of current search technologies, pose a substantial barrier to effective e-commerce in the Arabic-speaking region.

Solution

The solution fundamentally involves the development of an advanced search architecture capable of comprehending and processing the complexities of the Arabic language. This requires advancing past basic keyword matching to establish a system that accounts for morphology, diacritics, dialects, transliteration variations, context, and user intent.

The solution can be broken in following components:

1. Advanced Morphological Analysis:

i. **Stemming and Lemmatization:**Theseare natural language processing techniques used to reduce words to their base or root form. Stemming involves truncating words to their base form, while lemmatization considers the context and converts words to their dictionary form.

   a. **Algorithm Selection:**It is advisable to move beyond basic stemming algorithms and investigate more sophisticated options tailored for Arabic, including the Khoja stemmer and the Tashaphyne light stemmer. These algorithms are more adept at managing the intricate patterns of Arabic morphology [5].
   b. **Root Extraction:**The system must utilize a morphological analyzer, such as MADAMIRA, to extract the root of a word and provide additional morphological features, including part of speech, gender, and number.
   c. **Context-Aware Lemmatization:**Develop lemmatization tools that take into account the surrounding words to accurately identify the appropriate lemma. The term "عمل" may function as either a verb (worked) or a noun (work), with the context determining the appropriate lemma [6].
   d. **Dictionary-Based and Rule-Based Approaches:**Dictionary-based and rule-based approaches integrate dictionary lookups with rule-based systems to address out-of-vocabulary words and enhance accuracy [7].

ii. **Root-Based Search:**

   a. **Root Index:**Establish an independent index that catalogs the roots of all terms within the product descriptions.
   b. **Query Root Extraction:**Upon user input of a query, identify and extract the root(s) of the query terms.
   c. **Root Matching:**Conduct a search in the root index for the extracted roots and obtain all related words.

iii. **Handling of Prefixes and Suffixes:**

   a. **Affix Stripping:** Develop algorithms that can identify and strip common Arabic prefixes (e.g., "ال" - the, "و" - and, "ف" - so) and suffixes (e.g., "ات" - plural feminine, "ون" - plural

masculine).

b. **Affix-Aware Search:**Create algorithms to identify and remove common Arabic prefixes (e.g., "ال" - the, "و" - and, "ف" - so) and suffixes (e.g., "ات" - plural feminine, "ون" - plural masculine).

c. Partial Matching: Enable partial matching of prefixes and suffixes to account for slight variations in spelling or dialect.

2. Intelligent Handling of Diacritics:

i. **Diacritic-Insensitive Search:**

a. **Normalization:**Normalization involves generating a diacritic-normalized version of the product catalog by eliminating all diacritics.

b. **Query Normalization:**Query normalization involves the removal of diacritics from user queries prior to executing the search.

c. **Phonetic Matching:**Phonetic Matching: Utilize phonetic algorithms to match words according to their pronunciation, regardless of variations in diacritics [8].

ii. **Diacritic Restoration:**

a. **Statistical Machine Translation (SMT):**Statistical Machine Translation (SMT) involves training SMT models on parallel corpora containing both diacritic and non-diacritic text to effectively learn the restoration of diacritics [9].

b. **Recurrent Neural Networks (RNNs):**Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks or Gated Recurrent Units (GRUs), are employed to model the sequential characteristics of Arabic text for accurate diacritic placement prediction [10].

c. **Hybrid Approaches:**Integrate SMT and RNN models to utilize the advantages of each methodology.

d. **User Feedback:**User feedback should be solicited regarding the accuracy of diacritic restoration, which can subsequently inform enhancements to the models.

e. **Dictionary Lookup:**Utilize a diacritized lexicon to address ambiguities and enhance accuracy.

3. Dialectal Accommodation: It refers to the adjustments individuals make in their speech patterns to align with those of their conversational partners. This phenomenon often occurs in social interactions and can reflect various factors, including social identity and group dynamics.

i. **Dialect Identification:**

a. **Language Models:**Train language models on text data from various Arabic dialects to determine the dialect of a user's query through analysis of word usage and grammatical structures [11].

b. **Location Data:**Utilize available user location data to infer the probable dialect.

c. **User Profiles:**User profiles should include an option for users to specify their preferred dialect.

ii. **Dialectal Lexicons:**

a. **Crowdsourcing:**Crowdsourcing involves recruiting native speakers from various regions to develop and sustain dialectal lexicons.

b. **Web Crawling:**Create targeted web crawlers to gather textual data from websites and social media platforms tailored to various regions.

c. **Mapping:**Mapping involves establishing correspondences between dialectal terms and their Modern Standard Arabic (MSA) equivalents, as well as between various dialectal variations [12].

iii. **Machine Translation:**Machine translation refers to the automated process of converting text from one language to another using computer algorithms and models.

a. **Dialect-Specific MT Models:**Develop distinct machine translation models for each major dialect to facilitate the translation of product descriptions into various dialects [13].

b. **Adaptive MT:**Adaptive machine translation systems should be developed to learn from user interactions, thereby enhancing translation quality progressively.

iv. **Query Rewriting:**Transform user queries into various dialectal forms to enhance search capabilities and obtain results from products characterized in diverse dialects.

4. Standardized Transliteration and Romanization Support:

i. **Fuzzy Matching:**Fuzzy matching refers to techniques used to identify similarities between data entries that may not match exactly [14]. This approach is often employed in data cleaning and record linkage to enhance the accuracy of data analysis.

a. **Edit Distance:**Implement algorithms to calculate the edit distance, such as the Levenshtein distance, between the user's query and transliterated product descriptions.

b. **N-gram Matching:**N-gram Matching involves employing n-gram models to detect words that exhibit comparable character sequences, despite minor variations in their transliteration.

ii. **Phonetic Search:**Phonetic search refers to the process of locating words or phrases based on their phonetic representation rather than their exact spelling. This approach is particularly useful in applications such as speech recognition and information retrieval, where variations in pronunciation may occur.

a. **Soundex and Metaphone:**Adapt or develop phonetic algorithms such as Soundex or Metaphone that are specifically designed for the phonetic characteristics of Arabic.

b. **Arabic-Specific Phonetic Algorithms:**Investigate and develop phonetic algorithms that consider the distinct phonetic characteristics of Arabic, including emphatic consonants and guttural sounds [15].

iii. **Multiple Transliteration Indexing:**

a. **Common Schemes:**Utilize various established transliteration schemes for indexing product descriptions, including Buckwalter, ArabTeX, and SATTS.

b. **User-Defined Schemes:**User-defined schemes enable individuals to specify their preferred transliteration method, where feasible.

5. Contextual Understanding:

i. **Semantic Search:**Semantic search refers to the process of improving search accuracy by understanding the intent and contextual meaning of search queries. This approach enhances the relevance of results by considering the relationships between words and concepts.

a. **Word Embeddings:**Word embeddings can be enhanced by employing pre-trained Arabic word embeddings such as AraVec or FastText, or by training custom embeddings on an extensive corpus of Arabic e-commerce data. These embeddings represent semantic relationships among words [16].

b. **Sentence Embeddings:**Sentence embeddings utilize models to represent the meanings of

complete queries and product descriptions as vectors. Assess the similarity between query and product vectors to prioritize results.

c. **Contextualized Embeddings:**Investigate the application of contextualized word embeddings, including those produced by BERT and other transformer-based models, which effectively capture a word's meaning in relation to its surrounding context.

ii. **Knowledge Graphs:**

a. **Product Relationships:**Develop a knowledge graph illustrating the interconnections among products, categories, brands, attributes, and concepts.

b. **Query Expansion:**Utilize the knowledge graph to enhance user queries by incorporating related terms and concepts, thereby improving recall.

c. **Reasoning:**Utilize graph-based reasoning methods to deduce connections between products and user inquiries, resulting in enhanced search result relevance.

iii. **Query Expansion:**Query Expansion refers to the process of reformulating a search query to improve retrieval performance. This technique aims to enhance the relevance of search results by incorporating additional terms or synonyms that align with the user's intent.

a. **Synonym Expansion:**Employ a thesaurus or synonym database to systematically incorporate synonyms of query terms into the search.

b. **Related Term Expansion:**Employ techniques such as word embeddings or collaborative filtering to identify terms frequently associated with the query terms and incorporate them into the search.

c. **Contextual Expansion:**Analyze the surrounding words in the query to determine pertinent terms that are relevant within that specific context.

6. Improved User Interface:

i. **Right-to-Left Support:**Facilitate uninterrupted right-to-left text input, display, and navigation across the search interface.

ii. **Auto-Complete and Suggestion:**

a. **Morphological Awareness:**Create auto-complete algorithms capable of suggesting morphologically related words, considering factors such as stemming, lemmatization, and dialectal variations [17].

b. **Contextual Suggestions:**Offer recommendations that align with the user's ongoing search context, taking into account their prior queries and browsing behavior.

c. **Transliteration Support:**Provide auto-complete suggestions in Arabic script and Roman characters to accommodate users who prefer typing in transliterated Arabic.

iii. **Voice Search:**

a. **Automatic Speech Recognition (ASR):**These models are specifically trained on Arabic speech data, encompassing various dialects.

b. **Natural Language Understanding (NLU):**Natural Language Understanding (NLU) involves employing NLU techniques to analyze the output generated by the ASR system and to identify the user's intent [18].

c. **Noise Robustness:**Create ASR models that exhibit resilience to background noise and variations in user accents.

iv. **Visual Search:**Enable users to conduct searches utilizing product images. This is especially beneficial when users are unaware of the precise Arabic term for a product.

v. **Faceted Search:**Implement filters enabling users to narrow search results based on attributes including category, brand, price, color, and size.

7. Data Augmentation and Collection:

i. **Crowdsourcing:**
   a. **Data Labeling:**Data labeling involves utilizing Arabic-speaking individuals to annotate and categorize data for the training of NLP models. This includes tasks such as identifying parts of speech, correcting diacritization, and mapping dialectal terms.
   b. **Translation:**Utilize crowdsourcing for the translation of product descriptions into various dialects.

ii. **Synthetic Data Generation:** Synthetic data generation refers to the process of creating artificial data that mimics real-world data characteristics [19]. This technique is employed to enhance data privacy, augment datasets, and facilitate machine learning model training without compromising sensitive information.
   a. **Back-Translation:**Employ machine translation to convert text from various languages into Arabic and subsequently revert it to the original language, resulting in variations of the initial text.
   b. **Generative Adversarial Networks (GANs):**Generative Adversarial Networks (GANs) can be utilized to produce synthetic Arabic text data, thereby enhancing existing datasets.

iii. **Web Crawling:**
   a. **Targeted Crawling:**Create web crawlers that concentrate on gathering pertinent Arabic text data from e-commerce sites, forums, and social media platforms within the MENA region.
   b. **Data Cleaning:**Employ effective data cleaning methods to eliminate noise, inconsistencies, and extraneous information from the collected data.

8. Search Intent Recognition:

i. **Query Classification:**
   a. **Machine Learning Models:**Machine learning classifiers, including SVM, Naive Bayes, and Random Forests, are employed to categorize user queries into distinct intent types, such as informational, navigational, or transactional.
   b. **Rule-Based Systems:**Create rule-based systems that utilize keywords, patterns, and grammatical structures to discern user intent.

ii. **Personalized Search:**
   a. **User Profiles:**User profiles should be developed to retain data regarding user preferences, historical purchases, and browsing activity.
   b. **Collaborative Filtering:**Collaborative filtering techniques can be employed to recommend products by analyzing the behavior of analogous users.
   c. **Session-Based Personalization:**Session-Based Personalization involves customizing search results according to the user's ongoing browsing session, considering the products viewed and the queries submitted.

9. Flexible Word Order Handling:

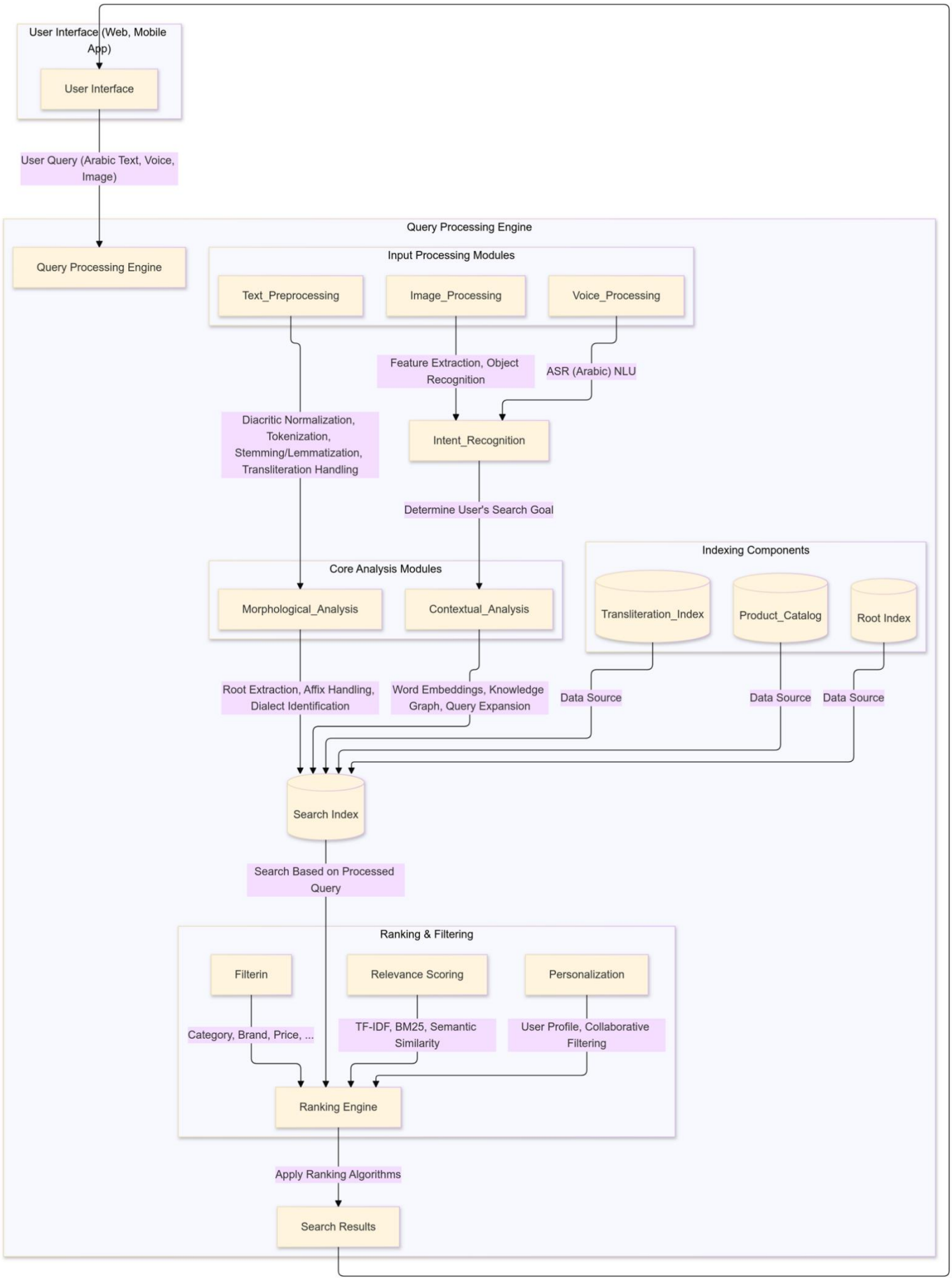
i. **Dependency Parsing:**
   a. **Dependency Trees:**Employ dependency parsing to construct dependency trees that illustrate the grammatical relationships among words in a query.

b. **Relationship Matching:**Align queries with product descriptions by focusing on the relationships among words instead of their precise sequence.

ii. **N-gram Models:**N-gram models are statistical language models that predict the likelihood of a sequence of words based on the occurrence of n-grams, which are contiguous sequences of n items from a given sample of text or speech. These models are widely used in natural language processing tasks, including speech recognition and machine translation.

a. **Statistical Language Modeling:**Statistical language modeling utilizes n-gram language models to identify statistical relationships among word sequences, enhancing the search engine's resilience to variations in word order [20].

b. **Smoothing Techniques:**Smoothing techniques are employed to address unseen n-grams and enhance the accuracy of the language model.


User Interface (Web, Mobile App)
User Interface
User Query (Arabic Text, Voice, Image)
Query Processing Engine
Query Processing Engine
Input Processing Modules
Text_Preprocessing
Image_Processing
Voice_Processing
Feature Extraction, Object Recognition
ASR (Arabic) NLU
Diacritic Normalization, Tokenization, Stemming/Lemmatization, Transliteration Handling
Intent_Recognition
Determine User's Search Goal
Core Analysis Modules
Morphological_Analysis
Contextual_Analysis
Indexing Components
Transliteration_Index
Product_Catalog
Root Index
Root Extraction, Affix Handling, Dialect Identification
Word Embeddings, Knowledge Graph, Query Expansion
Data Source
Data Source
Data Source
Search Index
Search Based on Processed Query
Ranking & Filtering
Filterin
Relevance Scoring
Personalization
Category, Brand, Price, ...
TF-IDF, BM25, Semantic Similarity
User Profile, Collaborative Filtering
Ranking Engine
Apply Ranking Algorithms
Search Results

Uses

The enhanced Arabic search functionality can be applicable for:

1. **Arabic-Speaking Consumers:**Arabic-speaking consumers will find products more easily and efficiently, resulting in a more satisfying online shopping experience.
2. **E-commerce Businesses:**E-commerce businesses will experience enhanced sales, greater customer satisfaction, and a more robust competitive advantage within the MENA market.
3. **Marketing Teams:**Marketing teams can utilize enhanced search data to gain insights into customer preferences, optimize product listings, and develop more effective marketing campaigns.
4. **Customer Service:**The implementation of an improved search function decreases inquiries concerning product discovery, thereby optimizing customer service resources.

Impact

The effective execution of these solutions will significantly influence multiple stakeholders:

1. **Improved User Experience:**Arabic-speaking users will navigate the online shopping process more intuitively, facilitating quick and effortless access to desired products.
2. **Increased Sales and Revenue:**Enhanced search result relevance will result in elevated conversion rates and increased sales for e-commerce enterprises.
3. **Enhanced Customer Loyalty:**A positive search experience enhances customer loyalty and promotes repeat business.
4. **Greater Market Penetration:**E-commerce enterprises can achieve enhanced market penetration in the MENA region by accessing a broader demographic of Arabic-speaking online consumers [21].
5. **Boost to the Digital Economy:** The expansion of e-commerce will enhance the overall development of the digital economy in the MENA region.
6. **Data-Driven Insights:**Enhanced search functionality will yield significant data regarding user preferences and behaviors, offering insights pertinent to product development, marketing, and business strategy.
7. **Competitive Advantage:**E-commerce platforms that effectively tackle the challenges associated with Arabic search will achieve a notable competitive advantage in the market [22].
8. **Job Creation:**The development and implementation of these solutions will generate new employment opportunities in fields such as natural language processing, software engineering, and data science.
9. **Cultural Preservation:**Enhancing online accessibility to products and information in Arabic will facilitate the preservation and promotion of the Arabic language in the digital era.
10. **Increased Accessibility:** Enhanced search will make online shopping more accessible to users with varying levels of digital literacy.

Scope

This study concentrates on improving search capabilities for e-commerce platforms aimed at the Arabic-speaking demographic. The principles and solutions presented are applicable to various domains, including online marketplaces and content platforms; however, the specific challenges and implementation details may differ. The geographical focus is predominantly the MENA region; however, the findings may also hold relevance for other areas with substantial Arabic-speaking

populations.

This study examines multiple facets of Arabic search, including:

1. **Text-based search:**Text-based search constitutes the primary focus, involving the analysis of user queries alongside product descriptions.
2. **Voice search:**The potential of voice search is examined as an emerging technology that may enhance user experience.
3. **Image search:**The potential for integrating image recognition with text-based search is recognized, although it is not the primary focus.

The study does not explore the complexities of payment gateways, logistics, or other elements of e-commerce that are not directly associated with search functionality.

Conclusion

Improving Arabic search functionality for e-commerce websites represents a significant technical challenge and is essential for accessing the extensive opportunities within the Arabic-speaking market. Addressing the linguistic complexities of Arabic and utilizing advanced NLP techniques enables e-commerce businesses to develop a more inclusive, user-friendly, and profitable online shopping experience. This necessitates a commitment to continuous research, development, and a thorough comprehension of the cultural and linguistic subtleties of the region.

This paper presents solutions that provide a roadmap for achieving the stated goal. These are not solely theoretical concepts; they represent practical strategies that can be implemented and refined over time. The evolving e-commerce landscape in the MENA region indicates that investment in optimizing Arabic search capabilities will enhance competitive positioning. The objective extends beyond simplifying product discovery; it involves establishing connections between businesses and consumers, promoting economic development, and empowering Arabic speakers in the digital era. The pursuit of effective Arabic search remains a work in progress; however, the potential benefits for those who engage with this challenge are substantial. The complexity of addressing issues related to language and culture is significant, and the solutions formulated today will influence the trajectory of e-commerce in the Arab world for the foreseeable future. The realization of contributing to the engagement of millions of users in the digital revolution is highly rewarding and renders this research trajectory genuinely fulfilling. This journey is valuable not only for its economic advantages but also for its significant positive effects on the lives of millions.

References

[1] M. Abdul-Mageed, M. T. Diab, and M. Korayem, “Subjectivity and sentiment analysis of modern standard Arabic and Arabic microblogs,” *Journal of the American Society for Information Science and Technology*, vol. 65, no. 10, pp. 2095–2110, 2014.

[2] M. N. Al-Kabi, A. A. Gigieh, K. Alsmearat, and M. A. Haidar, “An intelligent context-aware Arabic morphological analyzer for improving information retrieval,” *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 9, no. 2, pp. 50–77, 2013.

[3] A. Al-Thubaity, M. Al-Mulhem, and H. Al-Mubarak, “Arabic diacritization: A survey,” *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 13, no. 4, pp. 1–26, 2014.

[4] M. Attia, A. Toral, L. Tounsi, H. Sajjad, and L. Màrquez, “An automatically built lexicon for Arabic

morphological analysis and part-of-speech tagging," in *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*,[1] 2010, pp. 2471–2477.
[5] K. Darwish and W. Magdy, "Arabic information retrieval," *Foundations and Trends® in Information Retrieval*, vol. 7, no. 4, pp. 353–455, 2013.
[6] K. Darwish and W. Gao, "Simple effective Arabic language morphological stemming," *Journal of the American Society for Information Science and Technology*, vol. 65, no. 9, pp. 1876–1890, 2014.
[7] K. Darwish, "Building a shallow Arabic morphological analyzer in one day," in *Proceedings of the ACL-02 workshop on Computational*[2] *linguistics for the rapidly emerging varieties of Arabic-Volume 14*, 2002, pp. 1–8.
[8] A. De Roeck and W. Al-Fares, "A morphologically sensitive clustering algorithm for identifying Arabic roots," in *Proceedings of the*[3] *38th Annual Meeting of the Association for Computational Linguistics*,[4] 2000, pp. 199–206.
[9] M. Diab, K. Hacioglu, and D. Jurafsky, "Automatic tagging of Arabic text: From raw text to base phrase chunks," in *Proceedings of*[5] *the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*,[6] 2004, pp. 149–152.
[10] K. Duh, *Buckwalter Arabic Morphological Analyzer Version 1.0*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania, 2005.
[11] S. R. El-Beltagy and A. Rafea, "KP-Miner: A simple system for Arabic keyphrase extraction," *Journal of Computer Science*, vol. 5, no. 10, p. 772, 2009.
[12] A. Farghaly and K. Shaalan, "Arabic natural language processing: Challenges and solutions," *ACM Transactions on Asian Language Information Processing*[7] *(TALIP)*, vol. 8, no. 4, pp. 1–22, 2009.
[13] I. Guellil, H. Saâdaoui, F. Azouaou, N. Gueni, and S. Benyacoub, "Machine Learning-based Approach for Arabic Dialect Identification: A short survey," in *2020 7th International Conference on Computer and Communication Engineering (ICCCE)*, 2020, pp. 90–94.
[14] J. R. Hobbs, D. Appelt, J. Bear, D. Israel, M. Kameyama, M. Stickel, and M. Tyson, "FASTUS: A cascaded finite-state transducer for extracting information from natural-language[8] text," *Finite state devices for natural language processing*,[9] vol. 1, no. 383-406, 1997.
[15] S. Khoja and R. Garside, "Stemming Arabic text," Computer Science Department, Lancaster University, 1999.
[16] R. Kilany, "The Online E-Commerce in The Arab World," *International Journal of Online Marketing Research*, vol. 5, no. 1, pp. 1–7, 2019.
[17] L. S. Larkey, J. Allan, M. E. Connell, A. Bolivar, and C. Wade, "UMass at TREC 2002: Arabic/English CLIR," in *TREC*, 2002.
conditional random fields," *International Journal of Computer Processing of Oriental Languages*, vol. 25, no. 03, pp. 209–230, 2012.
[18] A. Pasha, M. Al-Badrashiny, M. Diab, A. El Kholy, R. Eskander, N. Habash, ... and O. Rambow, "Madamira: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic," in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, 2009.
[19] K. Shaalan, "A survey of Arabic named entity recognition and classification," *Computational Linguistics*, vol. 40, no. 2, pp. 469–510, 2014.
[20] O. Smrž and P. Smrž, "Functional morphology of Arabic: An approach to automatic parsing and generation," *The Prague Bulletin of Mathematical Linguistics*, no. 84, pp. 49–80, 2005.