# Comparative Analysis of Information Retrieval Models on Quran Dataset in Cross-Language Information Retrieval Systems

**AYMAN A. TAAN[1], SHAFIQ UR REHMAN KHAN[2], ALI RAZA[2], AYAZ MUHAMMAD HANIF[3], AND HIRA ANWAR[2]**

[1]Department of Computer Science, The Hashemite University, Zarqa 13133, Jordan
[2]Department of Computer Science, University of Sialkot, Sialkot 44000, Pakistan
[3]Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Haripur 22620, Pakistan

Corresponding author: Shafiq Ur Rehman Khan (shafiq.urrehman@uskt.edu.pk)

**ABSTRACT** English is an international language used for communication worldwide but still many cannot read, write, understand, or communicate in English. On the other hand, the World Wide Web has unlimited resources of information in different languages which English native find challenging to understand. To avoid such barriers, Cross-Language Information Retrieval (CLIR) systems are proposed, which refers to document retrieval tasks across different languages. This work focuses on the performance evaluation of different Information Retrieval (IR) models in CLIR system using Quran dataset. Furthermore, this work also investigated the length of query and query expansion models for effective retrieval. The results show that different length of queries has an impact on the performance of the retrieval methods in terms of effectiveness. Hence, after comprehensive experiments, an appropriate length of query for Arabic CLIR system is suggested along with the best query expansion and retrieval model.

**INDEX TERMS** Cross language information retrieval, query expansion, Arabic information retrieval, query length.

## I. INTRODUCTION

Individuals need the relevant information carved in their natural language, commonly in form of a query. The World Wide Web (WWW) offers a substantial amount of information in variegated languages. There are cumulative desires to the exploration of information in languages dissimilar to the query. For instance, retrieving documents written in the Arabic language with a query written in English. This creates a problem of Cross-Language Information Retrieval [1]–[3].

TThe representation issue is more obvious in CLIR and Multi-Lingual Information Retrieval (MLIR), in which the documents and the queries are defined in different languages. When the content is written in various languages, how can one build a similar inner depiction of queries when they ask for information? E.g., in what way we can identify the subsequent explanations define a similar portion of information? The important issue in CLIR is to implements a matching method in which terms similar to the other languages that define the similar sense. The translation model segment can be used in different ways such as a) representing the document into the query representation space, the method is known as Document Translation Approach DTA [4], b) b) representing the query into the document terms space, the method is known as Query Translation Approach QTA [5].

The Arabic language is unique to the six authorized languages of the United Nations (UN), and Arabic is spoken by 1.8 billion Muslims worldwide. Also, Arabic the mother language of over 400 millions. It is written right to left, and the elementary alphabet comprises 28 letters [6]. Classical Arabic is linguistic of Quran and is cast-off in modern Arabic text. Quran is the holy script [7], which comprises approximately 79,000 words making 114 chapters. Stimulating facts concerning the regular organization of the Quran are presented when Natural Language Processing (NLP) techniques took place. Foremost, a theory might be stated in different verses. For instance, the concept of Hell is deliberated in numerous verses and chapters. Furthermore, a single verse might comprise several themes. For example, verse 40 of Chapter 78 comprises solitary seven words telling five dissimilar thoughts like Allah, People, punishment, individual and, the judgment day [8], [9].

The associate editor coordinating the review of this manuscript and approving it for publication was Farhana Jabeen.

## A. MOTIVATION

These days, false information can lead a group of people to false beliefs, and thus hatred is spread all over the place. Religion is a very common practice in human society, and Islam has the second largest number of believers around the world. The misconception and misinterpretation about religion can cause distance among the people. Many want to know about the most authentic and undisputed religious resources, which indeed is the Quran, in Islam. The Quran is written in the Arabic language which is not a native language to a large population of the world. Thus, extracting information need from the holy Quran is a challenging task.

This study is driven by the same motivation that the IR system can help to fulfill the user information needs related to Quran. The main objectives of this study are; I) evaluating different retrieval models on the Quran dataset, ii) investigate the suitable length of user query that can extract relevant information in CLIR system and iii) this work also aims to carry out a systematic inquiry and examines the role of query expansion techniques for the performance enhancement of information retrieval models.

The motivation behind this study is presented in three research questions.

**RQ 1. Which ranking model achieves high effectiveness?**

**RQ 2. What is the appropriate query length for effective retrieval of Quran verses in the CLIR system?**

**RQ 3. Which query expansion technique can obtain the best query words?**

The rest of the paper is organized as, Section 2 discuss related work, Section 3 sheds light on the proposed methodology adopted to tackle the research questions. The experimental details and results are presented in Section 4 and finally, Section 5 concludes this research study.

## II. RELATED WORK

CLIR tackles the difference between the query language and document language. In order to resolve such differences, the system uses language translation techniques which remains a fundamental requirement. There are two basic approaches used i.e., Direct translation and Indirect translation. This section is comprised of some closely related work to translation and query expansion techniques.

### A. DIRECT TRANSLATION

Direct translation systems exploit bilingual dictionaries, parallel corpora, and machine translation algorithms to translate the source text. In such techniques, the text in one language is directly translated to another language without any mediator. The direct Translation (DT) technique can be categorized into three types, including dictionary-based translation, Machine-based translation, and corpora-based translation.

In the dictionary-based translation approach, bilingual dictionaries (machine-readable) are constructed and used in translation modules of CLIR systems [10], [11]. Various

public bilingual dictionaries are available free of cost. Despite its simplicity, dictionary-based method has two main problems associated with, i.e., a) ambiguity and b) lack of coverage. The former one (ambiguity) arises due to multiple translations of a single word, hence for a given query term. Selecting the appropriate translation terms is crucial for improving the retrieval effectiveness of the CLIR engine [12], [13]. Two types of solutions are proposed, a) single selection translators and b) multiple selection translators. In a single selection, the translator selects the single most suitable terms for individual query terms whereas, in the later solution more than one term is presented for each query term. The term co-occurrence method replaces this single method [14]–[16]. which states that correct translation of terms tends to co-occur as part of sublanguage whereas the incorrect translation is not. To improve the performance of multiple selector translations, bi-directional translation is used [17], [18].

As the name suggests, machine translation involve machines to translate the text from one language to another and becomes popular methods with the improvement of linguistic resources. Machine translations achieve up to 99% monolingual baseline on various language collections [19], [20], Google Translate API as one of the examples. Machine translation system achieves quite a remarkable improvement in language translation, yet these systems are still away from resolving language translation problems in CLIR. The effectiveness of the machine-based translation system varies from language to language. For a languages like Thai and Chinese, the MT system performs fifty percent below the monolingual baseline [20]–[22]. Dolomitic and using German query to search for a French document through Google Translate provide significantly poor results as compared to English language query. Ignoring OOV words affect significantly on the results of MT systems [23].

A corpus written in one language is translated into another language, such is called a parallel corpus can be acquired from different sources including United Nations (UN), Canadian parliament, Hong Kong Legislative Council and World Wide Web (WWW), etc. [24]. In such a technique, documents from different languages are analyzed side by side in order to generate a set of translation probabilities. The order of terms might be different in source and target text, so the order does not have any impact. The probabilistic model is proved to be more efficient as compared to alternative methods. Other resources can supplement such type of query translation. He and Wu (2008) combine parallel corpora and bilingual dictionaries techniques and observed positive results. One of the major drawbacks of the corpora-based technique is high time consumption [25]–[27]. Several alternatives have been proposed to overcome such drawbacks. These methods are proved to be successful but limited to some well-known languages. For an uncommon language, such a translation model needs more training material. Using comparable corpora might be one solution, which can be used to identify contexts containing similar expressions in two different languages [28], [29]. Another drawback of the

corpus-based translation technique that it ignores the terms that change over time and it gives new research direction in the mentioned domain.

## B. INDIRECT TRANSLATION

When there are no sufficient resources available for translation, indirect translation is an appropriate solution. As the name suggests, there is an intermediary source between the query source and the target document corpus [1], [30]–[32]. Such types are also called transitive translation. When both the query source and target collection are translated using the intermediate source, it is referred to as dual translation. In transitive translation approach, more than one intermediate source (language) is used and refer to as triangulated transitive approach where two pivot language used as intermediary sources [33]. In [33] the authors investigated the impact of pivot languages on probabilistic retrieval and revealed the lexical coverage of translation chain remains an important factor. The achieved results do not exceed the direct translation technique.

Latent Semantic Indexing (LSI) is among the earliest dual translation system [34] based on singular value decomposition using a linear algebra technique. LSI is used to extract important concepts from the text collection. Singular value decomposition presents a term-document matrix decomposed into two orthogonal matrices and a diagonal matrix. They found that LSI-based retrieval engines perform lower than the direct translation-based system regarding effectiveness. They further added LSI also contributed to high computation cost. Another approach named Latent Dirichlet Allocation (LDA) was proposed, featuring probabilistic approaches that also suffer from the same problem [35]. Alternatively, Explicit Semantic Analysis (ESA) was proposed for semantic similarity. A knowledge base (KB) is constructed as a first step and later ESA indexed documents based on term document association.

Document translation is a less popular CLIR technique but equally important. In such a method, the document is translated before indexing. The document translation process involved translating the document into the query language. Franz *et al.* [17], [36] had performed some experiments to compare the results of both query translation and document translation using a machine translation system. Franz *et al.* observe that the technique performed equally. Whereas McCarley (1999) emphasizes the relationship between the language rather than the translation systems.

## C. QUERY EXPANSION

Zhou et al [37]. studied the personalized CLIR problem through query expansion. They proposed a system where new features increase the user's original query with similar meaning. They used various query expansion techniques including pseudo-relevance feedback, simple personalized query expansion, penalty query expansion, and techniques based on various methods by computing the similarity between the user query and the user profile. They also compared query expansion techniques and studied the effects of frequency-based user model generation. They researched the fact that user models generated from historical usage information in one language can enhance the search in another language. Nwesri *et al.* [38] uses stemming to expanding queries, rather than to index the stemmed collection. They used stemming to generate the collection stem-word clusters, indexed the original collection, and expanded queries using the generated stem-word clusters. Their approach gives similar results to the typical approach where the search index is made of stems. They have shown that index size favors previous approaches, but the new technique will bring an improvement in flexibility in using different stemmers without touching the original index. They also proved that stem-word clusters could be manually refined or further combined with other approaches such as word co-occurrence and using synonyms from different sources.

The Arabic language is among the most spoken, written and read the language. Z. Yahya *et al.* [36] worked on query translation using concepts similarity based on Quran ontology for cross-language. They used dictionary-based approach and covered the issue of words that have more than one meaning which can decrease the retrieval performance if the query translation returns an incorrect translation. The proposed method is based on domain ontology using Quran concepts. Used to disambiguating translation of the query and improve the dictionary-based query translation. F. Ture *et al.* [39] proposed a method for learning optimal combination weights when building a linear combination of existing query translation approaches. From standard query-document relevance judgments, they constructed a set of classifiers, which produce a unique combination recipe for each query, based on a large set of features extracted from the query and collection. They showed with several experiments that the effectiveness of their method is significantly higher than state-of-the-art query translation methods and other combination strategies.

Pasha *et al.* [4] used a query expansion tool called DIRA (Dialectal Information Retrieval Assistant) [Arabic]. This tool generates search terms, comprising both lexical and morphological variants in modern standard Arabic and Egyptian Arabic when provided with queries in English or Modern Standard Arabic. No stemming decisions are made as part of DIRA in order to allow its output to be usable by a variety of Information retrieval systems with different stemming decisions. They showed that DIRA is the only system that addresses the problem of dialectal variation. Researchers proposed a method for lexical disambiguation associated with the query translation system (Arabic-French), based on the use of semantic networks. They built two semantic networks, one representing the ambiguous terms of the translated query and the second representing the knowledge base represented by a list of relevant sentences.

The work of Elayeb *et al.* [40] is essential in this regard who gave a review of Arabic CLIR. They reviewed existing approaches in the field of Arabic CLIR and their significant utility in the recent innovative research directions in the open
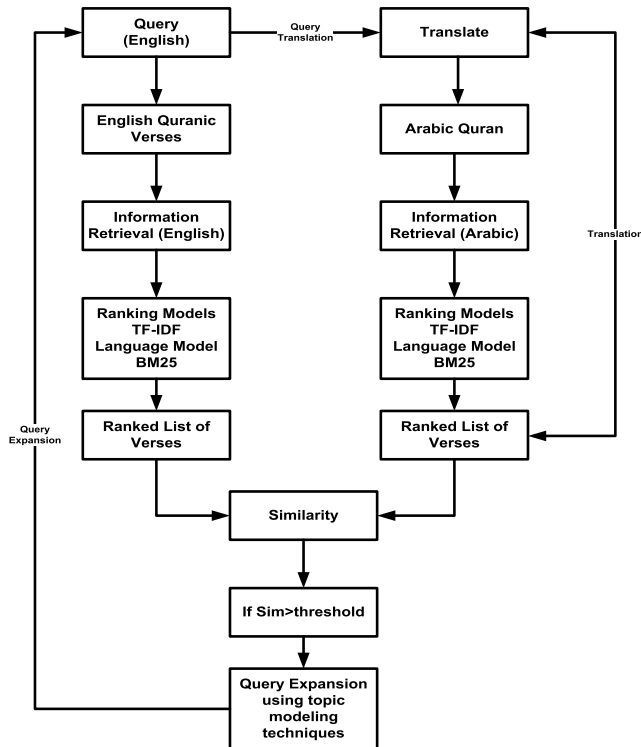
**FIGURE 1.** The proposed Quran CLIR framework.



**FIGURE 2.** A sample query sets associated concepts.

| Concept | Query English | Query Arabic |
|---|---|---|
| Artifact | Noah Ship | سفينة نوح |
| | Ship | سفينة |
| | Key | مفتاح |
| | Lamp | لمبة |
| | Mosque | مسجد |
| Events | Last day | اليوم الآخر |
| | Hajj | حج |
| | Sabbath | السبت |
| | Night of decree | ليلة القدر |
| Holy Books | Injeel | الإنجيل |
| | Qur'an | القرآن |
| | Torah | التوراة |
| | Zabur | زبور |
| Creation | Birds | طائر |
| | Fish | حوت |
| | Cow | بقرة |
| | Adam | آدَم |

area of information retrieval. Similarly, Alqudsi *et al.* [41] gave a survey for Arabic machine translation in which they recapitulated the most important methods used in machine translation from Arabic to English, identified and discussed their advantages and disadvantages.

## III. METHODOLOGY

Two types of Quran datasets are used in the experimentation process. The first is the original Quranic verses in Arabic where each verse of the Quran is separated into a single sentence with verse number and Surah number. In the second dataset, the same Quran verses are used, but this time, these are translated into the English language. In the English translated version of Quran verses, each line represents single verses with its verse number and Surah number. Separate IR systems are implemented for each version of Quran datasets, i.e., one for Arabic and one for English. The detailed methodology is illustrated in Figure 1, starting with query construction, query translation, data collection, IR systems implementation, similarity checking, and topic modeling for query expansion.

### A. QUERY SETS

There is no standard query benchmark available to evaluate the performance of the proposed architecture. For the evaluation of existing methods, a query set benchmark in English and Arabic has been developed. The query sets are extracted from 12 Quran concepts including artifacts, events, holy books, creation, body parts, location, religion, weather, false deity, substance/metal, astronomical bodies, food and

messengers. A sample query sets associated concepts is presented in Figure 2

### B. QURAN DATASET

The Quran dataset comprised of Quran verses downloaded from www.tanzil.net. Tanzil is a Quranic project to provide highly verified precise Quran text in Unicode. Tanzil Quran project has several features including accuracy, searchability, pause marks, compatibility, and flexibility. The Arabic and English language of the Quran text is downloaded from the website mentioned above.

### C. INFORMATION RETRIEVAL SYSTEM

As shown in Figure 1, two separate information retrieval systems are implemented. One for the English translation of the Quran and the other is for the Arabic version of the Quran. Standard information retrieval processes are adopted as presented in the following discussions.

#### 1) DOCUMENT PREPROCESSING

The computer cannot understand the documents written in natural languages like English or Arabic. A particular type of treatment is required to make them machine-understandable, such a process is known as document preprocessing which is comprised of several techniques include tokenization, stop words removal, stemming, and removing infrequent words from the text. Tokenization is the process of splitting the sentences of the document into single words. Stop words are those words in the text which carry no meaning when used individually. These include are, am, are, on, of, the, it, etc.
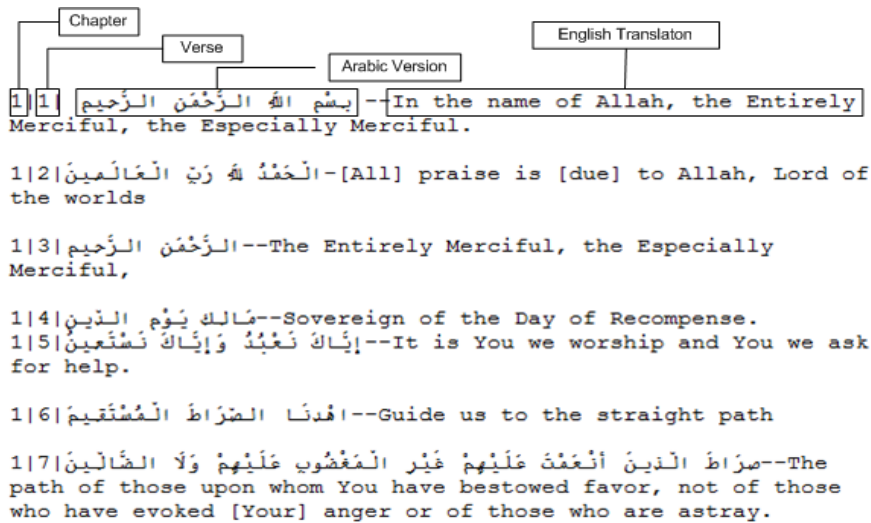
**FIGURE 3.** English translation dataset layout.

Stemming is the process of reducing the token/term to its root. For example, the term "expressions" can be stemmed to "express." Similarly, "explanation" can be reduced to "explain."

### 2) INDEXING
Indexing is a technique for organizing documents also called cataloging. Indexing is among the important part of any information retrieval system. After all the document preprocessing techniques, the next step is to index the terms to their corresponding documents.

### 3) QUERY PROCESSING
Queries are keywords or short texts which represent the user information need. Search engines or information retrieval systems use these keywords to retrieve the most relevant information from the web. User queries are also preprocessed before searching the indexed documents. We apply standard query processing procedures that are typically used in document preprocessing. These include spelling correction, tokenization, stop words removal, and stemming.

### 4) RANKING MODELS
After query processing, the terms of the query are searched in the index. Relevant documents are retrieved based on the user query keywords. The retrieved documents are then ranked by relevancy. Different retrieval models have been used which can be grouped into several classes including Boolean models, Vector Space models, Language models, Probabilistic models, and Machine learning models. The list of retrieval models used in this study is briefly discussed below.

#### a: DFR-BM25
DFR-BM25 stands for Divergence From Randomness -is the modified form of the BM25 ranking model which is used in search engines to rank the documents concerning the relevance. BM25 was initially proposed by Robertson *et al.* [42],

based on a probabilistic framework [43]. DFR-BM25 combines both the probabilistic retrieval framework and divergence from the randomness retrieval framework. The new version of BM25 is the addition to probabilistic and DFR frameworks.

#### b: DFRee
This weighting model ranks the document and stands for Divergence from randomness free (DFRee) implemented in the Terrier information retrieval platform [44]. DFRee computes the average of two probability measures. This model calculates the probability distribution of the query term in the collection of documents. The two possible situations are a) considering only the document having the query term and b) document is considered as a sample from the collection. The two sources of information have been averaged to rank the documents.

#### c: DIRICHLET LANGUAGE MODEL (DLM)
Dirichlet language model is a retrieval model that belongs to the family of language models for information retrieval. This phenomenon is first introduced by Ponte and Croft [45]. The language model of each document in the collection is constructed and ranked according to P (DQ), where D presents document and Q presents query. The generic steps involved in language models are a) extraction of terms from the documents, b) calculate the term frequency for each term, c) calculate the term count of each document, and d) assign term probability to each term.

#### d: HEIMESTRIA LANGUAGE MODEL
Ponte and Croft [46] were the pioneers introducing statistical language retrieval/ranking models for information retrieval. This language model was proposed by Hiemstra and Ponte Croft. It also belongs to the Language models family of ranking models for information retrieval.

### e: TF_IDF

This ranking model is the combination of two models: a) TF (term frequency) by Hans Peter Luhn [47] and b) IDF by Spark Jones' [48]. TF-IDF stands for Term frequency-inverse Document frequency. It is a statistical model that represents the importance of term or word in a document, where the document belongs to the collection of documents. It is one of the most effective and popular term weighting schemes. If a term appears in a single document more often, it means that the term is important. However, if the term appears more often in the collection of documents, this means it is the more general term. The TF-IDF weight model neglects the general term and favors specific but important terms. The mathematical formula is as below.

$$tfidf(t, d, D) = tf.idf \tag{1}$$

where tf is term frequency and can be represented as

$$tf(t, d) = 1 + log(f\_(t, d)) \tag{2}$$

Moreover, Idf present inverse document frequency and can be represented as

$$idf(t, d) = \frac{N}{|\{d\varepsilon D : t\varepsilon d\}|} \tag{3}$$

*JS_KLS* Proposed by G. Amati [49], JS_KLS is a hyper-geometric model based on the DFR approach. These models are parameter-free models where no parameter is required to tune the model. This model is based on two probability observations: a) probability of term in a document and b) probability of term within the entire document collection.

### D. QUERY EXPANSION

The query expansion is used to expand the query with similar terms in order to improve the performance of the ranking model. Most of the queries are single words that represent a concept in the Quran. For example, ''Paradis'' is a concept that represents an artifact or building that has very vital importance in Islam. In a case where a user is unfamiliar with the concept and does not know much about the topic, query expansion becomes a very useful tool for many information retrieval systems. Three query expansion models are used to expand the English query. These include a) Bose-Einstein (Bo1) [50], b) Kullback-Leibler (KL) [50] and c) Chi-Square divergence (CS) [44].

### E. EVALUATION

Two types of measures are used to evaluate the IR system'' a) effectiveness and b) efficiency. Effectiveness determines the retrieval performance, whereas efficiency is related to the implementation of the system. Most of the IR systems are evaluated using effectiveness measures [51], the same is adopted by the this study. These measures are listed below.

### 1) PRECISION

Precision is the fraction of relevant documents among retrieved documents. In simple words, we can say that how many documents were relevant among the retrieved documents.

$$precison = \frac{|retrieveddocuments \cap relevantdocuments|}{|retrieveddocuments|} \tag{4}$$

### 2) RECALL

The recall is a fraction of retrieved documents among relevant documents. In other words, how many relevant documents are retrieved from the whole relevant documents?

$$recall = \frac{|relevantdocuments \cap retrieveddocuments|}{|totalrelevantdocuments|} \tag{5}$$

### 3) F-MEASURE/F-SCORE

F-measure/F-score considers both precision P and recall R, which is the weighted harmonic mean of both precision and recall.

$$F = \frac{(1 + \beta^2)(precision)(recall)}{\beta^2(precision)(recall)} \tag{6}$$

$$F = \frac{(1 + \beta^2)PR}{\beta^2.P + PR} \tag{7}$$

## IV. RESULTS AND DISCUSSION
### A. EXPERIMENTAL SETUP

There is no benchmark available to evaluate the performance of the proposed CLIR system architecture for the Quranic dataset. For such purpose, a gold standard dataset is developed presented in the following section.

### 1) DATASET

The dataset used in this research study is comprised of Quran verses in Arabic and English languages. These are acquired from the freely available Quran translation (Maududi translation) website www.tenzil.com. The dataset is available in the.txt format. Each line of the text file presents a single verse with the chapter number as presented in Figure 3.

### 2) RELEVANCE JUDGMENT

The performance of information retrieval systems measures the effectiveness. For measuring such performance, three resources required:

a) A dataset, consisting of document collection.

b) The information needs usually referred to as queries.

c) A set of relevant judgments that represent a query-document pair. Such that, a document or set of relevant documents for each query.

The set of relevant documents for each query is typically decided by the human participants, where they assign relevant documents to each user query. For such cases, the term gold standard or ground truth is used. This requires enormous efforts and time to annotate the document according to the set of queries. As there is no gold standard or ground truth is freely available for the evaluation of the proposed CLIR architecture, it is decided to build our own. The processes involved to construct a gold standard for this study is depicted in Figure 4.
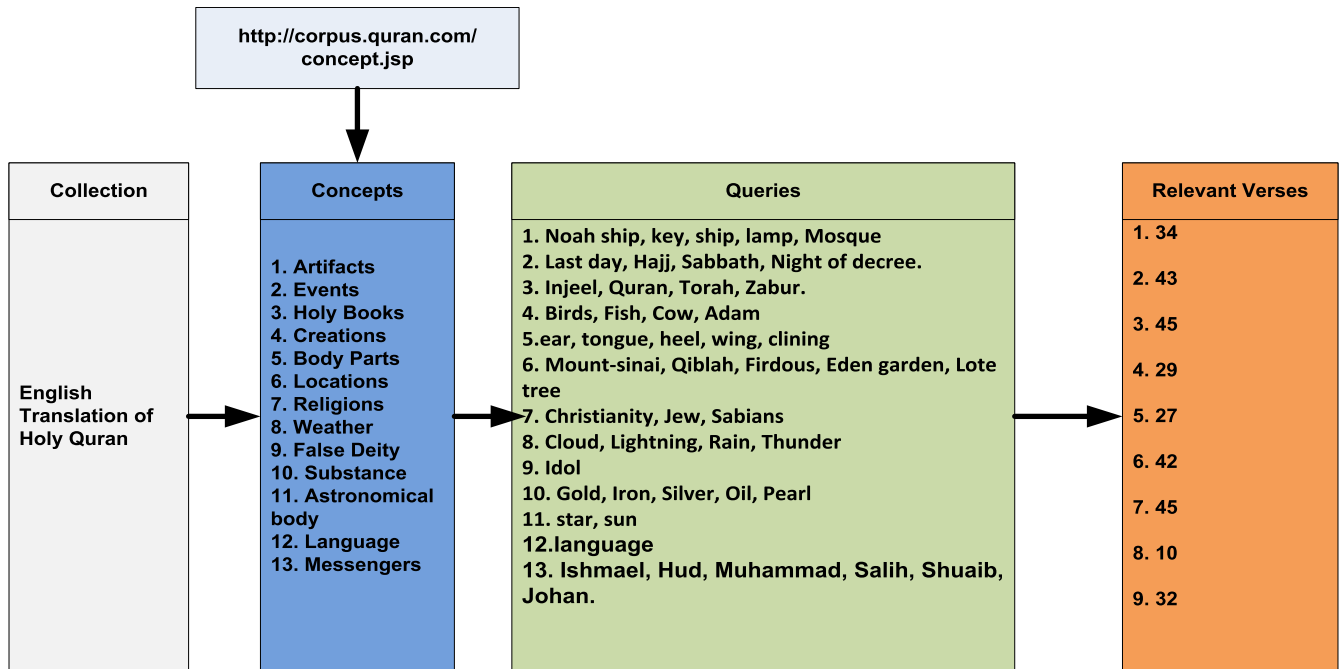
**FIGURE 4.** Ground truth or gold standard construction method.

To construct a gold standard, we need queries and document relevancy. Undoubtedly, providing relevant documents to queries is very hard for someone who is unfamiliar with the Quran. To develop a query set, we use the Quran concepts defined by Kais Dukes [52]. These are developed by a research group called "Language Research Group" at the University of Leeds. According to them: "The Quranic Ontology uses knowledge representation to define the key concepts in the Quran and shows the relationships between these concepts using predicate logic." They defined the Quran in 12 basic concepts and are further categorized into subcategories. The query set, comprised of 9 concepts, extracted from 12 concepts presented at the website shown in Figure 4. the query set containing queries like "ark of the covenant," "coin," "ink," "ladder," "mosque" "church" and "synagogue" is related to the concept "Artifacts." For each concept, we construct a random number of queries, and each query has a random number of documents (verses) (ranging from 1 to 25). The relevant document box (Figure 4) presents the total number of relevant verses for each concept. In such a way we construct the gold standard to evaluate the performance of our proposed architecture.

## B. EXPERIMENTAL RESULTS

The experimental results are evaluated using precision, recall, and f-score. This section discusses the results concerning the research questions outlined in Section 1.

*RQ1. Which Ranking Model Achieves High Performance in Terms of Effectiveness?*

Six retrieval models are used for comparison. These include DFR_BM25, DLM, HLM, Js_KLs, and DFRee. These models are evaluated using mean average

precision (MAP), Mean average recall (MAR), and Mean average f-score (MAF) at rank 5, 10, 15, and 20. These measures are presented in the below Equation.

$$MAP = \frac{sum of precisons ocre obtain by indivi dul query}{total number of queries} eq$$
(8)

$$MAR = \frac{sum of recall socre of individual query}{total number of queries} eq$$
(9)

### 1) PRECISION

The precision score achieved by the six above-mentioned retrieval models illustrated in Figure 5-a. A high MAP score is observed at @5. The MAP score gradually decreases at the lower rank as the number of retrieved documents increases whereas, the number of relevant documents does not increase similarly. Js_KLs achieve a high precision score and prove to be the best retrieval model as compared to other retrieval models. Js_KLs achieve MAP scores of 0.74, 0.564, 0.45, and 0.377 at @5, @10, @15, and @20, respectively. The DRFree perform slightly lower than Js_KLs in terms of MAP score and achieves 0.728, 0.552, 0.444, and 0.373 at @5, @10, @15, and @20, respectively. It is observed that the difference between the MAP score at lower ranks is getting smaller. The results also show that Js_KLs achieve higher precision than other retrieval models at higher ranks whereas; at lower ranks (MAP@15, MAP@20) DFRee achieve comparable performance. All the other retrieval models achieve MAP@5 > 0.7.

### 2) RECALL

Like Avg. P, average recall score is also absolved at the rank positions 5, 10, 15, and 20. Figure 5-b depicts the recall score
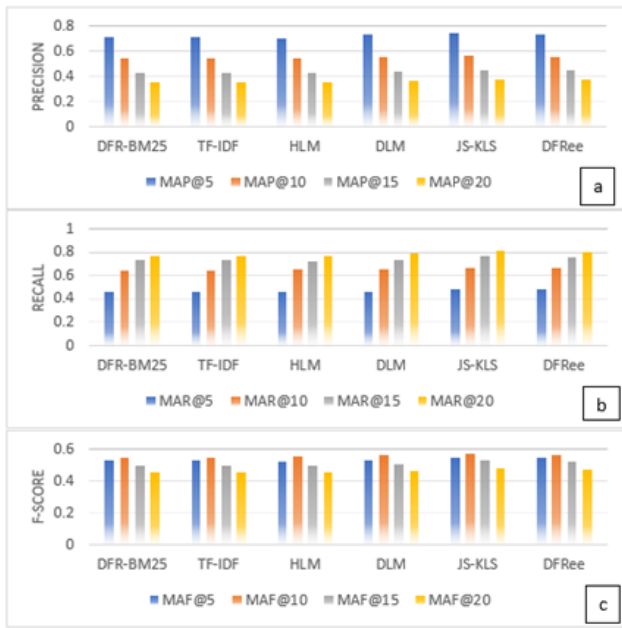
**FIGURE 5.** MAP, MAR and MAF score achieved by the retrieval models.

**TABLE 1.** The performance evaluation of Bo1, Chi-square, and KL query expansion with query length 3.

| | MAP | | | | MAR | | | | MAF | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 |
| DFR_BM25 | 0.72 | 0.54 | 0.43 | 0.36 | 0.47 | 0.65 | 0.71 | 0.76 | 0.53 | 0.55 | 0.50 | 0.46 |
| DLM | 0.68 | 0.52 | 0.42 | 0.36 | 0.43 | 0.60 | 0.69 | 0.74 | 0.49 | 0.52 | 0.49 | 0.45 |
| HLM | 0.68 | 0.50 | 0.40 | 0.32 | 0.44 | 0.60 | 0.68 | 0.71 | 0.51 | 0.51 | 0.47 | 0.41 |
| Js_KLs | 0.75 | 0.59 | 0.46 | 0.38 | 0.48 | 0.69 | 0.76 | 0.80 | 0.55 | 0.59 | 0.54 | 0.48 |
| DFRee | 0.72 | 0.57 | 0.46 | 0.38 | 0.47 | 0.68 | 0.75 | 0.80 | 0.53 | 0.58 | 0.53 | 0.48 |
| TF_IDF | 0.71 | 0.54 | 0.43 | 0.36 | 0.47 | 0.65 | 0.71 | 0.76 | 0.53 | 0.55 | 0.50 | 0.45 |

Bo1

| | MAP | | | | MAR | | | | MAF | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 |
| DFR_BM25 | 0.71 | 0.54 | 0.42 | 0.35 | 0.46 | 0.64 | 0.71 | 0.76 | 0.53 | 0.54 | 0.49 | 0.45 |
| DLM | 0.68 | 0.52 | 0.42 | 0.35 | 0.43 | 0.60 | 0.68 | 0.74 | 0.49 | 0.52 | 0.48 | 0.44 |
| HLM | 0.68 | 0.50 | 0.39 | 0.31 | 0.44 | 0.59 | 0.68 | 0.70 | 0.50 | 0.50 | 0.46 | 0.40 |
| Js_KLs | 0.75 | 0.58 | 0.46 | 0.37 | 0.48 | 0.69 | 0.76 | 0.79 | 0.55 | 0.59 | 0.53 | 0.48 |
| DFRee | 0.72 | 0.57 | 0.45 | 0.38 | 0.46 | 0.67 | 0.75 | 0.79 | 0.53 | 0.57 | 0.53 | 0.48 |
| TF_IDF | 0.71 | 0.54 | 0.43 | 0.35 | 0.46 | 0.64 | 0.71 | 0.76 | 0.52 | 0.54 | 0.49 | 0.45 |

Chi-Square

| | MAP | | | | MAR | | | | MAF | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 |
| DFR_BM25 | 0.71 | 0.5 | 0.42 | 0.35 | 0.46 | 0.64 | 0.71 | 0.76 | 0.53 | 0.54 | 0.49 | 0.45 |
| DLM | 0.67 | 0.52 | 0.42 | 0.35 | 0.42 | 0.60 | 0.68 | 0.74 | 0.48 | 0.52 | 0.48 | 0.44 |
| HLM | 0.66 | 0.49 | 0.39 | 0.31 | 0.43 | 0.58 | 0.67 | 0.70 | 0.49 | 0.49 | 0.46 | 0.40 |
| Js_KLs | 0.75 | 0.58 | 0.46 | 0.37 | 0.48 | 0.69 | 0.76 | 0.79 | 0.55 | 0.59 | 0.53 | 0.48 |
| DFRee | 0.72 | 0.57 | 0.45 | 0.38 | 0.46 | 0.67 | 0.75 | 0.79 | 0.53 | 0.57 | 0.53 | 0.48 |
| TF_IDF | 0.71 | 0.54 | 0.42 | 0.35 | 0.46 | 0.64 | 0.71 | 0.76 | 0.53 | 0.55 | 0.49 | 0.45 |

KL

by the retrieval models. Recall score at lower ranks increases. Js_KLs retrieval model attains high MAR scores of 0.82 and 0.76 at positions 20 and 15 respectively whereas; a Mean average recall score of 0.48 and 0.66 is achieved at position 5 are 10 respectively. DFRee, TF-IDF, and HLM proved to have the least efficient retrieval models when it comes to recalling.

### 3) F-SCORE

The retrieval models are also evaluated using the F-score measure. The F-score scores are presented in Figure 5-c. Js_KLs achieve slightly better performance than DFRee when it comes to MAF at rank 5. They achieve an almost similar score of 0.548 and 0.544 at rank 5. DRF-BM25, TF-IDF, HLM and DLM achieve MAP score of 0.528, 0.528, 0.522 and 0.531 at rank 5 respectively. Regarding RQ1, it is observed that Js_KLs attain high performance concerning Mean Average Precision, Mean Average Recall, and Mean Average F-score at rank 5, 10, 15, and 20.

### C. QUERY EXPANSION

Three different query expansion techniques have been deployed in order to measure the performance of retrieval models. This section of the document answers RQ2 and RQ3. The prior is related to the query expansion model that answers the best query expansion, model. The former will investigate the suitable length of a query to search for the Quranic datasets.

### D. QUERY LENGTH

The original query is expanded with different query lengths using the query expansion methods. The query length ranges from 3 to 10. The terms used to expand the original query

extracted from the top 10 retrieved documents. The implementation setup requires two parameters: a) the ranking model and b) the query expansion method. It works in such a way that the first documents are ranked using the ranking model, then query expansion methods are applied to extract the expansion terms. Finally, the expanded query is again searched using the indexed documents, and documents are ranked again according to the expanded query. The results of the retrieval model using query expansion are presented in the following discussion. MAP, MAR, and MAF at rank 5, 10, 15, and 20 are used for evaluation.

### 1) QUERY LENGTH THREE TERMS

In this experiment, the query length was extended to 3 terms. The best retrieval model is the Js_KLs retrieval model based on the experiments. The high improvement was observed at early precision where a moderate improvement is recorded in later ranks as presented in Table 1. We noticed that the language models approach for information retrieval degrades the performance. DLM and HLM reduce the MAP@5 from 0.72 to 0.68 and 0.70 to 0.68, respectively. The performance of TF-IDF is also degraded slightly, whereas the performance of Js_KLs, DFR_BM25 improves while the DFRee behaves similarly as with query expansion.

For Chi-Square, the best performing retrieval model is again Js_KLs. No significant change in the score is mentioned above observed for TF-IDF, DFRee, and DFR_BM25 as compared to the original query. A significant reduction in the scores is observed for language model approaches, i.e., DLM and HLM. Finally, using KL query expansion method, the HLM performance is further decreased whereas, the rest of

**TABLE 2.** The performance evaluation of Bo1, Chi-square, and KL query expansion with query length 4.

| | MAP | | | | MAR | | | | MAF | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 |
| DFR_BM25 | 0.71 | 0.54 | 0.42 | 0.35 | 0.46 | 0.64 | 0.71 | 0.76 | 0.53 | 0.54 | 0.49 | 0.45 |
| DLM | 0.68 | 0.52 | 0.41 | 0.35 | 0.44 | 0.61 | 0.69 | 0.75 | 0.50 | 0.52 | 0.48 | 0.45 |
| HLM | 0.68 | 0.49 | 0.39 | 0.32 | 0.44 | 0.58 | 0.66 | 0.71 | 0.50 | 0.49 | 0.45 | 0.41 |
| Js_KLs | **0.76** | **0.58** | **0.45** | 0.37 | **0.49** | 0.69 | 0.75 | 0.79 | 0.56 | 0.59 | 0.53 | **0.48** |
| DFRee | 0.74 | 0.57 | 0.45 | **0.38** | 0.48 | 0.67 | 0.75 | 0.79 | 0.54 | 0.58 | 0.52 | **0.48** |
| TF_IDF | 0.71 | 0.54 | 0.42 | 0.35 | 0.46 | 0.64 | 0.71 | 0.76 | 0.53 | 0.54 | 0.49 | 0.45 |

Bo1

| | MAP | | | | MAR | | | | MAF | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 |
| DFR_BM25 | 0.71 | 0.54 | 0.42 | 0.35 | 0.46 | 0.64 | 0.71 | 0.76 | 0.53 | 0.54 | 0.49 | 0.45 |
| DLM | 0.68 | 0.52 | 0.41 | 0.35 | 0.44 | 0.61 | 0.68 | 0.75 | 0.50 | 0.52 | 0.48 | 0.45 |
| HLM | 0.68 | 0.47 | 0.38 | 0.31 | 0.44 | 0.57 | 0.66 | 0.70 | 0.50 | 0.48 | 0.45 | 0.41 |
| Js_KLs | **0.76** | **0.58** | **0.46** | 0.37 | **0.49** | 0.69 | 0.75 | 0.79 | 0.56 | 0.58 | 0.53 | 0.48 |
| DFRee | 0.74 | 0.57 | **0.46** | **0.38** | 0.48 | 0.67 | 0.75 | 0.79 | 0.54 | 0.57 | 0.53 | 0.48 |
| TF_IDF | 0.71 | 0.54 | 0.42 | 0.35 | 0.46 | 0.64 | 0.71 | 0.76 | 0.53 | 0.54 | 0.49 | 0.45 |

Chi-Square

| | MAP | | | | MAR | | | | MAF | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 |
| DFR_BM25 | 0.71 | 0.54 | 0.42 | 0.35 | 0.46 | 0.64 | 0.71 | 0.76 | 0.53 | 0.54 | 0.49 | 0.45 |
| DLM | 0.68 | 0.52 | 0.41 | 0.35 | 0.44 | 0.61 | 0.68 | 0.75 | 0.50 | 0.52 | 0.47 | 0.45 |
| HLM | 0.68 | 0.47 | 0.38 | 0.31 | 0.44 | 0.57 | 0.66 | 0.70 | 0.50 | 0.48 | 0.45 | 0.41 |
| Js_KLs | **0.76** | **0.58** | **0.46** | 0.37 | **0.49** | 0.69 | 0.75 | 0.79 | 0.56 | 0.58 | 0.53 | 0.48 |
| DFRee | 0.74 | 0.57 | 0.46 | **0.38** | 0.48 | 0.67 | 0.75 | 0.79 | 0.5 | 0.57 | 0.53 | 0.48 |
| TF_IDF | 0.71 | 0.54 | 0.42 | 0.35 | 0.46 | 0.64 | 0.71 | 0.76 | 0.53 | 0.54 | 0.49 | 0.45 |

KL

**TABLE 3.** The performance evaluation of Bo1, Chi-square, and KL query expansion with query length 5.

| | MAP | | | | MAR | | | | MAF | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 |
| DFR_BM25 | 0.72 | 0.54 | 0.43 | 0.35 | 0.47 | 0.64 | 0.71 | 0.76 | 0.53 | 0.54 | 0.50 | 0.45 |
| DLM | 0.66 | 0.52 | 0.41 | 0.35 | 0.43 | 0.61 | 0.68 | 0.74 | 0.49 | 0.52 | 0.47 | 0.44 |
| HLM | 0.62 | 0.48 | 0.39 | 0.32 | 0.40 | 0.57 | 0.67 | 0.70 | 0.46 | 0.48 | 0.46 | 0.41 |
| Js_KLs | **0.75** | **0.58** | **0.46** | 0.38 | **0.48** | 0.69 | 0.76 | **0.80** | 0.55 | 0.59 | 0.53 | 0.48 |
| DFRee | 0.72 | 0.57 | 0.46 | **0.38** | 0.47 | 0.67 | 0.75 | **0.80** | 0.53 | 0.57 | 0.53 | 0.48 |
| TF_IDF | 0.72 | 0.54 | 0.43 | 0.35 | 0.47 | 0.64 | 0.71 | 0.76 | 0.53 | 0.54 | 0.50 | 0.45 |

Bo1

| | MAP | | | | MAR | | | | MAF | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 |
| DFR_BM25 | 0.72 | 0.54 | 0.43 | 0.35 | 0.47 | 0.64 | 0.71 | 0.76 | 0.53 | 0.54 | 0.50 | 0.45 |
| DLM | 0.66 | 0.52 | 0.41 | 0.35 | 0.43 | 0.61 | 0.68 | 0.74 | 0.49 | 0.52 | 0.47 | 0.44 |
| HLM | 0.62 | 0.48 | 0.39 | 0.32 | 0.40 | 0.57 | 0.67 | 0.70 | 0.46 | 0.48 | 0.46 | 0.41 |
| Js_KLs | **0.75** | **0.58** | **0.46** | 0.38 | **0.48** | 0.69 | 0.76 | **0.80** | 0.55 | 0.59 | 0.53 | 0.48 |
| DFRee | 0.72 | 0.57 | 0.46 | **0.38** | 0.47 | 0.67 | 0.75 | **0.80** | 0.53 | 0.57 | 0.53 | 0.48 |
| TF_IDF | 0.72 | 0.54 | 0.43 | 0.35 | 0.47 | 0.64 | 0.71 | 0.76 | 0.53 | 0.54 | 0.50 | 0.45 |

Chi-Square

| | MAP | | | | MAR | | | | MAF | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 |
| DFR_BM25 | 0.71 | 0.54 | 0.42 | 0.35 | 0.46 | 0.64 | 0.71 | 0.76 | 0.52 | 0.54 | 0.49 | 0.45 |
| DLM | 0.66 | 0.51 | 0.39 | 0.32 | 0.43 | 0.60 | 0.67 | 0.71 | 0.49 | 0.51 | 0.46 | 0.41 |
| HLM | 0.64 | 0.47 | 0.37 | 0.31 | 0.41 | 0.57 | 0.65 | 0.69 | 0.47 | 0.48 | 0.44 | 0.40 |
| Js_KLs | **0.74** | **0.58** | **0.46** | 0.37 | **0.48** | 0.68 | 0.76 | 0.79 | 0.54 | 0.58 | 0.53 | 0.48 |
| DFRee | 0.72 | 0.57 | 0.46 | **0.38** | 0.47 | 0.67 | 0.75 | 0.80 | 0.54 | 0.57 | 0.53 | 0.48 |
| TF_IDF | 0.71 | 0.54 | 0.42 | 0.35 | 0.46 | 0.64 | 0.71 | 0.76 | 0.52 | 0.54 | 0.49 | 0.45 |

KL

the retrieval models behave similarly to the previous query expansion methods.

### 2) QUERY LENGTH FOUR TERMS
These results present the MAP, MAR, and MAF at rank 5, 10, 15, and 20 of the six-retrieval models when the query is expanded to four terms using the Bo1 expansion method. The best performing retrieval model is again Js_KLs. The query with length 4 improves the results significantly as compared to the original query and query length 4.

Chi-square presents the MAP, MAR, and MAF at rank 5, 10, 15, and 20 of the six retrieval models when the query is expanded to four terms using the chi-square expansion method. The best performing retrieval model is again Js_KLs. No significant change in the above-mentioned score is observed for TF-IDF, DFRee, and DFR_BM25 as compared to the Bo1 query expansion method. Reduction in the scores is observed for language model approached, i.e., DLM and HLM but improved when compared to the query of length three terms. KL presents the results statistics of the KL query expansion method. All the retrieval models behave similarly to the previous query expansion methods.

### 3) QUERY LENGTH FIVE TERMS
The results presented in Table 3 the MAP, MAR, and MAF scores of six evaluation methods using query expansion method Bo1, Chi-square, and KL. The best retrieval model is the Js_KLs retrieval model. High improvement is observed at early precision when compared to the original query, whereas low MAP, MAR, and MAF scores are recorded as compared

to query length of 4 terms. We notice that the language models approach for information retrieval degrades the performance.

The performance of the Chi-square expansion method using 5 terms query is presented in Chi-Square. HLM retrieval model has the worst performance concerning all the evaluation parameters. Chi-square query expansion does not have any significant impact on the performance of the retrieval model while compared to the BO1 retrieval model. The KL table presents the results of the KL query expansion method. The KL method improves the results of HLM as compared to Bo1 and CS methods. No significant change in performance was observed for other retrieval models.

### 4) QUERY LENGTH TEN TERMS
Low evaluation score (MAP, MAR, MAF) score are achieved by all the retrieval models for a query of 10 terms except HLM as compared to query length of 5 terms. Bo1 shows the result of statistics query expansion method Bo1 while the query length was 10. Chi-Square shows no significant change recorded as compared to Bo1 when the CS query expansion method is used for expanding the original query. The performance of the HLM retrieval model is improved when the CS query expansion method is used. The KL table presents the results for the KL query expansion method. Almost similar evaluation scores are achieved by KL when compared to the CS method.

*RQ2. What Is the Appropriate Query Length for Effective Retrieval of Quran Verses in the CLIR System?*

The automatic query expansion provides a means to expand the query with some relevant terms. These relevant terms are extracted from the top K relevant documents which

**TABLE 4.** The performance evaluation of Bo1, Chi-square, and KL query expansion with query length 10.

| | MAP | | | | MAR | | | | MAF | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 |
| DFR_BM25 | 0.71 | 0.54 | 0.42 | 0.35 | 0.46 | 0.64 | 0.71 | 0.76 | 0.52 | 0.54 | 0.49 | 0.453 |
| DLM | 0.66 | 0.51 | 0.39 | 0.32 | 0.43 | 0.60 | 0.67 | 0.71 | 0.49 | 0.51 | 0.46 | 0.416 |
| HLM | 0.64 | 0.47 | 0.37 | 0.31 | 0.41 | 0.57 | 0.65 | 0.69 | 0.47 | 0.48 | 0.44 | 0.402 |
| Js_KLs | 0.74 | 0.58 | 0.46 | 0.37 | 0.48 | 0.68 | 0.76 | 0.79 | 0.54 | 0.58 | 0.53 | 0.480 |
| DFRee | 0.72 | 0.57 | 0.46 | 0.38 | 0.47 | 0.67 | 0.75 | 0.80 | 0.54 | 0.57 | 0.53 | 0.484 |
| TF_IDF | 0.71 | 0.54 | 0.42 | 0.35 | 0.46 | 0.64 | 0.71 | 0.76 | 0.52 | 0.54 | 0.49 | 0.453 |

Bo1

| | MAP | | | | MAR | | | | MAF | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 |
| DFR_BM25 | 0.71 | 0.54 | 0.428 | 0.356 | 0.46 | 0.64 | 0.70 | 0.75 | 0.52 | 0.54 | 0.49 | 0.45 |
| DLM | 0.66 | 0.50 | 0.396 | 0.321 | 0.42 | 0.59 | 0.67 | 0.71 | 0.48 | 0.51 | 0.46 | 0.41 |
| HLM | 0.64 | 0.47 | 0.372 | 0.308 | 0.42 | 0.57 | 0.64 | 0.69 | 0.47 | 0.48 | 0.43 | 0.39 |
| Js_KLs | 0.75 | 0.57 | 0.458 | 0.375 | 0.48 | 0.67 | 0.74 | 0.78 | 0.55 | 0.58 | 0.53 | 0.47 |
| DFRee | 0.72 | 0.55 | 0.458 | 0.379 | 0.46 | 0.65 | 0.74 | 0.78 | 0.53 | 0.56 | 0.52 | 0.47 |
| TF_IDF | 0.71 | 0.54 | 0.426 | 0.356 | 0.46 | 0.64 | 0.70 | 0.75 | 0.52 | 0.54 | 0.49 | 0.45 |

Chi-Square

| | MAP | | | | MAR | | | | MAF | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 |
| DFR_BM25 | 0.71 | 0.54 | 0.42 | 0.35 | 0.46 | 0.64 | 0.70 | 0.74 | 0.52 | 0.541 | 0.495 | 0.44 |
| DLM | 0.66 | 0.51 | 0.39 | 0.32 | 0.43 | 0.60 | 0.67 | 0.71 | 0.49 | 0.51 | 0.46 | 0.41 |
| HLM | 0.64 | 0.47 | 0.37 | 0.31 | 0.41 | 0.57 | 0.64 | 0.69 | 0.47 | 0.481 | 0.44 | 0.40 |
| Js_KLs | **0.74** | **0.57** | **0.45** | **0.37** | **0.48** | **0.67** | **0.74** | **0.78** | **0.54** | **0.58** | **0.53** | **0.47** |
| DFRee | 0.72 | 0.56 | 0.45 | 0.37 | 0.47 | 0.66 | 0.74 | 0.78 | 0.53 | 0.56 | 0.52 | 0.47 |
| TF_IDF | 0.71 | 0.54 | 0.42 | 0.35 | 0.46 | 0.64 | 0.70 | 0.74 | 0.52 | 0.549 | 0.49 | 0.44 |

KL



**FIGURE 6.** Bo1 method improves the performance of the retrieval model by expanding the query as compared to the original query. Query length of 4 terms achieves the high MAP@5 score.

are then added to a user query to improve the performance of the IR system. The top K relevant document means the top K retrieved documents. However, the performance of the IR system sometimes is detrimental when the top K documents are not relevant. It is important to note that if precision at early ranks is important, the automatic query expansion might not contribute to increasing the effectiveness of the system. The number of relevant documents also contributes to the performance of the query expansion method.

In order to answer research question 3, we evaluate the performance of each query expansion method concerning several query terms. Bo1 results present the performance of the Bo1 query expansion method that improves the performance of the retrieval model by expanding the query as compared to the original query. Query length of 4 terms achieves the high MAP@5 score. Chi-square depicts the MAP@5 values achieved by the six retrieval models using a query of different lengths. The Js_KLs retrieval model performs on the original query than the other retrieval models. As the query is expanded to 3 terms using the chi-square convergence method, the score of MAP@5 also increases for all the models except DLM and HLM. The best value is achieved when the query length is expanded to 4 terms. When the query length is expanded to 5, the MAP@5 value is decreased. The same pattern is observed for all the other evaluation measures. One of the important phenomena that are observed during this experiment is that the language model-based retrieval algorithms performance deteriorates as the length of the query increases. They perform best on short-length queries. The same trends are observed when using the KL query expansion method as presented in KL, the query
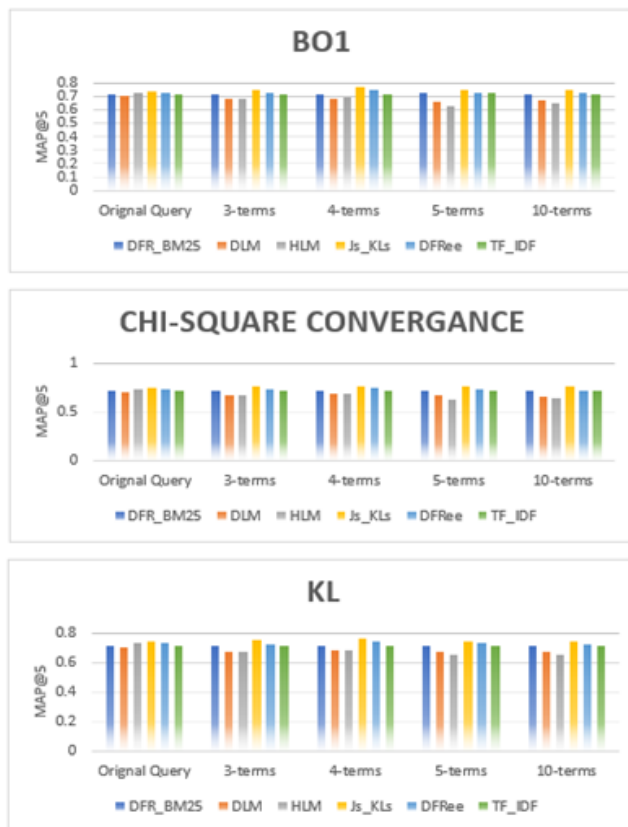
of length 4 terms achieves a high MAP@5 score when using Js_KLs, DFRee, TF-IDF, and DFR_BM25 retrieval models. DLM and HLM achieve low MAP@5 scores when the query length is increased. The same pattern is observed for the rest of the evaluation measures readings. Hence, the answer to research question 3 is that the best query length is 4 terms.

*RQ3. Which Query Expansion Technique Can Obtain the Best Query Words?*

The performance of the three-query expansion model evaluated for a query of 4 lengths concerning MAP@5 as presented in Figure 6. All the query expansion methods contribute towards the improvement of retrieval effectiveness of retrieval models. The different behavior is observed for different retrieval models. Query term expanded with KL achieves high MAP@5 for DFR_BM25 retrieval model. Bo1 query words attain a low MAP@5 score for the same model. The performance of all the three query expansion models for the Dirichlet language model is similar regarding MAP@5. For the Heimestria language model, Bo1 attain a high MAP@5 score. As mentioned earlier, Js_KLs outperforms all the other retrieval models; a similar improvement has been noted for the three-query expansion model. KL performs best for DFRee as compared to Bo1 and CS query expansion models. Finally, TF-IDF KL extracts more useful terms to improve the precision at higher ranks. In conclusion, no single query expansion model improves the performance of all retrieval
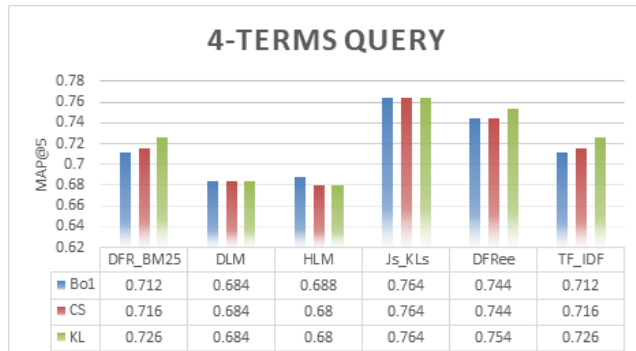
**FIGURE 7.** The best query expansion method is KL, as it increases the effectiveness of 4 ranking models.

models. Instead, these have a different impact on different retrieval models.

### 5) ARABIC QUERIES

The Arabic queries are first translated into English and then searched into the indexed English Quran translation. Figure 7 presents the results of retrieval models for translated Arabic queries. Js-KLs outperforms all the other retrieval models. Dirichlet language models perform better than the rest of the models. HLM, TF_IDF, and DFR_BM25 have the least performance regarding mean average precision. At MAP@20 almost all the retrieval models acquire a similar score.

### 6) ARABIC QUERY EXPANSION

Decrease in the score is observed for query expansion. On the other hand, a significant improvement is observed for the English queries. The reason for such a decrease is the translation of Arabic terms into English. As we translate the Arabic queries in the English language, some of the words do not match with the posting list in the index file.

## V. CONCLUSION

The two sets of queries (i.e., Arabic and English) are used to retrieve relevant verses from the parallel corpus of original Arabic and the English translation of the Quran. We observed that the performance of Arabic queries is not promising as compared to English queries. We change the scheme to translate the Arabic queries into English and then to search these queries in the English version of the Quran and record comparable improvement. The reason for low performance Arabic IR system is that the single Arabic terms may contain multiple meanings and thus it is challenging for the automatic translator to translate it appropriately. This study also investigates the query expansion models along with the retrieval models. Six different states of the art retrieval models are used in this study, whereas three state-of-the art query expansion models are also used.

The length of the query is a very important factor and can influence the performance of the IR systems. We investigate the different query lengths and suggest query length of 4 terms in CLIR systems. As far query expansion models are concerned, the English query Js-KLs model outperforms

DFR_BM25, TF_IDF, DLM, HLM, and DFRee models. The KL method of query expansion with query length of 4 terms attains promising results. In future, we are interested to propose a novel retrieval model that can improve the overall performance of CLIR systems.

## REFERENCES

[1] K. Kishida and N. Kando, "A hybrid approach to query and document translation using a pivot language for cross-language information retrieval," in *Proc. 6th Workshop Cross-Lang. Evaluation Forum, Accessing Multilingual Inf. Repositories*, vol. 4022, 2005, pp. 93–101.

[2] W. B. Croft, D. Metzler, and T. Strohman, *Search Engines: Information Retrieval in Practice*, vol. 520. Reading, MA, USA: Addison-Wesley, 2010.

[3] L. Boytsov and Z. Kolter, "Exploring classic and neural lexical translation models for information retrieval: Interpretability, effectiveness, and efficiency benefits," 2021, *arXiv:2102.06815*.

[4] A. Pasha, M. Al-Badrashiny, M. Altantawy, N. Habash, M. Pooleery, O. Rambow, R. Roth, and M. Diab, "DIRA: Dialectal Arabic information retrieval assistant," in *Proc. IJCNLP*, 2013, pp. 13–16.

[5] V. K. Sharma and N. Mittal, "Cross-lingual information retrieval: A dictionary-based query translation approach," in *Proc. Adv. Comput. Comput. Sci.*, 2018, pp. 611–618.

[6] C. D. Manning, P. Raghavan, and H. Schütze, "Probabilistic information retrieval," in *Proc. Introduction to Inf. Retr.*, 2009, pp. 220–235.

[7] M. H. Bashir, A. M. Azmi, H. Nawaz, W. Zaghouani, M. Diab, A. AlFuqaha, and J. Qadir, "Arabic natural language processing for Quranic research: A systematic review," 2021.

[8] F. Song and W. B. Croft, "A general language model for information retrieval," in *Proc. 8th Int. Conf. Inf. Knowl. Manage.*, 1999, pp. 316–321.

[9] T. Barik and V. Singh, "AQtpUIR: Adaptive query term proximity based user information retrieval," *J. Inf. Optim. Sci.*, vol. 41, no. 6, pp. 1479–1497, Aug. 2020.

[10] I. Moulinier and E. S. Lund, "Systems, methods, software, and interfaces for multilingual information retrieval," U.S. Patent 9 418 139, Aug. 16, 2016.

[11] G. Cao, J. Gao, and J.-Y. Nie. (Jul. 2007). *A System to Mine Large-Scale Bilingual Dictionaries From Monolingual Web Pages*. [Online]. Available: https://www.microsoft.com/en-us/research/publication/systemmine-large-scale-bilingual-dictionaries-monolingual-web-pages/

[12] L. Ballesteros and W. B. Croft, "Resolving ambiguity for cross-language retrieval," in *Proc. 21st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 1998, pp. 64–71.

[13] M. Aljlayl and O. Frieder, "Effective Arabic-english cross-language information retrieval via machine-readable dictionaries and machine translation," in *Proc. 10th Int. Conf. Inf. Knowl. Manage.*, 2001, pp. 295–302.

[14] J. Patel, K. Makvana, and P. Shah, "Cross-lingual information retrieval: Application and challenges for Indian languages," in *Proc. IEEE 5th Int. Conf. Converg. Technol.*, Dec. 2019, pp. 1–4.

[15] S. Zhuang, H. Li, and G. Zuccon, "Deep query likelihood model for information retrieval," in *Proc. 43rd Eur. Conf. Inf. Retr. (ECIR)*, 2021.

[16] J. Gao and J.-Y. Nie, "A study of statistical models for query translation: Finding a good unit of translation," in *Proc. 29th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2006, pp. 194–201.

[17] L. Zhang and X. Zhao, "An overview of cross-language information retrieval," in *Proc. Int. Conf. Artif. Intell. Secur.* Springer, 2020, pp. 26–37.

[18] N. Ferro and C. Peters, "Clef 2009 ad hoc track overview: Tel and Persian tasks," in *Proc. Workshop Cross-Lang. Eval. Forum Eur. Lang.*, 2009, pp. 13–35.

[19] E. Agirre, G. M. Di Nunzio, N. Ferro, T. Mandl, and C. Peters, "Clef 2008: Ad hoc track overview," in *Proc. Workshop Cross-Language Eval. Forum Eur. Lang.*, 2008, pp. 15–37.

[20] M. Adriani and I. Wahyu, "The performance of a machine translation-based English-Indonesian clir system," in *Proc. Workshop Cross-Lang. Eval. Forum Eur. Lang.*, 2005, pp. 151–154.

[21] L. Dolamic and J. Savoy, "Retrieval effectiveness of machine translated queries," *J. Amer. SoC. Inf. Sci. Technol.*, vol. 61, no. 11, pp. 2266–2273, Nov. 2010.

[22] M. Huang, S. Lu, and H. Xu, "Cross-language information retrieval based on weighted association patterns and rule consequent expansion," *Data Anal. Knowl. Discovery*, vol. 3, no. 9, pp. 77–87, 2019.

[23] A. Montazeralghaem, R. Rahimi, and J. Allan, "Term discrimination value for cross-language information retrieval," in *Proc. ACM SIGIR Int. Conf. Theory Inf. Retr.*, 2019, pp. 137–140.

[24] P. A. Chew, S. J. Verzi, T. L. Bauer, and J. T. McClain, "Evaluation of the Bible as a resource for cross-language information retrieval," in *Proc. Workshop Multilingual Lang. Resour. Interoperability*, 2006, pp. 68–74.

[25] B. Li and J. Yao, "Study on unknown term translation mining from Google snippets," *Information*, vol. 10, no. 9, p. 267, Aug. 2019.

[26] F. Huang, Y. Zhang, and S. Vogel, "Mining key phrase translations from web corpora," in *Proc. Conf. Human Lang. Technol. Empirical Methods Natural Lang. Process.*, 2005, pp. 483–490.

[27] B. Tianmei, "Research on English vocabulary translation technology based on dynamic bilingual corpus," in *Proc. Int. Conf. Comput., Inf. Process. Adv. Educ.*, Oct. 2020, pp. 288–291.

[28] M. Braschler and B. Ripplinger, "How effective is stemming and decompounding for German text retrieval?" *Inf. Retr.*, vol. 7, nos. 3–4, pp. 291–316, Sep. 2004.

[29] F. C. Gey, "Search between Chinese and Japanese text collections," in *Proc. NTCIR*, 2007, pp. 1–4.

[30] M. N. Mami, D. Graux, H. Thakkar, S. Scerri, S. Auer, and J. Lehmann, "The query translation landscape: A survey," 2019, *arXiv:1910.03118*.

[31] C. Chavula and H. Suleman, "Assessing the impact of vocabulary similarity on multilingual information retrieval for bantu languages," in *Proc. 8th Annu. Meeting Forum Inf. Retr. Eval.*, Dec. 2016, pp. 16–23.

[32] W. Kraaij, J.-Y. Nie, and M. Simard, "Embedding web-based statistical translation models in cross-language information retrieval," *Comput. Linguistics*, vol. 29, no. 3, pp. 381–419, 2003.

[33] M. N. Asim, M. Wasim, M. U. G. Khan, N. Mahmood, and W. Mahmood, "The use of ontology in retrieval: A study on textual, multilingual, and multimedia retrieval," *IEEE Access*, vol. 7, pp. 21662–21686, 2019.

[34] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[35] P. Sorg and P. Cimiano, "Cross-lingual information retrieval with explicit semantic analysis," in *Proc. Workshop Notes CLEF*, 2008, pp. 1–13.

[36] Z. Yahya, M. T. Abdullah, A. Azman, and R. A. Kadir, "Query translation using concepts similarity based on Quran ontology for cross-language information retrieval," *J. Comput. Sci.*, vol. 9, no. 7, p. 889, 2013.

[37] D. Zhou, M. Truran, T. Brailsford, and H. Ashman, "A hybrid technique for english-Chinese cross language information retrieval," *ACM Trans. Asian Lang. Inf. Process.*, vol. 7, no. 2, pp. 1–35, Jun. 2008.

[38] A. F. A. Nwesri and H. A. Alyagoubi, "Applying Arabic stemming using query expansion," in *Proc. 26th Int. Workshop Database Expert Syst. Appl. (DEXA)*, 2015, pp. 299–303.

[39] F. Ture and E. Boschee, "Learning to translate: A query-specific combination approach for cross-lingual information retrieval," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 589–599.

[40] B. Elayeb and I. Bounhas, "Arabic cross-language information retrieval: A review," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 15, no. 3, pp. 1–44, 2016.

[41] A. Alqudsi, N. Omar, and K. Shaker, "Arabic machine translation: A survey," *Artif. Intell. Rev.*, vol. 42, no. 4, pp. 549–572, Dec. 2014.

[42] S. E. Robertson and K. S. Jones, "Relevance weighting of search terms," *J. Assoc. Inf. Sci. Technol.*, vol. 27, no. 3, pp. 129–146, 1976.

[43] S. Robertson and H. Zaragoza, *The Probabilistic Relevance Framework: BM25 and Beyond*. Now, 2009.

[44] I. Ounis, G. Amati, V. Plachouras, B. He, C. Macdonald, and D. Johnson, "Terrier information retrieval platform," in *Proc. Eur. Conf. Inf. Retr.* Springer, 2005, pp. 517–519.

[45] X. Li and W. B. Croft, "Statistical language modeling for information retrieval," *Ann. Rev. Inf. Sci. Technol.*, vol. 39, no. 1, pp. 1–31, 2005.

[46] C. Zhai, "Statistical language models for information retrieval," *Synth. Lectures Hum. Lang. Technol.*, vol. 1, no. 1, pp. 1–141, Jan. 2008.

[47] H. P. Luhn, "A statistical approach to mechanized encoding and searching of literary information," *IBM J. Res. Develop.*, vol. 1, no. 4, pp. 309–317, 1957.

[48] K. Spärck Jones, "A statistical interpretation of term specificity and its application in retrieval," *J. Documentation*, vol. 60, no. 5, pp. 493–502, Oct. 2004.

[49] G. Amati, "Frequentist and Bayesian approach to information retrieval," in *Proc. Eur. Conf. Inf. Retr.* Springer, 2006, pp. 13–24.

[50] G. Amati and C. J. Van Rijsbergen, "Probabilistic models of information retrieval based on measuring the divergence from randomness," *ACM Trans. Inf. Syst.*, vol. 20, no. 4, pp. 357–389, Oct. 2002.

[51] H. Schutze, C. D. Manning, and P. Raghavan, *Introduction to Information Retrieval*, vol. 39. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[52] K. Dukes and N. Habash, "Morphological annotation of quranic Arabic," in *Proc. LREC*, 2010, pp. -1–7.

**AYMAN A. TAAN** was born in Jordan. He received the bachelor's degree in computer science and the master's degree in computer information systems from The Hashemite University, Jordan. His research interests includes learning about Information Retrieval and data mining field. He is a Professional Trainer in web development, motivator, and have passion in transferring his knowledge to others.

**SHAFIQ UR REHMAN KHAN** received the Ph.D. degree in time-aware information retrieval from the Capital University of Science and Technology, Islamabad, Pakistan. He is currently working as an Assistant Professor and the Head of the Department of the University of Sialkot, Pakistan. He is also a member of the PCN Research Group and 5th Dimension Research Laboratory. He has five years of experience in developing natural language processing, machine learning, and information retrieval systems. His research interests include information retrieval systems, temporal information retrieval systems, natural language processing, and machine learning.

**ALI RAZA** was born in Islamabad, Pakistan. He received the B.S. degree in computer science and the M.S. degree in software engineering. He is currently a Ph.D. scholar with the University of Sialkot. He is also working as an Assistant Director Academic and a Lecturer with the University of Sialkot. His research interests include natural language processing, information retrieval, and data mining.

**AYAZ MUHAMMAD HANIF** received the Ph.D. degree from the School of Design, Shanghai Jiao Tong University, China. He is currently an Enthusiastic Freelancer with UX/UI expertise. He is also working as an Assistant Professor and the Director of the Incubation Center, University of Sialkot. His research interests include the area of learning technology, user experience design, game-based learning, and affective technology in learning environment.

**HIRA ANWAR** was born in Islamabad, Pakistan, in 1997. She received the B.S. degree in information technology from the University Institute of Information Technology and the M.S. degree in project management from the Capital University of Science and Technology, in 2019. Since 2020, she has been working as a Research Assistant with the University of Sialkot. Her research interests include software project management, natural language processing, and software business models.