Research Journal of Information Technology, 2011
ISSN 1815-7432 / DOI: 10.3923/rjit.2011.


# A Novel Document Ranking Algorithm That Supports Mobile Healthcare Information Access Effectiveness

K. Agbele, A. Adesina, N.A. Azeez, A. Abidoye and R. Febba
Department of Computer Sciences, University of the Western Cape, Soft Computing and Intelligent Systems Research Group, Cape Town, South Africa

*Corresponding Author: K. Agbele, Department of Computer Sciences, University of the Western Cape, Soft Computing and Intelligent Systems Research Group, Private Bag X17, 7535, Bellville, Cape Town, South Africa*

## ABSTRACT

This study presented DROPT; an acronym for Document ranking Optmization algorithm approach, a new idea for the effectiveness of meaningful retrieval results from the information source. Proposed method extracted the frequency of query keyword terms that appears within the user context of Frequently Asked Questions (FAQ) systems on HIV/AIDS content related-documents. The SMS messages were analyzed and then classified, with the aim of constructing a corpus of SMS related to HIV/AIDS. This study presented a novel framework of Information Retrieval Systems (IRS) based on the proposed algorithm. The developed DROPT procedure was used as an evaluation measure. This "Term Frequency-Inverse Document Frequency (TFIDF)" method was applied to obtain the experimental result that was found promising in ranking documents not only the order in which the relevant documents were retrieved, but also both the terms of the relevant documents in feedback and the terms of the irrelevant documents in feedback might be useful for relevance feedback, especially to define its fitness function (mean weight).

**Key words:** Information retrieval, information retrieval system, ranking function, context awareness, relevance feedback, mobile information access, HIV/AIDS management

## INTRODUCTION

Now-a-days, increasing numbers of people use web search engines which enable them to access any kind of information from the Internet, in order to formulate better, well-informed decisions. However, the ability of search engines to return useful and relevant documents is not always satisfactory. Often users need to refine the search query several times and search through large document collections to find relevant information. Furthermore, with the emergent proliferation of mobile devices, users are increasingly using Internet services on the go. According to searchenginewatch.com, major search engines such as Google and Yahoo, take delivery of millions of search request per day. This fact obviously demonstrates the significance of search engines in our daily life (Glover *et al*., 2001).

As discussed by Agbele *et al*. (2010), access to information has important benefit that can be achieved in many areas including social-economic development, education and healthcare. In healthcare for example, access to appropriate information can minimize visits to physicians and period of hospitalization for patients suffering from chronic conditions, such as asthma, diabetes, hypertension and HIV/AIDS. Agbele,s method examines opening of health information system

based on ICT as one fundamental healthcare application area, especially within the context of the Millennium Development Goals to improve the management and quality of healthcare for development at lower cost.

Context awareness is, thus, the ability of an entity to be aware of the surrounding situations and use the information to perform some tasks. An entity can be a person, a place, or an object that is considered relevant to the interaction between a user and an application, including the user and application themselves (Fernando, 2004). Further, the first kind of context is defined as active context that influences the behaviours of an application and the second kind of context as passive context that is relevant but not critical to the application. This classification helps to understand the use of context in mobile applications.

Prasannakumari (2010) develops a very simple efficient method for contextual information retrieval from multimedia databases to meet any individual user information needs. Prasannakumari,s method combines learning by feedback approach and improved relevant ranking to build a better database. In this regards, context information can be environmental, application or device-oriented or user-related. Based on the contextual information acquired, a mobile system reacts, adapts and responds accordingly but only within the parameters that determine the perceived context.

Adesina *et al.* (2010) used SMS messages as a tool in a health provision environment in different forms of communication to form a set of pre-formed questions related to HIV/AIDS. The SMS were provided for all group participant of first year Computer Science Department, University of the Western Cape to form the SMS- Corpus. Therefore, an information retrieval system has its heart a collection database about certainty (Korfhage, 1997). In this regards, Information Retrieval System (IRS), is a system used to store items of information that need to be processed, searched and retrieved corresponding to a user's query.

According to relevant literatures of Nyongesa and Maleki-Dizaji (2006), Mauldin *et al.* (1987) and Chen *et al.* (2010), most IRSs suffer from keywords barriers to convey the semantic context meaning of retrieve documents. Further, the system first extracts keyword terms by using different approaches. As a consequence, such a system has two key problems; one is how to extract keywords specifically and the other is how to decide the weight of each keyword.

Bani-Ahmad and Al-Dweik (2011) proposed a new term-ranking approach that gives an approximation of the relative importance of the terms within the document where they are observed to improve similarity scores. This study presents DROPT algorithm procedure as relevant feedback from human assessment based on TFIDF method aiming to effectively adapt SMS-query keywords weights. Hence, user query reformulation applies by updating its profile. A user profile or model is a stored knowledge about a particular user. Simple model consists usually of keywords describing user's area of interest in context.

## CONCEPT OF THE PROPOSED DROPT ALGORITHM PROCEDURES

Based on Eq. 4, a ranking algorithm for documents retrieved from a corpus is developed with respect to document index keywords and the query vectors. This based on calculating the weight ($w_{ij}$) of keywords in the document index vector, calculated as a function of the frequency of a keyword $k_j$ across a document $d_i$.

Let a query vector, Q, be defined as:

$$Q = [q_1\ q_2\ q_3 \ldots q_1] \quad (1)$$

where, $q_i = (x_i, 1)$, $x_i$ being a term string with a weight of 1.

Let the indexed document corpus be represented by the matrix:

$$D=\begin{bmatrix} tf_{11}\ tf_{12}\ tf_{13}\cdots tf_{1l} \\ tf_{21}\ tf_{22}\ tf_{23}\cdots tf_{2l} \\ \vdots \\ \vdots \\ tf_{n1}\ tf_{n2}\ tf_{n3}\cdots tf_{nl} \end{bmatrix} \tag{2}$$

where, $d_{jk} = (y_{jk}, w_{jk})$, $y_{jk}$ being an index string, with weight $w_{jk}$.

Therefore, this leads to compute the convolution matrix, representing:

$$W = D\times Q = \begin{bmatrix} w_{11}\ w_{12}\ w_{13}\cdots w_{1l} \\ w_{21}\ w_{22}\ w_{23}\cdots w_{2l} \\ \vdots \\ \vdots \\ w_{n1}\ w_{n2}\ w_{n3}\cdots w_{nl} \end{bmatrix} \tag{3}$$

where, $w_{ij} = w_{kl}$ iff IsEqualStringIgnoreCase ($q_p$, $d_{rs}$); 0, otherwise; $|l| \leq |n|$, l being the number of terms in the query vector and n the number of retrieved documents that are indexed by at least one keyword in the query vector.

Salton (1970) studied weighted relevance of terms in a document by considering term frequency (tf) and term document frequency (idf). Term frequency is the number of times a given term occurs in a given document, while document frequency is the number of times the term occurs in all documents. The author argued that the more a term occurs in one document but less in other documents, the more relevant it is to that document. Consequently the relevance weight is proportional to the term frequency and inverse document frequency. In study of Salton and Buckley (1988), the relevance weight is given by:

$$w_{ij} = tf\times idf \tag{4}$$

$$\text{where } tf = \frac{freq_{ij}}{\text{total keywordcnt}},\ idf = \log\left(\frac{N}{n_k}\right)$$

where $freq_{ij}$ is the frequency of the $K_{ith}$ user in $D_{jth}$ query; totalkeywordcnt is the total keyword count in the document databases; $n_k$ is the number of documents indexed by the keyword $k_j$ and finally, N is the total number of documents containing keyword $k_j$.

To determine the overall fitness of all documents with respect to a given query, mean weight values for the term weight vectors are calculated as:

$$\varpi = \prod_{i=1}^{n} 1/l \sqrt{\sum_{j=1}^{l} w_{ij}^2} \tag{5}$$

For searching user weight of each vector term, a weighting approach (semantic process) of FAQ document collection based on TFIDF method is used. $tf_{ij}$ is defined as the number of occurrences of

keyword term $k_j$ in document $d_i$ and $idf_j$ defined as log ($N/df_j$) and $df_i$ is keyword the number of documents containing $k_j$ in which N is the total number of documents containing keyword $k_j$.

The relevance of a document will be measured according to the degree of fitness DF of the document with respect to the query vector with a small-operator defined as matrix G below:

$$G = [g_{ij}]_{n\times l}, \text{where } g_{ij} = \min\left(w_{ij}, \bar{q}_{ij}\right) \quad 1 \le i \le n, 1 \le j \le l \tag{6}$$

Therefore, any weight component of matrix G greater than the mean weight values will be retained to add to a matrix T is given by:

$$T = [t_{ij}]n\times l, \quad \text{where } \begin{cases} t_{ij} = g_{ij}, & \text{if } g_{ij} \ge \varpi \\ t_{ij} = 0, & \text{if } g_{ij} < \varpi \end{cases} \quad 1 \le i \le n, 1 \le j \le l \tag{7}$$

Based on matrix T, we calculate scores, $sco_i$, of all documents which are the largest weighting value of each corresponding vector is given by:

$$Sco_i = \max, \{t_{ij}\}, 1 \le i \le n \quad 1 \le j \le l \tag{8}$$

Document $d_i$ is retrieved if $sco_i$ is greater than zero ($sco_i > 0$) and added into the retrieved document set, D shown in Eq. 9.

So, average score ranging between 0 and 1 is computed for each document. Documents are sorted in ascending order of $Sco_i$, hence ranked and is given to the user:

$$D = \{d_i \mid \text{if } Sco_i > 0, 1 \le i \le n\} \tag{9}$$

The keyword set K provided by the documents and the weight values will be updated by the feedback of the users.

- Any new query term not belonging to K will be added and a new column of weight value will be computed and expanded for documents routinely
- If any retrieved document $d_i$ is retrieved by the users, the corresponding weight values with respect to the query keywords will be increased by Eq. 10. The default of $\beta$ is set to increase the corresponding weight values:

$$w_{ij} = (w_{ij})^{\beta}, \text{ where } 0 < \beta < 1, i \in \{i \mid d_i \in D\} \text{ and } j \in \{j \mid q^j = 1\} \tag{10}$$

The proposed DROPT algorithm procedures: an acronym for Document ranking Optimization will provide a limited number of ranked documents in response to a given query. It will also improve the ranking mechanism for the search results in an attempt to adapt the retrieval environment of the users and amount of relevant information according to each user's request. Finally, the proposed algorithm must be self-learning that can automatically adjust its search structure to a user's query behaviour.

**Issues to be resolved by the concept:**

- (i) The ability of the search engines to return useful and relevant documents is not always satisfactory. Often users need to refine the search query several times and search through large document collections to find relevant information. In this regards, these issues have been discussed in literature with the thought of using optimization techniques according to Glover *et al.* (1999, 2001). However, the necessary amount of relevant information is varied from diverse users. Erba *et al.* (2011) can enable the individual users to explore explicit relevance feedback to measure the variability in judgements and behaviour for the given query for ranking.

Erba *et al.* (2011) allows individual users to explore explicit relevance feedback to measure the variability in judgements and behaviour for a given query for ranking. The explicit relevance feedbacks give room to observe the consistency in relevance assessments across different individual users. The major challenge of this study includes how to gather satisfactory data and it is burdensome for users to provide explicit judgements. Thus, how to provide suitable amount of relevant information according to individual user information needs is what to be addressed in this study.

- (ii) It is important to lay emphasis on how to improve the ranking mechanism for the searching results of FAQ on HIV/AIDS content-related documents from the search engine. According to satisfying the users' preference, genetic algorithms have been helpful by many researchers to improve the search queries (Salton and Buckley, 1988; Yang and Korfhage, 1993). Though, their systems failed to offer a satisfactory evaluation to score and rank the retrieved information constantly.
- (iii) As discussed by Lin *et al.* (2006), Billerbeck *et al.* (2003) and Kim *et al.* (2001), query expansion afforded system users with relevant results from online users' feedback. However, highlighted below are the major flaws:

- Their system reformulate processes require users' additional preference based on the previous retrieved result
- Their system cannot make use of users' query experience to help the new users
- The existing search systems cannot change the search structure, whenever a user takes some actions, for instance, retrieving a correct relevant documents. Thus, self-learning IRS that can automatically adjust its search structure to user query behaviour is both valuable and essential

Hoque and Avery (2010) proposed and designed concept that support faster query execution. The results perform quicker and efficient having both time and space complexities reduced considerably. In this paper, a new method is proposed based on the three issues (i-iii) evaluated from the existing IRS of effectiveness, of ranking mechanism and self-adjustment of the users to improve mobile retrieval performance results in a health provision environment.

## THE PROPOSED DROPT ALGORITHM APPROACH

Based on the promise concepts described previously, here we proposed the procedure with the evaluation of the DROPT algorithm procedures effectiveness by a demonstrated example.

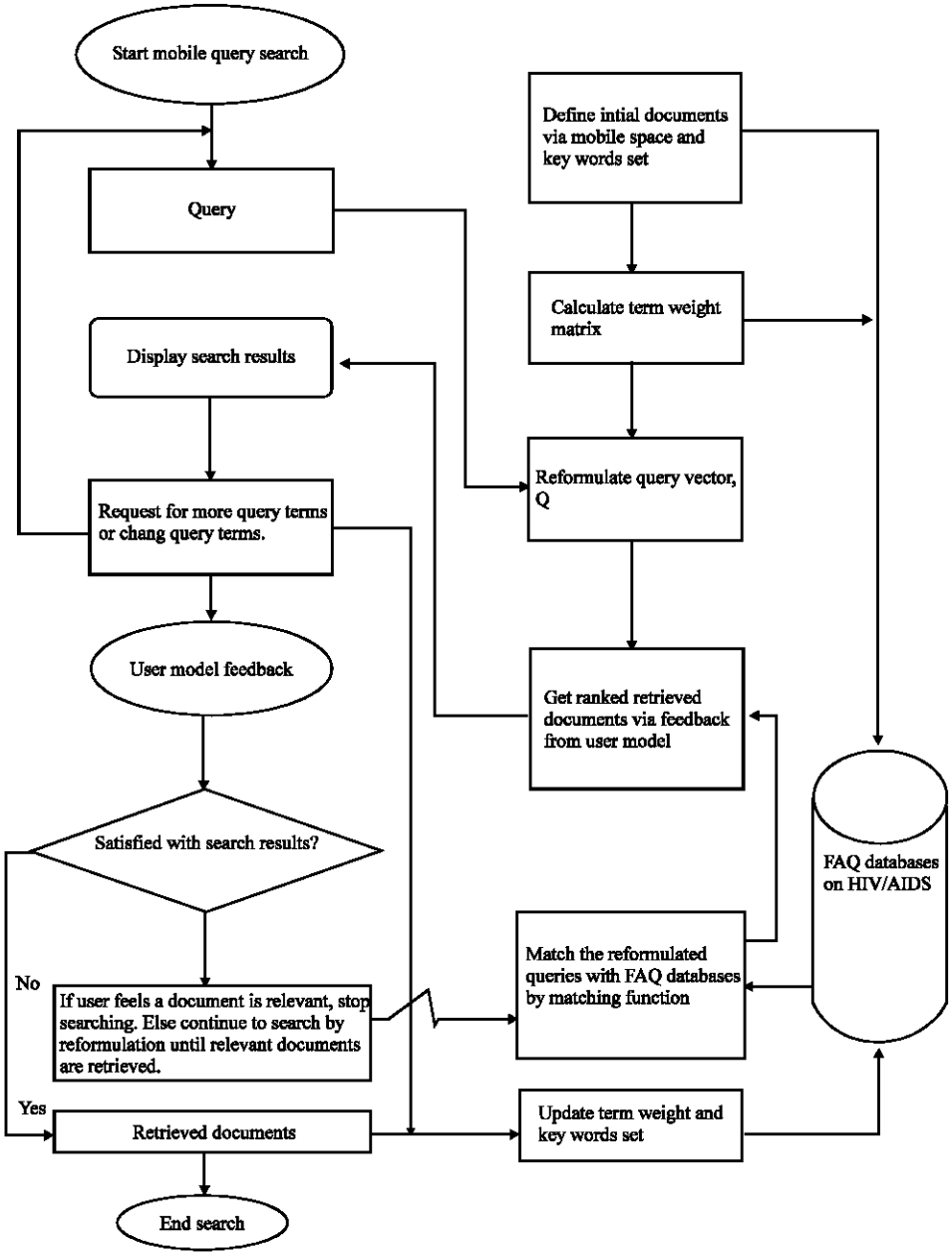


Fig. 1: Flowchart of the proposed algorithm procedures

**Present DROPT algorithm:** The DROPT algorithm is described below with its flowchart (Fig. 1).

**Stage 1: Initialization:**

- Set the initial index document corpus, $D_0 = \{d_1, d_2, d_3 \ldots, d_n\}$ and obtain the initial query keywords set, $K = \{k_1, k_2, k_3 \ldots, k_l\}$
- Define a set B, with the features of the documents as, $B = \{B_1, B_2\}$ where $B_1$ is the publishing or presentation year and $B_2$ is the properties of documents, including journal, thesis, conference, seminar, patent, textbooks, health technical reports, HIV/AIDS reports
- Set value for $\beta$

**Stage 2: Calculate Term weight matrix from the FAQ database:**

- Calculate the term frequency (tf)
  For each $k_j$ in K
  For each $d_i$ in $D_0$
  Find the number $tf_{ij}$ = ($freq_{i,j}$/total keywordcnt)
- Calculate the inverse document frequency (idf)
  For $k_j$ = 1 to $l$
  Put $n_j$ = 0
  For $d_i$ = 1 to n
  If $tf_{ij}$ > 0 Then $n_j$ = + 1
  Find the number idf = log ($N/n_j$)
- Compute convolution weight matrix
  For $q_i$ = $x_i$, 1
  For $d_{nl}$ = $y_{nl}$, $w_{nl}$
  Calculate as Eq. 3
- Get term set $w_{ij}$
  For $k_j$ = 1 to $l$
  For $d_i$ = 1 to n
  Calculate $w_{ij}$ as Eq. 4
- Calculate the mean weight $\varpi$ as Eq. 5

**Stage 3: Reformulate a query:**

- Formulate a query via mobile online interface
- If a user selects the features of B, filter n documents by $B_1$ and $B_2$ and obtain $\bar{n}$ documents
- Define query vector Q
  For each $k_j$ in K
  If ($k_j$ matches the query terms) Then $q_j$ = 1
  Else $q_j$ = 0

**Stage 4: Get feedback via user model from documents to be retrieved:**

- Create matrix G
  For i = 1 to $\bar{n}$
  For j = 1 to $l$
  $g_{ij}$ = min ($w_{ij}$, $\bar{q}_{ij}$) as Eq. 6
- Create matrix T by the mean weight $\varpi$
  For i = 1 to $\bar{n}$
  For j = 1 to $l$
  If $g_{ij}$ > = $\varpi$, then $t_{ij}$ = $g_{ij}$
  Else $t_{ij}$= 0 as Eq. 7
- Compute the scores and generate D, for the sets of retrieved documents
  For = 1 to $\bar{n}$
  $Sco_i$ = max ($t_{ij}$) for j = 1 to $l$

If $sco_i > 0$ Then
Add $d_i$ into D as Eq. 8

- Display the sets of retrieved documents according to the rank of the related scores i.e., Retrieval Status Value (RSV)
  For $d_i$ in D
  Sort $sco_i$ and display results as Eq. 9

**Stage 5: Match the reformulated queries with FAQ documents:**

- If a user feels that the document is relevant, he finishes the search. Then GO to Stage 4 to get $\varpi$ according to user's preference function
- Else, user continues to search in the database by reformulating the query, or stop querying until the relevant documents are retrieved
  GO to Stage 6

**Stage 6: Update term weight values and keywords set:**

- Update term weight values
  For $d_i = 1$ to $\bar{n}$ and $d_i \in D$
  If $d_i$ is retrieved
  For $j = 1$ to $l$ and $q_j = 1$
  Update $w_{ij}$ as Eq. 10
- Update keywords set, K
  For any query term $q_k$ not in K then
  Add $q_k$ into K
  For $d_i = 1$ to n, $k_j = l + 1$
  Calculate $w_{ij}$ as Eq. 4
- If user want to reformulate query Then
  GO to Stage 3

Else, Stop.

**Testing the validity of the proposed DROPT algorithm using TFIDF method:** This sub-section describes the effectiveness of document ranking terms procedure, including 10 document databases and 5 extracted SMS-query keywords set on HIV/AIDS content-related documents using TFIDF method.

**Stage 1:** The initial query keywords were first collected into the set K = {HIV, AIDS, symptoms, awareness, treatments} in the initial stage

**Stage 2:** The number of each keyword term occurred in each FAQ database was counted as keyword frequency and listed as shown in Table 1

**Stage 3:** Convolution weight matrix is computed as Eq. 3 to obtain Table 2

**Stage 4:** Therefore, the overall fitness of the entire documents with respect to a given query, mean weight values for the term weight vectors are obtained from Eq. 5 and listed in Table 3.

Table 1: Extracted significant keywords in each FAQs document

| Index documents | HIV | AIDS | Symptoms | Awareness | Treatment |
|---|---|---|---|---|---|
| $d_1$ | 1 | 2 | 2 | 1 | 0 |
| $d_2$ | 6 | 0 | 1 | 3 | 0 |
| $d_3$ | 3 | 3 | 0 | 0 | 2 |
| $d_4$ | 0 | 0 | 4 | 1 | 0 |
| $d_5$ | 1 | 6 | 1 | 0 | 0 |
| $d_6$ | 0 | 8 | 1 | 6 | 4 |
| $d_7$ | 3 | 0 | 0 | 0 | 12 |
| $d_8$ | 0 | 0 | 0 | 0 | 4 |
| $d_9$ | 5 | 0 | 1 | 1 | 1 |
| $d_{10}$ | 1 | 2 | 0 | 4 | 0 |

Table 2: Convolution weight matrix of Eq. 1 and 2

| | HIV | AIDS | Symptoms | Awareness | Treatment |
|---|---|---|---|---|---|
| $d_1$ | 0.866 | 0.163 | 0.076 | 0.000 | 0.076 |
| $d_2$ | 0.693 | 0.000 | 0.000 | 0.390 | 0.000 |
| $d_3$ | 0.433 | 0.488 | 0.000 | 0.000 | 0.306 |
| $d_4$ | 0.000 | 0.000 | 0.978 | 0.260 | 0.000 |
| $d_5$ | 0.144 | 0.976 | 0.153 | 0.000 | 0.000 |
| $d_6$ | 0.000 | 0.548 | 0.064 | 0.386 | 0.257 |
| $d_7$ | 0.231 | 0.000 | 0.000 | 0.000 | 0.978 |
| $d_8$ | 0.000 | 0.000 | 0.000 | 0.000 | 1.222 |
| $d_9$ | 0.722 | 0.000 | 0.153 | 0.163 | 0.153 |
| $d_{10}$ | 0.165 | 0.372 | 0.000 | 0.698 | 0.000 |

Table 3: Mean weight ($\varpi$) calculated for each document

| Index documents | Mean weight ($\varpi$) |
|---|---|
| $d_1$ | 0.179 |
| $d_2$ | 0.161 |
| $d_3$ | 0.144 |
| $d_4$ | 0.202 |
| $d_5$ | 0.199 |
| $d_6$ | 0.144 |
| $d_7$ | 0.201 |
| $d_8$ | 0.244 |
| $d_9$ | 0.154 |
| $d_{10}$ | 0.162 |

Table 4: Mean weight ($\varpi$)= 0.179 for overall fitness is compared with the weight of each document to determine their relevance for ranking

| Index documents | Mean weight ($\varpi$) of each document | Overall fitness mean weight |
|---|---|---|
| $d_1$ | 0.179 | ≥0.179 |
| $d_2$ | 0.161 | |
| $d_3$ | 0.144 | |
| $d_4$ | 0.202 | ≥0.179 |
| $d_5$ | 0.199 | ≥0.179 |
| $d_6$ | 0.144 | |
| $d_7$ | 0.201 | ≥0.179 |
| $d_8$ | 0.244 | ≥0.179 |
| $d_9$ | 0.154 | |
| $d_{10}$ | 0.162 | |

**Stage 5:** The relevance of a document is measured according to the degree of fitness with respect to the query vector as Eq. 6. So, the weight element of matrix G greater than the mean weight values is obtained and then matrix G is obtained from Eq. 7 and listed in Table 4

**Stage 6:** Based on matrix T, scores, $sco_i$, is calculated for the entire documents, which are the largest weighting value of each corresponding vector. Therefore, the retrieved set is D = {$d_8$, $d_4$, $d_7$, $d_5$ and $d_1$} from Eq. 8

According to Eq. 9, documents are sorted in ascending order of $sco_i$ and hence ranked and given to the user. The ranking of the retrieved set is D = {$d_8$ = 0.244, $d_4$ = 0.202, $d_7$ = 0.201, $d_5$ = 0.199 and $d_1$ = 0.179}. However, Eq. 10 can only be updated when a user makes a query including two terms. Hence, the weight value will increase according to the keywords provided by the two terms.

We then found that the ranking of the retrieved set is D = {$d_8$ = 0.244, $d_4$ = 0.202, $d_7$ = 0.201, $d_5$ = 0.199 and $d_1$ = 0.179} is sorted in ascending order which provides a limited number of ranked documents in response to a given query. It also improves the ranking mechanism for the search results in an attempt to adapt the retrieval environment of the users and amount of relevant information according to each user's request. Finally, the proposed algorithm is self-learning that routinely adjust its search structure to a user's query behaviour.

## AN INFORMATION RETRIEVAL SYSTEM-A PROPOSED FRAMEWORK BASED ON THE DEVELOPED DROPT ALGORITHM APPROACH

In the proposed framework for information retrieval as depicted in Fig. 2, user gives a mobile SMS-query (Raw Query) and the query is reformulated in order to improve the predicted relevance of the retrieved document. The reformulated query is searched against the databases. The proposed retrieval system incorporates the frequency of keyword terms that appear in FAQs databases related to HIV/AIDS content related-documents using term weighting TFIDF method by optimizing the ranking order of retrieved documents from the search engine. The information retrieval system searches for the matches in the document databases and thus retrieves search results of the matching process.

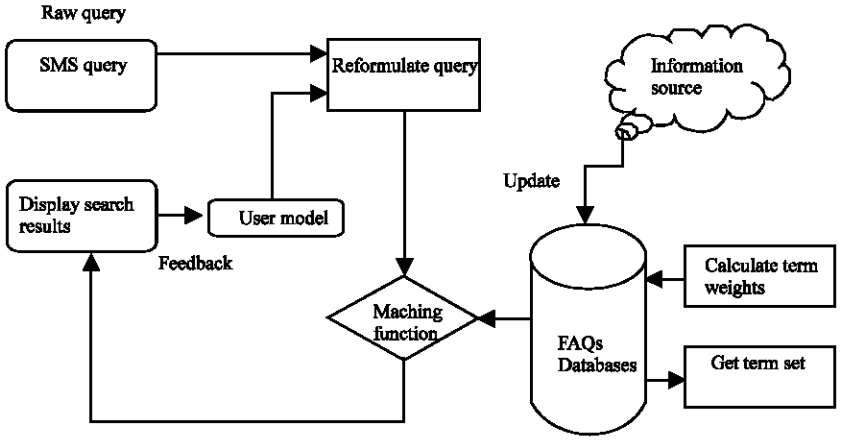


Fig. 2: Information retrieval system-proposed framework

Based on the relevance, the user will then display the search results. The relevance of the document is very important to the user. If the user feels that it is a relevant document, he finishes the search else user continues to search in the document database by reformulating the query until the relevant documents are retrieved that will satisfy users' information needs. Hence, user query reformulations will apply by updating its model. A user model is a stored knowledge about a particular user. Simple model consists usually of keywords describing user's area of interest. Sort those documents according to TFIDF method. The documents which have the high Retrieval Status Value (RSV) are considered as the top ranked documents.

The two main components in the proposed information retrieval system framework are document databases and reformulated query processing system. The document databases stores the databases related to documents and the representations of their information contents based on TFIDF method. A SMS-query keyword term is also associated with this component which automatically generates a representation for each document by extracting the frequency of the SMS-query keyword terms from the document contents. The reformulated query processing system consists of two subsystems: Searching-Matching Unit and Displaying-Ranking Unit.

Searching unit allows user to search the documents from the document database and matching unit does a comparison of all documents against the user's query. To improve the predicted relevance of the retrieved document, the reformulated query is searched against the databases. Searching-Matching unit does a thorough search and finds out which documents match the user query. This unit retrieves almost all the documents that match either part or whole of the entire query, that is, the unit retrieves relevant amid non relevant documents.

Displaying unit displays the search results based on relevance of the documents to user information needs and ranking unit ranks the document according to the relevance of the user query. Displaying-Ranking unit does a detailed display of search results and find out which documents have high RSV are considered as the top ranked documents. Therefore, Information Retrieval (IR) system ranks the documents according to the RSV between document and the query. If a document has got high RSV, that document is closer to the query. In other words the document is relevant to the query.

Generally IR system ranks the list of documents in the descending order. After processing the query effectively, the top most relevant documents are retrieved and it is given to the user. Though, relevance feedback is one of the processes in an information retrieval system that seeks to improve the system's performance based on a user's feedback. It modifies queries using judgments of the relevance of a few, highly-ranked documents and has historically been an important method for increasing the performance of information retrieval systems.

Specifically, the user's judgments of the relevance or non-relevance of some of the documents retrieved are used to add new terms to the query and to reweight query terms. For example, if all the documents, that the user judges as relevant contain a particular term, then that term may be a good one to add to the original query. It is made known by Salton (1970) that relevance feedback has improved the system's overall performance by 60 to 170% for different document collections. Given the apparent effectiveness of relevance feedback techniques, it is important that any proposed model of information retrieval include these techniques. In our proposed system, rather than modifying the matching function, we will modify the query vector using genetic algorithm to adapt the query vectors and to reflect a user's feedback about relevance.

## EXPERIMENTAL EVALUATION PERFORMANCE RESULTS

The values displayed in Table 4 shows the results of the evaluation based on the developed ranking algorithm for documents retrieved from a corpus of documents index keyword that are

expected to be found by the search with its associated mean weight and the query vector using keyword based IR. Average score ranging between 0 and 1 is computed for each document. Documents are sorted and were set as input in ascending order of Retrieval Status Values (RSV). Hence, ranked documents $d_8 = 0.244$, $d_4 = 0.202$, $d_7 = 0.201$, $d_5 = 0.199$ and $d_1 = 0.179$ is given to the user. The mean weight of each ranked document is greater than the overall fitness mean weight whose value is $(\varpi) = 0.179$ which satisfies set condition of Eq. 7. It demonstrated that irrespective of the retrieved document length, it gives response of the mean weight value of the user's document.

The satisfactory levels of the user were evaluated in Offline mode. While there is no information for analysis on precision and recall, testing system's effectiveness by self satisfaction was an alternative way adopted to include how relevant is the retrieved documents? And is the user satisfies with the function of adding personal new query keywords according to user's preference function? The proposed algorithm for document ranking optimization provides a limited number of ranked documents in response to a given query. It also improves the ranking mechanism for the search results in an attempt to adapt the retrieval environment of the users and amount of relevant information according to each user's request. Finally, the proposed algorithm is self-learning that routinely adjusts its search structure to a user's query behaviour.

## DISCUSSION

In our proposed method, the existing keyword set is collected from the FAQs databases as determined by the authors. The keyword set is extracted from all documents in conventional information retrieval which is time-consuming. The subsequent task of this research will focus on how to develop semantic information retrieval system that will overcome the drawback of keyword-based techniques by extracting useful semantics in mobile information for indexing and matching of content semantic. The GA will be used to adapt keywords' weights. The retrieval effectiveness will be evaluated in terms of recall and precision measurements and the proposed IRS is allied to mobile healthcare information access.

Though, this research project is at development and implementation stage. It is our strong belief that the full implementation and evaluation of the proposed information retrieval systems will assist users in documents ranking order according to their relevance. The approach retrieves limited number of ranked documents' identified keywords in response to a given query. It's easier to retrieve using keywords and this damage document retrieval performance. One solution to this is Eq. 10 that develops a relevant feedback mechanism such that keywords can be added or removed. Genetic algorithm will be used to adapt keywords' weights for optimal or near optimal solutions (Goldberg, 1989; Holland, 1975) in on-line mode using Java-script for implementation. Therefore, HIV/AIDS content-related documents with higher similarity query are to be judged more relevant to the query keyword terms and should be retrieved first to adapt the query vectors via feedback of the users. This will in turn help HIV/AIDS managements and lower the cost of healthcare provision.

Finally, investigation in the related works in the literature reveals that document ranking have not been sufficiently studied. Hence, the approach outlined in this study has better retrieval performance that requires less time than (Hoque and Avery, 2010; Bani-Ahmad and Al-Dweik, 2011; Prasannakumari, 2010) algorithm approaches does due to limited number of ranked documents. The DROPT algorithm approach guide the document to better retrieval effectiveness though limited, and can adjust the weights of keywords according to information from the indexed

documents. So performs better and support their findings. However, limited improvement was discovered. In the future, it is propose to design a good relevant feedback method such that performance of document retrieval can be improved.

## CONCLUSIONS

In this study, a new method is proposed based on three issues evaluated from the existing systems of effectiveness, self adjustment and improving ranking mechanism to the users. The proposed algorithm for document ranking optimization provides a limited number of ranked documents in response to a given query. It also improves the ranking mechanism for the search results in an attempt to adapt the retrieval environment of the users and amount of relevant information according to each user's request. Finally, the proposed algorithm is self-learning that routinely adjusts its search structure to a user's query behaviour.

The effectiveness of the system performance was evaluated numerically based on the self satisfaction of the feedback of the users' using TFIDF method. The algorithm has demonstrated the ability of providing satisfactory functions for users to add relevant feedback mechanism to improve document retrieval performance.

## ACKNOWLEDGMENT

We wish to thank Prof. Henry Nyongesa for his kind and helpful discussions and comments about the DROPT algorithm procedures. Also, we appreciate cooperative remarks from our colleagues in the research group.

## REFERENCES

Adesina, A.O., K.K. Agbele and H.O. Nyongesa, 2010. Text messaging: A tool in e-health services. Proceedings of the Southern Africa Telecommunication Networks and Applications Conference, Sept. 5-8, Stellenbosch, South Africa, pp: 1-15.

Agbele, K., H. Nyongesa and A. Adesina, 2010. ICT and information security perspectives in E-health systems. J. Mobile Commun., 4: 17-22.

Bani-Ahmad, S. and G. Al-Dweik, 2011. A new term-ranking approach that supports improved searching in literature digital libraries. Res. J. Inform. Technol., 3: 44-52.

Billerbeck, B., F. Scholer, H.E. Williams and J. Zobel, 2003. Query expansion using associated queries. Proceedings of the 12th International Conference on Information and Knowledge Management, Nov. 3-8, New Orleans, LA. USA., pp: 2-9.

Chen, M.Y., H.C. Chu and Y.M. Chen, 2010. Developing a semantic-enable information retrieval mechanism. Expert Syst. Applic., 37: 332-340.

Erba, F.G., Z. Yu and L. Ting, 2011. Using explicit measures to quantify the potential for personalizing search. Res. J. Inform. Technol., 3: 24-34.

Fernando, J., 2004. Factors that have contributed to a lack of integration in health information system security. J. Inform. Technol. Healthcare, 2: 313-328.

Glover, E.J., S. Lawrence, M.D. Gordon, W.P. Birmingham and C.L. Giles, 1999. Recommending web documents based on user preferences. Proceedings of the ACM Workshop on Recommender Systems: Algorithms and Evaluation, RSAE'99, Berkeley, CA., USA., pp: 1-9.

Glover, E.J., S. Lawrence, M.D. Gordon, W.P. Birmingham and C.L. Giles, 2001. Web search-your way. Commun. ACM, 44: 97-102.

Goldberg, D.E., 1989. Genetic Algorithms in Search Optimization and Machine Learning. Addison Wesley Publishing Co., Reading, Massachutes.

Holland, J.H., 1975. Adaptation in Natural and Artificial System. University of Michigan Press, Ann Arbor, USA.

Hoque, M.T. and V.M. Avery, 2010. Novel strategies to speed-up query response. Res. J. Inform. Technol., 2: 11-20.

Kim, B.M., J.Y. Kim and J. Kim, 2001. Query term expansion and reweighting using term co-occurrence similarity and fuzzy inference. Proceedings of the Joint 9th IFSA World Congress and 20th NAFIPS International Conference, July 25-28, Vancouver, BC Canada, pp: 715-720.

Korfhage, R., 1997. Information Storage and Retrieval. John Wiley and Sons, USA., ISBN-10: 0471143383, pp: 368.

Lin, H.C., L.H. Wang and S.M. Chen, 2006. Query expansion for document retrieval based on fuzzy rules and user relevance feedback techniques. Expert Systems with Appli., 31: 397-405.

Mauldin, M., J. Carbonell and R. Thomason, 1987. Beyond the keyword barrier: Knowledge-based information retrieval. Inform. Services Use, 7: 103-117.

Nyongesa, H.O. and S. Maleki-Dizaji, 2006. User modelling using evolutionary interactive reinforcement learning. Inform. Retrieval, 9: 343-355.

Prasannakumari, V., 2010. Contextual information retrieval for multi-media databases with learning by feedback using vector space model. Asian J. Inform. Manage., 4: 12-18.

Salton, G. and C. Buckley, 1988. Term-weighting approaches in automatic text retrieval. Inform. Process. Manage., 24: 513-523.

Salton, G., 1970. Automatic text analysis. Science, 168: 335-342.

Yang, J.J. and R.R. Korfhage, 1993. Query optimization in information retrieval using genetic algorithms. Proceedings of the 5th International Conference on Genetic Algorithms, (ICGA'93), Morgan Kaufmann Publishers Inc., San Francisco, CA., USA., pp: 603-613.