An Enhanced Concept based Approach for User Centered Health Information Retrieval to Address Presentation Issues

Ibrahim Umar Kontagora[1], Isredza Rahmi A. Hamid[2], Nurul Aswa Omar[3]

Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM) [1, 2, 3]
Batu Pahat, Johor Malaysia
Department of Computer Science, Niger State Polytechnic, Zungeru, Niger State, Nigeria[1]

Abstract—**The diversity of health information seekers signifies the enormous variety of information needs by numerous users. The existing health information retrieval systems failed to address the information needs of both medical expert and laymen patients. This study focused on designing an enhanced information retrieval approach using the concept based approach that would address the information needs of both medical experts and laymen patients. We evaluated and compared the performance of the proposed enhanced concept based approach with the existing approaches namely: concept based approach (CBA), query likelihood model (QLM) and latent semantic indexing (LSI) approach using Diagnosia 7, Medical Subject Heading (MeSH), Khresmoi Project 6 and Genetic Home Reference datasets. The experimental results obtained shows that the proposed enhanced concept based approach manage to score similarity scores of 1.0 (100%) in respect to maxSim values for all the runs in all the four datasets and idf weighting values of between 3.82 – 3.86 for all the runs in all the four datasets. While the existing approaches (CBA, QLM, LSI) scored the maxSim scores of 0.5 (50%) for all their runs in all the four dataset and idf weighting values of between 1.40 – 1.47 for all the four dataset, as a result of their inability to generate and display medical search results in both medical experts and layman's forms. These results shows that the proposed enhanced concept based approach is the best approach suited to be used in addressing presentation issues.**

Keywords—Enhanced concept based approach; existing concept based approach; medical discharge reports; medical expert form; layman's form

I. INTRODUCTION

The variety of clinical information seekers implies the high rate of information needs and subsequently a major requirement for the design of health information retrieval systems [2, 4, 6]. In order to satisfy the health information needs of laymen patients and their relatives [7, 11]. Querying of Health Information Retrievals for health advice has now become a general and noble task performed by individuals on the Internet [11. 14]. Previous researches on health information retrievals reveals that a huge percentage of search engine users perform web search for their health related information in the United States (US) [14, 17]. Health information retrieval systems need to be improved to effectively satisfy laymen patients' health information needs.

As health information retrieval systems are continuously used to improve the excellence of medical services in hospitals, the size and diversity of information is increasing and becoming compound [1, 4, 8]. Probing of clinical information retrieval systems for health advice has become an obligation today due to the huge growth in health related information over the Internet [9, 10, 14]. With the spreading awareness on the exploration of information extracted from medical discharge documents and clinical reports by laymen patients, searching online health related web-forums and other sources for health advice has become a common habit [17, 18, 19]. Adequate attention should be given to the information needs of laymen patients and their relatives, in order to avoid wrong interpretations of medical prescriptions and diagnosis from health expert, which could worsen their health conditions.

One major technique implored by previous researchers is the investigation of users query logs from viable search engines, which reveals that most of the structures in place have no reservations for users' information desires [2, 4]. However, there is a lack of focus to the development of user centered health information retrieval system that would generate and display medical discharge reports and medical search queries results in both medical expert and layman's forms online [5]. And this has made it very difficult for the existing information retrieval approaches to address the information needs of both medical expert and laymen patients [10, 12]. The query logs / medical search results of laymen patients revealed that, the medical texts contents are highly professional and hard to follow [12, 13]. As they still need to ask additional questions from medical experts regarding the content of their query/search results [14, 17]. Appropriate attention should be given to the information needs of laymen patients and their relations.

The main objective of this study is to design a proposed enhanced concept based approach for user centered health information retrieval that would address the information needs of both medical experts and laymen patients. It would generate and display medical search results in two separate forms (i.e. medical experts and layman's forms). The study has identified the major cause of the presentation issues to the ability of the current information retrieval systems concentrating on a specific group of people with expert health knowledge. The proposed enhanced concept based approach would be

designed to incorporate some special program modules that would generate and display both medical documents and medical search queries results in both medical expert and layman's forms. The remainder of this paper is organized as follows. Section 2 describes the study related works. Section 3 discusses the proposed enhanced concept based approach; Section 4 contains the performance analysis, experimental setup, datasets and performance metrics used, as well as result and discussions. And Section 5 concludes the work and gives direction for future work.

II. REVIEWED WORK

The inability of the existing information retrieval systems to address the information needs of different categories of end users (medical professionals and nonprofessionals) affects the system performance [2, 4, 7]. The none generation and display of search results in both medical experts and layman's forms, makes it difficult for the existing approaches to address the questions usually asked by laymen patients after reading through their medical search results [14, 17, 19]. In addition, the present information retrieval systems pays more attention on a specific group of people with expert health knowledge [18, 20, 22].The presentation issues faced by laymen patients in exploring information extracted from their medical discharge documents and Clinical reports online should be given adequate attention and addressed.

The use of web as a source for health related information has now become a widespread common practice performed by information seekers [1, 4, 7]. Search engines are widely used by information seekers as a means to access health information available on the internet [8, 9, 10]. However, while addressing users diverse information needs, such as searching information on a specific disease, preceding researches targeted only a specific group of users with expert health knowledge [17,18, 20]. Health-related content has become one of the most searched information on the Internet and also an important area of research in information retrieval.

In the recent time, health information retrieval algorithms have been widely designed to deal with the challenges of searching health related information from a diverse medical information sources, such as the general web, social media and hospital records [4, 10, 14]. Previous researches on medical information retrieval also disclosed how desperate patients are in apprehending the content of their medical discharge reports and clinical reports [14, 17, 18]. [19, 20, 21] in their work revealed how the existing information retrieval algorithms failed to address presentation issues, as they are unable to address the various information needs of a wide range of users, e.g. laymen patients and their relatives, researchers, clinicians, practitioners, etc. Robust algorithms should be developed to address the information needs of different categories of end users.

The essence of adopting the technique of labeling most specific concept terms during query expansion was to address similarities and presentation issues encountered by laymen patients in exploring information extracted from the web [3, 4]. In the medical domain, querying of the Internet for useful information has become increasingly important, owing to the huge amount of information available [17, 18]. Author in [21]

showed how the benefits of biomedical data retrievals could be rigorously restricted when users lack the know-how in creating effective search queries. In addition, the outcome of the study also shows that, the search engine's effectiveness was valued steadily higher when the query approval features are turned *on* vs. *off* [23]. Information retrieval systems should limit their search terms to most specific concept terms during query expansion in order to address similarities and presentation issues.

The work by [15] identified how suitable health advices, accessibility to relevant and reliable medical information could significantly reduce the mortality rates of epidemic diseases such as cancer diseases in nations. Additionally, the outcome of the study also shows that, the inability of national health authorities to readily make available health information on various web channels has significantly increased the mortality rates of epidemic diseases in several nations [17, 22]. The author in [23] identified the importance of patient and family centered method in pediatrics health care in making strategic decisions on their children's health care, as well as seeking the greatest attention of the teenager. The outcome of the study provided an improved understanding of the motives of using web channels in sharing information about a child's cancer knowledge, as well as spreading theoretical outlines for constructing additional knowledge in this regards.

III. PROPOSED ENHANCED CONCEPT BASED APPROACH

We proposed an enhanced concept based approach for user centered health information retrieval that would address the information needs of both medical expert and laymen patients. As it would generate and present medical discharge documents and medical search queries results in both medical expert and layman's forms. The proposed enhanced approach which consist of lines no. 1-14, comprised of four functions namely: Function for creating medical discharge reports in expert form, function for creating medical search queries in layman's form, function for generating medical discharge reports in layman's form designed and incorporated in it.

The novelty of our enhanced concept based approach is that, it generates and displays medical documents / medical search queries results in two separate forms (i.e. medical expert and layman's forms). Hence, it better addressed the questions usually asked by laymen patients and their relatives after reading through their medical discharge reports or medical search queries results online. This problem is usually caused due to the inability of the existing approaches to focus on addressing the information needs of different categories of end-users rather than focusing on medical experts' information needs.

In the enhanced concept based approach, two functions each are dedicated to generating and displaying search results in both medical expert and layman's forms. The functions for creating medical discharge reports and medical search queries results in expert forms are dedicated to displaying medical search results in experts form. While the functions for generating and displaying medical discharge reports and medical search queries results in layman's forms are dedicated to generating and displaying medical discharge reports and

medical search queries results in layman's form. With the design and incorporation of these four functions into our proposed enhanced concept based approach, the information needs of both medical experts and laymen patients are better addressed.

The working strategy of the enhanced concept based approach is as shown in Fig. 2, where enter "information to be searched" Z refers to the number of terms or concepts resulting from the main launched query. With the concept terms as T and extension or expansion terms as X. Search query is represented in the proposed enhanced approach as SearchQR and the z th concept term as TTz. The x th expansion terms are represented as Extension$_x$. MST means "most specific terms". The last expansion term in an expansion query is known as xth term and the symbol # represents space character (i.e., 0x20). The double quotation marks indicate that the string in it must appear consecutively as shown in Fig. 1:

A. Function for Creating Medical Search Query in Medical Expert Form - [Function A]

For the specific purpose of creating Medical Search Queries in Expert form, Fig. 2 was designed and incorporated into the proposed enhanced approach. Fig. 2 that comprised of lines no. 15 – 27, where line 15 declares the function module titled: "getTermsRecord (Topic)" in the main program module. And lines 16 to 17 checks and selects the terms record with such topic name from the pool of stored terms records. Lines 18 to 21 searches and extracts the topics, subtopics and their contents from the database of the selected dataset and displays it to the patient/their relatives. Lines 22 to 23 give the status of the launched search. Line 24 is the program module that creates medical search query in medical expert form. When no terms record are found for a particular search, lines 25 to 27 displays the message "No record found for the specified query". And finally, line 28 terminates the function module as shown in Fig. 2.

Fig. 3 is the sample interface view output of the medical search query in an expert form, for the enhanced concept based approach generated by Fig. 2. In order to perform a medical query search, the end-user would be required to click on the query button on the homepage and then enter the search term(s) in the search engine window. He then finally click on the medical search query button and automatically, the expert meanings for such medical term(s) would be searched for and displayed as shown in Fig. 3:

B. Function for Generating Medical Search Query in Layman's Form – [Function C].

For the specific purpose of generating Medical Search Queries in Layman's form, Fig. 4 which comprised of lines no. 29 - 40 was designed. Where line 29 declares the function module titled: "substitute-vocabulary (vocabulary)" in the main program module. And lines 30 to 31 fetches all the medical vocabularies found in the displayed search query results. Lines 32 to 33 assign Layman's translations for all the medical vocabularies found in the medical search query result. Lines 34 to 35 give the status of the launched search query. Line 36 is the program module that generates medical search

query in layman's form. When no replacement term is found for a particular search query, lines 37 to 39 displays the message "No replacement found for the specified grammar". Else replacement is made automatically for the particular vocabulary by lime 40. And finally, line 41 terminates the function module as shown in Fig. 4.

```
1    Enter "Information to be Searched"
2    For all Terms[MST] z ∈ [1,Z] do
3    Assign SearchQR = Launched SearchQR
4    SearchQR = "SearchQR[MST]" + "TTz[MST]"
5    For all the Extension Concepts[MST]
     y ∈ [1,Y], do
6    New
     SearchQR = SearchQR[MST] # "Extensionx[MST]"
7    Terminate;
8    [Module for creating Function A]
9    [Module for Creating Function B]
10   Present Output 1 (In Medical Professional
     form)
11   [Module for Generating Function C]
12   [Module for Generating Function D]
13   Present Output 2 (In Non-Medical
     Professional form)
14   End.
```

Fig 1.   The Proposed Enhanced Concept based Approach.

```
15   Declare: getTermsRecord(topic)
16   Check = Select[MST] " from Stored Patient
     Terms"
17   Where topic = @topic " Among stored topics "
18   If topic = @topic then
19   getTopic = "Search for Topics from stored
     Patient-Terms"
20   getContent= "Extract Content from stored
     Topics Contents"
21   getSubtopic ="Search for Subtopic from
     stored Patient-Terms"
22   Position = " 1 record found"
23   excode = 1
24   Create medical search query in medical
     expert's form
25   Else
26   Position = " No record found for the
     specified query"
27   Excode = 0
28   End.
```

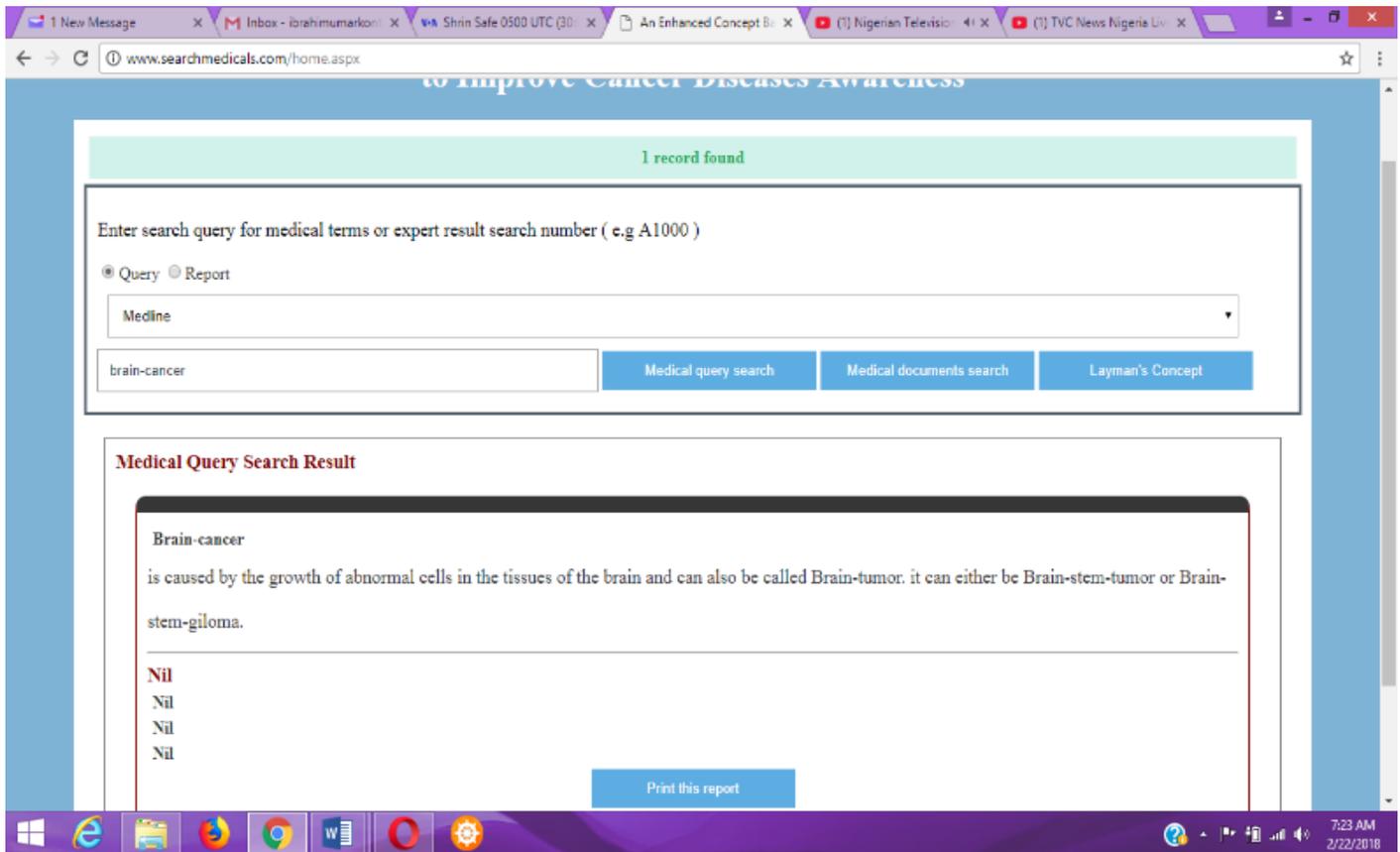Fig 2.   The Function for Creating Medical Search Query in Medical Expert's Form.

Fig 3.   The Interface View of Medical Query Search in Expert Form.

```
29   Declare: substitute-vocabulary(vocabulary)

30   Procure = Pick " among deposited layman's
     terms"

31   Where vocabulary = @vocabulary "from kept
     vocabulary"

32   If vocabulary = @vocabulary then

33    SubstituteVocabulary = Look "from
     deposited Layman's Terms"

34   Position = " 1 substitution dictated"

35   Excode = 1

36   Generate medical search query in layman's
     form

37   Else

38   Position = "No substitution dictated for
     the stated  vocabulary"

39   Excode = 0

40   Replace vocabulary + = (" + substitution
     +")

41   End.
```

Fig 4.   The Function for Generating Medical Search Query in Layman's
Form.

Fig. 5 is the sample interface view output of the medical search query in layman's form for the enhanced concept based approach. The layman's concept for all the medical term(s) found in the search result could be derived upon clicking on the layman concept command button on the homepage. And automatically, it generates the layman's translation for all the medical terms found in the medical search query results as shown in Fig. 5.

C. Function for Creating Medical Discharge Document in Expert Form – [Function B].

For the specific purpose of creating Medical Discharge Documents in Expert form, Fig. 6 which comprised of lines no. 42 - 53 was designed and incorporated into the enhanced concept based approach. In Fig. 6, line 42 declares the function module titled: "getPatientRecord (record-no)" in the main program module. While lines 43 to 44 checks and selects the patient record with such patient-no from the pool of stored patient records. Lines 45 to 46 extracts the patient record with such record-no from the database of the selected dataset and displays it to the patient/their relatives. Lines 47 to 48 give the status of the launched search. Line 49 is the module for creating medical discharge documents in expert's forms, designed and incorporated in the proposed enhanced concept approach. When no patient record is found for a particular search, lines 50 to 52 display the message "No record found for the specified query". And finally, line 53 terminates the function module as shown in Fig. 6.
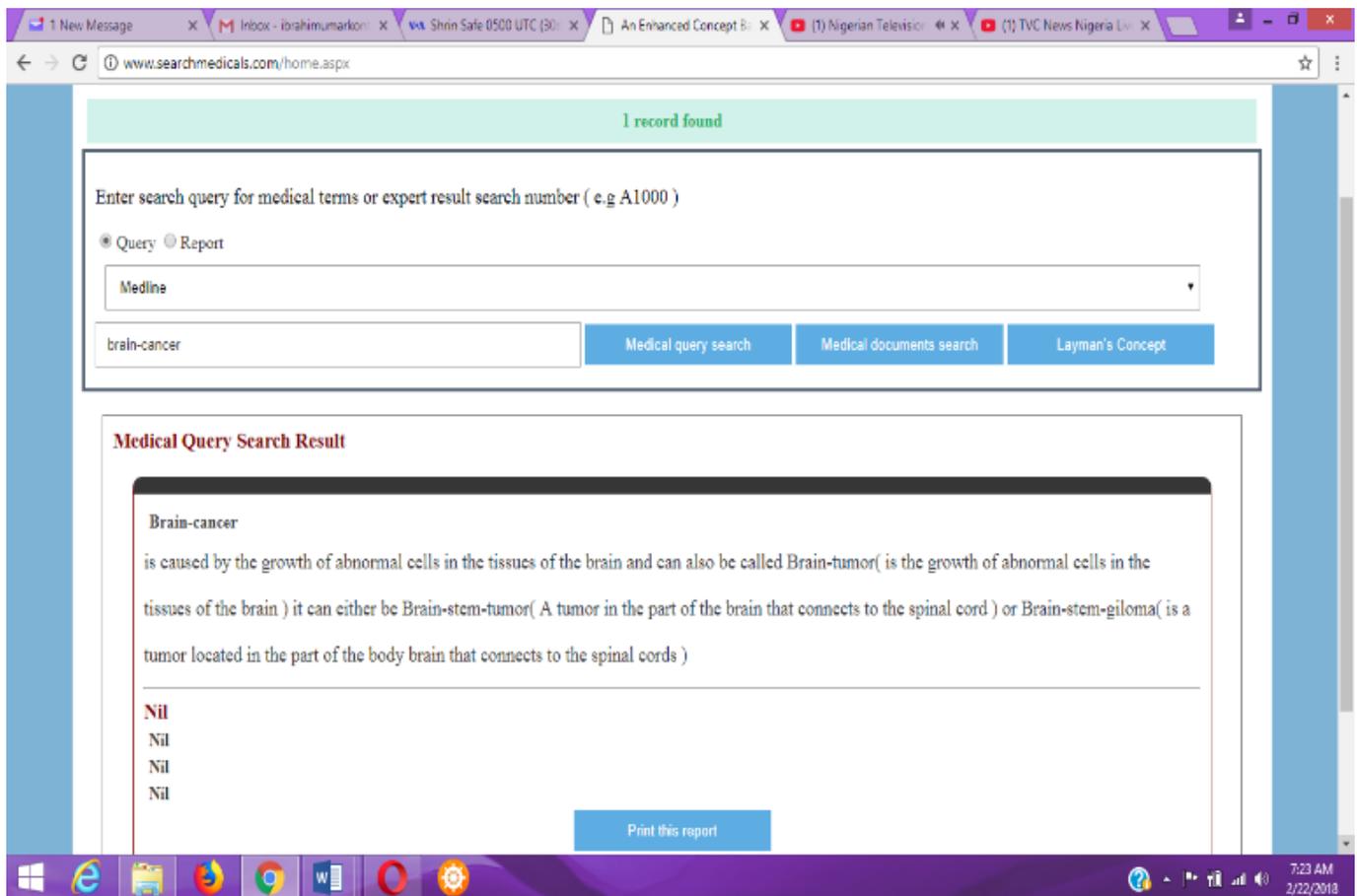
Fig 5.   The Interface View of Medical Search Query in Layman's Form.

```
42   Declare getPatientRecord (record-no)

43   Check = Select "from patient Record"

44   Where record-no = "@recno"

45   If record-no = @recno then

46   Record-no = " Extract from system
     database"

47   Position = " 1 record found"

48   excode = 1

49   Create Medical Discharge Document in
     Expert Form

50   Else

51   Position = "No record found for the
     specified query"

52 Excode = 0

53 End.
```

Fig 6.   The Function for Creating Medical Document in Expert Form.

Fig. 7 is the sample interface view output of the medical discharge document in expert form for the enhanced concept based approach. In order to perform a medical document search, the end-user would be required to click on the report button on the homepage. Then enters the patient record number in the search engine window and clicks on the medical document search button. Automatically, the medical discharge document with such medical record number would be retrieved and displayed as shown in Fig. 7.

D. Function for Creating Medical Discharge Document in Layman's Form – [Function D]

For the specific purpose of generating the layman's translation of all the medical terms found on the Medical Discharge Report in Fig. 7, Fig. 8 was designed and incorporated into the enhanced concept based approach. Fig. 8 is the function that generates Medical Discharge Reports in Layman's Forms and it comprise of lines no. 54 to 66. Where line 54 declares the function module titled: "fetchTranslationTerms (vocabulary)" in the main program module. And lines 55 to 56 checks for all the medical vocabulary contained in the patients displayed record. Lines 57 to 59 assign Layman's translations for all the medical vocabularies found on the medical discharge report. Lines 60 to 61 give the status of the launched search. Line 62 is the module for generating medical discharge documents in layman's forms, designed and incorporated into the proposed enhanced concept approach. When no patient record is found for a particular search, lines 63 to 65 display the message "No record found for the specified query". And finally, line 66 terminate the function module as shown in Fig. 8.
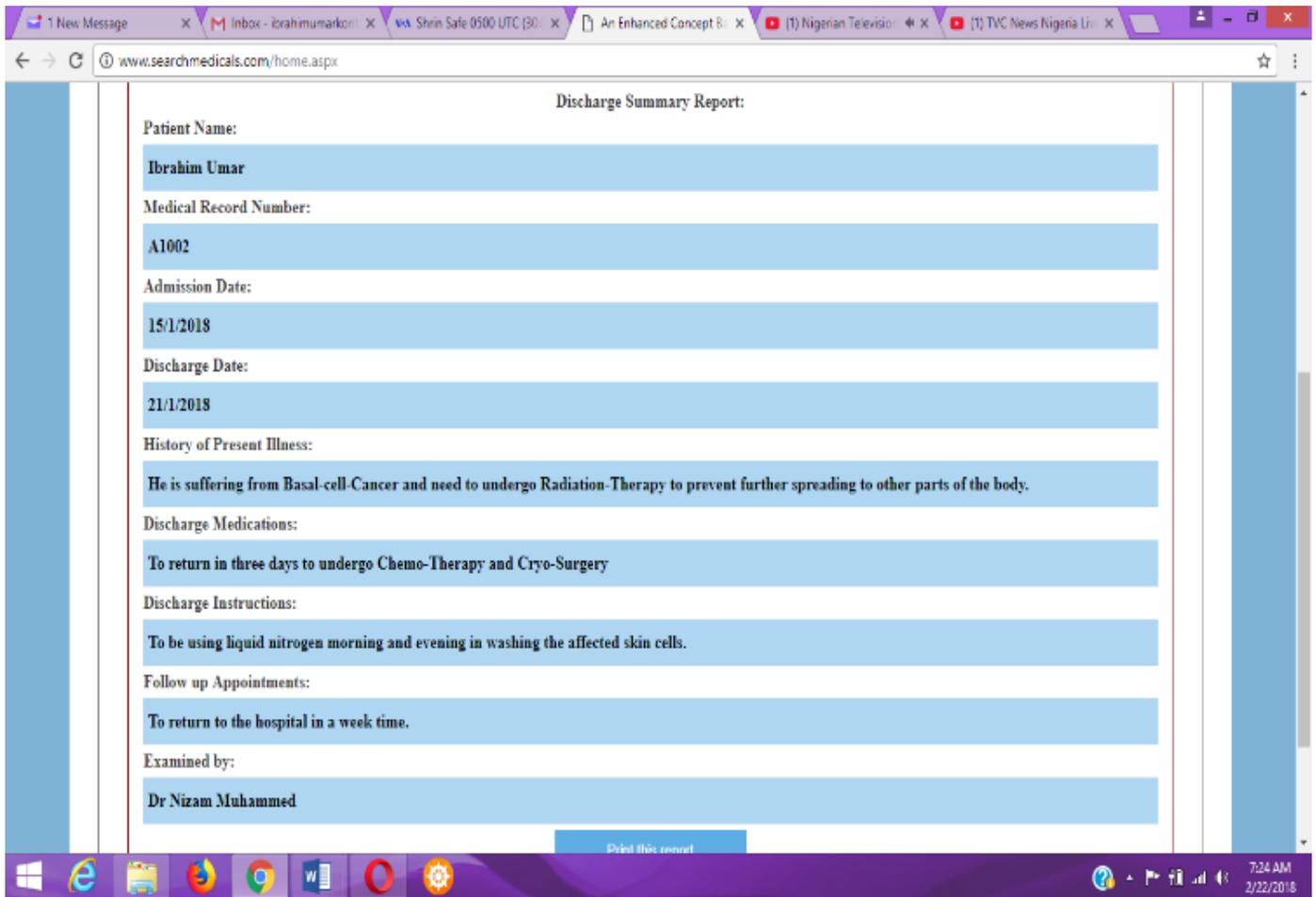
Fig 7.   The Interface View of Medical Discharge Document in Expert Form.

```
54 Declare: fetchTranslationTerms (vocabulary)

55 Examine = Pick "Among client Record"

56 Where vocabulary = @vocabulary

57 If vocabulary = @vocabulary then

58 vocabulary = "Retrieve from Launched Query"

59  Vmeaning  =  "Retrieve  from  the  dataset
dictionary"

60 Position = "1 patient-record found"

61 excode = 1

62 [Produce medical discharge reports in laymen
     Form]

63 Else

64 Position = "No Patient-record found"

65 Excode = 0

66 End.
```

Fig 8.   The Function for Generating Medical Discharge Reports in
Layman's Form.

Fig. 9 is the interface view of the sample output of a medical discharge report in layman's form for the enhanced concept based approach. The layman's concept for all the medical term(s) found in the medical discharge report would be provided upon clicking on the layman concept command button on the homepage. And automatically, it would be generated shown in Fig. 9.

The implementation of the enhanced concept based approach was done in two stages, where in the first stage, the two functions responsible for creating medical discharge reports and search queries in expert forms were designed and implemented. And in the second stage, the functions for generating and displaying medical discharge documents and search queries in layman's form were also created and implemented in the enhanced concept based approach. Hence, these four functions ensure that medical search results are generated and displayed in two separate forms (i.e. medical expert and layman's forms). By so doing, it has better addressed the questions that laymen patients and their relatives do come up with after reading through their medical discharge reports and medical search queries results online. These are usually caused as a result of the existing approaches targeting on specific group of people with expert health knowledge.
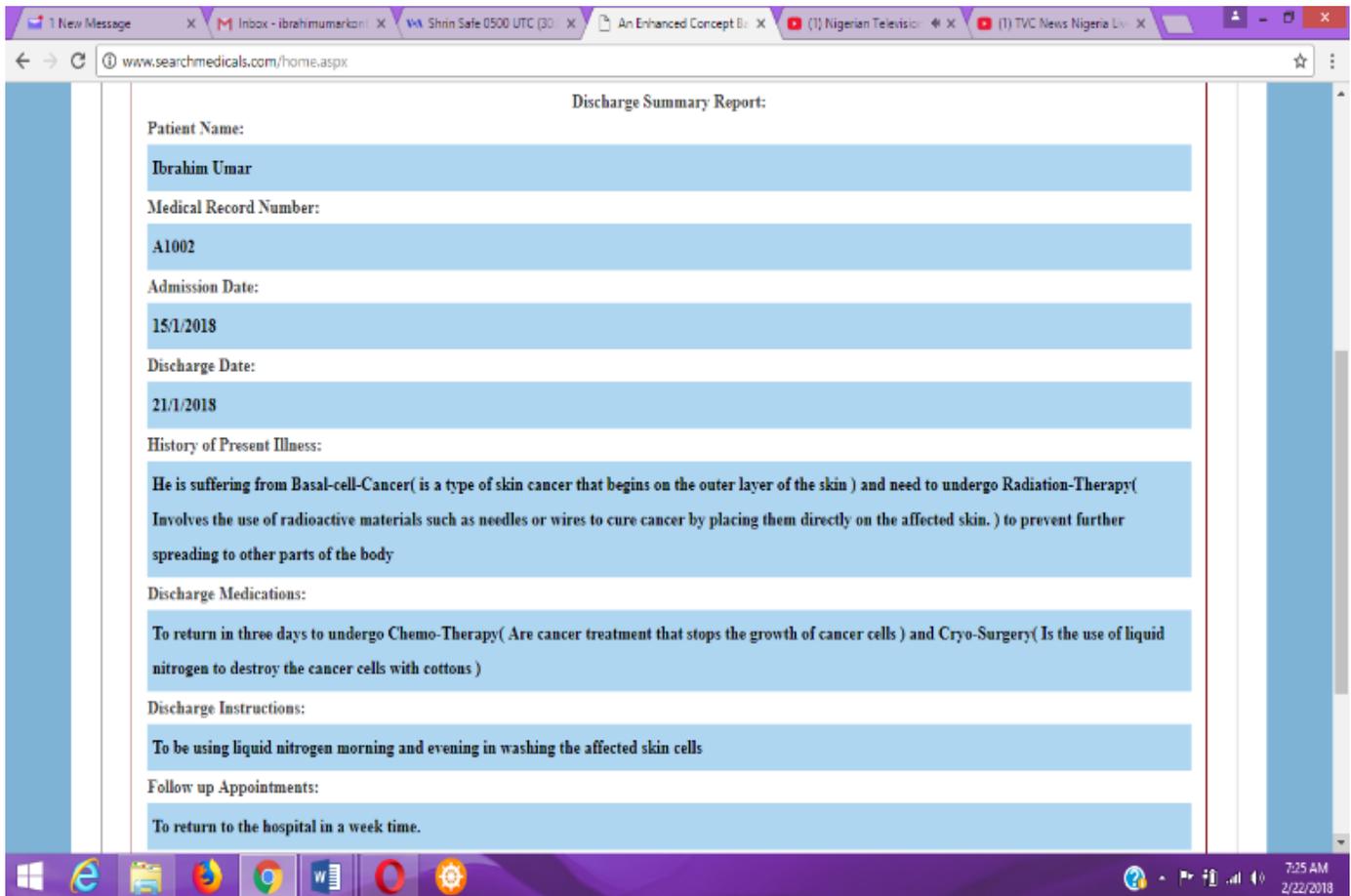
Fig 9.   The Interface View of Medical Discharge Reports in Layman's Form.

IV. PERFORMANCE ANALYSIS

In order to evaluate and compare the performance of our proposed enhanced concept based approach, with the existing approaches namely: concept based approach (CBA), query likelihood model (QLM) and latent semantic model (LSI), the text semantic similarity scoring function was used. Also presented in this section was the experiment setups, the dataset used in the research study, the detailed description of the text semantic similarity scoring function used in evaluating the performance of our proposed enhanced concept based approach with the existing approaches. Finally, is the discussion of the experimental results.

A. Experimental Setup

The experimental setup of the research study was conducted using the Hypertext Markup Language 5.0 (HTML 5.0), Cascading Style Sheet (CSS), the object oriented programming language C#.net, JavaScript and the windows 7 operating system. Also used are Intel (R) Core i7 processor, 3.40GHz and 4GB RAM memory. The creation of the website structure was made using HTML 5.0. The activation and creation of the various functions functionalities for the web application was done using the object oriented programming language called C#.net. The beautification of the web application through styles was made using the Cascading Style

Sheet (CSS). Finally, effects were added to the proposed web application structures using the JavaScript.

B. Dataset

The dataset used in this research work were downloaded from a pool of free large online medical subjects, issues, grammers and patient's database delivered by Diagnosia 7, Medical Subject Heading (MeSH), Metamap and Khresmoi Project 6 datasets. The coverage of these datasets were enormous as they covered a huge variety of patient's information, medical subjects and vocabularies. All medical vocabularies and subjects contained in these datasets were sourced from numerous internet foundations that include Clinical.gov, Genetics Home Reference and Health on the Net organization certified websites [1], [14].

C. Performance Metrics

The measurement of the performance of our enhanced concept based approach in relation to semantic similarity scores was carried out using the Text Semantic Scoring Function. Same was also used by previous researchers in their related work in measuring the similarity scores between two text segments. The illustration of how the function works and how the performance analysis was conducted is as shown below:

D. Text Semantic Similarity Scoring Function

For any two given input text segments T_1 and T_2, the semantic similarity scores of the concepts or terms enclosed within the two text segments T_1 and T_2 could be measured using the Text Semantic Scoring Function. It determines, scores and measures the semantic similarities between two text segments T_1 and T_2 in any given information retrieval system as shown in the semantic similarity scoring function below:

$$\text{Sim}(T_1, T_2) = \frac{1}{2} C \left(\frac{\sum_{w \in \{T1\}} (maxSim(w1T1) * idf(w))}{\sum_{w \in \{T1\}} idf(w)} + \right.$$

$$\left. \frac{\sum_{w \in \{T2\}} (maxSim(w2T2) * idf(w))}{\sum_{w \in \{T2\}} idf(w)} \right) \quad [15], [16] \quad (1)$$

The values of the semantic similarity scores ranges between 0 and 1, with score 0 representing the two text segments as the same or identical. And score 1 represents that the two text segments are the same or identical. The specificity of a word or concept is measured using the IDF weighting. When the specificity weighting is high, it signifies that the words are very precise and specific to that particular record. And when the specifity weighting is low, that signifies that the word is general, and common among several records. The IDF values for any given word Wi can be gotten using:

$$\text{IDF}(Wi) = \text{Log}(W1, W2) \quad [16], [15] \quad (2)$$

V. Result and Discussions

The evaluation and comparison of the performance of the enhanced concept based approach with the existing approaches: concept based approach (CBA), query likelihood model (QLM) and latent semantic indexing (LSI) was carried out using the performance metrics: text semantic similarity scoring function and IDF function. The enhanced concept based approach unlike the existing approaches focused more on addressing the information needs of both the medical experts and laymen patients. By generating and displaying results in two separate forms (i.e. in both medical expert and laymen forms). This has made the enhanced concept based approach to better address the questions that are usually asked by laymen patients after reading through their medical discharge reports and medical search queries results online.

Unlike the existing concept based approach and other existing approaches that pays more attention on a specific group of people with expert health knowledge, the enhanced concept based approach addresses the information needs of different categories of end users. As it provides layman's translation for every medical term(s) found on medical discharge report and medical search queries results online. The enhanced concept based approach was designed in two stages, where in the first stage, the two functions responsible for creating medical discharge documents and search queries in expert forms were designed and implemented. And in the second stage, the functions for generating and displaying medical discharge documents and search queries in layman's forms were also created and implemented in the enhanced concept based approach. Hence, the layman's translation for all the medical terms found on any medical discharge

documents and medical search queries results online were provided as shown from the results obtained in Tables 1 to 4.

Analysis using Text Semantic Similarity Scoring Function

For any two given input segments T_1 & T_2, the semantic similarities scores for the medical concepts and terms (W1, W2) contained in the two text segments T1 & T2 can be calculated using the text semantic scoring function in Equation 1 and 2. The resemblance among the words (W1, W2) in the two text segments T_1 & T_2 could be calculated using the similarity scoring function. Hence, from Equation 1 and 2, the similarity scores that exist between the two text segments could be determined by merging the text similarity scores and their corresponding specificity as tabulated in Tables 1 to 4. The text similarity scores indicates the interpretation of the connection between the words (W1, W2) in the two text segments using a threshold value of 0.50 as used in preceding research works. The semantic similarities scores values ranges between 0 and 1, with score 0 signifying that the words (W1, W2) in the two text segments as not the same or identical. And score 1, signifying that the words (W1, W2) in the two text segments as the same and identical.

For the specific purpose of validating the performance of our proposed enhanced concept based approach with the existing approaches (CBA, QLM, LSI), we made use of Medical Subject Heading (MeSH), Metamap, Diagnosia 7 and Khresmoi Project 6 datasets. We designed the enhanced concept based approach in such a way that, it implements the four functions incorporated in it. And this enables it to generate and display search results in both medical expert and layman's forms. With this development, two separate outputs are generated and displayed for every medical discharge report and medical search queries with high maxSim and idf values as shown from the result obtained in Tables 1 to 4.

TABLE I. Similarity Scores of the Enhanced Concept based Approach (ECBA) and Existing Approaches CBA, QLM, LSI in Respect to maxSim and IDF Values using Medical Subject Heading (Mesh) Dataset

Approaches	maxSim	idf
ECBA	1.0	3.82
CBA	0.5	1.46
QLM	0.5	1,42
LSI	0.5	1,42

TABLE II. Similarity Scores of the Enhanced Concept based Approach (ECBA) and Existing Approaches CBA, QLM, LSI in Respect to maxSim and IDF Values using METAMAP Dataset

Approaches	maxSim	idf
ECBA	1.0	3.86
CBA	0.5	1.44
QLM	0.5	1.43
LSI	0.5	1.40

We used Equation 1 and 2 to simulate the semantic similarity scores between the two displayed results T1 (Displayed Results 1) and T2 (Displayed Results 2), using 500 data extracted from Medical Subject Heading (MeSH) dataset. We evaluated and compared the words (W1, W2) contained in the two texts segments T1 (Displayed Results 1) and T2 (Displayed Results 2) for the specific purpose of obtaining their maxSim and idf values. The outcome of the comparison shows that the medical terms/concepts (W1, W2) contained in the two texts segments (T1-Dispalyed Results 1 and T2-Displayed Results 2) by the enhanced concept based approach are exactly thesame and identical for all the displayed results. This has made it to have high similarities scores (maxSim Values) of 1 in all the displayed results and high idf weighting values compared to the existing approaches. The existing approaches (CBA, QLM, LSI) displayed only one form of result (Medical Expert Form) with some terms not specific to the search term(s), making them to score low values of maxSim and idf weighting as shown in Table 1.

We used Equation 1 and 2 to simulate the semantic similarity scores between the words (W1, W2) contained in the two displayed results T1 (Displayed Results 1) and T2 (Displayed Results 2). And using 500 data extracted from the dictionaries of Metamap dataset. The outcome of the comparison and evaluation shows that the medical terms/concepts (W1, W2) contained in the two texts segments (T1-Dispalyed Results 1 and T2- Displayed Results 2) by the enhanced concept based approach are exactly thesame and identical for all the displayed results. This has made it to have high similarities scores (maxSim Values) of 1 in all the displayed results and high idf weighting values compared to the existing approaches as shown in Table 2.

Equations 1 and 2 were used to simulate the semantic similarity scores between the words (W1, W2) contained in the two displayed results T1 (Displayed Results 1) and T2

(Displayed Results 2). And using 500 data extracted from the dictionaries of Diagnosia 7 dataset. We evaluated and compared the words (W1, W2) contained in the two texts segments T1 (Displayed Results 1) and T2 (Displayed Results 2) for the specific purpose of obtaining their maxSim and idf values. The outcome of the comparison shows that the medical terms/concepts (W1, W2) contained in the two texts segments (T1-Dispalyed Results 1 and T2- Displayed Results 2) by the enhanced concept based approach are exactly thesame and identical for all the displayed results as shown in Table 3.

TABLE III. Similarity Scores of the Enhanced Concept based Approach (ECBA) and Existing Approaches CBA, QLM, LSI in Respect to maxSim and IDF Values using Diagnosia 7 Dataset

Approaches	maxSim	idf
ECBA	1.0	3.84
CBA	0.5	1.43
QLM	0.5	1.44
LSI	0.5	1.47

TABLE IV. Similarity Scores of the Enhanced Concept based Approach (ECBA) and Existing Approaches CBA, QLM, LSI in Respect to maxSim and IDF Values using Khresmoi Project 6 Dataset

Approaches	maxSim	idf
ECBA	1.0	3.84
CBA	0.5	1.45
QLM	0.5	1.42
LSI	0.5	1.43

Similarity Scores of the Proposed Enhanced Concept Based Approach ECBA) and the existing approaches (CBA, QLM, LSI) in respect to maxSim and idf Values in Tables 1 to 4
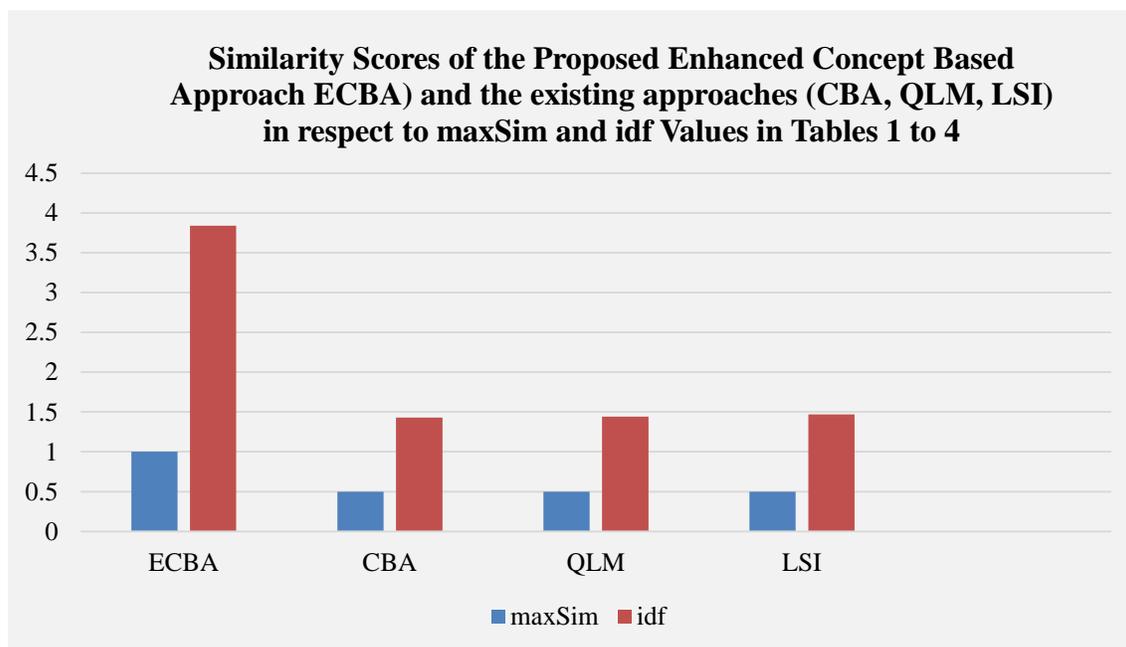
Fig 10. Graphical Representation of the Similarity Scores of the Enhanced Concept Based Approach (ECBA) and existing approaches CBA, QLM, LSI in respect to maxSim and idf Values in Tables 1 to 4.

Equations 1 and 2 were used to simulate the semantic similarity scores between the words (W1, W2) contained in the two displayed results T1 (Displayed Results 1) and T2 (Displayed Results 2). And using 500 data extracted from the dictionaries of Khresmoi Project 6 dataset. We evaluated and compared the words (W1, W2) contained in the two texts segments T1 (Displayed Results 1) and T2 (Displayed Results 2) for the specific purpose of obtaining their maxSim and idf values. The outcome of the comparison shows that the medical terms/concepts (W1, W2) contained in the two texts segments (T1-Displayed Results 1 and T2- Displayed Results 2) by the enhanced concept based approach are exactly thesame and identical for all the displayed results. This has made it to have high similarities scores (maxSim Values) of 1 in all the displayed results and high idf weighting values compared to the existing approaches. The existing approaches (CBA, QLM, LSI) displayed only one form of result (Medical Expert Form) with some terms not specific to the search term(s), making them to score low values of maxSim and idf weighting as shown in Table 4.

Fig. 10 is the graphical representation of the average similarity scores for the Enhanced Concept Based Approach **(ECBA)** and existing approaches CBA, QLM, LSI in respect to maxSim and idf Values from Tables 1 to 4.

With the design and full incorporation of these four special functions modules namely: function for creating medical discharge documents in expert form, function for creating medical search query in expert form, function for creating medical discharge document in layman's form and finally, function for creating medical search query in layman's form into the enhanced concept based approach, the questions that laymen patients and their relatives do ask after reading through their medical discharge documents and medical search queries results online were better addressed. As layman's translations for every medical term(s) found in the search results are provided. Hence, the challenge of the inability of the existing approaches (CBA, QLM, LSI) to generate search results in both medical expert and layman's forms has been better addressed.

The outcome of the experimental results obtained in Tables 1 to 4 and Fig. 10 using the Text Similarity Scoring and IDF Functions shows that, the Enhanced Concept Based Approach manage to score a similarity score of 1.0 (100%) in respect to maxSim values in all the runs for all the four datasets. And idf weighting values of between 3.82 – 3.86 in all the runs for all the four datasets used while the existing approaches scored a maxSim scores of 0.5 (50%) in all the runs and idf weighting values of between 1.40 – 1.47 for all the four datasets. This low result scored by the existing approaches was due to their inability to generate and display search results in both medical experts and layman's forms. These generated outputs have clearly shown that, the Enhanced Concept Based Approach better addressed the presentations issues, as it is able to generate and display medical discharge documents and medical search queries results in both medical expert and layman's forms online.

The inability of the existing approaches (CBA, QLM, LSI) to incorporate functions modules that would focus on generating and displaying medical discharge documents and medical search queries online in both medical expert and layman's forms had affected their performances in respect to addressing the information needs of different categories of end-users. This is contained in the results displayed in Tables 1 to 4 and Fig. 10. Additionally, that had affected the rate of medical information communication between the medical experts and their respective laymen patients. Also, the failure of the three existing approaches to include addressing the information needs of both medical expert and layman's patients right from the algorithm design stage, had significantly affected their performances in that regard.

The scientific reasons behind these better results obtained by the enhanced concept based approach compared to the three existing approaches (CBA, QLM, LSI), could be better explained in respect to the outcome in Tables 1 to 4 and Fig. 10. The enhanced concept based approach concentrated more on designing and implementing the four special functions modules incorporated into it, which generates and displays medical discharge documents and medical search queries results in both medical expert and layman's forms. Also, the enhanced concept based approach did not limit its scope to addressing the information needs of medical experts alone, as it addresses the information needs of both medical experts and laymen patients/ their relatives. This has made the enhanced concept based approach to obtain better results that better addressed the information needs of both medical experts and laymen patients compared to the three existing approaches used in this experiment.

VI. Conclusion

The aftermath of the experimental results obtained in Tables 1 to 4 and Fig. 10 using Text Similarity Scoring and IDF Functions shows that, the Proposed Enhanced Concept Based Approach manage to score a similarity scores of 1.0 (100%) in respect to maxSim values in all the runs in all the four dataset. And idf weighting values of between 3.82 – 3.86 in all the runs in all the four datasets while the existing approaches (CBA, QLM, LSI) scored the maxSim scores of 0.5 (50%) in all the runs in all the four datasets and idf weighting values of between 1.42 – 1.47. These poor results obtained by the three existing approaches were due to their inability to generate and display search results in both medical experts and layman's forms. This also shows that the Proposed Enhanced Concept Based Approach is the best approach suited to be applied in addressing presentations issues encountered by the existing approaches. Because, the existing approaches are unable to generate and display medical discharge documents and medical search queries results online in both medical expert and layman's forms.

The novelty of our proposed enhanced concept based approach is that, it generates and displays medical search results in two separate forms (i.e. in both medical expert and Layman's users' forms). By so doing, it better addressed the questions that are usually asked by laymen patients and relatives after reading through their medical discharge reports and medical search queries results online. This problem is usually caused due to the inability of the existing approaches to concentrate more on addressing the information needs of

different categories of end-users (both medical professional and nonprofessional) rather than concentrating on those users with medical expert knowledge e.g. clinicians, medical experts, nurses, medical researchers, etc.

Additionally, the obtained results in Tables 1 to 4 and Fig. 10 also shows that, the proposed enhanced concept based approach is the best suited approach to be used in tackling the presentations issues faced by the existing approaches (CBA, QLM, LSI). As they failed to address the information needs of both medical expert and laymen patients, due to their inabilities to generate and display medical discharge documents and medical search queries results in two separate forms. However, this singular act of generating search results in both medical and layman's forms by the proposed enhanced concept based approach has helped to better answered the questions that are usually asked by laymen patients regarding the meaning of the content of their medical discharge documents and medical search queries results online. Future work on this research study should address presentation issues in respect to retrieved audios, videos and images.

REFERENCES

[1] Z. Xiaoshi, X. Yunqing, X. Zhongda, N. Sen, H. Qinan and H. Yaohai, "Concept-based medical document retrieval: THCIB at CLEF eHealth lab 2013 task 4." In Proceedings of the ShARe/CLEF eHealth Evaluation Lab, 2013.

[2] H. Suominen, L. Kelly, L. Goeuriot, L. Hanlen, A. Névéol, C. Grouin and G. Zuccon, "Overview of the CLEF eHealth evaluation lab 2015," In International Conference of the Cross-Language Evaluation Forum for Languages (pp. 429- 443). Springer, Cham. 2015.

[3] L. Kelly, L. Goeuriot, H. Suominen, A. Névéol, J. Palotti and G. Zuccon, "Overview of the CLEF eHealth evaluation lab 2016.

[4] N. Ksentini, M. Tmar and F. Gargouri, "Miracl at CLEF 2014: eHealth information retrieval task," In: Proceedings of the ShARe/CLEF eHealth Evaluation Lab. 2014.

[5] R. White and E. Horvitz, "Cyb erchondria: Studies of the escalation of medical concerns in web search," In Technical report, Microsoft Research. 2015.

[6] E.M. Voorhees and R.M. Tong, "Overview of the TREC 2011 medical records track," In: Proceedings of TREC, NIST. 2011.

[7] L. Goeuriot, L. Kelly, W. Li, J. Palotti, P. Pecina, G. Zuccon, A. Hanbury, G. Jones and H. Mueller, "ShARe/CLEF eHealth Evaluation Lab 2014, Task 3: User centered health information retrieval," In Proceedings of CLEF 2014 (2014).

[8] L. Kelly, L. Goeuriot, H. Suominen, A. Névéol, J. Palotti and G. Zuccon, "Overview of the CLEF eHealth evaluation lab 2016.

[9] C. J Kalpathy, H. Muller, S. Bedrick, I. Eggel, A. G.S de Herrera and T. Tsikrika, "The CLEF 2013 medical image retrieval and classi_cation tasks," In: Working Notes of CLEF 2013, Cross Language Evaluation Forum. 2013.

[10] H. Thakkar, G. Iyer, K. Shah, and P. Majumder, "Team IRLabDAIICT at ShARe/CLEF eHealth 2014 Task 3: User-centered Information Retrieval system for Clinical Documents," In: Proceedings of the ShARe/CLEF eHealth Evaluation Lab, 2014.

[11] S. Fox, "Health topics: 80% of internet users lo ok for health information online," In Technical Report, Pew Research Center, 2011.

[12] C. Bellegarda, T.J. Leone, D.H. Hickam and W.R. Hersh, "Concept-based medical document retrieval: THCIB at CLEF eHealth lab 2013 task 3," In: Proceedings of the ShARe/CLEF eHealth Evaluation Lab. 2013.

[13] W. Shen, J.Y. Nie, X. Liu and X. Liui, "An investigation of the effectiveness of concept-based approach in medical information retrieval GRIUM @ CLEF2014eHealthTask 3.," In: Proceedings of the ShARe/CLEF eHealth Evaluation Lab. 2014.

[14] H. Thakkar, G. Iyer, and P. Majumder, "A comparative study of approaches in user-centered health information retrieval". ArXiv preprint arXiv: 1505.01606. 2015.

[15] S. Xie and Y. Liu, "Using Corpus and Knowledge Based Similarity Measure in maximum marginal relevance for meeting summarization," In Acoustics, speech and signal proceeding, 2008.

[16] M. Rada, C. Courtney and S. Carlo, "Corpus-based and knowledge-based Measures of Text Semantic Similarity," In American Association for Artificial Intelligence (www.aaai.org). All right reserved @ 2006.

[17] I.U Kontagora and I.R.A Hamid, "Comparative Studies of Information Retrieval Approaches in User-Centered Health Information System. In: Ghazali R., Deris M., Nawi N., Abawajy J. (eds) Recent Advances on Soft Computing and Data Mining. SCDM 2018. Advances in Intelligent Systems and Computing, vol 700. Pp 171 – 180. Springer, Cham. 2018.

[18] L. Goeuriot, L. Kelly, W. Li, J. Palotti, P. Pecina, G. Zuccon and H. Mueller, "Share/clef eHealth evaluation lab 2014, task 3: User-centered health information retrieval," In *Proceedings of CLEF 2014*.

[19] C. Boyer, M. Gschwandtner, A. Hanbury, M. Kritz, N. Pletneva, M. Samwald and A. Vargas, "Use case definition including concrete data requirements (D8.2)," In Public deliverable, Deliverable of the Khresmoi EU project. 2012.

[20] J. Leveling, L. Goeuriot, L. Kelly and G.J.F. Jones, "DCU@TRECMed 2014: Using ad-hoc baselines for domain-specific retrieval," In: Proceedings of TREC 2014, NIST, 2014.

[21] D.A. Hanauer, D. T. Wu, L. Yang, Q. Mei, K.B. Murkowski-Steffy, V.V. Vydiswaran, and K. Zheng, "Development and empirical user-centered evaluation of semantically-based query recommendation for an electronic health record search engine," In Journal *of biomedical informatics*, *67*, 1-10. 2017.

[22] D. Novillo-Ortiz, T. Hernández-Pérez and F. Saigí-Rubió, "Availability of information in Public Health on the Internet: An analysis of national health authorities in the Spanish-speaking Latin American and Caribbean countries," In International *journal of medical informatics*, *100*, 46-55. 2017.

[23] S. Rehman, K. Lyons, R. McEwen and K. Sellen, "Motives for sharing illness experiences on Twitter: conversations of parents with children diagnosed with cancer," In Information*, Communication & Society*, *21*(4), 578-593. 2018.