Information Retrieval Approach based on Recursive Query Shifting

Mawloud Mosbah
University 20 Août 1955 of Skikda, Algeria
Faculty of Sciences
Department of Informatics
mos_nasa@hotmail.fr

Abstract: In this paper, we introduce an information retrieval approach based on a Recursive Query Shifting. The idea is coming from the observation that precision value, in Precision/Recall curves, begins very high before downgrading and so no matter what the considered parameters: feature, matching measure, threshold, etc. In other words, the first results are commonly better than the other results coming later. Considering then the first returned result, recursively, as a query seems to contribute very well for improving the system accuracy. The idea is adopted firstly in the case of content based image retrieval and generalized for the case of documentary retrieval. The simple specificity is that the first returned image is considered as the new query, in the case of image retrieval, while the new query is extracted from the first returned document, in the case of text retrieval, for the reason that there is no documentary retrieval system with an entire document as a query. The proposed approach falls in the purview of mechanisms for pseudo relevance feedback with long-term learning helping to improve the interrogation protocol. It consists to reformulate the original submitted query recursively as an attempt for being closer to the user requirement as well as to the other collection elements. The approach is materialized into two algorithms namely *QRM1* and *QRM2*. The experiments conducted on returned *Google Scholar* results as a text collection and *COREL-1K* images benchmark yield very promising results.

Key-words-- Recursive Query Shifting, Information Retrieval, Content based-image retrieval, Text Retrieval.

I. INTRODUCTION

Information retrieval is qualified as a complex and not well defined problem. Indeed, unlike database systems, information retrieval one addresses semi-structured and non-structured data. Moreover, no matter what the information nature being retrieved (text, image or video), the users are still globally not satisfied. This problem, stilling open within scientific research area, requires then more improvement. Improving information retrieval system accuracy consists of enhancing its components that of indexing stage, interrogation protocol, and matching process. The purpose of the indexing stage is to index the collection elements through encoding them into a significant and compact form known as a features index. Owing to its overload in terms of run time, this stage is established off line. How to encode the elements of the collection being retrieved is designated through the adopted model. There is a large spectrum of models: Boolean model relying on ensemble theory, vector model based on algebra and probabilistic model relying on probabilistic theory. Contributions, done in the indexing stage, focusing in characteristics or features extraction, attempt to propose an effective signature. Many efforts have been put for proposing an effective indexing stage, either within text [1], image [2], [3] or video [4]. For matching process, aiming to compare between the query representation and the established index, a large spectrum of matching measures (distances, quasi-distance, similarities and divergences) have been considered for attaining a high retrieval quality [5], [6], [7], [8]. For interrogation protocol, interesting in query shifting/reformulation, many schemes have been considered such as query expansion/reformulation [9], [10], [11], (pseudo)relevance feedback scheme [12], [13], [14].

In this paper, we focus on interrogation protocol through introducing a new approach based on recursive query shifting. The considered approach takes advantage of three previously proposed schemes namely: query reformulation [10], (pseudo) relevance feedback [12], and collective contribution of retrieved information materialized in incremental-KNN [13] and multi-queries approach [15], [16], [17], [18], [19].

The remained of the paper is arranged as follows: in section 2, we give some background through talking about interaction, pseudo relevance feedback, and some re-ranking algorithms. Section 3 presents our introduced method based on a recursive query shifting. We conclude the paper with a conclusion.

II. BACKGROUND

a) Interaction and Interrogation Protocol

Interaction module includes interrogation protocol and results' visualization through system graphical user interface. Its protocol delimits the interrogation language in which the query has to be formulated. Interaction module is very important compound owing to its direct relation with the user query and in consequence with the user requirement being satisfied. Efforts, done in this direction, focus on the query through attempting to reformulate and expand it [10] in a way that leads to good results. Another direction is to re-rank results [20] before the final visualization to the user through soft computing [21] and data mining [22], [23] tools. The second basic task of the interaction module is to display results. The exploration of the elements collection being retrieved is another retrieval way ensured by some information retrieval graphical user interface such as Yahoo [24]. Actual trend for overcoming the gap between the user query and his/her need is the appealing of multi-queries scheme [15], [16], [17], [18], [19].

b) Relevance Feedback

Relevance is an interesting notion within information retrieval field. This complex concept [25] is defined as the correspondence, in terms of information, between document and query. There are primarily two kinds of relevance: system relevance and user relevance. System relevance [26] is referred to the score attributed by the system as a relevance evaluation for the document content with respect to that of the submitted query. Such relevance is determinist. User relevance [16] consists of the relevance judgement of the user for the results answered by the system. Such relevance is non-determinist and subjective. One way to improve information retrieval system quality is to reduce the angle between the both notions namely: system relevance and user relevance through relevance feedback scheme. Relevance feedback is appeared firstly within text retrieval field and generalized later to other information natures such as image [13] and video [27]. Relevance feedback mechanism, coming from text information retrieval [12], consists of deeming by the user some documents among those returned by the system. Relevance feedback mechanism, which proved its performance in the case of text with different models [28], [29], seems to be more suitable for the case of image retrieval where it has been applied later [30], [31], [32]. Indeed, contrary to text and video, a whole fraction look is enough to inject relevance feedback information in the case of image retrieval.

Unfortunately, it is not always possible that the user injects his/her relevance feedback information. Pseudo relevance feedback, aiming to look for some correlations and regularities between the first returned results, has been adopted then for replacing the user judgement [33], [20], [34].

Another essential point for the (pseudo) relevance feedback is the collective judgement of the results. Indeed, the encouraged results obtained by the algorithms, proposed previously in our works such as Majority Voting Re-ranking Algorithm (MVRA) [9], [35], [36], [37] and Incremental-KNN [13], pertain primarily to the collective judgement rather than for the relevance feedback scheme itself.

c) Re-ranking based on (Pseudo) Relevance Feedback: Query Reformulation and Query Expansion, Clustering and Classification, Parameterization, and Majority Voting Re-ranking Algorithm

In [37] and [38], we have classified re-ranking algorithms into two classes: algorithms based on pseudo relevance feedback and algorithms based on relevance feedback. The two categories are classified into three other classes: query reformulation or expansion, clustering or classification, and parameterization.

The mental user requirement, in terms of information, is expressed relatively employing a user query allowing, more or less, getting access for the relevant documents [39]. A basic way to improve the results of an information retrieval system is then adopting query reformulation/expansion mechanism [10]. For reformulation techniques, there are commonly two large classes: global techniques working in the entire collection and local techniques working only in some returned results [40].

Among re-ranking algorithms based on reformulation, we can quote: *Query Point Movement* [41], *Standard Rocchio Formula* [42], *Bayesian Relevance Feedback* [43], and *Adaptive Query Shifting* [44].

Many algorithms of clustering and classification have been used as a post-processing aiming to improve information retrieval results through using respectively pseudo and relevance feedback. We can cite: K-means [13], [45], Hierarchical Agglomerative Clustering Algorithm (HACM) [13], [45], Self Organizing Map [46] for clustering and K-nearest Neighbours (KNN) [13], [47], [48], [49] Incremental-KNN [13], Neural Networks [50], Support Vector Machine (SVM) [51], and decision tree [52] for classification.

Parameterization algorithms allow to increase the weight of relevant features such as in *Feature Weighting* algorithm [53] or to optimize the parameters in the case of many matching measures such as in '*Optimizing of the Parameters of the Similarity Metric*' method [13] for enabling to better rank the images deeded as relevant by the user.

Other improvement class of algorithms is materialized by Majority Voting Re-ranking Algorithm (MVRA), we have introduced in [36], which re-ranks the first returned images answered by an image retrieval system through considering sub-set of images as candidates and the others as electors. On the basis of these two sets (candidates and electors), a vote operation is organised for generating a new better rank. This algorithm is validated again in [35], [37] through its comparison with other pseudo relevance feedback algorithms of the literature. The effectiveness of the algorithm is validated later into documentary retrieval through a combination with query reformulation [9]. Even the supervised version of the algorithm (SMVRA: Supervised Majority Voting Re-ranking Algorithm) is considered in [20].

III. OUR PROPOSED APPROACH BASED ON RECURSIVE QUERY SHIFTING

Our first proposed method, known as *Query Reformulation Method 1* (*QRM1*), is inspired from some essential personal review of information retrieval problems and systems. This review is summarized into the following points:

Amelioration is mandatory: Comparing to database systems attacking structured data, information retrieval is qualified as a not well defined problem for the reason that the information being addressed is semi-structured or unstructured one. Such systems, built for addressing not well defined problems, need surely amelioration, improvement and enhancement through post-processing steps.

Interrogation protocol has to receive more attention: Although the fact that representation is on the core of any information retrieval system (which makes indexing stage more valuable), the interrogation protocol and query representation have to receive more attention for the reason that the system is built for satisfying the user requirement expressed by the query. Moreover, submitting the query is the first operation into the system which influences much more the entire system quality with respect to the other operations coming later.

Pseudo relevance feedback is better as post-processing: it is a mechanism helping to enhance the results of information retrieval system before any results visualization and without any user assistance. The pseudo relevance feedback is which we need then for integrating enhancement within the information retrieval system.

Collective decision: Some algorithms we have previously introduced into image retrieval such as *MVRA* [9], [35], [36] and *incremental-KNN* [13] sustain the idea that the collective decision for designating the relevant elements is better than the single decision based commonly on the submitted query. The idea is not really novel. Indeed, multi-queries approach [15], [16], [17], [18], [19] and meta-engine approaches [54] adopted the same idea.

First results are good: Our previous experimentations reveal that the first results of any query are relevant. The performance downgrades increasingly. The idea of the algorithm is to keep only the first result of any query that we have not yet keep and considering it as the new query. Its first result will be kept and considered as the new query again and so on.

The first result, returned by the retrieval system as an answer for a query, is technically (according to the considered setting: indexing method and the adopted matching measure) more close to the elements of the asked database than the query itself. So, it seems to be more effective to keep only this first result and consider it as a new query for asking again the database and so on. According to that, the system needs much iteration for asking the data base. Each new query requires iteration for getting only one result. However, we hope to encounter this demerit by the remarkable upgrading of effectiveness.

For the second algorithm, known as *Query Reformulation Method 2* (*QRM2*), we proceed to fuse *QRM1* with the idea of *Query Point Movement* technique [41] through considering the centroid of the initial query and the first returned result as the new query.

According to the pre-cited analysis, we introduce then a intuitive information retrieval approach based on a Recursive Query Shifting. The approach is materialized into two algorithms: *QRM1* and *QRM2*. The architecture of QRM1 and the associated pseudo code are given respectively in Fig.1 and Fig.2. The architecture of QRM2 and the associated pseudo code are given respectively in Fig.3 and Fig.4.

The difference between *QRM1* and *QRM2* is into computing the query to be considered per iteration (step 5). *QRM1* presumes each new retrieved result as the new query to be considered while *QRM2* considers the centroid of the all retrieved results including the new retrieved result as the new query.

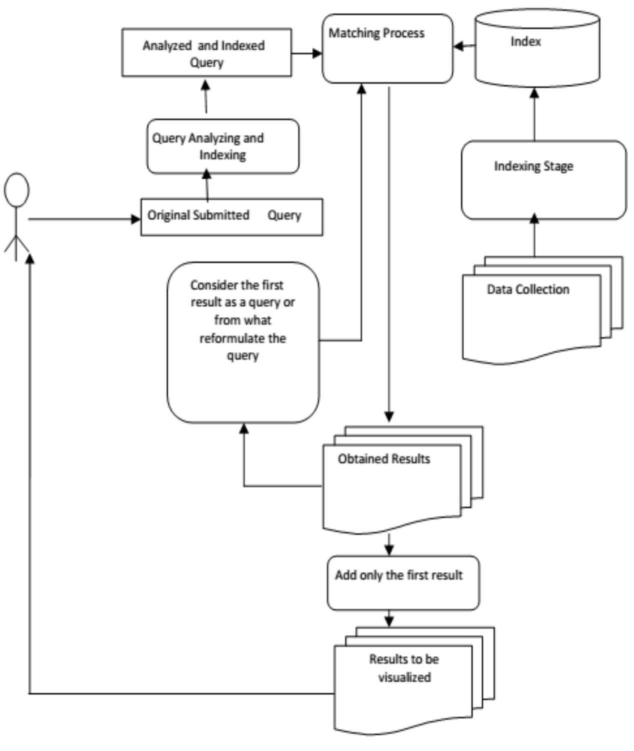

Fig 1. The General Architecture of the System taking into account QRM1.

Initialization

Results_to_be_visualized←Φ

Images_collection={image$_1$, image$_2$, .., image$_N$}

COUNT←0

Step 1:The user submits the *query* according to the interrogation language: an image query in the case of image retrieval, and some key-words in the case of text retrieval.

Step 2: The query image is encoded according to the considered indexing signature for the case of image retrieval, and text query is sent to Google Scholar in the case of text retrieval.

Step 3: Retrieve the closest image through comparing the indexed_*query* to the indexed_*images* according to the considered matching measure in the case of image retrieval, and receive documents results from Google Scholar in the case of text retrieval.

Step 4*: Results_to_be_visualized*←*Results_to_be_visualized* **plus** *closest element* that does not exist into results_to_be_visualized

COUNT←COUNT+1

Step 5: new_Query←Indexed_closest_image recently added to results_to_be_visualized //in the case of images retrieval

new_Query←old_Query+ one word from the first returned document recently added to results_to_be_visualized //in the case of text retrieval

Step 6: if COUNT<N then go to ***Step 2***

Step 7: END.

Fig 2. The pseudo code of QRM1.

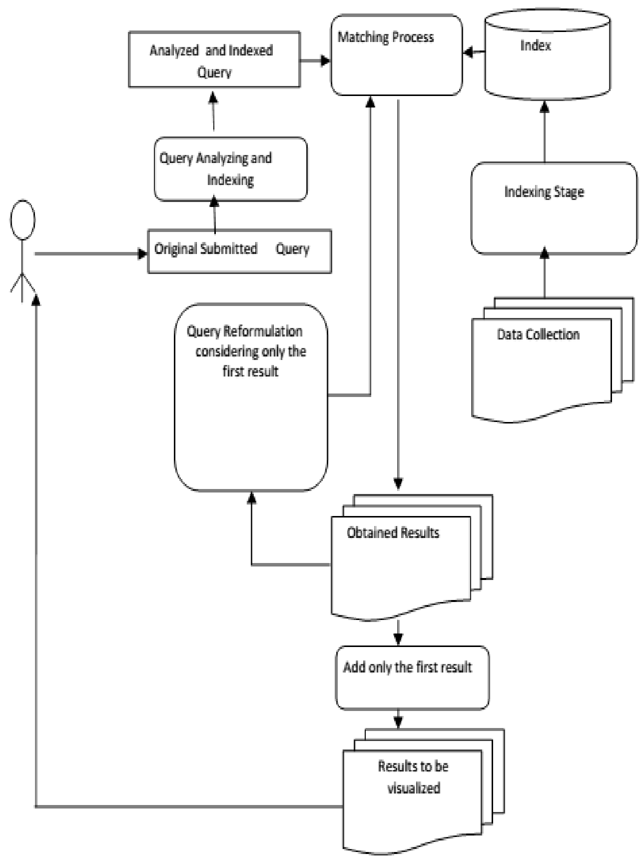

Fig 3. The general Architecture of the System taking into account QRM2.

Initialization

Results_to_be_visualized🡐Φ

Images_collection={$Image_1$, $Image_2$, .., $Image_N$}

COUNT🡐0

Step 1: The user submits the *query* according to the interrogation language: an image query in the case of image retrieval, and some key-words in the case of text retrieval..

Step 2: The query image is encoded according to the considered indexing signature for the case of image retrieval, and text query is sent to Google Scholar in the case of text retrieval.

Step 3: Retrieve the closest image through comparing the indexed_*query* to the indexed_*images* according to the considered matching measure in the case of image retrieval, and receive documents results from Google Scholar in the case of text retrieval.

Step 4*: Results_to_be_visualized*🡐*Results_to_be_visualized* **plus** *closest element*

COUNT🡐COUNT+1

Step5:new_Query🡐centroid of Indexed_Results_to_be_visualized //in the case of image

$new_{Query} = old_{query} + one\ word\ having\ the\ highest\ score\ from$ $\left(\frac{\sum DVR}{NDV}\right)$ //in the case of documents

Step 6: if COUNT<N then go to ***Step 2***

Step 7: END.

Fig 4. The pseudo code of QRM2.

Where: DVR are documents to be visualized representation, and NDV is the number of documents to be visualized.

The assumption of *QRM1* relies on the idea that the new retrieved result is closer to the rest elements of the collection being retrieved than the original query. It seems then to be a good idea to retrieve elements with another closer element. For *QRM2*, relying on the centoid of the retrieved elements is inspired from *query point movement* technique. The difference is that here the mechanism is an automatic ranking operation done on all collection elements rather than semi-automatic re-ranking one done just on subset of the collection.

The idea behind the proposed approach is coming from the fact that a system such as information retrieval system, addressing not well defined problem where there is no a formal description, requires always an improvement step. This enhancement step is performed usually as a post-processing task utilizing relevance or pseudo relevance feedback. In our proposed approach, we proceed to shift the original query after any retrieved result not visualized to the user. In other words, we do not wait for the proceeding of the initial search for launching shifting query; we have to launch it from the beginning. Another characteristic of the proposed approach is that there is a collective participation into the query shifting which not the case is before.

It is so worthy to note that the effectiveness of *QRM1* and *QRM2* are not really related to the global performance of the considered setting but it is related to the performance with respect to the first returned result apart from the other results coming later that will be ignored.

IV. MATERIAL AND EXPERIMENTS

1) Materials

For testing the effectiveness of QRM1 and QRM2, we conduct the following experiments into *COREL-1K* images collection [55] and into some documents returned by Google Scholar [56] as results for some submitted queries. For the considered matching measure, we have chosen for Ruzicka similarity for images as advised within [5] and cosinus similarity [57] for the case of text. As signature in the case of visual information (images), we have utilized the three first colour moments. For the evaluation metrics, we have utilized precision [58], recall [58], utility value [59] and precision and recall adapted for the case of the web [60].

Fig 5. Some images representing th 10 classes of *COREL-1K* images database.

2) Experimental Results on COREL-1K

Besides the proposed algorithms namely: QRM1 and QRM2, algorithms considered for the case of images retrieval are: Incremental-KNN [13], (pseudo) Query Point Movement [41], (pseudo) Standard Rocchio Formula [42], (pseudo) Adaptive Shifting Query [43], Supervised Majority Voting Re-ranking Algorithm [9], [20],[35], [37], KNN [13], [47], [48], [49], Feature Weighting [53], optimization of the parameters of the similarity metrics [13], (pseudo) Bayesian Relevance Feedback [43], k-means [13], [45], and HACM [13], [45].

Fig.6 and Fig.7 show respectively some CBIR results and the indexing process using color moments.

Fig 6 Snapshot screen of CBIR results.

As depicted in Fig.7, QRM2 outperforms clearly the considered algorithms of the literature based both on pseudo and relevance feedback, while QRM1 is better than all the considered algorithms except Incremental-KNN.

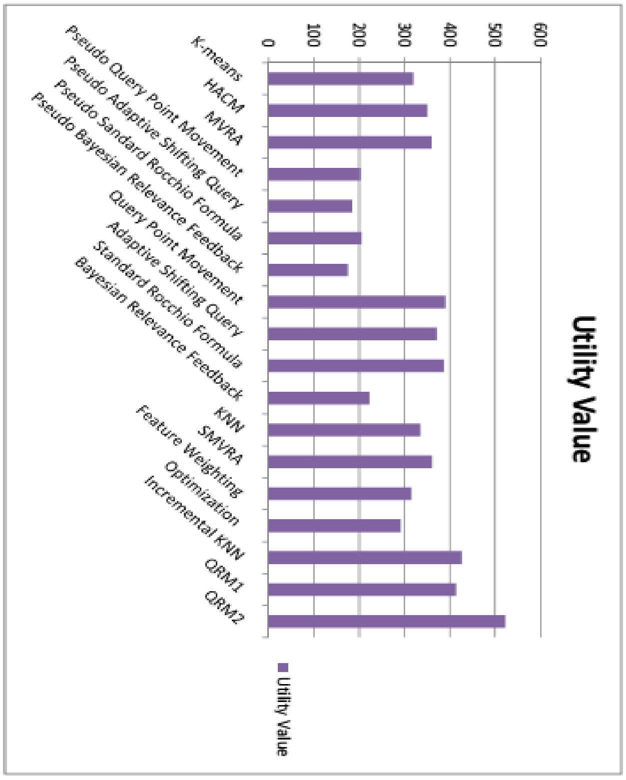

Fig 7. A Basic CBIR system vs. Shifting Query Method (QRM1 and QRM2) vs. some techniques of literature in terms of Utility Concept.

3) Experimental on Google Scholar

Table1 and Table2 show the submitted queries and their associated shifted queries (added terms) generated respectively by QRM1 and QRM2 for the case of document retrieval. Table3 shows the accuracy of Google Scholar, QRM1, and QRM2 over the considered queries.

4) Discussion

The method proposed in this paper is sustained theoretically for the reason that it is usually difficult, for the common user, to submit the appropriate query which really reflects his/her information need and leading to good performances. Practically, the proposed method, combining shifting query, pseudo relevance feedback, collective participating of the retrieved results (especially for QRM2 algorithm), improves results whether the initial results are not so bad and also not very good. Indeed, it is so difficult to improve good results, and it is not possible to extract relevant queries, leading to better performances, from results of low quality. The proposed method is also comparable to some shifting query techniques of the literature. Using two different experimental collections, that of a local database and some results answered by Google Scholar engine, allow testing the proposed method into two directions: image and text.

a) For the case of CBIR

The proposed algorithms namely: QRM1 and QRM2 are very promising especially QRM2 which outperforms all experimented literature techniques. For QRM1, it is relatively comparable with the other re-ranking techniques of literature.

b) For the case of Google Scholar

QRM1 algorithm yields an improvement comparing with the accuracy of Google Scholar and QRM1 algorithm which is effective only for some queries. The results may be justified as follows:

- Google Scholar is an effective research engine difficult to be improved.
- The precision considered for the case of the web reflects much more homogeneity rather than effectiveness.
- We meet some technical problems related to the used materials. These problems may be summarized as follows: (a) some PDF documents returned by Google Scholar are formatted as images from where text is difficult to be extracted; (b) some returned documents are not free. The both document cases have been overtook which influences surely accuracy computation and comparison.

V. CONCLUSION

In this paper, we introduced two algorithms based on query shifting considering collection decision, reformulation, and implicitly multi-queries scheme. The first algorithm, called QRM1, considers iteratively the first returned result as a new query. QRM2 combines the idea of QRM1 with Query point Movement technique. QRM1 and QRM2 have been tested into image and text through using respectively COREL-1K images collection and results returned by Google Scholar. QRM2 outperforms many shifting query techniques of literature in the case of image retrieval while QRM1 improves relatively the accuracy of Google Scholar for the case of text retrieval.

REFERENCES

[1] Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schutze. *Introduction to information retrieval*. Cambridge: Cambridge University press, 2008, pp. 1-519.

[2] Meryem Salmi and Bachir Boucheham. "Content Based Image Retrieval Based on Cell Color Coherence Vector (Cell-CCV)". In Proc. *ISKO-Maghreb: 4th International Symposium on Concepts and Tools for Knowledge Management (ISKO-Maghreb)*, 2014, pp. 1-5.

[3] Bougueroua, Salah, and Bachir Boucheham. "GLI-Color: Gradual Locality Integration of Color Features for image retrieval." In Proc. *International Conference on Intelligent Information Processing, Security and Advanced Communication*, 2015, pp. 26.

[4] Smoliar, Stephen W., and HongJiang Zhang. "Content based video indexing and retrieval." *IEEE multimedia* Vol. 1 No. 2, 1994, pp. 62-72.

[5] Mosbah. M. & Boucheham. B. "Matching Measures in the Context of CBIR. A Comparative Study in terms of Effectiveness and Efficiency". In Proc. *World Conference on Information Systems and Technology*, 2017, pp. 245-258.

[6] Mosbah. M., & Boucheham. B. Distance Selection based on Relevance Feedback in the Context of CBIR using the SFS meta-heuristic with One Round. *Egyptian Informatics Journal*, Vol. 18, No. 1, 2017, pp. 1-9.

[7] Tao, Tao, and ChengXiang Zhai. "An exploration of proximity measures in information retrieval." In Proc. *30th annual international ACM SIGIR conference on Research and development in information retrieval*, 2007, pp. 295-302.

[8] Shechtman, Eli, and Michal Irani. "Matching local self-similarities across images and videos." In Proc. *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.

[9] Mosbah, M. Improving the Results of Google Scholar Engine through Automatic Query Expansion Mechanism and Pseudo Re-ranking using MVRA. *Journal of Information and Organizational Sciences*, Vol. 42, No. 2, 2018, pp. 219-229.

[10] Carpineto, C. And Romano, G. A survey of automatic query expansion in information retrieval. *ACM Comput. Surveys*. Vol. 44, No. 1, 2012, pp. 1-50.

[11] Zhai, Yun, Jingen Liu, and Mubarak Shah. "Automatic query expansion for news video retrieval". In Proc. *IEEE International Conference on Multimedia and Expo*, 2006, pp. 965-968.
[12] Rocchio, J. J. *Relevance Feedback in information retrieval*, The SMART retrieval system: experiments in automatic document processing, 1971, pp. 313-323.
[13] Mosbah, M., & Boucheham, B. Relevance Feedback within CBIR. *International Journal of Computer, Information Science and Engineering*, Vol. 8, No. 4, 2014, pp. 19-23.
[14] Wang, Roy, Milind R. Naphade, and Thomas S. Huang. "Video retrieval and relevance feedback in the context of a post-integration model." In Proc. *IEEE Fourth Workshop on Multimedia Signal Processing*, 2001, pp. 33-38.
[15] Belkin, N. J., Kantor, P., Fox, E. A., & Shaw, J. A. (1995). Combining the evidence of multiple query representations for information retrieval. *Information Processing & Management*, Vol. 31, No. 3,1995, pp. 431-448.
[16] Rieh, S. Y. (2006). Analysis of multiple query reformulation on the web: The interactive information retrieval context. *Information Processing & Management*, Vol. 42, No. 3, 2006, pp. 751-768.
[17] Belkin, N. J., Cool, C., Croft, W. B., & Callah, J. P. (1993, July). The effect multiple query representations on information retrieval system performance. In Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 339-346). ACM.
[18] Akahin, H. C., & Guran, M. N. Content-based microscopic image retrieval system for multi-image queries. *IEEE transactions on information technology in biomedicine*, Vol. 16, No. 4, 2012, pp. 758-769.
[19] Jin, X., & French, J. C. Improving image retrieval effectiveness via multiple queries. *Multimedia tools and applications*, Vol. 26, No. 2, 2005, 221-245.
[20] Mosbah, M. & Boucheham, B. Re-ranking in the Context of CBIR: A Comparative Study. In Proc. *World Conference on Information Systems and Technology*, 2017, pp. 297-307
[21] Crestani, Fabio , and Gabriella Pasi, eds. Soft Computing in Information Retrieval: Techniques and Applications. 2000, Physica, pp. 1-393.
[22] Srinivasan, P., Ruiz, M. E., Kraft, D. H. & Chen, J. Vocabulary mining for information retrieval: rough sets and fuzzy sets. *Information Processing & Management*, Vol. 37, No. 1, 2001, pp. 15-38.
[23] Singh, B., & Singh, H. K. (2010, December). Web data mining research: a survey. In Proc. *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 2010, pp. 1-10.
[24] https://fr.yahoo.com/ visited in January 14th 2018.
[25] Mizzaro., S. Relevance: The whole history. *Journal of the American Society for Information Science*, Vol. 48, No. 9, 1997, pp. 810-832.
[26] Cleverdon, C. Evaluation tests of information retrieval systems. *Journal of Documentation*, Vol. 26, No. 1, 1970, pp. 55-67.
[27] Yan, R., Hauptmann, A. G., & Jin, R. Negative pseudo-relevance feedback in content-based video retrieval. In Proc. *The eleventh ACM international conference on Multimedia*, 2003, pp. 343-346.
[28] Lavrenko, V., & Croft, W. B. Relevance based language models. In Proc. *The 24th annual international ACM SIGIR conference on Research and development in information retrieval*, 2001, pp. 120-127.
[29] Lv, Y., & Zhai, C. A comparative study of methods for estimating query language models with pseudo feedback. In Proc. *The 18th ACM Conference on Information and Knowledge management*, 2009, pp. 1895-189). ACM.
[30] Zhou, X. S., & Huang, T. S. Relevance feedback in image retrieval: A comprehensive review. *Multimedia systems*, Vol. 8, No. 6, 2003, pp. 536-544.
[31] Rui, Y., Huang, T. S., Ortega, M., & Mehrota, S. Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Transactions on circuits and systems for video technology*, Vol. 8, No. 5, 1998, pp. 644-655.
[32] Zhou, X. S., & Huang, T. S. Relevance feedback in content-based image retrieval: some recent advances. *Information Sciences*, Vol. 148, No. 1, 2002, pp.129-137.
[33] Lv, Yuanhua, and ChengXiangZhai. "Positional relevance model for pseudo-relevance feedback." Proc. *33rd international ACM SIGIR conference on Research and development in information retrieval*, 2010, pp. 579-586.
[34] Yan, Rong, Alexander G. Hauptmann, and Rong Jin. "Negative pseudo-relevance feedback in content-based video retrieval." Proc. *The eleventh ACM international conference on Multimedia*, 2003, pp. 343-346.
[35] Mosbah, M., & Boucheham, B. (2015). Majority Voting Re-ranking Algorithm for Content Based-Image Retrieval. In Proc. *Metadata and Semantics Research* 2015, pp. 121-131.
[36] Mosbah, M., & Boucheham, B. New algorithm for re-ranking based on the vote operation in the context of CBIR. In Proc. *World Symposium on Computer Applications & Research (WSCAR)*, 2014, pp. 1-7.
[37] Mosbah, M., & Boucheham, B. Pseudo Relevance Feedback based on Majority Voting Re-ranking Mechanism. International *Journal of Web Science*, Vol. 3, No. 1, 2017, pp. 58-81.
[38] Mosbah, M. Mesures de Distance dans le Contexte de la Recherche d'Images par le Contenu (CBIR). Dissertation Thesis, University 20 Août 1955 of Skikda, Algeria, 2017.
[39] Salton, G., & McGill, M. L. Introduction to Information Retrieval. McGraw-Hill Int. Book Co, 1984.
[40] Xu, J., & Croft, W. B. Query expansion using local and global document analysis. In Proc. *The 19th annual international ACM SIGIR conference on Research and development in information retrieval*, 1996, pp. 4-11.
[41] Porkaew, K., & Chakrabarti, K. Query Refinement for Multimedia Similarity Retrieval in MARS. In Proc. *The seventh ACM International Conference on Multimedia*, 1999, pp. 235-238.
[42] Rui, Y., Huang, T. S., Mehrotra, S. Content-based image retrieval with relevance feedback in MARS. In Proc. *International Conference on Image Processing*, 1997, pp. 815-818.
[43] Giacinto, G., & Roli, F. Bayesian Relevance Feedback for Content-based Image Retrieval. *Pattern Recognition*, Vol. 37, No. 7, 2004, pp. 1499-1508.
[44] Giacinto, G., Roli, F., & Fumera, G. Adaptive Query Shifting for content-based image retrieval. In Proc. *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, 2001, pp. 337-346
[45] V.S.V.S Murthy, E. Vamsidhar, J.N.V.R Swarup Kumar and P. Sankara Rao. Content based Image Retrieval using Hierarchical and K-means Clustering Techniques. *International Journal of Engineering Science and Technology* Vol. 2, No. 3, 2010, pp. 209-212.
[46] Chan, C. H., & King, I. Using Biased Support Vector Machine to Improve Retrieval Result in Image Retrieval with Self-Organizing Map. Proc. *International Conference on Neural Information Processing*, 2004, pp. 714-719).
[47] Ray-I Chang, Shu-Yu Lin, Jan-Ming Ho, Chi-Wen Fann, and Yu-Chun Wang. A Novel Content based Image Retrieval System using K-means/KNN with Feature Extraction. *Computer Science and Information Systems*, Vol. 9, No. 4, 2012, 1645-1661.
[48] Giorgio Giacinto, A Nearest-Neighbour Approach to Relevance Feedback in Content based Image Retrieval. In Proc. The 6th ACM International Conference on Image and Video Retrieval, 2007, pp. 456-463.
[49] T. Dharani, I. Laurence Aroquiaraj. Content based Image Retrieval System using Feature Classification with Modified KNN Algorithm. Preprint arXiv, 2013, pp. 1-6.
[50] Muneesawang, P., & Guan, L. A Neural Network Approach for Learning Image Similarity in Adaptive CBIR. In Proc. *IEEE Fourth Workshop on Multimedia Signal Processing*, 2001, pp. 257-262.
[51] Wang, L., Gao, Y., Chan, K., L., Xue, P., & Yau, W. Y. Retrieval with Knowledge-driven Kernel Design: An Approach to Improving SVM-based CBIR with Relevance Feedback. In Proc. *Tenth IEEE International Conference on Computer Vision*, 2005, pp. 1355-1362.
[52] Shy, C. R. Relevance Feedback Decision Trees in Content-based Image Retrieval. In Proc. *IEEE Workshop on Content-based Access of Image and Video Libraries*, 2001, pp. 68-72.
[53] M. L. Kherfi, D. Ziou, Relevance Feedback for CBIR: A New Approach based on Probabilistic Feature Weighting with Positive and Negative Examples. *IEEE Transactions on Image Processing* Vol. 15, Vol. 4, 2006, pp. 1017-1030.
[54] Maroj, M., and Elizabeth Jacob. Information Retrieval on Internet using meta-search engines: A Review. 2008.
[55] http://Wang.ist.psu.edu/docs/related.shtml visited in March 5th 2014.
[56] http://scholar.google.com/ visited in January 14th 2018.
[57] J. R. Smith. Integrated Spatial and Feature Image System: Retrieval, Analysis and Compression. PHD Thesis, Columbia University, USA,1997.
[58] Babu, G.P., Mehtre, B.M., Kankanhalli, M. S.: Color indexing for efficient image retrieval. *Multimedia Tools Appl*. Vol. 1, No. 4, 1995, pp. 327-348.
[59] Fishburn, P.: Non-linear Preference and Utility Theory. Johns Hopkins University Press, Baltimore, 1998.
[60] Menczer, F. Complementing search engines with online web mining agents. *Decision Support Systems* Vol. 35, No. 2, 2002, pp. 195-212.

TABLE I. SUBMITTED QUERIES AND THEIR TERMS ADDED BY QRM1 OVER DIFFERENT ITERATIONS

Initial query	Term added for QRM1 Algorithm								
	1st Round	2nd Round	3th Round	4th round	5th round	6th round	7th round	8th round	9th round
Information retrieval	University	Data	Structure	Term	Recall	Document	Query	Database	client
Query Expansion	Local	Analysis	Term	Retrieval	Image	Spatial	Search	Similarity	Measure
Software maintenance	Statement	Knowledge	Research	Visualization	Evolution	Activity	Performance	Development	Design
Verification and validation	Data	Model	Simulation	System	Expert	Performance	Scientific	References	transformation
Object design	Visualization	Data	Chapter	Principles	Learning	Screen	Affordance	Constraint	Project
Functional programming	State	System	Representation	Static single assignment (SSA)	Learning	Game	Aggregate	Data	Analysis
Remote object access	Distributed	System	Information	Image	Data	Classification	Feature	Tracking	Remote Method Invocation(RMI)
Business component resources	Model	Value	Change	Strategic	Human	Management	Family	Work	Diversity
Object modelling with UML	State	Use	Relation	Access	Security	Case	Design	Analysis	Aspect
Code performance optimization	Channel	Fading	Trellis	Source	Energy	Cache	Grid	Node	Check

TABLE II SUBMITTED QUERIES AND THEIR TERMS ADDED BY QRM2 OVER DIFERENT ITERATIONS.

Initial query	Term added for QRM2 Algorithm								
	1st Round	2nd Round	3th Round	4th round	5th round	6th round	7th round	8th round	9th round
Information retrieval	University	Web	chapter	Document	References	Reading	Classification	Clustering	text
Query Expansion	Local	Analysis	Context	document	Feedback	Concept	Passage	Term	performance
Software maintenance	Slice	Statement	Program	Variable	output	Soft	Techniques	Set	method
Verification and validation	Use	Data	Techniques	system	Accuracy	Test	Computer	Method	simulation
Object design	Chapter	notes	Visualization	Data	References	Class	Smart	Work	internet
Functional programming	Data	Use	Value	State	Section	Term	Function	Language	error
Remote object access	Distributed	system	Communication	Replication	Example	Consistency	Summary	Architecture	security
Business component resources	Model	Value	Firm	Research	Strategy	Component	Level	Element	Proprietary
Object modelling with UML	State	Use	Case	Developer	Problem	Analysis	Diagram	Sequence	design
Code performance optimization	Channel	LDPC	Fading	Degree	Density	Distribution	Threshold	Rayleigh	block

TABLE III THE ACCURACY OF GOOGLE SCHOLAR, QRM1, AND QRM2 OVER THE CONSIDERED QUERIES.

Queries	Precision		
	Google Scholar	QRM1	QRM2
Information retrieval	83.19%	90.01%	85.66%
Query expansion	82.29%	87.00%	81.78%
Software maintenance	95.66%	84.01%	87.92%
Verification and validation	78.08%	74.78%	88.20%
Object design	66.04%	83.15%	85.93%
Functional programming	92.85%	89.45%	90.41%
Remote object access	89.48%	92.22%	70.24%
Business component resources	84.46%	78.65%	80.59%
Object modelling with UML	94.49%	95.22%	86.80%
Code performance optimization	82.08%	79.15%	87.84%
Global Added accuracy Percentage		+5.02%	-3.25%
Number of queries where there is an added accuracy		5 from 10	4 from 10