Volume 6, Issue 5, May 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# Application of Vector Space Model to Query Ranking and Information Retrieval

**E. E. Ogheneovo***
Dept. of Computer Science, University of Port Harcourt, Port Harcourt, Nigeria

**R. B. Japheth**
Dept. of Maths/Computer Science, Niger Delta University, Yenagoa, Nigeria

***Abstract:*** ***Retrieving information from the Internet and large databases is quite difficult and time consuming especially if such information is unstructured. Several algorithms and techniques have been developed in the area of data mining and information retrieval yet retrieving data from large databases continue to be problematic. A typical large Information Retrieval application, such as a Book Library System or Commercial Document Retrieval Service, will change constantly as documents are Added, Changed, and Deleted. This constrains the kinds of data structures and algorithms that can be used for Information Retrieval. In this paper, the Vector Space Model (VSM) technique of information retrieval was used. First, we computed the similarity scores using the weighted average of each item. The cosine measure is then used to compute the similarity measure and to determine the angle between document's vector and the query vector since VSMs are based on geometry whereby each term has its own dimension in a multi-dimensional space, queries and documents are points or vectors in this space. The cosine measure is often used. We then found out that it is easier to retrieve data or information based on their similarity measures and produces a better and more efficient technique or model for information retrieval.***

***Keywords: Databases, information retrieval, query ranking, vector space model, data mining***

## I. INTRODUCTION

Retrieving information [1] [2] [3] from the Internet and from large database is quite difficult and time consuming especially unstructured information [4] [5]. A lot of algorithms and techniques have been developed in the area of data mining and information retrieval yet retrieving data from large databases continue to be problematic. In this research work, we used the vector space model to rank information based on computed cosine. First, we computed the similarity scores using the weighted average of each item. The cosine measure is then used to compute the similarity measure and to determine the angle between document's vector and the query vector since Vector Space Models [6] [7] [8] are based on geometry whereby each term has its own dimension in a multi-dimensional space, queries and documents are points or vectors in this space. The cosine measure is often used.

Another feature that Information Retrieval Systems [9] [10] share with DBMS is Database Volatility. A typical large Information Retrieval application, such as a Book Library System or Commercial Document Retrieval Service, will change constantly as documents are Added, Changed, and Deleted. This constrains the kinds of data structures and algorithms that can be used for Information Retrieval [11]. A typical Information Retrieval System meets the following functional and non-functional requirements. It must allow a user to add, delete, and change documents in the database [12]. It must provide a way for users to search for documents by entering queries, and examine the retrieved documents. It must accommodate databases in the megabyte to gigabyte range, and retrieve relevant documents in response to queries interactively-often within 1 to 10 seconds [13].

Information Retrieval [14] is the multidisciplinary science of data search. Information Retrieval is now a mature field in computer science, and its paradigms are actively applied in diverse areas such as search engines, file systems, digital libraries, and video surveillance. The data repositories can store anything from text to digital video, and the collection size is often immense [15][16]. These massive collections pose significant challenges for the algorithms which manipulate the data. A major example of a massive dataset is the World Wide Web. There is considerable debate over exactly how much information is available on the World Wide Web. The dynamic nature of the web makes it difficult to determine the exact size [17]. The goal of information retrieval (IR) is to provide users with those documents that will satisfy their information need. We use the word "document" as a general term that could also include non-textual information, such as multimedia objects. Users have to formulate their information need in a form that can be understood by the retrieval mechanism [18]. There are several steps involved in this translation process that we will briefly discuss below .Also, the contents of large document collections need to be described in a form that allows the retrieval mechanism to identify the potentially relevant documents quickly. In both cases, information may be lost in the transformation process leading to a computer-usable representation [19] [20]. Hence, the matching process is inherently imperfect. Collection size is often immense. These massive collections pose significant challenges for the algorithms which manipulate the data [21]-[23].

 

## II. RELATED WORK

Berry et al. [24] propose a vector space model called orthogonal factorization matrix for retrieving information in a large database. The technique uses mathematical concepts from linear algebra for determining the weighted value of terms in documents using matrix approach whereby terms are represented in rows and documents are represented in columns each document is then checked against the query to determine the frequency of occurrence of each term. These are then represented in a matrix form and the cosine score of each ranked document is then computed. The resulting matrix of the term-by-document is then normalized and then transposed by ranked reduction by using QR factorization followed by SVD to identify and remove redundant information in the matrix representation of the database. By using ranking of the term-by-document matrix, a geometric interpretation of the vector space model is formed.

Singh and Dwivedi [25] discuss the various approaches of vector space model to compute similarity score of hits in information retrieval. These approaches are Term Count Model, TF-idf model and the vector space model based on normalization. Based on the similarity function between vector document and query term, the similarity function is computed using database collection of retrieved documents, query and index term. The term frequency-inverse document frequency (Tf-idf) is used to determine how important a word is in a document based on weighting factor in information retrieval and converts textual representation of information into a vector space model, term-count model gives better results for long documents when compared with small documents. Thus long documents have a score. The Tf-idf method uses weight to show the importance of words in the document especially for stop-words (e.g., a, an, the etc.) filtering that is common to give weights to meaningful terms. This is because stop-words are known to have low weight. The three approaches of vector space all perform well for long documents where the frequency term in documents is high.

Hiemstra and De Vries [26] propose the language model by the retrieval algorithms to the widely accepted traditional algorithms for information retrieval: the Boolean model, the vector space model, and the probabilistic model. The proposed algorithm is used to match terms when computed for efficient information retrieval. The language models for information retrieval are somehow similar to both the tf.idf term weighting in vector space model and relevance weighting in probabilistic model. Thus the work proposed a strong theoretical approach of the language modeling approach by showing that the approach performs better than the weighting algorithms developed in traditional models. Therefore, the language modeling approach result in tf.idf term weighting, the tf component and the idf component are both logarithmic thus making it a tf + idf algorithm and not tf.idf algorithm as formerly claimed. Furthermore, they use collection frequency instead of document frequencies.

Lv and Zhai [27] proposed adaptive feedback approach to information retrieval. It is a learning approach to adaptively predict the optimal balance coefficient for each query and each collection. These include discrimination of query, discrimination of feedback documents, and divergence between query and feedback documents. They used three heuristics to characterize the balance between query and feedback information by determining a number of features and combining them using linear regression to predict their coefficient.

Tsatsaronic and Panagiotopoulous [28] propose the Generalized Vector Space model for retrieving semantic information for word thesauri like WordNet. In this technique, they incorporated semantic information by modifying the standard vector space model. From the experimental evaluation, a test was conducted on the performance of the semantic relatedness measure (SR) for a pair of words using three benchmark data sets. The semantic relatedness measure considers all of the semantic links in WordNet such as a graph, weight edges based on type and depth by computing the maximum relatedness between any two nodes, connected via one or more paths. Then a performance evaluation was conducted. The correlation for the three data sets shows that SR performance is better than any other measure of semantic relatedness. Using three TREC collections, the experimental result shows that semantic information can boost text retrieval performance.

Wong et al. [29] modeled information by using vector spaces. The model proposed a generalized Vector Space Model (GVSM). Considering the limitations associated with Boolean model of information retrieval due to its sound generalization of the traditional vector space model for computing the correlation of relevant terms. First, the authors explained how the elements of Boolean algebra can be modeled as vectors in a vector space and by representing terms as Boolean expression by showing whether two vectors are identical or orthogonal. They show that if to vectors are identical (i.e., not orthogonal), then the corresponding Boolean expressions have at least one minterm in common. Using GVSM, they generalized the term vector representation such that the coefficients are not binary such that the representation of a document is taken to be a sum of term vectors. The vector sum operator and the document is hypothesized as a vector sum of the associated term vectors.

## III. METHODOLOGY

The Vector Space Model (VSM) is a technique used to represent documents and queries as vectors in multidimensional space, whose dimensions are the terms used to build an index to represent the documents [30]. It is the most widely used technique for information retrieval due to its simplicity; efficiency over large document collections and it is very appealing to use. The effectiveness of the VSM depends mostly on the term weighting applied to the term of the document vectors. The VSM has three phases: (1) the document indexing phase, where content bearing terms are extracted from the document text; (2) the weighting of the indexed terms to enhance retrieval of document relevant to the user; and (3) ranking the document with respect to the query based on similarity measure. Figure 2.x shows a typical example of a Vector Space Model for two documents, three terms and a query.

 

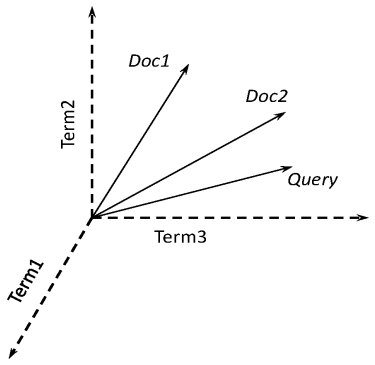


Fig. 1: A Vector Space Model for two documents, 3 terms, and a query

Calculations in the VSM are based on geometry whereby each term has its own dimension in a multi-dimensional space, queries and documents are points or vectors in this space. The cosine measure is often used to computes the similarity measure and to determine the angle between documents vector and the query vector. This cosine measure formula is:

$$Cosine(D_i, Q) = \frac{\sum_{j=1}^{t} d_{ij} \cdot q_j}{\sqrt{\sum_{j=1}^{t} d_{ij}{}^2 \cdot \sum_{j=1}^{t} q_j{}^2}}$$

where $D_i$ = number of documents

Q = query

ij = rows and columns

The angle between two vectors is a measure of divergence between the vectors in a V-dimensional Euclidean space, where V represents the size. Thus VSM is used to rank documents by the similarity between the query and each document (Salton and McGill, 1983).

The similarity measure of two documents is determined by:

sim(q, d) = $\frac{1}{WqWd} \sum wq, t$ x $W_{q,t}W_d = \sqrt{\sum_t W2d, t}$

where $W_q$ is the same for all documents

$w_{q,t}$ and $w_{d,t}$ can be accumulated as we process the inverted list

$W_d$ can be pre-computed

Using the vector space model for retrieving information, a user queries the database to find relevant documents, by using the vector representation of the documents involved. Using query matching, it is possible to find document most similar to the query used and also to determine the weighting of terms using the cosine measure. Using a collection of terms and documents, we illustrate the query matching process based on six-dimension vector space using six (6) terms and five (5) documents. In this case, a user is searching for information about how chickens produce their young ones. Based on this term-by-document query search process, we write the corresponding query as a vector using one (1) to represent the query term that appears in each document and zero (0) to represent non-terms in each document. We then search for relevant documents by computing the angles between the query vector and the document vectors. A threshold value of 0.5 is set to determine if a document will be returned as relevant or not. Using this threshold, a document is returned as relevant only if the cosine of the angle it makes with the query vector is greater than the threshold value we already set. Thus all the documents having the term 1 ($T_1$): chickens(s) are returned as relevant.

Using query comparison to determine the document that is most relevant, we use the cosine of the angle between the query vector and the document vectors. The equation used is:

Thus using VSM, document collection can be represented by a matrix of term weights.

$D_i = (D_{i1}, D_{i2}, \ldots, D_{it})$ $Q = (q_1, q_2, \ldots, q_t)$

| | **$Term_1$** | **$Term_2$** | **…** | **$Term_t$** |
|---|---|---|---|---|
| $Doc_2$ | $d_{11}$ | $d_{12}$ | … | $d_{1t}$ |
| $Doc_2$ | $d_{21}$ | $d_{22}$ | … | $d_{2t}$ |
| ⋮ | ⋮ | ⋮ | | ⋮ |
| $Doc_n$ | $d_{n1}$ | $d_{n2}$ | | $d_{nt}$ |

The weight based on count and idf value, we computed the count $TF_{i,j}$ and the WEIGHT and based on the resultant values, we computed the similarity scores using Pythagoras theorem and applied the dot product by first determining the cosine angles between the term vector and document vectors. We then ranked the document based on the similarity scores computed to determine the relevance by which these documents can be retrieved. Table 3.x shows the result computed based on the terms and the documents.

**Query**: How chickens produce their young ones

How chickens lay eggs and latch them

$D_1$: Chickens are in the class of birds

$D_2$: Chickens latch their young ones

 

$D_3$: Chickens lay eggs before they latch them

$D_4$: The eggs are incubated before being latched to chickens. $D_5$: The incubation period takes 21 day D = 5, IDF = $\log\left(D/dfi\right)$ dfi = number of documents containing term j.

Table: Weight based on term count and idf value

| | | Count TFij | | | | | | | | WEIGHTS $W_i$ = Tfi=j = IDfj | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Terms** | **Q** | **$D_1$** | **$D_2$** | **$D_3$** | **$D_4$** | **$D_5$** | **Dfi** | **D/dfi** | **IDFj** | **Q** | **$D_1$** | **$D_2$** | **$D_3$** | **$D_4$** | **$D_5$** |
| Chicken (5) | 1 | 1 | 1 | 1 | 1 | 1 | 5 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Lay (5) | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 5 | 0.6990 | 0.5 | 0 | 0 | 0.699 | 0 | 0 |
| Eggs | 1 | 0 | 0 | 1 | 1 | 1 | 3 | 1.67 | 0.2227 | 0.2227 | 0 | 0 | 0.2227 | 0.2227 | 0.2227 |
| Incubate (ia) | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 2.5 | 0.3979 | 0 | 0 | 0 | 0 | 0.3999 | 0.3999 |
| Latch (ed) | 1 | 0 | 1 | 1 | 1 | 1 | 4 | 1.25 | 0.0969 | 0.0969 | 0 | 0.0969 | 0.0969 | 0.0969 | 0.0969 |
| Bird (5) | 0 | 1 | 1 | 0 | 0 | 0 | 2 | 2.5 | 0.3979 | 0 | 0.3979 | 0.3979 | 0 | 0 | 0 |

**COMPUTING SIMILARITY SCORES**

Using Pythagoras theorem, we compute the magnitude of the vector as

$|D_i| = (a_{12+} a_{22} + a_{32} + \ldots\ldots a_{n2})^{½}$

$|D_1| = \sqrt{12+12} = \sqrt{2} = 1.4142$

$|D_2| = \sqrt{12+12+12} = \sqrt{3} = 1.7321$

$|D_3| = \sqrt{12+12+12+12} = \sqrt{4} = 2.0$

$|D_4| = \sqrt{12+12+12+12} = \sqrt{4} = 2.0$

$|D_5| = \sqrt{12+12+12+12} = \sqrt{4} = 2.0$

$|Q| = \sqrt{12+12+12+12} = \sqrt{4} = 2.0$

## IV. RESULTS AND DISCUSSION

Based on the computation, we applied the dot product which is Q. Di

$Q.D_1$ = 1x1+ 1 x 0 + 1 x 0 +1 x 0 + 1 x 0 + 0 x 1 = 1

$Q.D_2$ = 1x1+ 1 x 0 + 1 x 0 +1 x 1+ 0 x 1 = 2

$Q.D_3$ = 1x1+ 1 x 1 + 1 x 1 +1 x 1 = 4

$Q.D_4$ = 1x1+ 1 x 0 + 1 x 1 +0 x 1+ 1 x 1 = 3

$Q.D_5$ = 1x1+ 1 x 0 + 1 x 1 +0 x 1+ 1 x 1 = 3

But Dot product = (magnitudes produce) cosine angle

Cosine Q = sin (Q, Di) = $\dfrac{q\ .\ Di}{|Q|\ x\ |Di|}$

$$\therefore\ Cosine\ Qd_2 = \frac{Q.D_1}{|Q|x|D_2|} = \frac{1}{2.0x\ 1.4142} = 0.3536$$

$$Cosine\ Q\,d_2 = \frac{Q\ .\ D_2}{|Q|\ x\ |D_2|} = \frac{2}{2.0\ x\ 1.7321} = 0.5773$$

$$Cosine\ Q\,d_3 = \frac{Q\ .\ D_3}{|Q|\ x\ |D_3|} = \frac{4}{2.0\ x\ 2.0} = 1$$

$$Cosine\ Q\,d_4 = \frac{Q\ .\ D_4}{|Q|\ x\ |D_4|} = \frac{3}{2\ x\ 2} = 0.75$$

$$Cosine\ Q\,d_5 = \frac{Q\ .\ D_5}{|Q|\ x\ |D_5|} = \frac{3}{2\ x\ 2} = 0.75$$

From the computation, $d_3$ is ranked 1, $d_4$ and $d_5$ are ranked 2, $d_2$ is ranked 4 while $d_1$ even though it is not ranked, it also has some terms in the document. First, we computed the similarity scores using the weighted average of each item. The cosine measure is then to compute the similarity measure and to determine the angle between documents vector and the query vector since VSMs are based on geometry whereby each term has its own dimension in a multi-dimensional space, queries and documents are points or vectors in this space. The cosine measure is often used. We then found out

 

that it is easier to retrieve data or information based on their similarity measure and produces a better and more efficient technique or model for information retrieval. From these results, it was seen that query $d_3$ is ranked most and will appear first when retrieving information from the Internet using a search engine.

## V. CONCLUSIONS

Retrieving information from the Internet and from large database is quite difficult and time consuming especially if such information is unstructured. A lot of algorithms and techniques have been developed in the area of data mining and information retrieval yet retrieving data from large databases continue to be problematic. In this research work, we used the vector space model for retrieving information on the Internet. First, we computed the similarity scores using the weighted average of each item. The cosine measure is then to compute the similarity measure and to determine the angle between documents vector and the query vector since VSMs are based on geometry whereby each term has its own dimension in a multi-dimensional space, queries and documents are points or vectors in this space. The cosine measure is often used.

We then found out that it is easier to retrieve data or information based on their similarity measures and produces a better and more efficient technique or model for information retrieval. This research work is very significant in that it aim to design a tool that will enable users to retrieve information from the Internet more efficiently and effectively. Another feature that Information Retrieval Systems share with DBMS is Database Volatility. A typical large Information Retrieval application, such as a Book Library System or Commercial Document Retrieval Service, will change constantly as documents are Added, Changed, and Deleted. This constrains the kinds of data structures and algorithms that can be used for Information Retrieval.

## REFERENCES

[1] R. Baeza-Yates and B. Ribeiro-Neto, " Modern Information Retrieval," ACM Press.

[2] T. Y. Liu, J. Xu, T. Qin, W. Xiong and H. Li, "LETOR: Benchmark Dataset for Research on Learning to Rank for Information Retrieval," In Proceedings of the Learning to Rank workshop in the 30th annual International ACM SIGIR Conference (SIGIR'07) on Research and Development in Information Retrieval.

[3] J. Xu and H. Li. , "Adarank: A Boosting Algorithm for Information Retrieval," In Proceedings of the 30th Annual International ACM SIGIR (SIGIR'07) Conference on Research and Development in Information Retrieval, pp. 391–398, New York, NY, USA, 2007.

[4] P. Castells, M. Fernandez, and D. Vallet,." An Adaptation of the Vector Space Model for Ontology-Based Information Retrieval," Knowledge and Data Engineering. IEEE Transactions, Vol. 19, No. 2, pp. 261-272, 2007.

[6] A. B. Manwar, H. S. Mahalle, K. D. Chinchkhede and V. Chavan, " A Vector Space Model for Information Retrieval: A MATLAB Approach," Indian Journal of Computer Science and Engineering (IJCSE), Vol. 3, No. 2, pp. 222-229, 2012.

[7] C. Zeng, Z. Lu and J. Gu, "A New Approach to Email Classification Using Computer Vector Space Model," In Proceedings of the Future Generation Communication and Networking Symposia, 2008, FGCNS'08, pp. 162-166.

[8] I. R. Silva, J. N. Souza and K. S. Santos , " Dependence Among Terms in Vector Space Model," Proceedings of the Database Engineering and Applications symposium, pp. 97-102, 2004.

[9] S. M. Beitzel, E. C. Jensen, A. Chowdhury and O. Frieder, " Varying Approaches to Topical Web Query Classification," In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 783–784, New York, NY, USA, 2007.

[10] M. Chau, "Teaching Key Topics in Computer Science and Information Systems Through A Web Search Engine Project," ACM Journal of Educational Resources in Computing, Vol. 3, No. 3, pp. 1 – 4, 2003.

[11] J. Laerty and Zhai, C., "Document Language Models, Query Models, and Risk Minimization for Information Retrieval," In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval , pp. 111–119, New York, NY, USA.

[12] M. Fontoura, E. J. Shekita, J. Y. Zien, S. Rajagopalan and A. Neumann, " High Performance Index Build Algorithms for Internet Search Engines," In VLDB, Proceedings of the 13th International Conference on Very Large Data Bases, pp. 1158 – 1169, 2004.

[13] P. K. C. M. Wijewickrema and A. R. M. M. Ratnayake , "Enhancing Accuracy of a Search Output: A Conceptual Model for Information Retrieval, Journal of the University Librarians Association of Sri Lanka, Vol. 17, Issue 2, pp. 119-135, 2013.

[14] E. Xing, A. Ng, M. Jordan and S. Russell, " Distance Metric Learning, with Application to Clustering with Side-Information," In Advances in NIPS, Vol. 15, 2003.

[15] U. Lee, Z. Liu and J. Cho, "Automatic Identification of User Goals in Web Search," In Proceedings of the 14th International Conference on World Wide Web (WWW'05), pp. 391–400, New York, NY, USA, 2005.

[16] A. Singhal, "Modern Information Retrieval: A Brief Overview," Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, Vol. 24, Issue 1, pp. 35-42, 2001.

[17] D. Gruhl, L. Chaver, D. Gibson, J. Mayer, P. Pattanayan, A. Tomkins and J. Zien , "How to Build a Webfountain: An Architecture for Very Large-ScaleText Analysis," IBM Systems Journal, Vol. 43, No. 1, pp. 256 – 272, 2004.

 

[18] J. M. Pontea and W. B. Croft, "A Language Modeling Approach to Information Retrieval," In Research and Development on Information Retrieval, pp. 275–281, 1998.

[19] J. Kang and G. Kim, " Query Type Classification for Web Document Retrieval.," In Proceedings of the 27th Annual Int'l ACM SIGIR Conference on Research and Development in Information Retrieval, 2003.

[20] D. S. Guru and H. S. Nagendraswamy , "Clustering of Interval-Valued Symbolic Patterns Based on Mutual Similarity Value and the Concept of Mutual Nearest Neighbor," In ACCV(2), pp. 234-243, 2006.

[20] X. Geng, T. Y. Liu, T. Qin, A. Arnold, H. Li and H.-Y., Shum , "Query Dependent Ranking Using K-Nearest Neighbor," SIGIR'08, July 20-24, 2008, Singapore.

[21] P. Fraternali, "Tools and Approaches for Developing Data Intensive Web Applications: A Survey," ACM Computing Survey, Vol. 31, No. 3, pp. 227 – 263, 1999.

[22] P. Debraand R. Post, "Information Retrieval in World Wide Web: Making Client-Based Searching Feasible," In Proceedings of the 1st International World Wide Web Conference (Geneval, Swizerland), 1994.

[23] R. Nallapati, "Discriminative Models for Information Retrieval," In Proceedings of the 27th Annual International ACM SIGIR conference (SIGIR'04) on Research and development in information retrieval, pp. 64–71, New York, NY, USA, 2004. ACM.

[24] M. W. Berry, Z. Drmač and E. R. Jessup, " Matrices, Vector Spaces, and Information Retrieval," Society for Industrial and Applied Mathematics, Vol. 41, No. 2, pp. 335-362, 1999.

[25] J. N. Singh and S. K. Dwivedi, "Analysis of Vector Space Model Information Retrieval. In Proceedings of the National Conference on Communication Technologies and its Impact on Next Generation computing (CTNGC'12), Int'l Journal of computer Applications (IJCA), 2012.

[26] D. Hiemstra and A. P. De Vries, "Relating the New Language Models of Information Retrieval to the Traditional Retrieval Models," CTIT Technical Report TR-CTIT-00-00, http://www.ctit.utwente.nl, pp. 1-14, 2000.

[27] Y. Lv and C. Zhai, "Adaptive Relevance Feedback in Information Retrieval,". In Proceedings of CIK'09, November 2 – 6, Hong Kong, China, 2009.

[28] G. Tsatsaronic and V. Panagiotopoulous, "A Generalized Vector Space Model for Text Retrieval Based on Semantic Relatedness," In Proceedings of the EACI 2009 Student Research Workshop, Athens, Greece, 2nd April, 2009, Association of Computational Linguistics, pp. 70-78, 2009.

[29] S. K. Wong, W. Ziarke, V. V. Raghaven and P. C. N. Wong ," On Modeling of Information Retrieval concepts in Vector Space, ACM Transactions on Database System, Vol. 12, No. 2, pp. 299-321, 1987.

[30] G. Salton, E. A. Fox. and H. Wu, " Extended Boolean Information Retrieval," Communications of the ACM, Vol. 26, No. 11, pp. 1022-1036, 1983.

 