

Elucidating Complex Queries Based on Curtailing Technique for Effective Green Communication

Anshy Singh¹, Himanshu Sharma²

¹Department of Computer Engineering and Application, GLA University, Mathura, India

²Department of Computer Engineering and Application, GLA University, Mathura, India
Mathura.anshy.singh@gla.ac.in¹, Mathura.himanshu.sharma@gla.ac.in²

Article History: Received: 10 November 2020; Revised: 12 January 2021; Accepted: 27 January 2021;
Published online: 05 April 2021

Abstract: A system has been developed to improve on the quality of search for which we have replaced the present searching methods by our procedure Explore Scheme, which greatly increases the utility of Web, over what is available today. None of the Search Techniques is able to deal effectively and efficiently with the huge volume of information posted on World Wide Web. In the current searching methodology, our proposed system emphasizes on enhancing the searching method by our proposed Explore-Scheme procedure, which provides solution to the complex queries based on contextual expansion. It is our observation that in investigating relevance feedback questions (both existing and new), we will implement the Explore-Scheme algorithm on web in this way we can enhance the ability which will give rise to the information which is good, of some use and relevant. Currently, we are implementing the system that is able to demonstrate important properties of the presented approach. Our method reduces the word to its atomic part thus curtailing the size of the dictionary by saving Storage space and processing time enhancing the processing speed of query. We can finally say that our proposed algorithm will definitely improve the information retrieval capabilities and are able to deal effectively and efficiently with the huge volume of information posted on World Wide Web.

Keywords: Information Retrieval, Indexing, Stemming, Content-based Search, Data Clustering

1. Introduction

Each human being after surfing the web or utilizing the corporate intranet forms an opinion regarding the ways and the means which develops a “good” Web founded systems as well as applications [1]. The perception of each human being differs [2]. There are the people who enjoy vibrant graphics, and then there are the others who wish to have lucid text [3,4,5]. There are few people who ask for the ample information while there are the few who aim at a presentation which his abbreviated [6]. In reality, the point of view of the people who use “goodness” might be more important than any technical discussion of Web applications quality [7,8,9]. Retrieval of information, utility, effectiveness as well as functionality gives a good ground to assess the merit of system which is web based [10,11,12]. The steep development of the database in most of the avenues of human action has designed a want for renewed tools which will turn data into concrete knowledge [13,14,15]. The scholars from several technological avenues like machine learning, statistical data analysis, pattern recognition, information retrieval and information extraction as well as neural network have come forward to meet the demand have been searching formulae and ideas [16 -19]. Such efforts have pioneered us to a research avenue which is generally addressed as date mining [20]. It may be defined as the concrete extraction of non-obvious rooting out the underlined which is unknown before and has the potential of beneficial information from the data that is given [21, 22]. In fact till now, maximum task in the discovery of knowledge in the databases has been related with data which is structured [23]. Nevertheless, maximum information can be accessed in the form of given text as in the documents, emails, web presentations and manuals etc [24]. This form of information comprises the restricted internal structure obviously different from the tabular information stored in Conventional Databases [25].

Information retrieval consequences are generally given to the users like a documents based on rank which is in the decreasing order of relevance. This system incorporates the feedback which is based on the relevance of the user where he or she can avail the opportunity to evaluate the list of ranks and figure out the documents which are important to answer the query of the users and the one which do not [26]. Afterwards the information is utilized by the feedback algorithm and it causes a new rank document which is displayed to the user and this process repeats itself.

The effectiveness of relevance feedback is explored in helping the users to explore a target document which is decided already. Moreover, an innovative approach is devised by utilizing a fact that, as the number of the options of the user is many, it is still found to be limited. In this way it is possible to devise and peruse the whole space of feasible options of the user for a cycle of feedback on the importance of the document which is presented and achieve the upper bound the efficiency of the executed relevant feedback algorithm. We can interpret the upper bound as the end result obtained by an ideal user. In fact, the options of this ideal user enable

the system to collect maximum information for feedback which is important [27]. We have the benefit of this approach ahead as it permits the study of feedback which is important sans the ambiguities of the studies of the user.

2. WebSearch Technologies

The study of Web Search technologies classifies into several categories like - Hyperlink Exploration, Meta Searches, Content-Based Multimedia Searches, XML Content Retrieval, SQL Approaches, Information Retrieval and others [28].

There are the various Search Engines that have been developed, on basis of Web Search Technologies. Most Search Engines provides Web Search services as well as portals which the bases of web home out of which the users can explore several services as e-commerce, searches as well as news of chat rooms etc [29]. The major Search Engines like Clever on Hyperlink, Grouper on Clustering, Huskey Search on MetaSearch, Inquires on MetaSearch, MetaSEEK and PicToSeek on Image, that he commercial search engines which have not implemented the advanced Search Technologies These Search Engines are the snapshots of the current situation.

3. Architecture of Search Engine

Mainly there are two methods which one applies to Web Search services as Directories and Genuine Search Engine.

Search engines, like HotBot, generate their listing on their own. Web gathering information is crawled by Search engines, afterwards the users seek through the database generated by them.

Human beings play a pivotal role in listing for a directory, like Yahoo which is like he advertisements of newspaper-classified [30]. A small description of site to directory is submitted by the web author. Matches are sought after by the search engine in accordance with the elaboration reserved. Architecture of Search Engine is shown in Figure 1.

Instead of two above services, a Hybrid Search Engine is also developed to maintain the associated directory. Generally Search Engine consists of three factors i.e.

- Spider
- Indexes
- Query Engine

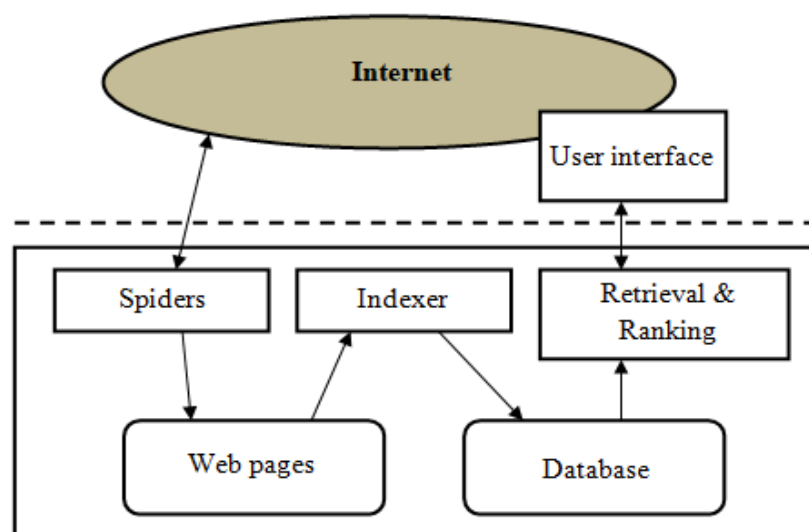


Figure 1. Architecture of Search Engine

3.1. Spider

Search engine sends out a Spider which is a unique piece of Software. It is sent out to index Web site pageso be incorporated in Search Engine of database information. To put it differently a Spider is basically program

which is automated and it downloads the documents which crawler passes to it. Its function is similar to that of browser's, the time it is connected to a Web site as well as download pages. There are three basic tools usefully used to implement an experimental Spider i.e.–

- LINUX
- JAVA.Net
- CPAN

3.2. Indexer

All the words are extracted by the indexer module through every page, as well as it records the URL's where every word struck. Its result is quite big "look up table" which could get all the URL's which indicate to pages where a particular word takes place. In fact, the table is constrained to the pages which were covered in the process of crawling [31]. Algorithmically examining information is how this process is known where the items to build a data structure which could be quickly searched. To devise an index of the Web is basically an initiation towards the execution of quick and precise Search Engine. The size and its fast changing rate process unique problems in the text indexing of the Web.

Index quality shows the significance of the search engine database to retrieve the appropriate and comprehensive information. The indexing scheme should process hundreds of gigabytes of information effectively. Queries should be tackled rapidly, at a speed of hundreds to thousands per second. The indexing function is executed through the indexer and the sorter. It executes an amount of functions. It shares out these hits into a set of "barrels", building a partly sorted forward index. It carries out a different significant function. It parses out all the links in every web page and stores significant data about them in an anchors file. This file includes sufficient information to establish where every link points from and to, and the text of the link.

3.3. Query Engine

This engine holds accountability to take and fill Search entreaties from users. Engine rests largely on indexes as well as often on the repository of the page [4]. Query processing activity does the analysis of a question as well as compares it to indexes so that it may get things which are important in this regard. One or two keywords are used by the user together with Boolean modifiers into a Search Engine as well the result sets are often big. Afterwards the ranking module acts the works of finding out the results which are close to the top and are the most probably to be what the user looks for. The leading Search Engine manifested that exploring for a particular word, went through multiple pages [32]. A research revealed that maximum users turn anxious if they are to read more than one pages to search for the result they aim at. Therefore there must be some ranking and system which minimizes the number of pages for the user to search his or her result.

The field of search engine quality research describes its significance not only from a common interest of data science in the presentation of search systems and also from a wider discussion on search engines over general society. It is the practical relevance of data retrieval methods to large-scale text collections. A web search engine is the noticeable example, it is declared, and search engines can be found in many different applications, such as desktop search or enterprise search. Search engines have been around for many years.

Search engines can be used with small collections, such as a few hundred emails and documents on desktop or particularly huge collection, like the entire Web. There might be only some customers of a specified function. Scalability is obviously an imperative problem for search engine plan. Designs that work for a specified function must carry on working as the total degree of data and the amount of consumer grow.

4. Information Extraction

Mapping of natural language texts into something which is defined before, templates or structured representation, that at the time filled, manifest an extract of important information out of original text is termed as Information Extraction. Information relates to the entities of interest in the execution domain which could be companies or the people or the bonds between such entities, generally in the form of acts where entities participate like the takeovers of the company and the successions of the management. The extracted information might be treasured in databases so that it can be queried, as well as data mined, as a gist in natural language and so forth. First required initiation is the linguistic pre-processing which comprises the tools or linguistics like tokenization, part of speech tagging and so forth so that it might feed the system of information extraction. It is

aimed to extract from the documents, financial words as well as interesting financial events. There is interest in us for particular events between specific companies/persons. Moreover, a specific event like taking over between two companies will be utilized as a characteristic to name the document and the same event of taking over between various companies.

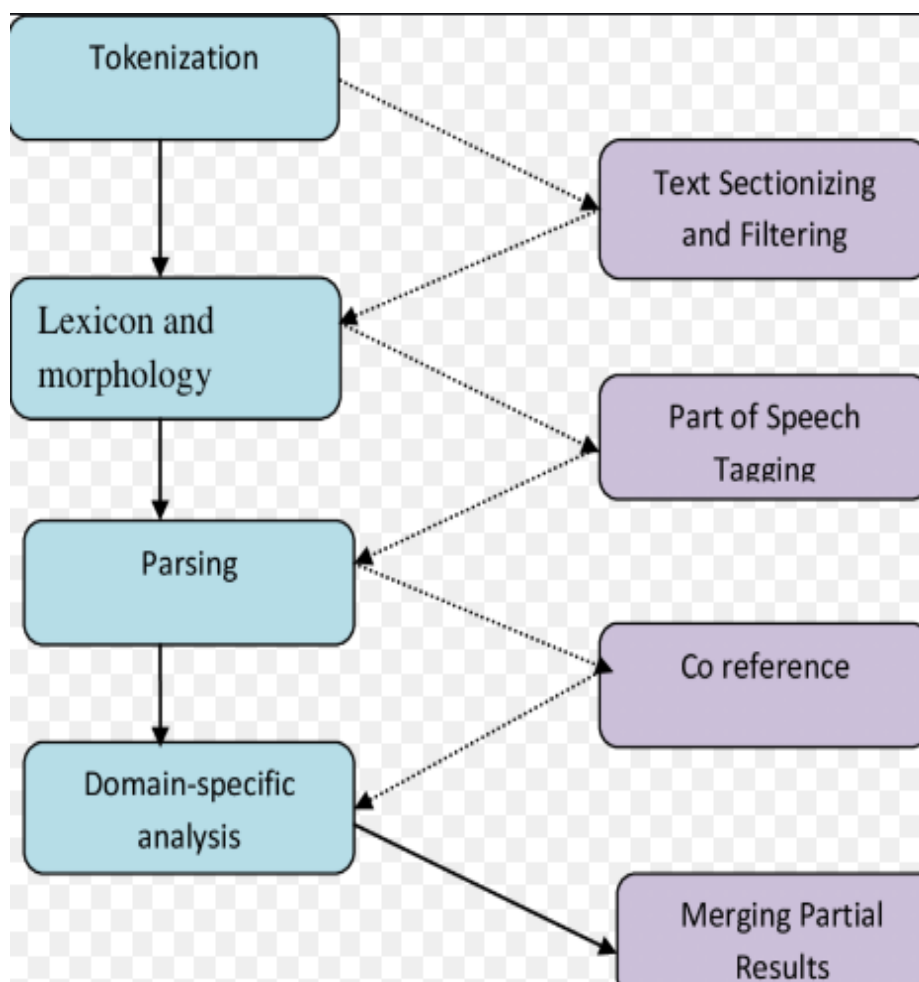


Figure 2. General diagram of information extraction system

General diagram of information extraction system is shown in Figure 2. Information extraction is the mission to recognize construction and unite information from natural language texts over the web. Specified a field of interest, it need to generate a data base on this area. As data combining from ordered sources is in universal and highly consistent than the usage of unstructured texts, web information extraction is principally interesting for the subsequent information needs. The data that could not mined from structured or semi-structured sources, like XML documents, single web sites or tables, yet is widen diagonally different web pages. The information that is predictable to be nearby on the web. Therefore, it can say in common that web information extraction is suitable for all subjects that public inscribe about.

5. Information Retrieval

Information Retrieval techniques are widely used in Web document search. Information Retrieval is basically demarcates which documents out of many, would be retrieved to meet the user's search for some information. The user's information need is represented by....A query represents the information of the user, as well as comprises some Search tags as well as probably the extra information such which is important. Therefore, decision to retrieve is taken by comparing the words of a query with the index terms coming up in the document. It might be a binary decision, or it might incorporate the degree of relevance of the query. The most talked about technique employed by Search engines in several Information Retrieval techniques are the Relevance Feed Back, Data Clustering as well as Stemming are most popular techniques used by Search Engines. Figure 3 shows the example of information retrieval system

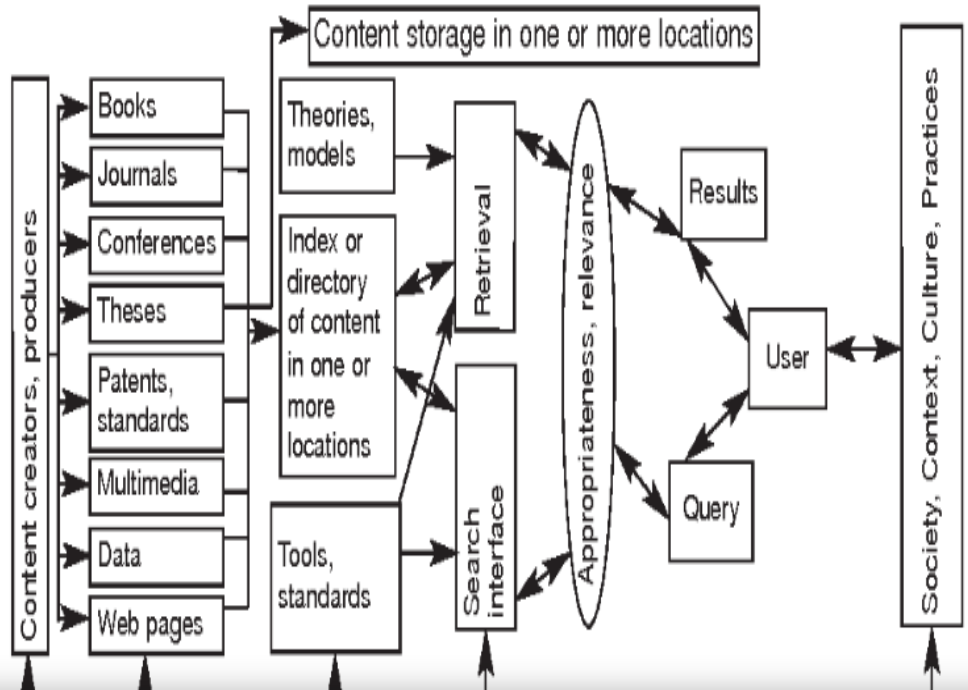


Figure 3. Example of Information Retrieval System

Apart from the above major Search techniques, some ad hoc methods include –

- Various enhanced Spiders can be formed in the Literature. Some Spiders can be extended, customized by someone, and can be relocated, scaled and are particular about web site.
- Task with an aim at causing the components to be required for Web Search more effective, like better ranking algorithm as well as more efficient Spiders.
- Artificial Intelligence might be utilized to gather and suggest Web pages.
- A natural language interface designed to make the system easier to use.
- Representation, comparison, interaction also sometimes modification and judgment are the processes significant in Information Retrieval systems. Figure 4 illustrates the example of filtering diagram.

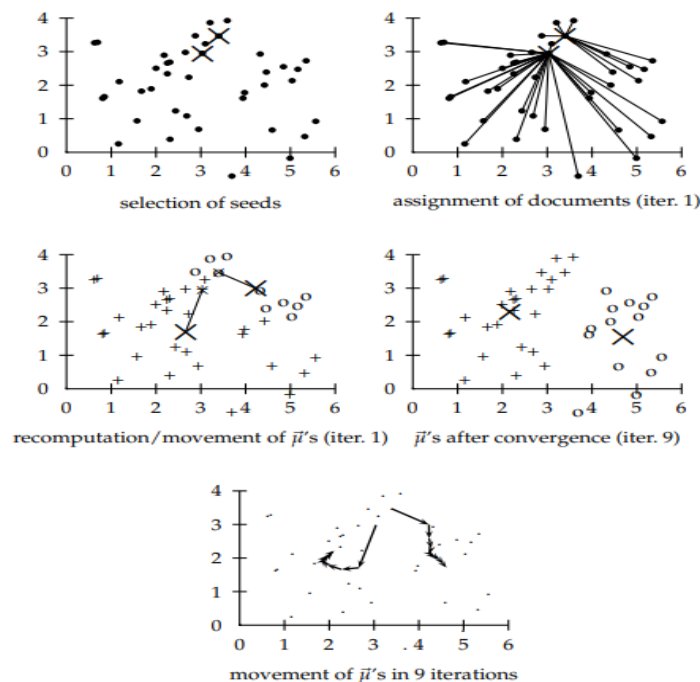


Figure 4. Example of filtering diagram

Algorithm 1:

1. $(s_1, s_2, \dots, s_k) \leftarrow \text{select random words}(x_1, \dots, x_N)$
2. For $n \leftarrow 1$ to N
3. do $\mu_k \leftarrow s_k$
4. while stopping criteria is not meet
5. do for $k \leftarrow 1$ to K
6. do $\omega_k \leftarrow \{ \}$
7. for $n \leftarrow 1$ to N
8. do $j \leftarrow \text{argmin}_j | \mu_j - s_n |$
9. reassign words
10. for $k \leftarrow 1$ to K
11. re-computation of words
12. return the result

6. Information Filtering

A more recent field in information science is Information filtering which came into being due to increasing volume of online transient data. In Information Retrieval, the data-set is viewed as being relatively static; in Information Filtering it is as a dynamic 'stream of information. In Information Retrieval, user queries represent short-term information needs; in Information Filtering, user queries represent long-term information needs. Focus of filtering information upon the effectiveness of filtering, making an effort to give more fine-grained filtering utilizing relational, rule-based, retrieval of information as well as artificial intelligence methods. A diagrammatic representation of an Information Retrieval/Information Filtering system is given in Figure 5.

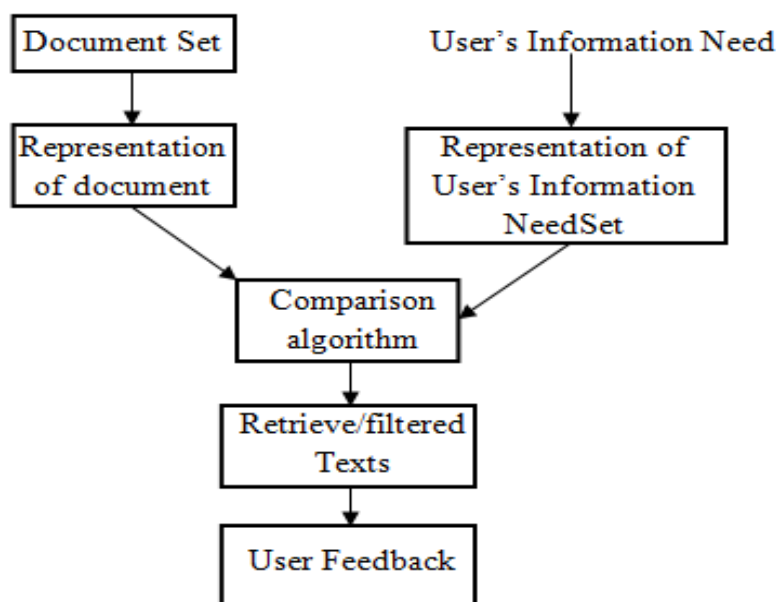


Figure 5. Architecture of retrieval/filtering system

Some of the mechanisms arising in Information Retrieval/Information Filtering systems are as follows –

Representation: The information need of the user as well as the document set should be presented in such a way as the computer might affect the comparison. Techniques of representation vary from using indexes, matrices or vector representation, to the connectionist network, semantic networks and more modern representations- neural networks.

Comparison: Comparison mechanisms utilized are basically dependent on the representations which are underlying utilized in the system. The first primitive systems incorporate using simple string matching algorithms; a bit advanced system employ statistical operations which involve vector and matrix computations, and some contemporary systems use manipulation of the network (semantic/neural) via spreading activation search mechanisms or propagation techniques.

Feedback: To improve the performance of the Information Retrieval/Information Filtering system, a feedback mechanism is usually incorporated. It generally incorporates the utility stating the satisfaction or the otherwise with documents that returned. After such feedback, the query is generally transformed to achieve better results moreover, the filtering process resumes. This process can be automatic or manual. Searching the web using many of the common search engines involves a manual modification- by adding more terms etc. Whereas other systems involve automatic expansion/modification of profile/query using thesauri, etc. again.

Another recognizable methodology when recommender system is content-based sifting. Content-based filtering techniques depend on a profile of the user's inclination and on a portrayal of the thing. In this filtering to portray the things, keywords are utilized and afterward a client profile is worked to show the sort of thing this client likes. In brief, these calculations attempt to prescribe things that are identified with those that a client preferred before. An assortment of applicant things is contrasted and things before evaluated by the client and the best-coordinating things are used.

To create client profile, the framework often centers around two sorts of data: (I) a model of the client's inclination (ii) a record of the client's connection with the recommender framework. Generally, these techniques utilize a thing profile (for example a bunch of particular credits and highlights) describing the thing inside the framework. Depends on the weighted vector of thing highlights the framework makes a substance based profile of clients. The weights show the importance of every component to the client and could be calculated from autonomously evaluated content vectors utilizing an assortment of procedures.

7. Our Approach

In our approach, the key terms of a query or document have been put forward in parts instead of through pristine words. Over here the suffixes are removed from the base words. As far statistical analysis is concerned, it is found that it has its importance to compare the text so that the can be identified with common meaning and form for it is similar e.g., the terms stopped and stopping will be counted as the same and would be assumed as taken from stop. The process places such common forms.

For instance - the terms "searching", "searched" and "searches" are stemmed from "search". In this research, all the terms are stemmed out before Indexing. Moreover, all query words are stemmed before being retrieved. Hence a query for the term "searching" will be similar with a document which comprises the term "searches". It means that different variants of a word can be conflated to one representative form -moreover, it also reduces the size of dictionary, which means that the number of distinct words required for representing a number of documents. A dictionary of smaller size ends up a saving the processing time as well as storage. Our proposed algorithm Explore-Scheme works as follows.

8. Explore-Scheme Algorithm

- 1) Establish the word for search
- 2) While not at end of file do
 - (a) Read the text and go to label 1

If mismatched, fail then go to label2

if matched and condition is not matched then fail and go to label3

if matched and condition met then fire and go to label 4,5,6,7 and 8

- (b) Return stem word and compare with database for searching the result

1: Step 1a

SSES ->SS

2: Step 1b

(m>0) EED ->EE

3: Step 1c

(*v*) Y -> I

4: Step 2

(m>0) ATIONAL -> ATE

(m>0) TIONAL -> TION

5: Step 3

(m>0) ICATE -> IC

- 6: Step 4
(m>1) AL ->
7: Step 5a
(m>1) E ->
(m=1 and not *o) E->
8: Step 5b
(m>1 and *d and *L) ->single letter

9. Conclusion

To improve the search quality, it has been proposed by us an algorithm for which we have replaced the present searching methods by our procedure Explore-Scheme, which greatly increases the utility of Web, over what is available today. None of the Search Techniques is able to deal effectively and efficiently with the huge volume of information posted on World Wide Web. Currently, we are implementing the system that is able to demonstrate important properties of the presented approach. It is our observation that in investigating relevance feedback questions (both existing and new), Explore-Scheme algorithm will be executed upon Web as well and hence improve every one's capability to search good, useful, as well relevance information. Our proposed system emphasizes on enhancing the Searching methodology by procedure Explore-Scheme, which provides solution to the complex queries based on contextual expansion. It reduces the word to its atomic part thus curtailing the size of the dictionary by saving storage space and processing time enhancing the processing speed of query. We can finally say that our proposed algorithms will definitely improve the information retrieval capabilities and are able to deal effectively and efficiently with the huge volume of information posted on World Wide Web. A simply searching methodology algorithm can perform as well as more sophisticated ones and our proposed method is robust and remarkably good in determining the poor systems. Our approach is effective and outperforms the already retrieved results.

References

1. C.J.C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, G. Hullender, "Learning to Rank using Gradient Descent", in Proceedings of the International Conference on Machine Learning, 2005.
2. D. Kelly and J. Teevan, "Implicit feedback for inferring user preference: A bibliography", In SIGIR Forum, 2003.
3. Goyal, M., Shape, size and phonon scattering effect on the thermal conductivity of nanostructures. *Pramana*, 91(6): p. 87.2018.
4. Goyal, M. and B. Gupta, Study of shape, size and temperature-dependent elastic properties of nanomaterials. *Modern Physics Letters B*, 33(26): p. 1950310. 2019.
5. D. Oard and J. Kim, "Modeling information content using observable behavior", In Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology, 2001.
6. David Hawking, Nick Craswell, Francis Crimmins and Trystan Upstill, "How valuable is external link evidence when searching enterprise webs?", In Proc. ADC, 2004.
7. Goyal, M. and B. Gupta, Analysis of shape, size and structure dependent thermodynamic properties of nanowires. *High Temperatures--High Pressures*, 48.2019.
8. PK Singh, K Sharma, Mechanical and Viscoelastic Properties of In-situ Amine Functionalized Multiple Layer Graphene/epoxy Nanocomposites, *Current Nanoscience* 14 (3), 252-262
9. Singh PK, & Sharma K, Molecular Dynamics Simulation of Glass Transition Behaviour of Polymer based Nanocomposites, *Journal of Scientific & Industrial Research*, 77 (10) 592-595. (2018).
10. Goyal, M. and M. Singh, Size and shape dependence of optical properties of nanostructures. *Applied Physics A*, 126(3): p. 1-8.2020.
11. David Hawking, Trystan Upstill and Nick Craswell, "Towards better weighting of anchors (poster)", In Proc. SIGIR, 2004.
12. E. Agichtein, E. Brill, S. Dumais, and R. Ragno, "Learning User Interaction Models for Predicting Web Search Result Preferences", In Proceedings of the ACM Conference on Research and Development on Information Retrieval (SIGIR), 2006.
13. A Kumar, K Sharma, AR Dixit A review of the mechanical and thermal properties of graphene and its hybrid polymer nanocomposites for structural applications, *Journal of materials science* 54 (8), 5992-6026.

14. K Sharma, M Shukla, Three-phase carbon fiber amine functionalized carbon nanotubes epoxy composite: processing, characterisation, and multiscale modeling, *Journal of Nanomaterials* 2014
15. A Yadav, A Kumar, PK Singh, K Sharma, Glass transition temperature of functionalized graphene epoxy composites using molecular dynamics simulation, *Integrated Ferroelectrics* 186(1), 106-114
16. PK Singh, K Sharma, A Kumar, M Shukla, Effects of functionalization on the mechanical properties of multiwalled carbon nanotubes: A molecular dynamics approach, *Journal of Composite Materials* 51(5), 671-680
17. A Kumar, K Sharma, AR Dixit, Carbon nanotube-and graphene-reinforced multiphase polymeric composites: review on their properties and applications, *Journal of Materials Science*, 1-43
18. MK Shukla, K Sharma, Effect of carbon nanofillers on the mechanical and interfacial properties of epoxy based nanocomposites: A review, *Polymer Science, Series A* 61(4), 439-460
19. A Kumar, K Sharma, AR Dixit, A review on the mechanical and thermal properties of graphene and graphene-based polymer nanocomposites: understanding of modelling and MD simulation, *Molecular Simulation* 46(2), 136-154
20. K Mausam, K Sharma, G Bharadwaj, RP Singh, Multi-objective optimization design of die-sinking electric discharge machine (EDM) machining parameter for CNT-reinforced carbon fibre nanocomposite using grey relational analysis, *Journal of the Brazilian Society of Mechanical Sciences and Engineering* 41 ...
21. J. Allan. HARD Track Overview in TREC 2003, "High Accuracy Retrieval from Documents", 2003.
22. K Sharma, KS Kaushalyayan, M Shukla, Pull-out simulations of interfacial properties of amine functionalized multi-walled carbon nanotube epoxy composites, *Computational Materials Science* 99, 232-241
23. Jansen, B. J., Spink, A., & Saracevic, T., "Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web", *Information Processing and Management*, 36(2), 207-227, 2000.
24. Kim M., Raghavan V., "Adaptive concept-based Retrieval Using a Neural Network", In *Proceedings of ACM SIGIK Workshop on Mathematical/Formal Methods in Information Retrieval*, Athens, Greece, 2000.
25. MK Shukla, K Sharma, Improvement in mechanical and thermal properties of epoxy hybrid composites by functionalized graphene and carbon-nanotubes, *Materials Research Express* 6(12), 125-323
26. K Kumar, K Sharma, S Verma, N Upadhyay, Experimental Investigation of Graphene-Paraffin Wax Nanocomposites for Thermal Energy Storage, *Materials Today: Proceedings* 18, 5158-5163
27. S. Fox, K. Karnawat, M. Mydland, S. T. Dumais and T. White, "Evaluating implicit measures to improve the search experience", In *ACM Transactions on Information Systems*, 2005.
28. Kumar, Manoj, and Ashish Sharma. "Mining of data stream using "DDenStream" clustering algorithm." 2013 IEEE International Conference in MOOC, Innovation and Technology in Education (MITE). IEEE, 2013.
29. Sharma, Ashish, Ashish Sharma, and Anand Singh Jalal. "Distance-based facility location problem for fuzzy demand with simultaneous opening of two facilities." *International Journal of Computing Science and Mathematics* 9.6 590-601 (2018).
30. Ram, Anant, et al. "A density based algorithm for discovering density varied clusters in large spatial databases." *International Journal of Computer Applications* 3.6 1-4 (2010).
31. Kulshrestha, Jagrati, and Anant Ram. "An Analytical Study of the Chain Based Data Collection Approaches." 2019 4th International Conference on Information Systems and Computer Networks (ISCON). IEEE, 2019.
32. Agarwal, Rohit, A. S. Jalal, and K. V. Arya. "A review on presentation attack detection system for fake fingerprint." *Modern Physics Letters B* 34.05 (2020): 2030001.