International Journal of Engineering Science Invention
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org Volume 2 Issue 10‖ October 2013 ‖ PP.56-61

An Expressive Multiple Query Processing For The Patented Medical Database in Handling the Temporal Domain Event

[1] T.Yogameera, [2] Dr.D.Shanthi Saravanan

[1] *PG Student,* [2] *Professor,*
[1,2,]*Department of Computer Science and Engineering, P.S.N.A College of Engineering and Technology, Tamilnadu.*

ABSTRACT: *The Intellectual query processing has become mandatory for efficient information retrieval .The traditional approaches such as Try and See approach, Prior-Art- Search, As You Type approach, Fuzzy Approach, Filtering algorithms, Graph Methods were not sufficiently proven worth in the upcoming temporal events when implied in a context of data mart or an enterprise database. This paper suggest a innovative methodology with three streamlined techniques namely Automatic Error Correction, Topic Relevant Query Suggestion with extended Query Augmentation to enhance the functionality of patented data search in high dimensional databases. The patented data from the sources are first clustered into topics and classes, when given a query the highly coherent cluster partitions are recovered. The upshots in each coherent cluster are combined generating top K relevant answers for the examiners from the database. After going through a detailed study about the different literatures on search and retrieval of information, it is decided to propose a new novel approach that amplifies the user's intention contour and enhances the retrieval time with more efficient memory management of the database. Further this technique can be implemented in patented medical databases which would give better results with an economical Query processing in accurate and proficient electronic data systems.*

KEY WORDS*: Automatic Error Correction, Query Augmentations, Query Analysis, Referential Medical database, Patented Structure.*

I. INTRODUCTION

The Patented Medical Databases have now been used for referential report generations with detailed and analyzed metadata structure. The Research is underway to implement the referential analysis with the automated machines and with human robots, so that the process can be with accurate analyzation and speed up the further treatment after observations recorded. In accurate syndrome cases and other immediately diagnostic needed cases like echo cardio problem, cancer etc. We are in need of proper less time consuming appraised information retrieval tool to be designed. Thus relating this module to the domain of data mining the query processing becomes the core of attention needed for automated/expert referential activity. The common approaches like As You Type, Prior Art Relevancy, Graph Method, Type Ahead Search, Click through Data, SVM Ranking have not proven worth efficient in this era of medical treatments. Seeking the problem, the proposed 3 proven streamlined techniques Automatic Error correction, Topic Relevant Query suggestion with extended query augmentation helps in précised query contour and quick retrieval of referential data from the patented medical databases in the hospital data mart which holds big data of a medical forum. More over when implemented in cloud the multiple queries processing with the foreign enhanced metadata analysis provides the best way in making proper decisions for the expert doctors.

1.https://www.google.com/patents.

2.www.bmj,com/content..
3.www.nationalarchieves.gov.uk/informati on-management.

4. www.annauniv.edu/ipr.

II. OVER VIEW ON RELATED TECHNIQUES:

2.1 Query search techniques: Click through data [1]: It finds the subset of the surveying data, the Boolean operators used for scaling has only three criteria (Disagree, Neutral, and Agree). The technique consumes much time and effort due to lack of understanding of functionality of search context. In As You Type[9]

approach, letter by letter query suggestion with topic relevancy is provided, the user gratification found by the trie structure currently by suffix part of keyword explores a huge search space thus it need to be dealt with, moreover the error correction becomes a trial approach only. SVM ranking [12] of top k answers, concerned with all context of information retrieval performed with the citation edges in the graph, word net when easy to maintain in the homogeneous database. The performance is not sensitive to heterogeneous database with tunable parameter (α). More over the complex conceptual indexing based on large scale database and backend algorithms with AND/OR semantics need to be concentrated. Prior-Art-Search [11] Presence of mismatch and vague terms was found by the pseudo relevance feedback and automatically select better match, but there is a need for enhancement of this mile stone approach with extended query augmentation in statistical distribution which now deals only with less skewed retrieval.

2.2 Query processing techniques: Pattern Matching [1] NFA computation for dealing with temporal events must concentrate on shared buffer and database with current version states and points to recent events, for future edge evaluation with both logical and temporal decision making. Regular expression matching [3] the queries are converted into regular expressions with NFA binary logic, The Field Programmable Gate Array (FPGA) extended to self reconfigurable GA for the configuration bit generation reduces the number of state traversals thus speed up the row and column traversal and search operations.

2.3 Information Retrieval strategies for the given Processed Query: Backtracking algorithm [10] the processing of the query using selection- join-aggregation was enhanced with run time efficiency with this algorithm, It is an apt logical programming algorithm for the constrain based satisfaction. It finds all the possible solutions within the time bound as search space has been pruned with the invalid branch optimization. Trie Structure Analysis [2] the current retrieval based on prefix part in the sub-linear search algorithm need to reconsider in the temporal and structured high dimensional data marts. The length matching and loop processing when taken into account may result in fast pruning of search space that needs to be dealt with inverted indexing Multiple Query Optimizations[4] (MQO) The spanning of multiple events with parameterized scalability needs unambiguous indexing. In addition the query rewriting must be performed to speed up subscriptions and publish the notifications of the related events after finding the filtered commonalities and merging them for the efficient retrieval.

III. RELATED WORKS AND DISCUSSION ON FUTURE WORKS

Larkey[11] has studied the problem of patent classification but neglected the Prior Art search which is present milestone. X.Xue & W.B Croft [5] discusses about the query generation in the patent for finding the referential answer. Our problem constrained here with the relevancy of the retrieved results. Azzorpadi[6]surveyed 8 patents(including Medical DB) to obtain preference and functionality oriented finding for the given query with two approaches prior art search & sophisticated method of content analysis proving better retrieval of results with citation analysis with indexed SVM ranking[12]. Yan Cao, Jufan & Gu.Li[7] has discussed about the automatic error correction approach with the partial or full keyword relevancies and retrieving top K answers from the partitions Our proposed system differs from existing system when implemented in the medical data analysis keeping all the advantages of the above approaches with enhanced user friendly and expressive processing of high dimensional data analysis in temporal events.

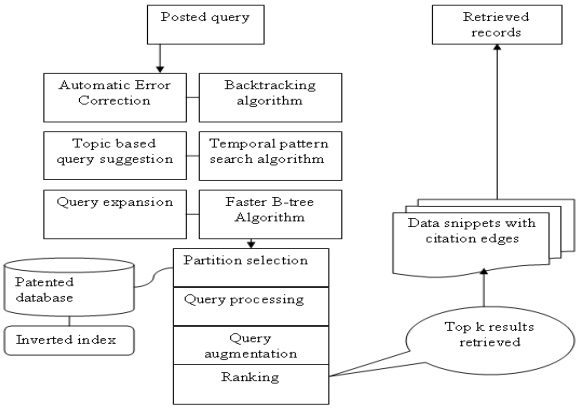


FIG 1: A low level architecture, describes the information management system in the hospital enterprise with efficient query processing.

The information management structure in the medical database used in appraised query processing has been modeled. The posted query by the expert will first be verified and automatic error correction is performed to enter into correct consolidated record links then the topic based suggestion such as echo, cancer, liver disorder etc based on the temporal events updated is provided. Further the query is expanded and rewritten by the system which uses inverted index to capture the upshots. The top k answers are retrieved based on the citation edges and query keyword, and visualized to the expert.The existing technologies discussed in the related works need to be further concentrated with the temporal high dimensional database events. The clustering and classification technique is the perfect management strategy to handle the upcoming Big data that are topically related. The query when proposed to the system must be deeply processed with both topic relevancy and query key relevancy with automated error correction techniques made possible.Though the recently available trends mines the query results based on the content, it is indeed necessary to step into the smart mining with improved vision on intelligent information retrieval from the engineered database. Thus the points to concentrate on include the retrieval of query results that are highly relevant are the following

- The relevant results within accurate user's contour.
- The speedup of appraisal in the big data analyzation
- Clouding the structured(or) patented databases that are clustered based on the topics and related based on their inner class partitions
- Finding of positive partition classes for the given query.
- Data visualization in ranked order with pattern matching made more efficient.

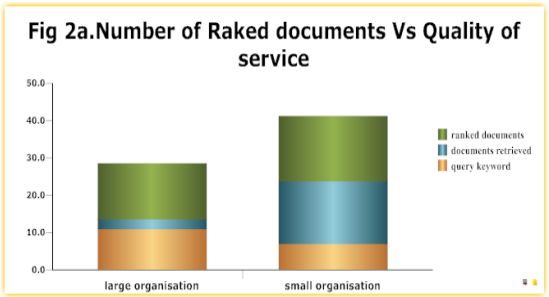
Our paper concentrates in these points. The qualified technique Automatic Error correction now not mostly available in medical database is quintessential for accurate result retrieval. The Backtracking algorithm is used for a quick test whether the partial solution can be a valid solution or not. The recursive depth first search strategy in our tree helps in pruning the irrelevant search space and in determining whether the branch is valid or not thus ongoing with nano unit time operations. Topic based query suggestion can be achieved by our proposed temporal pattern search algorithm that arrays the events of same type with the time stamp value, this is useful to skip unnecessary histories and encapsulate the users recall view of records. The recently enhanced Faster B-Tree algorithm is used to make a sorted data storage and perform the updating and retrieval of the temporal events in logarithmic time, thus helps in query expansion and suggestion that makes a friendly interface for the user.

IV. EXPERIMENTAL RESULTS:

We have implemented our proposed techniques. We compared with the prior art technique and SVM ranking in the retrieval of information in the simulated medical database. The obtained results were satisfying in the advanced information retrieval in respect of relevancy, quality and ranking within the bounded time factor.

4.1 Relevancy of documents:

We evaluate the effect on k (the number of selected partitions). We partitioned the records into topics and classes and evaluated the effectiveness and quality by varying the value of k. To evaluate the result quality, we used the milestone technique p@k, where the precision is the ratio of the number of retrieved relevant results to the number of retrieved results, and p@k is the precision of the top-k results.



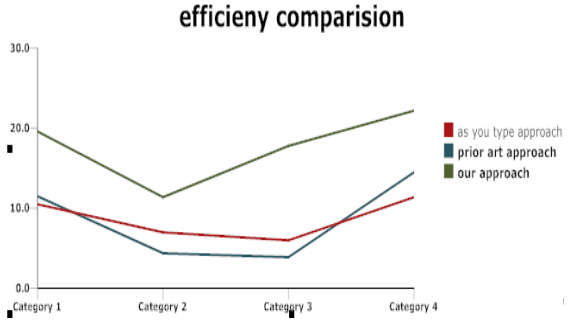


Fig2b.Comparision with different approaches (with p@k)

Fig. 2a shows the results. The main reason is as follows: First, the more records used to answers a query, the more relevant answers, and thus the higher precision. Second, as each query usually belongs to limited number of topics, only several partitions are relevant to the query. Thus, the precision is stable when 'k' is large enough. For example, where k '>' 10, they achieved nearly the same precision. Then, we evaluated the efficiency. Fig. 2b shows the results. We see that with the increase of k, the elapsed time increased. This is because the more the categories used to answer a query the efficiency increases.

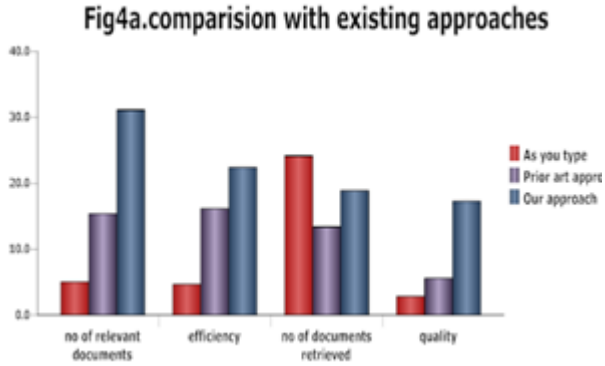
4.2 Precision Comparison: In this section, we compare the result quality. Table 1 shows the experimental results. We can see that our three techniques can improve the result quality. For example, for p@50, error correction improved to 0.83, query suggestion increases the precision to 0.82, and query expansion can improve the precision to 0.88. More importantly, our method by combing the three methods can improve the precision to 0.88, and the improvement ratio is about 0.84, achieved by query rewriting. The main reason is as follows: First, the automatic error correction can provide users accurate keywords based on users inputs. Second, the query expansion can suggest relevant keywords. Third, topic-based query suggestion can provide users topic- relevant keywords.

Approaches compared	p@ k	Query Suggestion	Query Expansion	Automatic Error Correction	Overall efficiency	Ranking of techniques
As you Type	10	0.77	0.66	0.62	0.6833	3
Prior art search	15	0.79	0.70	0.78	0.7566	2
Our approach	25	0.82	0.88	0.83	0.8433	1

Table 1: Quality comparison

4.3 Efficiency Comparison : In this section, we compare the efficiency. We partitioned the data into 24 partitions. We used three computers to manage the data and each node managed three partitions. For each partition, we built the corresponding inverted indexes. We first compared the two methods by varying the number of keywords. We can see that for different numbers of keywords, our method always outperformed the existing method SVMPR, with the speedup ratio about 8. This reflects that our method achieved high efficiency since we employ an effective partition-based method. We then evaluated the two methods by varying the number of returned answers. We can see that for different values, our method always outperformed SVMPR. This is because, we partition the data into eight partitions and each partition was inversely indexed and searched by different cores. More importantly, our partition-based method can prune the search space and thus can improve the performance significantly.

4.4 Scalability : We also evaluate the scalability of our method. Fig. 4a shows the experimental results by varying the number of patents. Fig. 4a shows the scalability on quality and we can see that with the increase of the number of records, the precision of query suggestion and query expansion increased slightly. This is because we can utilize



more data to select the topic of each record and find more relevant keyword pairs. The more data used the higher precision of the topic model. For query correction, the precision nearly kept the same as we only used the trie structure to correct the keywords but the quality increased as we prposed the prefix based depth first search. On the other hand, our method scaled very well.

V. CONCLUSION:

The user friendly expert analysis query processing techniques discussed so far had made milestones in the domain of information retrieval and management. But our proposed suggestion may make a perfect benchmark. The patented data from the sources are first clustered into topics and classes, when given a query the highly coherent cluster partitions are recovered. The upshots in each coherent cluster are combined generating top K relevant answers for the examiners from the database. The ongoing implementation of our method had shown high efficiency and increased quality of the partitioned based search strategy with the simulated results. The techniques streamlined namely query suggestions, Topic relevancy, query augmentation when handshake with existing prior art relevancy may prove marvelous enhancements in knowledge engineering.

VI. Acknowledgement:

It is my pleasure to acknowledge Mr.J.Venkatesan Prabhu, Head, Kaashiv- InfoTech, Chennai for his support in implementation part and for his guidance throughout this course of work.

REFERENCES

[1] J. Agrawal, Y. Diao, D. Gyllstrom, and N. Immerman, "Efficient Pattern Matching over Event Streams," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 147-160, 2008.

[2] R.S. Boyer and J.S. Moore, "A Fast String-Searching Algorithm," Comm. ACM, vol. 20, no. 10, pp. 762-772, 1977. [3] R. Cox, "Regular Expression Matching Can Be Simple and Fast," http://swtch.com/rsc/regexp/regexp1.html, 2007.

[4] A. Demers, J. Gehrke, M. Hong, M. Riedewald, and W. White, "Towards Expressive Publish/Subscribe Systems," Proc. 10th Int'l Conf. Extending Database Technology (EDBT), pp. 627-644, 2006

[5] X. Xue and W.B Croft, "Automatic Query Generation for Patent Search," Proc. ACM Conf. Information and Knowledge Management (CIKM), pp. 2037-2040, 2009.

[6] L. Azzopardi, W. Vanderbauwhede, and H. Joho, "Search System Requirements of Patent Analysts," Proc. 33rd Int'l ACM SIGIR Conf. Researchand Development in Information Retrieval (SIGIR), pp.775-776, 2010.

[7] A User-Friendly Patent Search Paradigm Yang Cao, Ju Fan, and Guoliang Li IEEE Transaction on knowledge and data engineering, vol. 25, no. 6, 2013.

[8] L.S. Larkey, "A Patent Search and Classification System," Proc. Fourth ACM Conf. Digital Libraries, pp. 179-187, 1999.

[9] J. Fan, H. Wu, G. Li, and L. Zhou, "Suggesting Topic-Based Query Terms as You Type," Proc. Int'l Asia Pacific Web Conf. (APWEB), pp. 61-67, 2010.

[10] G. Li, J. Feng, and C. Li, "Supporting Search-As-You-Type Using SQL in Databases," IEEE Trans. Knowledge and Data Eng., vol. 25, no. 2, pp. 461-475, Feb. 2013.

[11] S. Bashir and A. Rauber, "Improving Retrievability of Patents in Prior-Art Search," Proc. European Conf. Information Retrieval (ECIR), pp. 457-470, 2010.

[12] Y. Guo and C.P. Gomes, "Ranking Structured Documents: A Large Margin Based Approach for Patent Prior Art Search," Proc. Int'l Joint Conf. Artificial Intelligence (IJCAI), pp. 1058-1064, 2009.

AUTHOR DETAILS:

Ms.T.Yogameera completed her B.E Computer science and Engineering in 2012 with first class from R.V.S College of Engineering and Technology, Dindigul, TamilNadu and pursuing M.E. Computer Science and Engineering in P.S.N.A College of Engineering and Technology, Dindigul, TamilNadu. She has secured best academician award 2 times during her school days in Shri Maharishi Vidya Mandir(CBSE), Dindigul, TamilNadu, she has grabbed the first best project award in the international science and technology contest during her 9TH, has secured state 2ND rank during her 11TH , School second rank in her 12TH . She has participated in two international symposiums presenting her own paper and won second prize in both, she has attended 3 national level workshops and 1 international workshop, 2 national seminars, and 15 days internship program on mobile operating system and has been acknowledged as the class representative for 4 continuous semesters during her B.E. Her domains of interest are data mining, information retrieval, query processing, database management, currently concentrating on information management in medical systems.

Dr.D.Shanthi Saravanan received her B.E Computer Science and Engineering in 1992 from Thiagaraja College of Engineering, Madurai, TamilNadu and M.E Computer Science and Engineering from Manonmaniam Sundaranar University, TamilNadu and PhD in Soft Computing from Brila Institute of Technology, Ranchi. She is currently working as a Professor in Department of Computer Science and Engineering, PSNA College of Engineering and Technology. She has more than 21 years of Teaching and Programming Experience. She is a member of various professional societies like IEEE, CSTA, IAENG and IACSIT. Her research interest includes Genetic Algorithm, Neural Networks, and Intelligent Systems, Image processing, Embedded System, Machine Learning and Green Computing. She has published more than 25 papers in international journals and conferences and also 5 books in computing and applications. She is a reviewer of various international journals.