Ranking of Complex Data with the Use of Fuzzy Logic Reasoning

[1]Prof. Akshay Agrawal, [2]Snehal T. Thimbare, [3]Yogita D. Gawde, [4]Indu Jayendran
[1]Asst. Professor, [2,3,4]BE Student, [1,2,3,4]Comp. Engg. Dept, SSJCET, Asangaon, India.
[1]akshay16661@gmail.com, [2]snehal0720@gmail.com, [3]yogitagawde20@gmail.com, [4]indujndrn18@gmail.com

Abstract - Page Ranking is one of the best tool for sorting the data thus we have used the paging technique to introduce it with the rising technology of fuzzy logic .In this paper we define a proper solution to the data mining by using fuzzy logic .fuzzy logic is a method of reasoning which resembles with human reasoning as human has only two choices either yes or no. other than these two choices a fuzzy logic can give more options that means certainly yes, possibly yes, cannot say, cannot predict etc. In this paper we study a better approach for finding a proper solution search through two types of data. This gives a idea that how fuzzy logic can be useful in search engine optimization. We study how to unify immediacy information in fuzzy logic search to make our search prediction perfect and more faster. fuzzy search logic works on two kinds of database and its working is similar as compared to paging and indexing the sentences .we get a comparative graph which shows that vast data can give more similar words and phrases so the possibility of better solution to a query can easily found. In this study we show the graphical approach which can give a clear idea how the fuzzy logic works in search engine optimization. we conducted a experimental study on both data bases which shows an interchange between space and time complexity.

Keywords — Fuzzy logic reasoning , search engine optimization, ,fuzzy data management, unified data, immediacy ranking.

I. INTRODUCTION

Data mining- It is a staple process which is useful in discovering interesting knowledge from a vast data base. Data mining is a term coined to describe the process of sifting through large and complex databases for identifying valid, novel, useful, and understandable patterns and relationships. Data mining involves the inferring of algorithms that explore the data, develop the model and discover previously unknown patterns. The model is used for understanding phenomena from the data, analysis and prediction. The accessibility and abundance of data today makes knowledge discovery and data mining a matter of considerable importance and necessity.

Fuzzy logic is an emerging technique which resembles to the probability of getting more answers to a query. Fuzzy logic has two different meanings. In a narrow sense, fuzzy logic is a logical system, which is an extension of multi valued logic. However, in a wider sense fuzzy logic (FL) is almost synonymous with the theory of fuzzy sets, a theory which relates to classes of objects with un sharp boundaries in which membership is a matter of degree. In this perspective, fuzzy logic in its narrow sense is a branch of FL. Even in its more narrow definition, fuzzy logic differs both in concept and substance from traditional multi valued logical systems. fuzzy logic includes more options to user for taking more closer sequential answers. Binary logic is either 1 or 0. Fuzzy logic is a continuum of values between 0 and 1. This may also be thought of as 0% to 100%. An example is the variable OLD PERSON. We may say that

age 75 is 100% OLD, 50 is 50% OLD, and 30 is 0% OLD. In the binary world everything below 18 would be 100% OLD, and everything above would be 0% OLD.

In this paper we study the following problem : *How to enhance the search of a query to get a efficient and highly true in probability from a large data base?* The problem observed in finding a similar or a synonymous phrase from a database is can be solved by using the immediacy technique with the use of fuzzy logic algorithm. fuzzy variables captures the measurement uncertainty ,as a part of experimental data we can get a crisp output from a complex data base. The main objective of this study is to determine the use of fuzzy logics in optimization of searching of a query in different databases. The comparative study of space and time complexity for both the database. a graphical analysis has been done which shows how better the system works on use of fuzzy logic.

II. LITERATURE SURVEY

Despite the fact that the bulk of the information we assimilate every day is fuzzy, most of the actions or decisions implemented by human or mechanical machines to give crisp or binary output. Following are the reference studies used while implementing this study.

a) Most of the existing approaches focuses on Instant search and page ranking principles as in case done by Cetindil, J. Esmaelnezhad, C. Li, and D. Newman [1] the problem of analyzing the query log of an instant-search system. They propose a classification scheme for user typing behaviors. The identified typing behaviors to estimate the success rate of such a system in the absence of click-through data and also compare the log of an instant-search system and that of a traditional search system on the same data. The results show that on a people directory search system, instant search can typically save 2 seconds per search, reduce the typing effort by showing the results with fewer characters entered, and increase the success rate.

b) The top-k query processing algorithm also includes the various suggestions for getting an optimal solution to a query search but the disadvantage is of to k query is the time and space required for this execution is more. As in case done by G. Li, J. Wang, C. Li, and J. Feng [4] They study how to answer top-k queries in this paradigm, i.e., as a user types in a query letter by letter, and how to efficiently find the k best answers. Instead of inventing completely new algorithms from scratch, this study challenges when adopting existing top-k algorithms in the literature that heavily rely on two basic list-access methods random access and sorted access. They represent two algorithms to support random access efficiently. The novel techniques to support efficient sorted access using list pruning and materialization. The extending techniques to support fuzzy type-ahead search which allows minor errors between query keywords and answers. The report of experimental results on several real large data sets to show that the proposed techniques can answer top-k queries efficiently in type-ahead search.

c) The study done by the author [6] shows detail study and drawbacks of QAC and shows the results of query completion in less time complexity. Query Auto-Completion (QAC) is a popular feature of web search engines that aims to assist users to formulate queries faster and avoid spelling mistakes by presenting them with possible completions as soon as they start typing. They present the first large-scale study of user interactions with auto-completion based on query logs of Bing, a commercial search engine. The results confirm that lower-ranked auto-completion suggestions receive substantially lower engagement than those ranked higher. observation that users are most likely to engage with auto-completion after typing about half of the query, and in particular at word boundaries. Using auto-completion varies with the distance of query characters on the keyboard.

d) Since the popularity of business process outsourcing is increasing, vendor selection has become a widely studied topic in the literature of supply chain procurement . Because of its inherent complexity, supplier selection is considered as a multi-criterion problem which has both qualitative as well as quantitative factors. Performance, technical capability, financial stability, and quality of the supplier are identified as main criteria in supplier selection.

e) The study done by author [5] shows the information searching using proximity ranking and the fuzzy methods Difficulty in finding relevant information in unstructured text documents. The methods of fuzzy search, instant search and proximity ranking and how they can be used in the process of annotation of documents. These various methods can be integrated to give better search results and to achieve efficient space and time complexities.

f) Proposed a novel alternative approach which facilitates the generation of the structured metadata automatically using OpenNLP, methods of Instant-fuzzy search and Proximity ranking. It is done by identifying documents which are likely to contain the information of interest. And this information will be subsequently useful for querying the database.

g) Fuzzy set theory has been used to model systems that are hard to be defined precisely. It incorporates imprecision and subjectivity of human decision making into the model formulation and solution process. The fuzzy logic and results of the fuzzy approach are much better than traditional approach due to its ability to capture difference and ambiguity in view of linguistic variables by different evaluators. This is especially important in assessing vendor risks because assessors' risk attitude may be quite different from each other.

III. AIM AND OBJECTIVE

The aim is to get efficient space and time complexity by using fuzzy logic from two different data pages with different size of memory which shows a graphical representation of both database which requires less time (in ms). Fuzzy logic provides varying solution to construct all possible probability of getting a precise output for a user with better peculiarity.

Objective -

- More accurate the data more accurate the result.
- Automation using data mining is most widely used
- As the ordinary language which uses symbolic conventions and not uses any logical contents.
- Fuzzy logic provides more probable solution to query.
- Each database is collective heavier than other which contain less similar or more probable synonym words which provides more relevant answers to a user. Data should be more clearly visible to user.
-Fuzzy logic provides more classified outcomes from heavy database.
-Time complexity will be compared for both the database of a single query.
- The lesser and more relevant answer can get.

IV. PROPOSED SYSTEM

Defining the powerful database for creating a relevant solution to the input query. Both the database are to be clearly different from each other. The relativity from both database should be vary from each other. The fuzzy algorithm then applied to both database. Testing of buffer, for both database. Analysis of query result shows that data is bigger and solution n time required is less. Also chances of Getting best relevant answer is more.

1. Query input

a query is a request for an information. A *query* is an inquiry into the database using the `SELECT` statement. A query is used to extract data from the database in a

readable format according to the user's request. example In our database there are two types of database of different size. In which if we have any information about a disease like BOOK then we can insert a keyword P. L. DESHPANDEY BOOKS then algorithm shows the related search . The inserted output then can called as a query.

2. Database information

Database is an information bank in which any kind of data can be inserted so that any system can work efficiently. In this case we are using two types of database. The information of both DB is given below.

DATABASE 1:

In database 1 we have taken two kind of information that of books and cars which have the attributes like information about the author , publication , content ,pages etc. and the attributes of cars are car name ,model no. launching time and date ,brand etc. this DB is of 512MB and this is applied for GENERAL SEARCH.

DATABASE 2:

In this DB we have the information similar to DB1 but it contains more information than DB1. so the time required for the query search through this DB is more for the GENERAL SEARCH. We applied our main FUZZY LOGIC for this DB2. and the result obtained from Fuzzy logic is given in result. The graphical analysis is also shown so that comparison is done easily.

Above steps can be shown as follows:

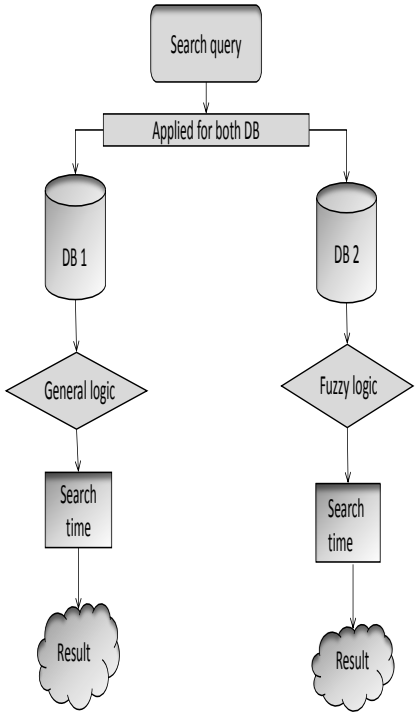

Fig 1: System Architecture

V. ALGORITHM AND IMPLEMENTATION

Query Detecting Validation Algorithm:

Algorithm 1[q(i),V]
Input : query q(i) = h<w1,w2, . . . ,wli>
// where wi is a last keyword;
//w is the inserted query;
Output: a valid-phrase vector V ;
initials
step-1
x<- number of keywords found in db1;
y<- number of keywords found in db2;
(qc, x) ← FindLongestCachedPresegmented(*db1, C*)
(qc, y) ← FindLongestCachedPresegmented(*db2, C*)
// for both db similar input query is applied;
m ← number of keywords in qc {input query}

step -2
if m > 0 **then** // m number of related keys found
for i ← 1 **to** m − 1 **do** // Copy the valid valid-phrase vector
V [i] ← Vc[i]
// phrase validation testing;
if wm == qc[m] **then** // The last keyword of qc is a complete keyword in q

V [m] ← Vc[m]
else // Incremental computation for the last keyword retrieved from cache

step-3
V [m] ← ? //related query is not found;
foreach (start,S) *in* Vc[m] **do**
//repeat above steps 1 and 2;
newS ← compute active nodes for wm

incrementation process-
incrementally from S
if newS == ? **then**
V [m] ← V [m] ∪ (start, newS)
foreach (start,S) *in* V [m] **do**
// Incremental computation for the phrases partially cached
for j ← m + 1 **to** l **do**
newS ← compute active nodes
from S by appending wj
if newS == ? **then** break
V [j] ← V [j] ∪ (start, newS)
S ← newS
// new related key found;
for i ← m + 1 **to** l **do** // Computation of non-cached phrases
S ← compute active nodes for wi
V [i] ← V [i] ∪ (i,S)
for (j ← i + 1)
to do
newS ← compute active nodes
//show active nodes;
from S by appending wj
if newS == ? **then** break
V [j] ← V [j] ∪ (i, newS)
S ← newS
cache (q, V) in C
return V // return to base query which matches to related input;

VI. ADVANTAGES OF FUZZY LOGIC

1) Fuzzy logic is a great solution to complex problem.

2) It gives more processing ways to any kind of logic related data.

3) Simple to use, any logical experiment is done by using fuzzy logic.

4) Fuzzification, defuzzification techniques makes it simple to give a high probable output.

5) It is easy to understand.

VII. RESULT

The above table gives a example result which is done and obtained result is shown for the database.

Sr	Website name	Key	Meta key	Imk	Result (GB)	
					DB1	DB2
1.	computersofts.c	Yes	203	4	19	15
2.	javea-computer suppliers	Yes	106	9	19	10
3.	Mercedes newcomers	Yes	124	8	23	16
4.	pzcomputers.com	Yes	108	10	26	20
5.	Pl deshpandey	Yes	200	4	14	09

Table 1: queries search and query found in(ms)

VIII. CONCLUSION

It has been observed that as data mining is Finally, to achieve the final ranking results, for applying the multi-criteria models .we obtain the relevant less space and time complexity to achieve the comparison from both database and we get the efficient space and time complexity. The main objective of this study is to determine the use of fuzzy logics in optimisation of searching of a query in different databases. The comparative study of space and time complexity for both the database. a graphical analysis has been done which shows how better the system works on use of fuzzy logic.

REFERENCES

[1] I. Cetindil, J. Esmaelnezhad, C. Li, and D. Newman, "Analysis of instant search query logs," in WebDB, 2012, pp. 7–12.

[2] R. B. Miller, "Response time in man-computer conversational transactions," in Proceedings of the December 9-11, 1968, fall joint computer conference, part I, ser. AFIPS '68 (Fall, part I). NewYork, NY, USA: ACM, 1968, pp. 267–277. [Online]. Available: http://doi.acm.org/10.1145/1476589.1476628.

[3] C. Silverstein, M. R. Henzinger, H. Marais, and M. Moricz, "Analysis of a very large web search engine query log," SIGIR Forum, vol. 33,no. 1, pp. 6–12, 1999.

[4] G. Li, J. Wang, C. Li, and J. Feng, "Supporting efficient top-k queries in type-ahead search," in SIGIR, 2012, pp. 355–364.

[5] R. Schenkel, A. Broschart, S. won Hwang, M. Theobald, and G. Weikum, "Efficient text proximity search," in SPIRE, 2007, pp. 287–299.

[6] Bhaskar Mitra, Milad Shokouhi, Filip Radlinski, Katja Hofmann,"On User Interactions with Query AutoCompletion," Microsoft Cambridge, UK.

[7] M. Zhu, S. Shi, N. Yu, and J.-R. Wen, "Can phrase indexing help to process non-phrase queries?" in CIKM, 2008, pp. 679–688.

[8] A. Jain and M. Pennacchiotti, "Open entity extraction from web search query logs," in COLING, 2010, pp. 510–518.

[9] Joachims, T. (2002) 'Optimizing search engines using clickthrough data', Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. Edmonton, Alberta, Canada

[10] Frydenberg, M., Miko, J. (2011) 'Taking it to the Top: A Lesson in Search Engine Optimization', Information Systems Education Journal, 9(1), pp. 24-40. ISEJ [Online]. Available at: http://isedj.org/2011-9.

[11] Radhika Goel, Anjali Sardana, and Ramesh C. Joshi ―Parallel Misuse andAnomaly Detection Model‖ International Journal of Network Security, Vol.14, No.4, PP.211-222, July 2012

[12] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. GhorbaniA Detailed Analysis of the KDD CUP 99 Data Set.

[13 Thales Sehn Korting,‖ C4.5 algorithm and Multivariate Decision Trees‖

[14] Mrutyunjaya Panda1 and Manas Ranjan Patra2, IJCSNS International Journal of Computer Science and Network Security,VOL.7 No.12, December 2007‖ Network Intrusion Detection Using Naïve Bayes‖

[15] Shaik Akbar, Dr.K.Nageswara Rao, Dr.J.A.Chandulal, International Journal of Computer Applications (0975 –8887) Intrusion Detection System Methodologies Based on Data Analysis

[16] Nathan Einwechter, ―An Introduction To Distributed Intrusion Detection Systems.

