Proposed Generic Full Text Searching Algorithm: A Database Approach

Zain Ul Hassan [#1], Muhammad Naeem [#2], Muhammad Khalid [#3]
Department of Computer Science, University of Agriculture
Faisalabad, Punjab, Pakistan

Abstract**—***This paper shows the different full text searching algorithms techniques. Apply in different databases with different searching techniques. This paper reviews different implementation techniques and proposed a generic full text searching algorithm. This generic algorithm will implement in any data base for full text searching. This paper provides a simple and easier full text searching algorithm.*

Keywords — *Full text searching, databases, algorithms*

I. INTRODUCTION

Database systems are commonly uses in database applications. Database management system is serving as back end for traditional database applications, large websites and web services. A database is a collection of data. Database is design, builds and populated with data for a specific propose. Conceptual model provides a concept for database. Defining a database consists of sending queries and retrieves data from database. Databases use Data manipulation language and data definition language and data control language statements. Texts have a structure. The text can be logically divided into document and each document into fields. The text may have a complex hierarchical or graph structure .full text search lets users and application and full text queries. Again character based data. Before full text search runs on data the programmer creates a full text index on the table queries perform linguists search against text data in full text such as English (US) or English (UK). Full text search is a search for documents, which satisfy query and optionally return them in same order. Full text search operate with collection of documents where documents is considered as a "bag of words". Text search operators in database existed for years. There is no linguistic support. The FTS idea came from the idea to pre-process document. Pre-processing includes

1. Parsing document to lexemes.
2. Applying linguistics rules.
3. Store preprocessing documents.

This paper consists of four section first one is material and methods, second is result, third is discussion and fourth is conclusion.

II. MATERIAL & METHOD

A full text searching technique is used to search a document in computer. Full text search is based on Metadata searching. So it is different from simple searching techniques. Full text search is consisting of the following function [1] given below:

1. Natural language Full Text Search
2. Boolean Full Text Search
3. Full Text Search with Query Expansions
4. Full Text Stop word
5. Full text Restrictions
6. Fine Turning Full Text search
7. Adding a Collation for full text indexing

A fast string searching algorithm is define in [6].which uses string and delta 1 and delta 2 and match string length. A simple fast hybrid pattern matching algorithm defined in [7]. This is based on Horspool algorithm, Sunday's algorithms, reverse Colussi and Turbo BM algorithm. Most algorithms are string matching algorithms and do not supports full text searching or matching.

Most databases support full text searching but in different ways. We proposed algorithm that is generic and implement in every database.

A-FULL TEXT SEARCH MySql

My Sql Support's full text searching. MySql use two functions MATCH and AGAINST for full text Searching [2].

B-FULL TEXT SEARCH Sql Server

In sql server to run a full text search query the database administrator first create full text index on table containing one or more characters [3]. Full text search consists of user tables, full text generator, Thesaurus files, stop list objects, sql server query processor, full text engine, indexer and filter daemon manager [4].

C-FULL TEXT SEARCH USING ORACLE TEXT

In oracle database the oracle text query uses the select statement and uses contains or CATSEARH in the where clause [5]. Oracle text uses CONTEXT and CTXCAT indexes [4, 5].

III.PROPOSED GENERIC ALGORITHM

Form the above literature review we proposed a generic full text searching algorithm the below is a generic full text search.

1. START
2. INITIALIZE→ DB
3. CONNECT→DB
4. LOAD TABLE→DB
5. SEARCH TABLE→DB
6. CREAT INDEX→TABLE
7. CREATE FTS on INDEX
8. SPECIFY INDEX for FTS
9. INPUT→ FT
10. FTS →TABLE
11. OUTPUT →FT
12. EXIT

The fig.1 shows Flowchart diagram of full text searching and show the execution of the full text searching algorithm and flow of the full text searching in a database. This searching is based on the file text or paragraph.

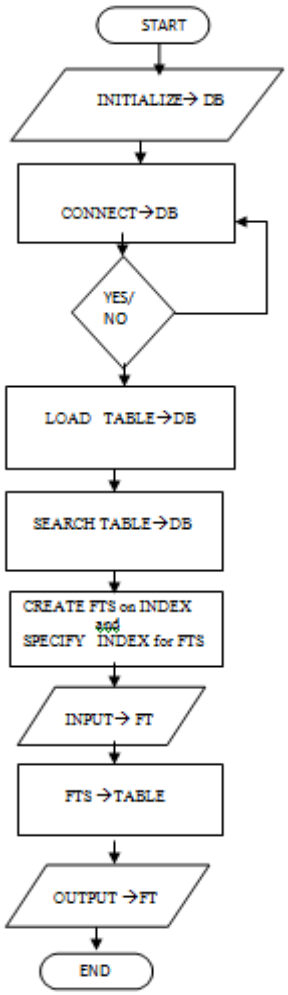

Fig .1 Flow chart Diagram of FTS

IV. RESULT AND DISCUSSION

Form the proposed algorithm first set is start and then second step is initialization of database. The third step is concoction of database this contain a connection string or connection through wizard the fourth step is searching tables in a database. Fifth and six steps is search database table in database and create index of the table. Fifth and sixth steps are interchangeable depend on the database after creating the table Full Text Searching (FTS) created on the index. When full text search created then index is specify for the full text searching input full text for searching in table. The specify index search the given value of string in form of output. The last step is exit.

V. CONCLUSION

The full text search is not a simple search. And simple searching is difficult to apply on metadata. This paper proposed a new and easy search technique that is easily apply on metadata. And it support string searching and paragraph but simple searching algorithm just support limited number of string. The future work is to implement this generic algorithm on database such as mysql, sqlserver.tsql.dbsql and oracle databases.

ACKNOWLEDGMENT

This work is made possible through support provided by Mr. Muhammad Haroon and all my teachers, friends and family.

REFERENCES

[1] "Full Text Search."Internet: https: //dev.mysql.com/doc/refman/5.0/en/fulltext-search.html.[jun. 19,2015].

[2] "Full Text Search."Internet::https://dev.mysql.com/doc/refman/5.6/en /innodb-fulltext-index.html.[Jun.19, 2015].

[3] "Full Text Search." Internet: https://msdn.microsoft.com/en-us/library/ms142571.aspx.[Jun.20, 2015].

[4] "Sql Server .Process."internet:https://msdn.microsoft.com/en-us/library/ms142571.aspx#sqlprocess.[Jun.22, 2015].

[5] "Querying with Oracle Text."Internet:http://docs.oracle.com/cd/ B28359_01/text.111/b28303/query.htm#g1016054[Jun.28, 2015].

[6] R.S.Boyer and J.S.Moore." A Fast String Searching Algorithm".Communications of the ACM.vol.20, pp.762-772.October, 1977.

[7] F.Franek,C.G.Jennings and W.F.Smyth."A Simple Fast Hybrid Pattern-Matching Algorithm". *A. Apostolico, M. Crochemore, and K. Park (Eds.): CPM 2005*, pp. 288–297, 2005

[8] N.Singla and D.Garg. "String Matching Algorithms and their Applicability in various Applications" Internati.onal Journal of Soft Computing and Engineering.Vol.I,pp.218-222,January,2012