isrjournals

International Journal of Advanced Research in Computer Science Engineering and Information Technology

Volume: 4, Issue: 3,Special Issue: 1 ,Mar,2016 ,ISSN_NO: 2321-3337

Efficient Sort Search on Massive Data

Ishwarya.M[1], Kalaiarasi.S[2], Mrs.Revathy[3]

Student, Dept. of Computer Science and Engineering, Agni College of Technology, India.[1, 2]
Asst. Professor, Dept. of Computer Science and Engineering, Agni College of Technology, India. [3]

ABSTRACT- *Efficient top-N retrieval of records from a database has been an active research field for many years. We approach the problem from a real-world application point of view, in which the order of records according to some similarity function on an attribute is not unique. Many records have same values in several attributes and thus their ranking in those attributes is arbitrary (based on random choice). For instance, in large person databases many individuals have the same first name, the same date of birth, or live in the same city. Existing algorithms (Table-scan bassed T2S algorithm) are ill-equipped to handle such cases efficiently. We experimentally show that our method outperforms Dynamic Sorting Algorithm (DSA) for top-k retrieval in those very common cases where we used with dynamically scheduling the resources based on the data which are provided with, this efficient short search algorithm along with the massive data retrieval on a very fine tuple data's can be of a different dataset.Here in this Project we are going to use these logics for the need of solution in the field of medical research. Where there are many manageable databases that are been used in a common path for the end of healthy need and the retrieval of solution for the cause of illness to a human being.*

Keywords:-Massive data, dynamic sort searcher, selection sort searcher, selective retrieval

1. INTRODUCTION

The System Development Lifecycle framework is designed to outline a complete development and implementation process suitable for developing complex applications. SDLC is a process followed for a software project, within a software organization. It consists of a detailed plan describing how to develop, maintain, replace and alter or enhance specific software. We experimentally show that our method outperforms Dynamic Sorting Algorithm (DSA) for top-k retrieval in those very common cases where we used with dynamically scheduling the resources based on the data which are provided with, this efficient short search algorithm along with the massive data retrieval on a very fine tuple data's can be of a different dataset. The life cycle defines a methodology for improving the quality of software and the overall development process.

- Business – legislation regulatory requirements, policy, SOP's, guidelines etc.
- Process – how the business is implemented
- Data – the core business data elements collected for the business
- Application – the gate to the business collecting
- Infrastructure- the servers, network, workstations, etc.

Volume: 4, Issue: 3,Special Issue: 1 ,Mar,2016 ,ISSN_NO: 2321-3337

This project has been deployed based on a data mining algorithm technique and the algorithm used in this project is known as pattern gathering.The user location is found by tracking the IMEI number of his/her mobile.The data set has been mined from the database.A profile will be created in the database for each new user. The admin is able to track or gather the information of a moving object in a particular location by monitoring or getting instructions from the user.This gathering pattern process is done through online so, that the updating process of the moving object will be faster.

2. THE DATA MINING PROCESS MODEL

2.1 Define the business problem

This model used to define the problem of the existing algorithm to overcome in the proposed system.

2.2 Build a data-mining database

This model is used to modify the data from the database.The tasks in building a data mining database are:

a. Data collection

b. Data description

c. Selection

d. Data quality assessment and data cleaning

2.3 Explore the data

The goal is to identify the most important fields in predicting an outcome, and determine which derived values may be useful.

2.4 Prepare data for modeling

This is the final data preparation step before building models. There are four main parts to this step: a. Select variables b. Select rows c. Construct new variables d. Transform variables

2.5 Data mining model building

The most important thing to remember about model building is that it is an iterative process.The process of building predictive models requires a well-defined training and validation protocol in order to insure the most accurate and robust predictions. This kind of protocol is

Volume: 4, Issue: 3,Special Issue: 1 ,Mar,2016 ,ISSN_NO: 2321-3337

sometimes called supervised learning. The essence of supervised learning is to train (estimate) Training and testing the data-mining model requires the data to be split into at least two groups: one for model training (i.e., estimation of the model parameters) and one for model testing.

2.6 Evaluation and interpretation

After building a model, you must evaluate its results and interpret their significance. Remember that the accuracy rate found during testing applies only to the data on which the model was built.

2.7 Deploy the model and results

Once a data mining model is built and validated, it can be used in one of two main ways. The first way is for an analyst to recommend actions based on simply viewing the model and its results. The second way is to apply the model to different data sets.

3. SYSTEM ANALYSIS

3.1 Existing System

The existing application provides query based table scan to retrieve the results.They also use indexes with specific attributes to build the performance on view. It also has bound based fashion which will consist of lower-bound and upper-bound scores.The cost of pre-computing the data structures and update process is a major cost.Database of all hospital cannot be maintained by governmentSearch efficiency is slow.Survey is taken manually at the end of the year and it has human errors.

4. MAIN FEATURES OF DATA MINING

4.1 Automated prediction of trends and behaviors:

Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data — quickly.

4.2 Automated discovery of previously unknown patterns:

Data mining tools sweep through databases and identify previously hidden patterns in one step.

4.3 Databases:

Databases can be larger in both depth and breadth.

4.3.1 More columns:

Analysts must often limit the number of variables they examine when doing hands-on analysis due to time constraints.

4.3.2 More rows:

Larger samples yield lower estimation errors and variance, and allow users to make inferences about small but important segments of a population.

5. TECHNIQUES IN DATA MINING

5.1 Artificial neural networks:

Non-linear predictive models that learn through training and resemble biological neural networks in structure.

5.2 Decision trees:

Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

5.3 Genetic algorithms:

Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.

5.4 Nearest neighbor method:

A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k ³ 1). Sometimes called the k-nearest neighbor technique.

5.4 Rule induction:

The extraction of useful if-then rules from data based on statistical significance.

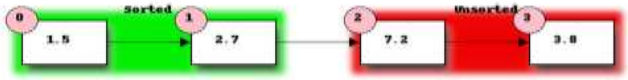
6. SELECTION SORT SEARCH ALGORITHM IMPLEMENTATION

Divide the list to be sorted into a sorted portion at the front (initially empty) and an unsorted portion at the end (initially the whole list).Find the smallest element in the unsorted list:

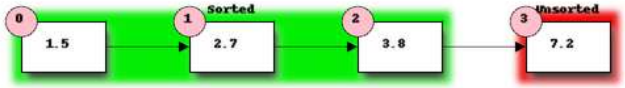
Volume: 4, Issue: 3,Special Issue: 1 ,Mar,2016 ,ISSN_NO: 2321-3337

1. Select the first element of the unsorted list as the initial candidate.

2. Compare the candidate to each element of the unsorted list in turn, replacing the candidate with the current element if the current element is smaller.



3. Once the end of the unsorted list is reached, the candidate is the smallest element.



4 Swap the smallest element found in the previous step with the first element in the unsorted list, thus extending the sorted list by one element.

5 Repeat the steps 2 and 3 above until only one element remains in the unsorted list.

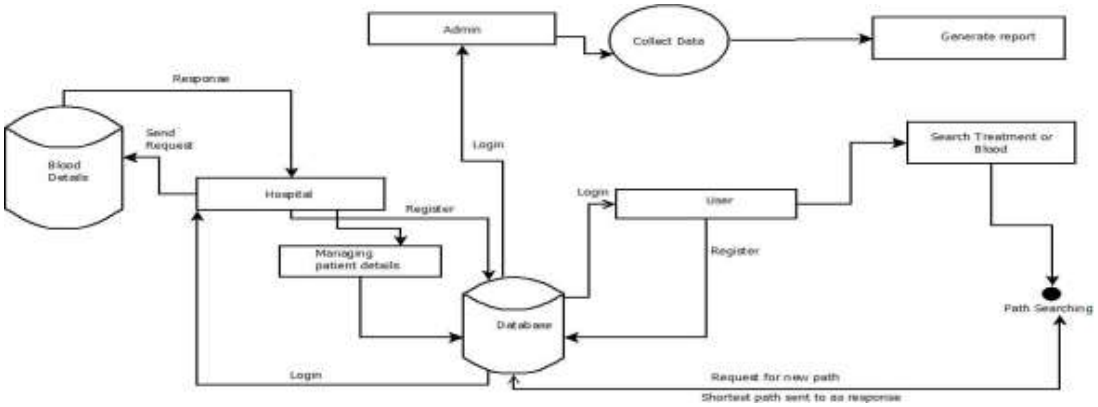


Fig 1 Massive data architecture

7. CONCULSION AND FUTUREWORK

In the proposed system we use Dynamic Sorting Algorithm to search and retrieve data very easily from huge databases.It reduces time complexity. Efficiency is very high. The top results only shown when we search any kind of data.And the database of the each hospital is maintained by the government. This database will helps to identify the patient affected by which diseases And this information will help to take the remedy of the government An early termination will helps to terminate the repeated terms in the table and we can save the data in an efficient manner in the database.

Collections of documents, images, videos, and networks are being thought of not merely as bit strings to be stored, indexed, and retrieved, but as potential sources of discovery and knowledge, requiring sophisticated analysis techniques that go far beyond classical indexing and keyword counting, aiming to find relational and semantic interpretations of the phenomena underlying the data.

REFERENCES

[1] R. Akbarinia, E. Pacitti and P. Valduriez , "Best position algorithms for top-k queries" , *Proc. 33rd Int. Conf. Very Large Data Bases* , pp.495 -506 , 2007.

[2] H. Bast, D. Majumdar and R. Schenkel , "Io-top-k: Index-access optimized top-k query processing" , *Proc. 32nd Int. Conf. Very Large Data Bases* , pp.475 -486 , 2006

[3] Y. Chang, L. Bergman and V. Castelli , "The onion technique: Indexing for linear optimization queries" , *Proc. ACM SIGMOD Int. Conf. Manag. Data* , pp.391 -402 , 2000

[4] G. Das, D. Gunopulos, N. Koudas and D. Tsirogiannis , "Answering top-k queries using views." , *Proc. 32nd Int. Conf. Very Large Data Bases.* , pp.451 -462 , 2006

[5] R. Fagin, R. Kumar and D. Sivakumar , "Efficient similarity search and classification via rank aggregation" , *Proc. ACM SIGMOD Int. Conf. Manage. Data* , pp.301 -312 , 2003

BIOGRAPHY

Ishwarya.M,pursuing her B.E. in CSE department in Agni College of Technology, Anna University in Chennai, India, in May 2016.

Kalaiarasi.S,pursuing her B.E. in CSE department in Agni College of Technology, Anna University in Chennai, India, in May 2016.

Revathy.M received her B.E. in CSE department from Karpagam Engineering College,Coimbatore, Anna University in Chennai, India, in May 2011, and got her M.E. in CSE department from Konkunadu Engineering College,Erode, Anna University in Chennai, India in August 2013.