

Information Retrieval using Distributed Event Based System

Prof N. Prasanna Balaji¹ and Susarla Venkata Prasad Ravi Teja²

¹(HOD IT&CSE-2) GuruNanak Institutions Technical Campus
 Kanapur, Hyderabad, Telangana

²GuruNanak Institutions Technical Campus Kanapur, Hyderabad, Telangana.
 E-mail: ¹hodit.gnec@gniindia.org, ²ravisusarla@ymail.com

Abstract: This is a paper that which was going to say about new way of retrieving the information that was stored in the different kinds of databases, normally we will use the information retrieval systems (IRS) to retrieve the information and the Distributed Event Based System(DEBS) is to make the work done in the form of different events.

Now in this paper I want to combine both these technologies to get the most efficient retrieval system that which can retrieve the information that was in databases by creating the distributed events that which will reduce the complexity of understanding of retrieval of any kind of information from any kind information. The overhead of storage and the time consumption in searching may be reduced and this combination helps to effectively work. This may result in an efficient system while adopting the DEBS

1. INTRODUCTION

An information retrieval process begins when a user enters a query into the system. Queries are formal statements of information needs, for example search strings in web search engines. In information retrieval a query does not uniquely identify a single object in the collection. Instead, several objects may match the query, perhaps with different degrees of relevancy. An information retrieval system is a hardware or software module that which was used to retrieve the information from the databases or from the large amount of data.

The distributed event based system is concept of dividing the specific task into the events and to make the work done easy for the tasks that which are in big amount execution process as we are combing the both IRS and DEBS for it will be complex to design the events for the information retrieval but it will be easy for the next enhancements to get the good IR systems for the huge amounts of databases such as the multimedia databases, etc.

2. FACTORS INFLUENCING IRS & DEBS

The factors that are influencing the IRS are:

1. Precision
2. Recall

$$\text{Precision} = \frac{\text{Number_Retrieved_Relevant}}{\text{Number_Total_Retrieved}}$$

$$\text{Recall} = \frac{\text{Number_Retrieved_Relevant}}{\text{Number_Possible_Relevant}}$$

Fig. 1: Source: Information Retrieval Systems Theory and Implementation (The Information Retrieval Series, 1)

We can consider the required and unrequired information was retrieved for a particular query can be categorized as shown below

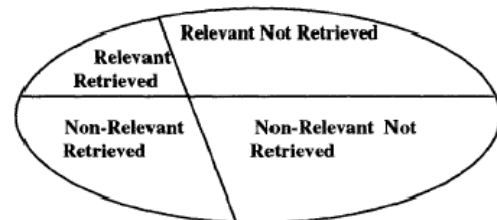


Fig. 2: Source: Information Retrieval Systems Theory and Implementation (The Information Retrieval Series, 1)

The factors that are influencing the DEBS are:

1. How to divide
2. How to manipulate
3. To extract results
4. To combine the results extracted

How to divide: this is that which defines divide specific task into discreet events by considering the different factors.

How to manipulate: this factor describes that how to implement the events that are be implemented

To extract results: to get the results that are related to the query that was given.

To combine the results extracted: this is the most difficult to perform that is way because each event that which will have different results for different events to combine them is a very difficult and we have to combine them in a sequential manner without fail.

3. HOW TO RETRIEVE INFORMATION USING DEBS

The steps that are involved to retrieve the information using DEBS are as follows:

1. Input the query
2. Divide the query into nested manner
3. Implementing the process
4. Reading all the results
5. Combining all results
6. Validation
7. Verification
8. Give the results back

4. ALGORITHM TO IMPLEMENT THE INFORMATION USING THE DEBS

S1: Take the input query from the user as per his requirement.
S2: Divided the query into the nested manner i.e., identify the number of tasks can be performed and how they are identical are not.

S3: we can implement the any kind of algorithm to get the sub task to be get executed for the retrieving of data.

S4: we should gather the results from the each sub task that was divided all collect all the information that was required for the retrieval of information and discard the results that which are irrelevant to the query given.

S5: After discarding the unwanted results or irrelevant information we should combine all the results those which are found relevant to the user query.

S6: Now we should give the information to the user before that that we should perform some validations as per the user query was given such as duplicate data, data loss, etc.

S7: as the algorithm is new to the system we should compare with the other efficient systems that we are getting the relevant data for the user query or not in the form of the verification with the existing systems or algorithms.

S8: Finally we should give the information that was retrieved from the data base using the DEBS concept.

5. FLOW OF IMPLEMENTATION

The following diagram will explain the flow of execution of the information retrieval using the DEBS

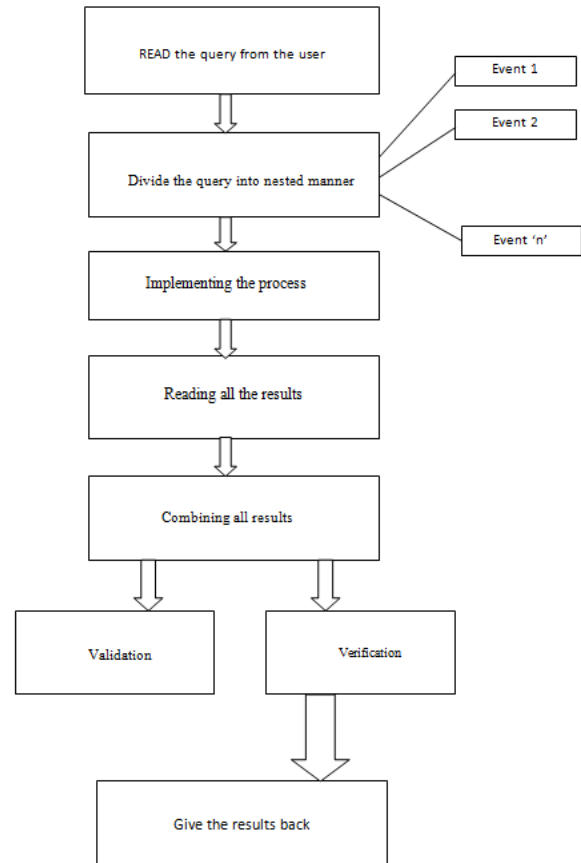


Fig. 3: flow of Implementation to Information Retrieval using DEBS

6. SIMPLE EXAMPLE TO DEMONSTRATE INFORMATION RETRIEVAL USING DEBS

Let us consider that we had a tiny database that which has the all kinds of alphabets, numbers and some other small values in it. Now we have to retrieve the information from the t-db (tiny database).

Let us assume that we have to retrieve the information that was related to the alphabets I.e., a, b, c, d.....etc .we can four different ways of alphabets such as lower case, upper case, and other.

Now we will see that how the information will be retrieved from the database according to the algorithm given.

At first we have to read the query in the form that which is having the correct grammar related to the system. Now divide the query into the events such as for alphabets retrieval one event and for the numerical other one and some other events as we already seen above in alphabets there will be upper and lower and cursive, etc. For numerical values there will be integers, fractions, decimals, mixed, etc

For each kind of value there will be a specific event and we have to monitor them according to their occurrence of execution and to maintain them in a specific manner of execution.

The following hierarchy will explain about the different events that are presented to retrieve the information using the DEBS.

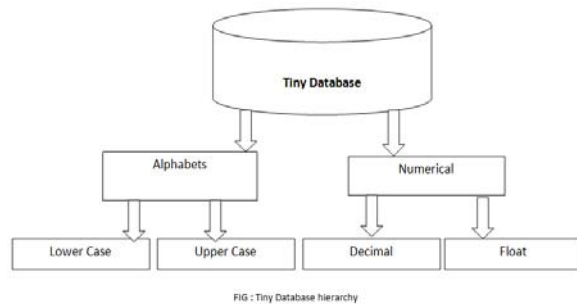


Fig. 4: Tiny Database Hierarchy

Now we will divide the events for the given query according to the hierarchy for this application we will get the events as follows

1. Event_alphabet()
 - a) Event_lowercase()
 - b) Event_uppercase()
2. Event_numerical()
 - Event_decimal()
 - Event_float()
 - Event_integer()

Let us assume that we have to retrieve the value 'Z' that which was in alphabets in t-db after posting the query the following events will be executed in sequential manner to get the information that was explained by the following diagram.

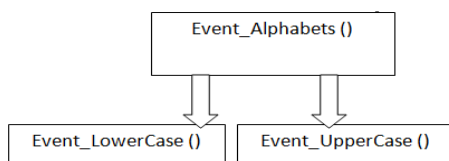


Fig. 5: flow of Events to Information Retrieval using DEBS

Now we will get the two kinds of results as in the query itself it was mentioned that the alphabet 'Z' should be retrieved from the t-db as of that the two events are executed parallel instead of searching for one at one time and the other by iteration we can simplify the process of execution using the DEBS for the information retrieval.

7. APPLICATIONS

We can apply this algorithmic concept to any kind of application that in which the information has to be retrieved from the databases as it will make easy to understand the concept of the information retrieval from the huge and complex databases.

8. MERITS

- Less complex
- Applicable to any kind of database
- We can apply to the any information retrieval algorithm
- High efficiency

9. DEMERITS

- Events should be designed accurately
- Categorization of relevant and irrelevant information that was retrieved.

10. PRESENT WORK

At present we are working on the small databases to retrieve the information from the databases such as college databases and also with the help of the small 1Tb HDD's

11. FUTURE WORK

We want to implement the algorithm on the large and huge amount of database to provide the good and efficient information retrieval to the current day world

We also want to join the information and data together and want make distinguish between them in only one database

REFERENCES

- [1] Information Retrieval Systems Theory and Implementation (The Information Retrieval Series, 1)
- [2] Introduction to Information Retrieval Christopher D. Manning *Stanford University* Prabhakar Raghavan *Yahoo! Research* Hinrich Schütze *University of Stuttgart*
- [3] Test Collection Based Evaluation of Information Retrieval Systems By Mark Sanderson
- [4] Methods for Evaluating Interactive Information Retrieval Systems with Users By Diane Kelly
- [5] Chu, H. (2003). Information representation and retrieval in the digital age. Medford, NJ:Information Today.
- [6] Blair, D. C. (1990). Language and representation in information retrieval. Amsterdam: Elsevier Science
- [7] Soergel, D. (1985). Organizing information: Principles of database and retrieval systems. Orlando, FL: Academic Press.
- [8] Baeza-Yates, R. and Ribeiro-Neto, B. Modern Information Retrieval, Addison Wesley, New York, 1999.

-
- [9] Gorry, G.A. and Morton, M.S.S. A framework for management information systems, *Soloan Management Review*, 13, 55-70,1971.
- [10] Zadeh, L.A. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems*, 19, 111-127, 1997.
- [11] Arasu, A., Babcock, B., Babu, S., Cieslewicz, J., Datar, M., & Ito, K. (2008). STREAM: The Stanford Data Stream Management System. In Garofalakis, M., Gehrke, J., & Rastogi, R. (Eds.), *Data Stream Management: Processing High-Speed Data Streams*. Berlin, Germany: Springer.
- [12] Leitner, P., Michlmayr, A., Rosenberg, F., & Dustdar, S. (2008). End-to-End Versioning Support for Web Services. In *Proceedings of the International Conference on Services Computing SCC'08*. Honolulu, HI: IEEE Computer Society.
- [13] Curbera, F., Doganata, Y. N., Martens, A., Mukhi, N., & Slominski, A. (2008). Business Provenance - A Technology to Increase Traceability of End-to-End Operations. In *Proceedings of the 16th International Conference on Cooperative Information Systems (CoopIS'08)* (S. 100-119). Monterrey, Mexico: Springer.
- [14] Zadeh, L.A. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems*, 19, 111-127, 1997.
- [15] Li, G., Cheung, A., Hou, S., Hu, S., Muthusamy, V., Sherafat, R., et al. (2007). Historic Data Access in Publish/Subscribe. In *Proceedings of the Inaugural International Conference on Distributed Event-Based Systems (DEBS'07)* (pp. 80- 4). Toronto, Canada: ACM Press.
- [16] Michlmayr, A., Rosenberg, F., Leitner, P., & Dustdar, S. (2008). Advanced Event Processing and Notifications in Service Runtime Environments. In *Proceedings of the 2nd International Conference on Distributed Event-Based Systems (DEBS'08)* (pp. 115-125). Rome: ACM Press.
- [17] Mühl, G., Fiege, L., & Pietzuch, P. (2006). *Distributed Event-Based Systems*. Berlin, Germany: Springer-Verlag.
- [18] Michlmayr, A. (2010). Event Processing in QoS-Aware Service Runtime Environments, PhD Thesis, Vienna University of Technology
- [19] Antonio Carzaniga, David S. Rosenblum, and Alexander L. Wolf. Design and evaluation of a wide-area event notification service. *ACM Transaction on Computer Systems*, 2001.
- [20] Peter R. Pietzuch and Jean M. Bacon. Hermes: A Distributed Event-Based Middleware Architecture. Submitted to the Workshop on Distributed Event-Based Systems (DEBS), 2002.