INTERNATIONAL JOURNAL FOR TRENDS IN ENGINEERING & TECHNOLOGY
VOLUME 32 ISSUE 1 – AUGUST 2018 - ISSN: 2349 – 9303

User Based Retrieval System

M. Mohanapriya[1],
Department of Computer Science and Engineering,
Periyar Maniammai Institute of Science and Technology,
priyamanikandan271@gmail.com

T. Kavitha[2]
Assistant Professor,
Department of Computer Science and Engineering,
Periyar Maniammai Institute of Science and Technology-,

Abstract: Now-a-days large amount of heterogeneous event data is increasingly generated in the systems. Owing to the difference between event data and traditional relational data, the matching of heterogeneous events is highly non-trivial. The existing structure-based matching techniques for relational data also fail to perform owing to the poor discriminative power of dependency relationships between events. A generic pattern based matching framework is proposed, which is compatible with the existing structure-based techniques. Finally, extensive experiments demonstrate the effectiveness of our pattern-based matching compared with approaches adapted from existing techniques, and the efficiency improved by the bounding, pruning and heuristic methods. This application help to store the log files in cloud and retrieve data based on the events. This is web-based application and developed in Eclipse and MySQL. This application will help to rectifying existing problem and effective data retrieval based on events.

Keywords:Bigdata, MySQL, Generic pattern matching
.

I. INTRODUCTION

Information systems (e.g., OA and ERP systems) of different divisions or branches in large corporations keep on generating heterogeneous event logs. It is strongly desired to integrate the event data, e.g., for finding steps leading to same data (provenance analysis) in multiple sectors, identifying similar complex procedures (complex event processing) in different branches, or obtaining a global picture of business processes (workflow views) in various divisions. Without exploring the correspondence among heterogeneous events, query and analysis on the event data (simply merged together) may not yield any meaningful result. Unfortunately, directly applying existing schema matching techniques may fail to obtain the right mapping of heterogeneous events. Owing to the independent encoding systems in different sources, the widely used methods based on typographic similarity (e.g., string cosine similarity) or linguistic similarity (using dictionary of ontology like WordNet) of event names are often unlikely to perform. To solve the matching problem with “opaque” names, graph based matching approaches exploit the structural information among attributes (events in our case). It relies on the statistics of dependency relationships, e.g., how often two events appear consecutively. The more similar the dependency relationship is, the more likely the corresponding events can be mapped with each other. The matching problem is to find a “best” mapping that can maximize the similarity of dependency relationships between two datasets.

II RELATED WORK

2.1. Approximate Event Matching

One of the early works on approximate event matching is ATOPSS which defines an approximate matching model based on fuzzy membership functions that specify the degree that a value in an event matches a value in a subscription. A-TOPSS does not consider schema approximation. Another work is STOPSS which considers schema and value semantic matching. It proposes the use of agreed-upon onto logs and a system architecture that generates events other than the original ones by replacing concepts with taxonomic or ad-hoc related concepts.

2.2. Event Ranking

Previous work has considered ranking events according to range predicates, preference, diversity and freshness, probability of occurrence, fuzzy membership of attribute values or focused on efficient ranking in sliding windows rather than the ranking functionality. All of these works do not use semantic relatedness of events as a factor for ranking. FOMatch considers scoring based on semantic matching and evaluation was conducted using thresholds, however a precision recall tradeoff was not investigated.

2.3. Schema Matching

The schema matching problem at the most basic level refers to the problem of mapping schema elements (for example, columns in a relational database schema) in one information repository to corresponding elements in a second repository. While schema matching has always been a problematic and interesting aspect of information integration, the problem is exacerbated as the number of information sources to be integrated, and hence the number of integration problems that must be solved, grows. Such schema matching problems arise both in “classical” scenarios such as company mergers and in “new” scenarios such as the integration of diverse sets of query able information sources over the web.

2.4. Provenance Analysis

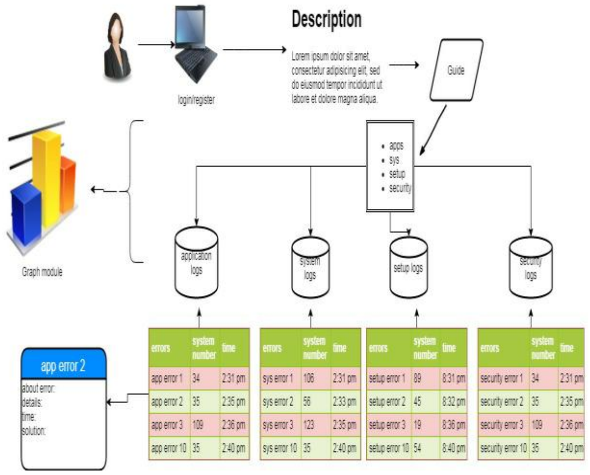
Technological advances have enabled the capture of massive amounts of data in many different domains, taking us a step closer to solving complex problems such as global climate change and uncovering the secrets hidden in genes. Workflow management systems are therefore increasingly used for managing and analyzing this data, allowing users to specify complex, multi-step, “in-silicon” experiments or analyses. To ensure reproducibility and verifiability of results, many workflow systems are now providing support for provenance.

2.5. Text Joins in an RDBMS for Web Data Integration

The integration of information from heterogeneous web sources is of central interest for applications such as catalog data integration and warehousing of web data (e.g., job advertisements and announcements). Such data is typically textual and can be obtained from disparate web sources in a variety of ways, including web site crawling and direct access to remote databases via web protocols. The integration of such web data exhibits many semantics and performance-related challenges.

II.III. SYSTEM ARCHITECTURE

The figure (fig 3.1) depicts the working of the system and thereby shows the view of each module in the entire process.



errors	system number	time
app error 1	34	2:31 pm
app error 2	35	2:35 pm
app error 3	109	2:36 pm
app error 10	35	2:40 pm

errors	system number	time
sys error 1	106	2:31 pm
sys error 2	56	2:33 pm
sys error 3	123	2:35 pm
sys error 10	35	2:40 pm

errors	system number	time
setup error 1	89	8:31 pm
setup error 2	45	8:32 pm
setup error 3	19	8:36 pm
setup error 10	54	8:40 pm

errors	system number	time
security error 1	34	2:31 pm
security error 2	35	2:35 pm
security error 3	109	2:36 pm
security error 10	35	2:40 pm

Fig 3.1 Architecture diagram for the proposed system

III. IV. PROPOSED SYSTEM

It is a web-based application. All events are separated with different patterns. We proposed with a graph to monitor how much events occurs in an application. A generic pattern based matching framework is proposed, which is compatible with the existing structure-based techniques. In this application can be consist guideline for how to access the logs. To improve the matching efficiency, we devise several bounds of matching scores for pruning. Recognizing the NP-hardness of the optimal event matching problem with patterns, we propose efficient heuristic. This proposed system can be easy for accessing and retrieving data in the database based on the events.

Advantages:

- ✓ It can easily access the data from database.
- ✓ It improves efficiency in retrieving data.
- ✓ It is secured because of cloud-based storage

IV. V. METHODOLOGIES

MODULES:

- Authentication
- Creating Search History
- Data Grouping
- Pattern Matching
- Monitoring

5.1 AUTHENTICATION

In this module the user has to register first, then only he/she has to access the database.After registration the user has to login to the site. The Authorization and Authentication (fig 5.1) process facilitates the system to protect itself and besides it protects the whole mechanism from unauthorized usage. The Registration involves in getting the details of the users who want this appication.

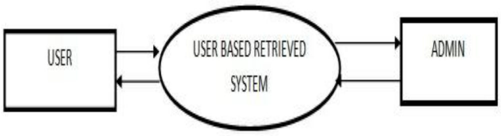


Fig 5.1. Authentication for Proposed Sysytem

5.2 ADMIN AUTHENTICATION

In Admin Authentication (fig 5.2), admin is managing the Database and Severs in the proposed System. He uploads the required Data in the backend and monitors the data that give a proper result to user.

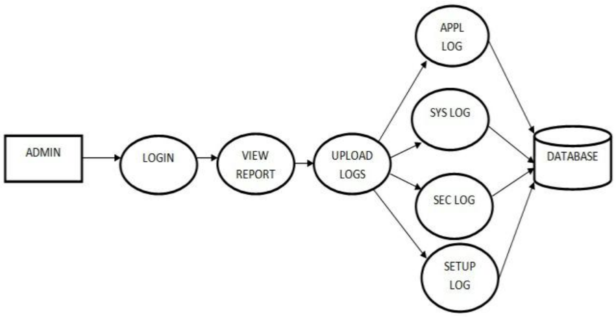


Fig 5.2. Admin Authentication for Proposed System

5.3 USER AUTHENTICATION

The Fig 5.3 explains the User Authentication of the Proposed System. The User Authentication is used to verify the valid user and provides the Database access in Web Application. User can view the event in the Web Application.

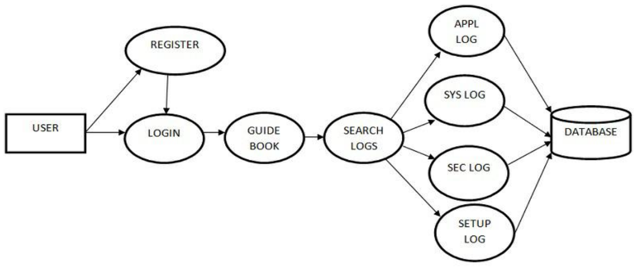


Fig 5.3 User Authentication for Proposed System

5.4 CREATING SEARCH HISTORY

Any personal documents such as browsing history and emails on a user's computer could be the data source for user profiles. This focus on frequent terms limits the dimensionality of the document set, which further provides a clear description of user's interest.

This module allows the search engine to better understand a user's session and potentially tailor that user's search experience according to their needs. Once query groups have been identified, search engines can have a good representation of the search context behind the current query using queries and clicks in the corresponding query group.

5.5 DATA GROUPING

User's queries can be classified into different query clusters. Concept-based user profiles are employed in the clustering process to achieve personalization effect.

The most similar pair of concept nodes, and then, merge the most similar pair of query nodes, and so on. Each individual query submitted by each user is treated as an individual node and each query with a user identifier. we perform the grouping in a similar dynamic fashion, whereby we first place the current query and clicks into a query group

5.6 PATTERN MATCHING

To ensure that each query group contains closely related and relevant queries and clicks, it is important to have a suitable relevance between the current query groups.

We assume that users generally issue very similar queries and clicks within a short period of time. The search history of a large number of users contains signals regarding query relevance, such as which queries tend to be issued closely together.

This captures the relationship between queries frequently leading to clicks on similar URLs. Query reformulation graph and the query click graph from search logs, and how to use them to determine relevance between queries or query groups within a user's history.

5.7 MONITORING

Monitoring is to first treat every query in a user's history as a query group, and then merge these groups in an iterative fashion. However, this is impractical in our scenario for two reasons.

First, it may have the undesirable effect of changing a user's existing query groups, potentially undoing the user's own manual efforts in organizing her history.

Second, it involves a high-computational cost, since we would have to repeat many query group similarity computations for every new query.

VI. CONCLUSION

Event-based systems are coupled, via event subscriptions and patterns, to the semantics of the underlying event schema and values. Approximate semantic matching of heterogeneous events has been discussed in this paper in order to address event semantic coupling. Event semantic of types, properties and values has been considered as a dimension of decoupling required to scale event-based systems out to high heterogeneous environments such as the sensor web.

A general model has been proposed with a hybrid instantiation based on both thesauri and distributional semantics-based semantic similarity and relatedness measures. Experiments have been conducted on real-world events extracted from Freebase. Results show that the proposed hybrid matcher outperforms matchers based on a single semantic similarity or relatedness measure.

REFERENCES:

1.Aumueller, D., Do, H.-H., Massmann, S., and Rahm, E.Schema and ontology matching with COMA++. Proceedings of the 2005 ACM SIGMOD international conference on Management of data, ACM (2005), 906-908.

2.Belkin, N.J. and Croft, W.B. Information filtering andinformation retrieval: two sides of the same coin? Commun.ACM 35, 12 (1992), 29-38.

3.Drosou, M., Stefanidis, K., and Pitoura, E. Preference-aware publish/subscribe delivery with diversity. Proceedings of the Third ACM International Conference on Distributed Event- Based Systems, (2009), 6:1--6:12.

4.Wasserkrug, S., Gal, A., Etzion, O., and Turchin, Y. Complex event processing over uncertain data. Proceedings of the second international conference on Distributed eventbased systems, (2008), 253-264.

5.Liu, H. and Jacobsen, H.-A. A-TOPSS: a publish/subscribe system supporting approximate matching. Proceedings of the 28th international conference on Very Large Data Bases, VLDB Endowment (2002), 1107-1110.

6.Erhard Rahm, Philip A. Bernstein: A survey of approaches toautomatic schema matching. VLDB Journal 10(4) (2001).

7.Jayant Madhavan, Philip A. Bernstein, Erhard Rahm:Generic Schema Matching with Cupid. VLDB 2001: 49-58.

8.Silvana Castano, Valeria De Antonellis, Sabrina De Capitani di Vimercati: Global Viewing of Heterogeneous Data Sources. TKDE 13(2): 277-297(2001)