

# Random Query Formation for Novel Database

Gajendra A. Shinde<sup>1</sup>, Rutuja R. Kukawalkar<sup>2</sup>, Puja K. Bachhav<sup>3</sup>, Govind M. Podar

Graduate Student, Department of Computer, Gangamai College of Engineering, Maharashtra, India<sup>1</sup>

Graduate Student, Department of Computer, Gangamai College of Engineering, Maharashtra, India<sup>2</sup>

Graduate Student, Department of Computer, Gangamai College of Engineering, Maharashtra, India<sup>3</sup>

Head of Department, Department of Computer, Gangamai College of Engineering, Maharashtra, India<sup>4</sup>

**ABSTRACT:** Data Mining is an analytic process designed to explore data (usually large amounts of data - typically business or market related - also known as "big data") in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns. These real-world databases contain hundreds or even thousands of relations and attributes. Traditional predefined query forms are not able to satisfy various ad-hoc queries from users on those databases. In this paper, we mainly study the ranking of query form components and the dynamic generation of query forms. We will use the indexing technique. It is query optimization method and reduces the execution time also enhances the efficiency. The ranking of form components is based on the captured user preference. A user can also fill the query form and submit queries to view the query result at each iteration.

**KEYWORDS:** DQF, Query Forms, Banking, Query Execution, Admin.

## I. INTRODUCTION

Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD), an interdisciplinary subfield of computer, is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, datapre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. The term is a misnomer, because the goal is the extraction of patterns and knowledge from large amount of data, not the extraction of data itself. It also is a buzzword and is frequently applied to any form of large-scale data or information processing (collection, extraction, warehousing, analysis, and statistics) as well as any application of computer decision support system, including artificial intelligence, machine learning, and business intelligence. The popular book "Data mining: Practical machine learning tools and techniques with Java" (which covers mostly machine learning material) was originally to be named just "Practical machine learning", and the term "data mining" was only added for marketing reasons. Often the more general terms "(large scale) data analysis", or "analytics" – or when referring to actual methods, artificial intelligence and machine learning – are more appropriate.

Query form is one of the most widely used user inter-faces for querying databases. Traditional query forms are designed and predefined by developers or DBA in various information management systems. With the rapid development of web information and scientific databases, modern databases become very large and complex. In natural sciences, such as genomics and diseases, the databases have over hundreds of entities for chemical and biological data resources. Many web databases, such as Freebase and DBpedia, typically have thousands of structured web entities. Therefore, it is difficult to design a set of static query forms to satisfy various ad-hoc database queries on those complex databases. Many existing database management and development tools, such as EasyQuery, Cold Fusion, SAP and

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 3, March 2017

Microsoft Access, provide several mechanisms to let users create customized queries on databases. However, the creation of customized queries totally depends on users' manual editing. If a user is not familiar with the database schema in advance, those hundreds or thousands of data attributes.

## II. LITERATURE REVIEW

### A. Combining Keyword Search and Forms for Ad Hoc Querying of Databases

In this paper, we investigate combining the two with the hopes of creating an approach that provides the best of both. Specifically, we propose to take as input a target database and then generate and index a set of query forms offline. At query time, a user with a question to be answered issues standard keyword search queries; but instead of returning tuples, the system returns forms relevant to the question. The user may then build a structured query with one of these forms and submit it back to the system for evaluation. In this paper, we address challenges that arise in form generation, keyword search over forms, and ranking and displaying these forms. We explore techniques to tackle these challenges, and present experimental results suggesting that the approach of combining keyword search and form-based interfaces is promising.

#### Disadvantage:

They did not consider recognizing synonyms of both data terms and form terms, and more generally, extending our implementation to handle more sophisticated query expansion.

### B. BioGuideSRS: Querying Multiple Sources with a user-centric perspective:

Biologists are frequently faced with the problem of integrating information from multiple heterogeneous sources with their own experimental data. Given the large number of public sources, it is difficult to choose which sources to integrate without assistance. When doing this manually, biologists differ in their preferences concerning the sources to be queried as well as the strategies, i.e. the querying process they follow for navigating through the sources. In response to these findings, we have developed BioGuide to assist scientists search for relevant data within external sources while taking their preferences and strategies into account. In this paper, we present BioGuideSRS, a user-friendly system which automatically retrieves instances of data by using BioGuide on top of the SRS system.

#### Disadvantage:

They did not BioGuide SRS may benefit from text-mining tools to automate the mapping between the graphs.

### C. Automated Creation of a Forms-based Database Query Interface

In this paper, we seek to maximize the ability of a forms-based interface to support queries a user may ask, while bounding both the number of forms and the complexity of any one form. Given a database schema and content we represent an automated technique to generate a good set of forms that meet the above desiderata. While a careful analysis of real or expected query workloads are useful in designing the interface, these query sets are often unavailable or hard to obtain prior to the database even being deployed. Hence generating a good set of forms just using the database itself is a challenging yet important problem.

#### Disadvantage:

The results matching all five queries were available in all five websites, not all of these queries could be issued using the forms provided.

### D. Similarity Measures for Categorical Data: A Comparative Evaluation

Measuring similarity or distance between two entities is a key step for several data mining and knowledge discovery tasks. The notion of similarity for continuous data is relatively well-understood, but for categorical data, the similarity computation is not straightforward. Several data-driven similarity measures have been proposed in the literature to compute the similarity between two categorical data instances but their relative performance has not been evaluated. In this paper we study the performance of a variety of similarity measures in the context of a specific data mining task: outlier detection. Results on a variety of data sets show that while no one measure dominates others for all types of problems, some measures are able to have consistently high performance.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 3, March 2017

## Disadvantages:

They did not consider how a similarity measure handles the different characteristics of a categorical data set, and this needs to be explored.

## III. PROPOSED WORK

### 3.1 Dynamic Query Form:

Traditional databases can't give output as user's satisfaction i.e. some normal users can't understand such programming languages. To overcome the drawback of previous technique, the dynamic query form is generated. Dynamic query form handles the database without having knowledge of query languages. It is not required to have expertise in any database query language to deal with data by using this system.

### 3.2 Proposed Method:

The system proposes DQF, a novel database query form interface, which is able to dynamically generate query forms. The system is proposed to have the following modules[3]-

In "Query Form Enrichment", query form elements for the admin/end user DQF suggests an indexed list and the user choose wanted attributes into present query form.

In "Query Execution", admin/end user completes the present query form into the present query form and DQF runs the present query and display its query result and finally admin /end user returns it feedback.

In "Customized Query Forms",the proposed system returns visual interfaces for creators to generate or customize query forms. The problem of those different tools is that, this is not for admin/end-users, for the professional creators who are well known with their databases/data.

In "Database Query Recommendation",current invention develops collaborative techniques to suggest database query elements for database exploration. Its treat SQL queries as items in the collaborative filtering approach, and recommend similar queries to related users. C

### 3.3 Methodology:

Dynamic Query form will actually into two steps, first one is Query Execution and second one is Query enhancement [3].

In "Query Execution", admin/end user completes the current query form with wanted form attribute, and present query form will be put in result by the system and after output of the query shown by the system. If admin/end user receive wanted output from submitted query form and then admin/end user will index the query form in future work. For future use the query form will be restored [3].

In "Query Enhancement",as we discuss above in query execution method, admin/end user again select the query attribute if admin/end user is not satisfied with query results then he again he will again complete the query form and submit it and obtain the desired output[3]. After admin/end user is satisfied with the obtained output then take help of the query from ranking list, otherwise he will select another form attribute. Because of query enhancement methodology, attribute enhancement based on end user's choice and improvement of query form and that's why efficiency of query form is increased [3].

In "User Action", to decide whether a query form is desired or not, a user does not have time to go over every data instance in the query results. In addition, many database queries output a huge amount of data instances. In order to avoid this "Many-Answer" problem, we only output a compressed result table to show a high-level view of the query

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 3, March 2017

results first. Each instance in the compressed table represents a cluster of actual data instances. Then, the user can click through interested clusters to view the detailed data instances [3].

## IV. SYSTEM ARCHITECTURE

Framework engineering is comprised of module, for example, client, make yearly preparing timetable customer CIRT framework and database.

- **Admin/End-user:-**

In this module client used our system with admin's authenticates userid and authenticate password. System just checked client will have an authorization to access into system.

- **Query Form:-**

In this stage validate client i.e. admin take some effective decision after analyzing data by using different query forms. Query forms consists of different number of attributes i.e. components of database and also using several criteria.

- **Query Result:-**

This is our desired framework in which all data with respect to admin.

- **Database:-**

Database is utilized/used for to save different data required by our DQF for Banking Sector framework i.e. Query Forms, Query Result, Ranked based Query Result, Analysis report of database and Final Report.

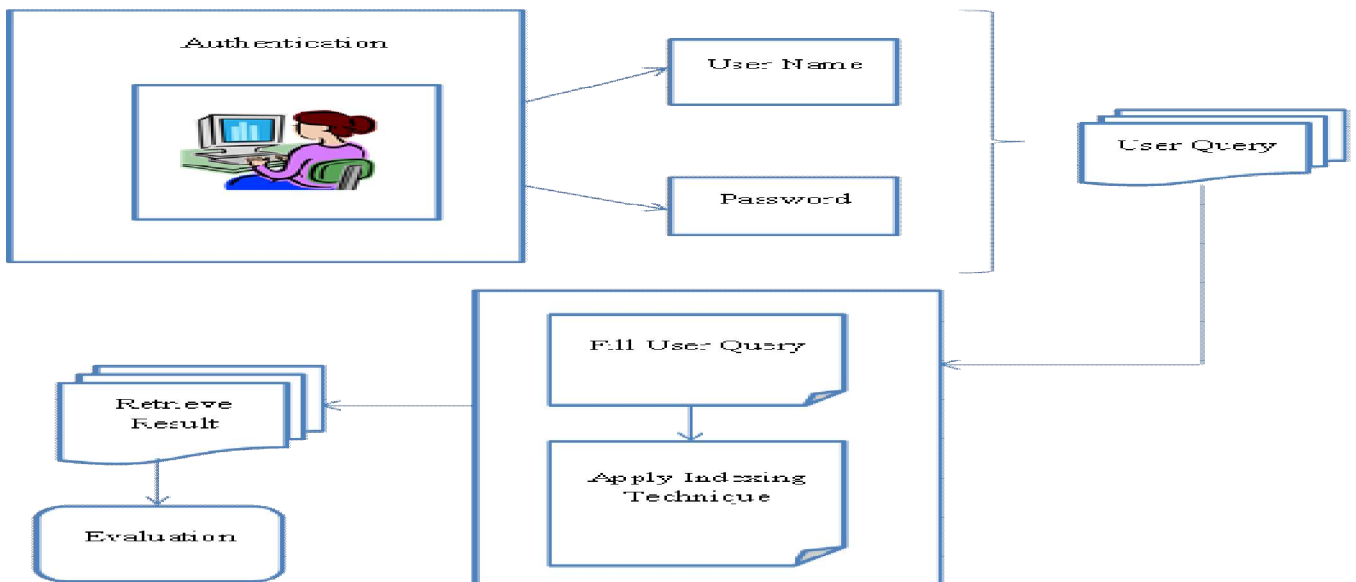


Fig: System Architecture

## V. RESULTS AND COMPARISON

- **User Login:**

The user login which is user entry level here the authentication which is done based on the username and password. A login or logon refers to the credentials required to obtain access to a computer system or other restricted area. Logging in or on and signing in or on is the process by which individual access to a computer system is controlled by identifying and authenticating the user through the credentials presented by the user. If the user is new member he/she should register the login form to database query result and retrieve.

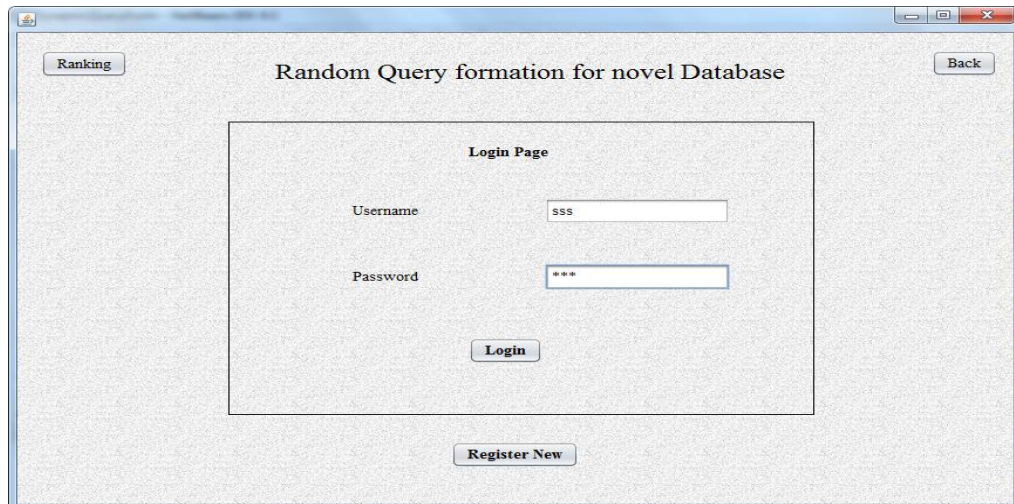


## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

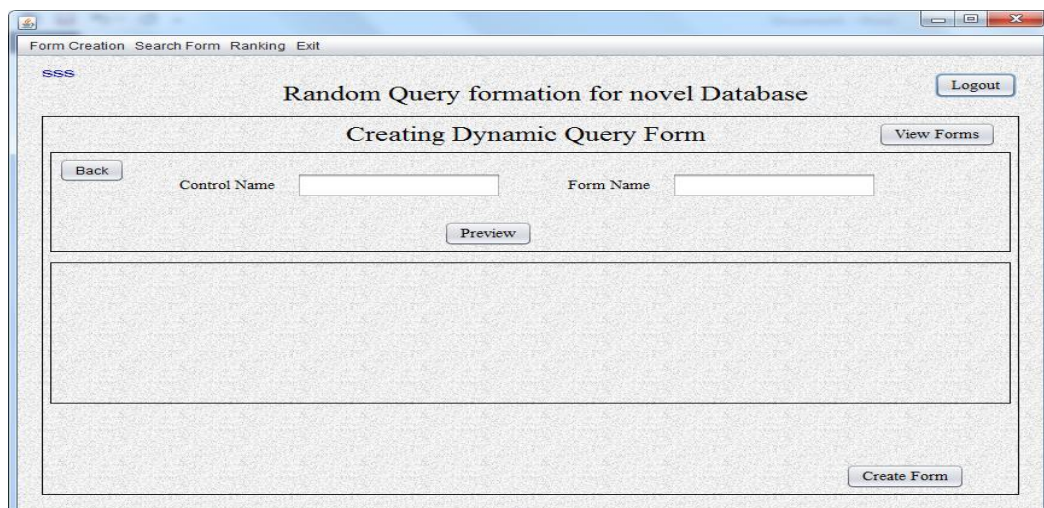
Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 3, March 2017



- **Query Generation & Execution:**

The dataset which is gathered from UCI machine learning repository which is loaded and preprocess the data for cleaning the input. By considering a general search engine for particular duration from that they have to collect information like Anonymous ID, query, query time, URL. From this dataset they have to extract required information related to our next step process. The user fills out the current query form and submits a query. DQF executes the query and shows the results. The user provides the feedback about the query results.



- **Database Query Recommendation:**

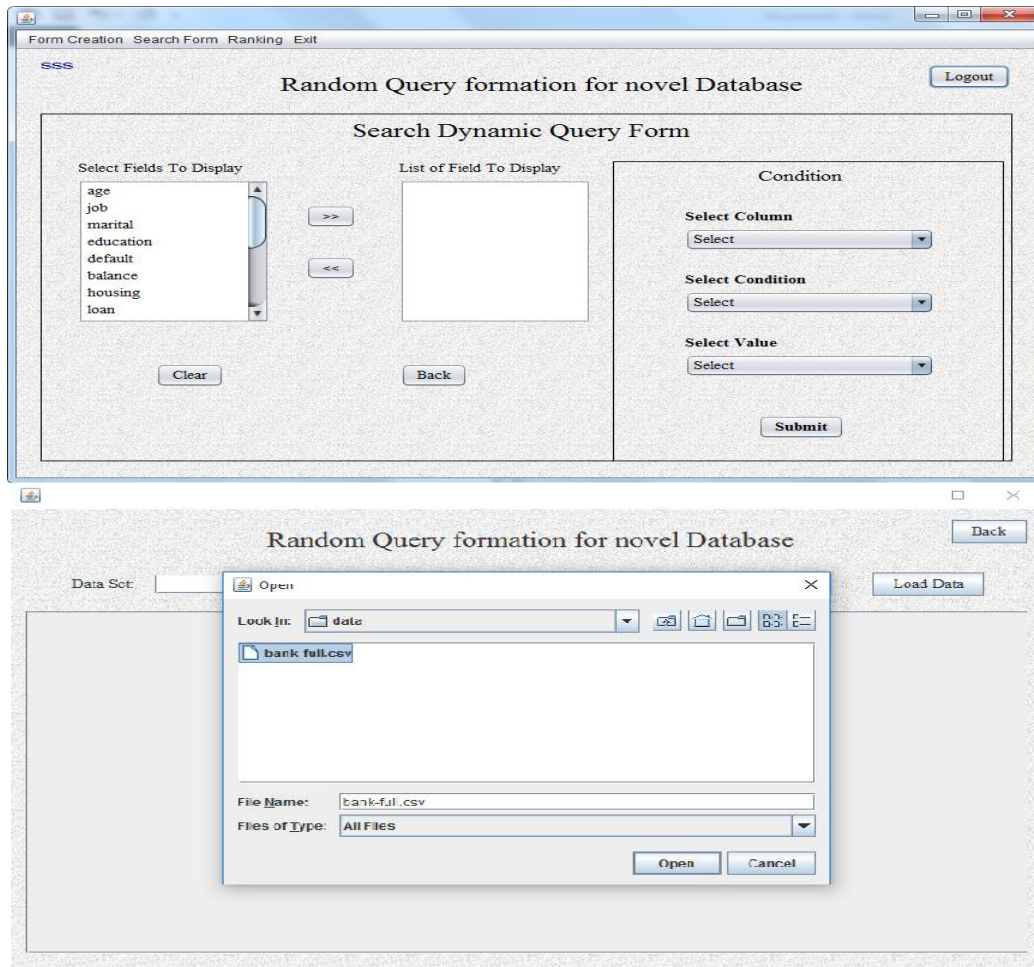
Recent studies introduce collaborative approaches to recommend database query components for database exploration. To assist users in this context, we draw inspiration from Web recommender systems and propose the use of personalized query recommendations. The idea is to track the querying behavior of each user, identify which parts of the database may be of interest for the corresponding data analysis task, and recommend queries that retrieve relevant data. They treat SQL queries as items in the collaborative filtering approach, and recommend similar queries to related users.

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 3, March 2017



- **Indexing Technique Retrieval:**

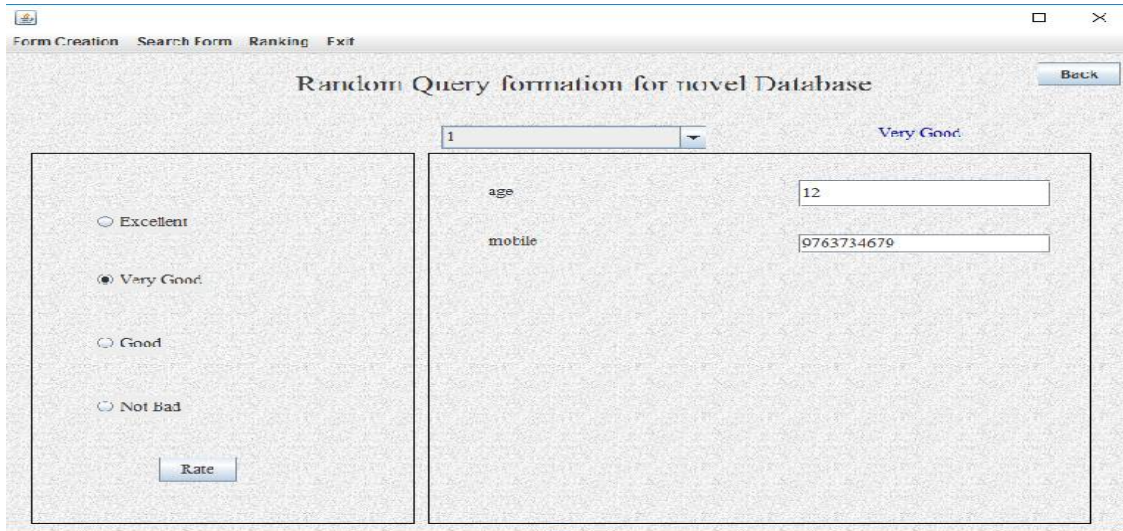
The indexing technique is a query optimization method and reduces the execution time also enhances the efficiency. Our experimental results show that our sampling approach never misses any true patterns when the number of feature instances. That improves the speed of data retrieval operations on a database table at the cost of additional writes and storage space to maintain the index data structure. Based on the query form the database engine which retrieves the result to the user. The retrieved result which is ranked based on the user query selection.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 3, March 2017

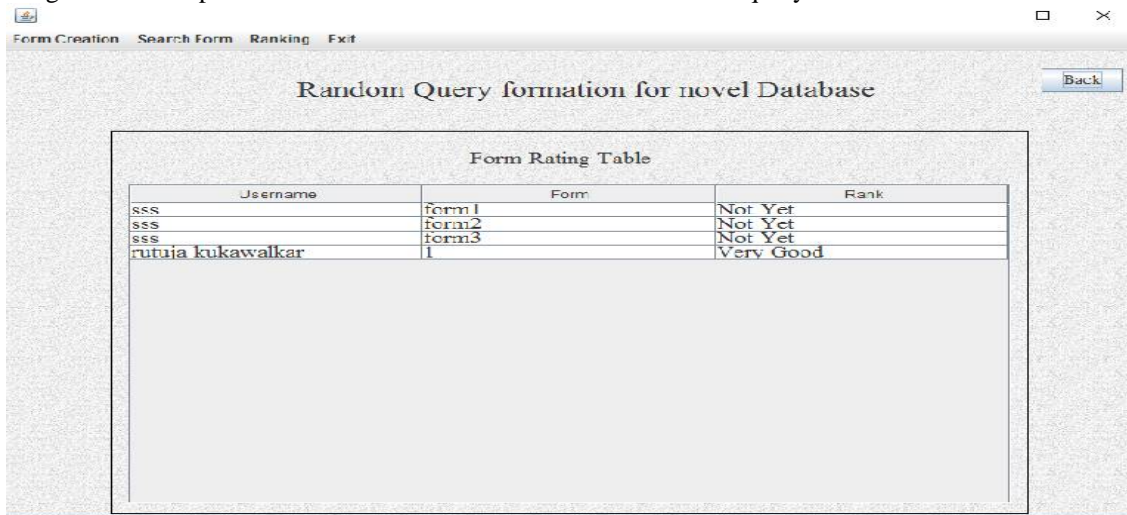


- Performance Analysis:**

Experimental results show that the dynamic approach often leads to higher success rate and simpler query forms compared with a static approach. The performance metrics which shows the graphical comparison of existing and proposed approach.

## VI.CONCLUSION

In this paper we propose a dynamic query form generation approach which helps users dynamically generate query forms. The key idea is to use a probabilistic model to rank form components based on user preferences. We capture user preference using both historical queries and run-time feedback such as click-through. Experimental results show that the dynamic approach often leads to higher success rate and simpler query forms compared with a static approach. The ranking of form components also makes it easier for users to customize query forms.



Username	Form	Rank
sss	form1	Not Yet
sss	form2	Not Yet
sss	form3	Not Yet
rutuja kukawalkar	1	Very Good



