

# Incorporating Visual Data Mining in Online Test Analysis

<sup>1</sup>T Anuradha, <sup>2</sup>J. Vikram Datta, <sup>2</sup>M. Shabaaz, <sup>2</sup>K. Lokesh, <sup>2</sup>T. Yamini

<sup>1</sup>Associate professor –Department of Electronics and computers

<sup>2</sup>B.tech- Department of electronics and computers

<sup>1,2</sup>Koneru Lakshmiah College of Engineering

Green fields, vaddeswaram, Guntur, India

E.Mail: [jvdatta009@gmail.com](mailto:jvdatta009@gmail.com)

**Abstract**— The proposed system allows the test conductor to monitor important aspects of the test taker during the process of the online examination; the user interactions with the system are recorded and stored. The recorded information will then be used to dynamically generate visual representations which can easily be analyzed by the test conductors. The process of Visual data mining makes the result analysis process fast and easy when the number of test takers is very large. The system takes into consideration several factors other than the response provided by the test taker.

**Keywords:** Visual Data Mining; Graphs; Knowledge Discovery Techniques; jfree Charts; Monitor Test Taker's Behaviour

## 1. Introduction

In recent days more and more educational systems are opting for E-testing rather than the traditional paper based test, this method is also very famous with various companies and corporations which use E-testing while recruiting new employees. We know that the number of students/aspirants appearing for the E-tests is very large; so the process of analysing the results becomes very time consuming and burdensome. Data mining techniques are used to analyse the results to a particular extent, but incorporating Visual data mining opens new horizons for extracting data that was previously left unnoticed<sup>[1]</sup>. Visual data mining requires sufficient computation power and coding the required parameters that are to be observed during the test<sup>[1]</sup>.

Due to the large number of people attempting the test, it is very much a possibility that more than one test taker will obtain the same scores in the test<sup>[5]</sup>. So, further segregation of the test takers can be done by using the “test takers interaction” with the system as a vital parameter. These interactions are of several types which can be used according to the level of segregation required.

## 2. Approach

The aim of the system is to refine the large amounts of data and present them in a format that can be easily interpreted; this calls for the use of Visual data mining.

The system is designed to record various user interactions such as: (1) time taken to answer each question, (2) number of options changed for each question, (3) total time taken for completing the exam, (4) correct answers from various levels. These options are coded into the software which will be provided to the client (test conductor).

The software is designed so the test conductor can host the tests via Tomcat Web Server and all the recorded

information is stored in data bases handled by ORACLE SQL

The test conductor, who is our client, will have to initially provide the standard data-bases condensed in a dump file which will then be incorporated into the software code or he may have to create the database if there is no dump file. The dump file must contain all the questions that the test conductor wants in the exam, however the system also provides a method to add or remove questions from the software interface through the administrator.

The entire system is built using JDK1.6 and can be run on internet explorer 6.0 on a computer running windows XP.

The testing system is stored in the server machine. By using Java Database Connectivity, the responses from each client machine are communicated and stored in the databases present in the server machine. The system can be termed as a data exploration approach exploiting information visualization.

## 3. Data Visualisation

Data mining (the analysis step of the knowledge discovery in databases process,<sup>[1]</sup> or KDD), a relatively young and interdisciplinary field of computer science<sup>[2][3]</sup> is the process of discovering new patterns from large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems.<sup>[2]</sup> The overall goal of the data mining process is to extract knowledge from a data set in a human-understandable structure<sup>[2]</sup> and besides the raw analysis step involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of found structure, visualization and online updating.<sup>[2]</sup>

Data Mining<sup>[4]</sup> evolved from a simple extraction of raw data to an analytical process of exploring large amount of

data in order to cite the common denominators or patterns. Basically, data mining involves three processes  
 Step 1: Exploration - involves data preparation  
 Step 2: Model building and Validation - involves choosing the ones that are best suited to be used

Step 3: Deployment - involves using the chosen data to proceed with the generation of the outcome

Data visualization is the core aspect of our system. The main goal of Data Visualization is to communicate information clearly and effectively through graphical

means. It doesn't mean that data visualization needs to look boring to be functional or extremely sophisticated to look beautiful. To convey ideas effectively, both aesthetic form and functionality need to go hand in hand, providing insights into a rather sparse and complex data set by communicating its key-aspects in a more intuitive way. Yet designers often fail to achieve a balance between form and function, creating gorgeous data visualizations which fail to serve their main purpose — to communicate information<sup>[5]</sup>

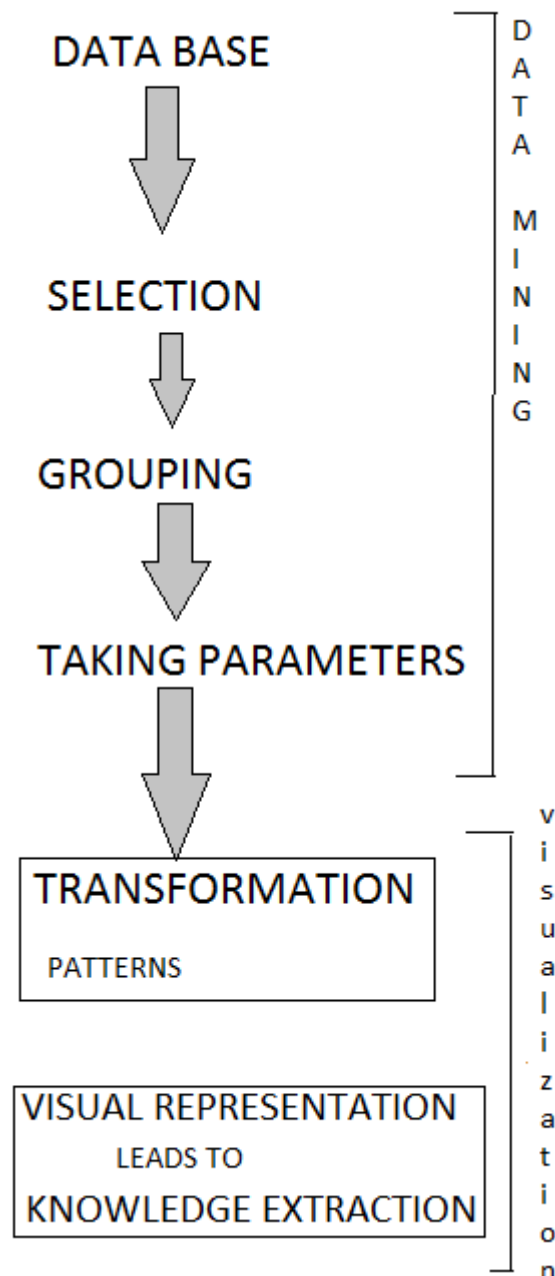


Fig 1: process of data mining.

Although there are several traditional data mining methods already in use for this particular purpose, they face several shortcomings and limitations. Also extraction of knowledge from such primitive methods is a daunting task. The efficient extraction of hidden information

requires skilled application of complex algorithms and visualization tools, which must be applied in an intelligent and thoughtful manner based on intermediate results and background knowledge. The whole KDD process is therefore difficult to automate, as it requires high-level

intelligence. By merging automatic and visual mining, the flexibility, creativity, and knowledge of a person are combined with the storage capacity and computational power of the computer.

#### 4. The System

Our system allows two types of users namely, the administrator and the test taker. The client must before hand specify the number of administrators that (s)he wishes to employ. The client must also provide information regarding the registration information for each test taker. Each new test taker will be given a user name and password after registration. The proposed system facilitates the testing on various subjects and not just a single subject, the subject selection is done once the user completes his/her registration. The test taker is shown his/her performance after the completion of the test.

To the test taker, our system appears as any other test taking software, but the real purpose of our software is to facilitate the test conductor in making the procedure of result analysis easier. Further more by using data visualization we can unearth several hidden aspects of the test taker, like for example if the test taker is giving fluke answers or if (s)he is following a premeditated answering pattern.

There are several third part softwares that specialize in converting given data into our preferred graphical representation. But such softwares are not cheap. Our project uses JFreechart which is a third Party API used to maintain Data Visualization Techniques. Although the system may not be high end software, it however does the same job at a lower cost.

The system treats each web page as an event (each web page contains one question) and monitors/records the operations performed by the test taker on each event. The data that is recorded is then used for dynamically generating the visual interpretation for the administrator.

The system is designed to be the cross platform supportable. The system is supported on a wide range of hardware and any software platform which is having JVM built into the system. This application is being developed using J2EE; hence it is extremely portable.

The system is designed with good user interface so as to provide the tutors with the ability to generate graphs from data and adjust the graph parameters dynamically.

#### 5. Modules

- Admin
- Users

The main functionality of Admin is deriving Behavioural Patterns of the users and monitoring those patterns through visualization Technique. The sub module s of Admin are

**Add Questions** Here Admin can perform adding questions of various subjects like c, c++, java, oracle etc.

**Subject Selection** Here Admin can choose various subjects to add questions for E-Testing system for Learners.

**Modify Questions** Here Admin can modify the questions as well as answers of various subjects.

**Delete Questions** To maintain test quality Admin has a facility to delete the questions of various subjects.

**Monitor Users Through Data Visualization Technique** Here Admin is monitoring the users Behavioural Patterns through visualization technique

**Behavioural Patterns Monitored By Admin**

Type of questions performed by users. Time taken for each question while attempting the exam. How many times the users change the options in each question. Total time taken for E-Testing completion.

**Change Password**

Admin can change the password .

**Logout**

Here Admin is navigated to Homepage.

The user has registered to attempt E-Testing system. The sub modules of the user are

**Login**

User should Login to perform the E-Testing system.

**Register**

User should submit details while registration time.

**Subject Selection**

Before going to E-Testing system page user should select any subjects like java, c, c++, oracle or any other subject that the test centre is supposed to conduct the test for

**Attempting E-Testing System**

User should attempt the E-Testing system. Here users' Behavioural Patterns are captured and those are stored in Database for future use.

**Logout**

Here user is navigated to Homepage.

#### 6. JDBC CONNECTIVITY

The JDBC provides database-independent connectivity between the J2EE platform and a wide range of tabular data sources. JDBC technology allows an Application Component Provider to: Perform connection and authentication to a database server, Manages transactions, Move SQL statements to a database engine for preprocessing and execution, Execute stored procedures, Inspect and modify the results from Select statements

The generated application is the first version upon the system. The overall system is planned to be in the format of distributed architecture with homogeneous database platform. The major objective of the overall system is to keep the following components intact, System consistency, System integrity, Overall security of data, Data reliability and Accuracy, User friendly name both at administration and user levels, considering the fact of generality and clarity.

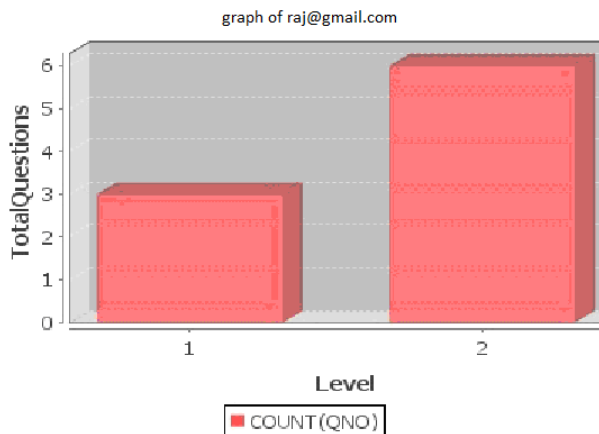
#### 7. Knowledge Discovery Through The generated Graphs

Once all the test takers have completed taking the test, the process of knowledge discovery starts. The test

conductor has to decide which candidates can be selected for the limited amount of seats available.

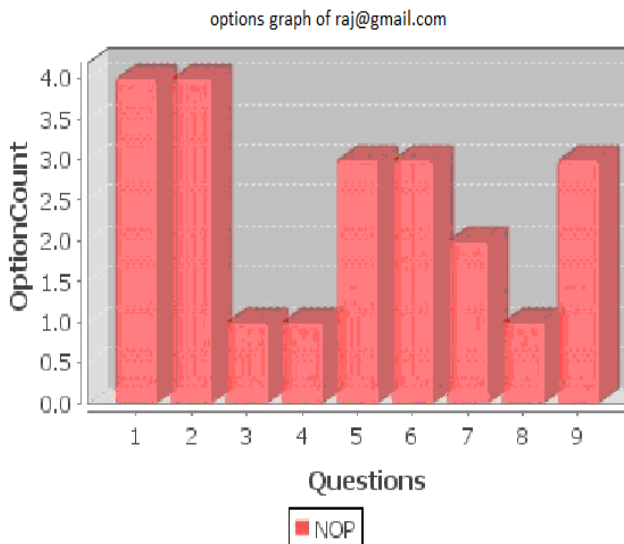
The following few graphically generated charts will help him in doing so.

The test conductor will be able to check various questions (belonging to different levels of difficulty) that a particular test taker answered. Answering more questions from the “difficult level” means he is potentially the better candidate. As it can be see from the following graph:



Another important aspect that our system monitors is the number of times the test taker has changed the option for a particular question. This is a vital aspect because it helps the test conductor to identify which candidate is confident in giving answers. More number of option changes define a test taker as a weak candidate.

Although the test taker will not be known about this aspect, or else the whole idea will be annulled.



The graph shows the change of options for each question for a particular test taker. It can be deduced from the graph that the person is not so confident about answering the first and second questions while he has no problems with the third and fourth questions.

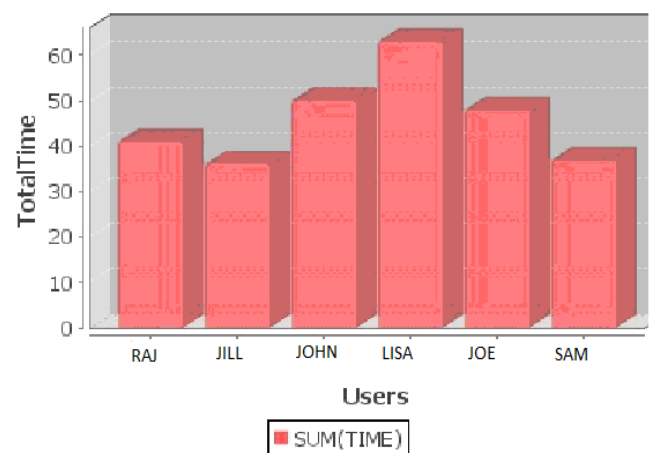
The next aspect that our system considers is the time taken for answering each question. The test designer predefines an amount of time that will be required to solve

a particular question, if a candidate is done with that particular question before that time, it can be considered that (s)he might be guessing the answer. This can be a very important consideration when short listing the candidates.

Graphs are also generated for the total time taken by each candidate to complete the entire examination.

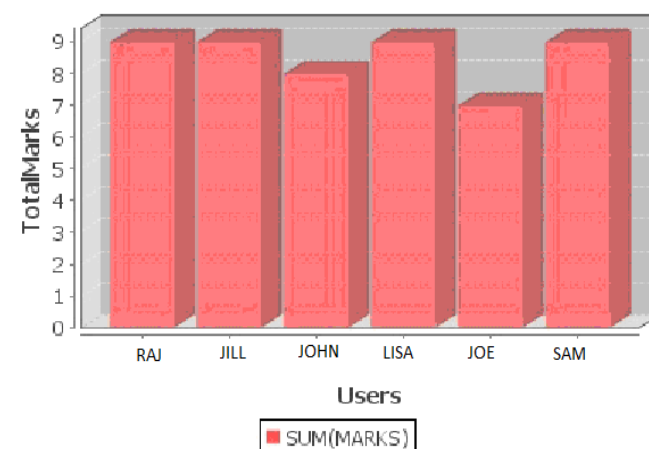
The graphs can be compared with each other to obtain data that will help the test conductor further in selecting the best candidates from the attempting lot. While analysing the time graph, it is important to keep in mind that a candidate taking the least amount of time to complete the exam is not necessarily the best candidate and similarly the candidate who takes the most time is not necessarily the worst candidate, for all we know that the candidate taking the least amount of time may be guessing all the answers.

### Multi Graph of Time Lines



Finally the system also generates a graph that collectively shows the marks of all the candidates together thus giving the test conductor a broad idea of the performance of each candidate. This graph is very important because it shows the total marks of each student, which is the main criterion for selection.

### Multi Graph of Marks



## 8. Conclusion

This system presents an approach to let tutors monitor learners' strategies during online tests. The approach exploits data visualization to draw the data characterizing the learner's test strategy, in order to trigger the tutor's attention and to let him/her discover previously unknown behavioural patterns of the learners and conceptual relationships among test items. The time taken for each question, level of questions answered, options changed for each question and the total time taken for answering the test are the behavioural patterns taken into consideration in this project. In this way, the tutor is provided with a powerful tool that let him/her review the whole process and evaluate in possible improvements.

## 9. Advancements

Aspects There is a lot of scope for modifying this system to achieve further levels of abstraction, to separate the most closely contested candidates. Aspects such as number of questions left un-answered also can be incorporated into the system

## REFERENCES

- [1] Fayyad, Usama; Gregory Piatetsky-Shapiro, and Padhraic Smyth (1996). "From Data Mining to Knowledge Discovery in Databases". Retrieved 2008-12-17.
- [2] "Data Mining Curriculum". ACM SIGKDD. 2006-04-30. Retrieved 2011-10-28.
- [3] Clifton, Christopher (2010). "Encyclopædia Britannica: Definition of Data Mining". Retrieved 2010-12-09.
- [4] <http://www.the-data-mine.com/Misc/DataMining>
- [5] **Monitoring On Line Tests Through Data. Visualization.** Gennaro Costagliola, Vittorio Fuccella, Massimiliano Giordano, Giuseppe Polese.
- [6] Dietel and Dietel, 3rd edition," internet and world wide web".
- [7] D.A. Keim, "Information Visualization and Visual Data Mining,"IEEE Trans. Visualization and Computer Graphics, vol. 8, Jan.-Mar. 2002.
- [8] J. Bath, "Answer-Changing Behavior on Objective Examinations,". Educational Research, no. 61,.

## Vitae

<sup>2</sup>**J.Vikram Datta:** He is a student of fourth year B.tech of Electronics And Computers Branch from Koneru Lakshmaiah college of engineering.



<sup>2</sup>**M.Shabaaz:** is a student of fourth year B.tech of Electronics And Computers Branch from Koneru Lakshmaiah college of engineering



<sup>2</sup>**K.Lokesh** is a student of fourth year B.tech of Electronics And Computers Branch from Koneru Lakshmaiah college of engineering

