International Journal of Information Technology and Knowledge Management
January-June 2011, Volume 4, No. 1, pp. 69-76

ISSUES INVOLVED IN DESIGNING OF USER INTERFACES TO FETCH QUALITY DATA FOR AUTOMATION OF RESULTS IN A TYPICAL UNIVERSITY EXAMINATION SYSTEM

Vikas Sharma[1] & Anjana Sharma[2]

Data or Information quality is a critical issue in majority of information intensive organisations especially for the universities who have to deal with huge volume of students' personal data and awards for compilation of results. The whole credibility of University Examination System is dependent on accuracy and timely declaration of results in addition to the information conveyed to the students well in time. The appearing students' data is the base data for automation of results in addition to ensure excellence in support services which is further evaluated using different quality yard sticks- accuracy, completeness, timeliness, reliability and believability. Gathering of students' information on well documented format can minimize the risks of data loss, inaccuracy, incompleteness, etc. In this paper, an attempt is made to analyse the data quality of traditional examination form being used to gather information of undergraduate students of Himachal Pradesh University, Summer Hill, Shimla followed by designing of Intelligent Character Recognition (ICR) compatible examination form and development of an electronic interface "Student Management Information System (SMIS)" software prototype to maintain data quality at source rather than to do later stage manual efforts of data cleaning.
Keywords: Data Quality, Information Quality, ICR, Accuracy, Completeness, Data Cleaning, Reengineering

1. Introduction

Due to the impact and the advents in Information and Communication Technology (ICT), nowadays more emphasis is being given on information or data quality by the data intensive organisations. Data is raw facts whereas the information is described as processed form of data and useful to make decisions. The conversion of data into information is a continuous process because processed form of data may be information for someone to make decisions but for others a figure or raw fact. The users are the real observers to differentiate between these two terms as per their requirements. In this paper, the information quality (IQ) or data quality (DQ) terms have been used interchangeably until mentioned explicitly in some special context. The high-quality data has emerged as a new basis for competition in this world. Today's business strives to offer agile, highly customized, and globally sustainable solutions by producing high quality information as a baseline for managing strategic corporate capacity and assets [4]. The growth of data warehouses and direct access of information from various sources have also increased the necessity of awareness about high-quality information [5]. The data quality (DQ) is a comprehensive term and has different meanings to the different users. It goes beyond accuracy and includes other dimensions such as completeness, timeliness, accessibility, security, etc. Assuring the quality of information is both important and difficult. But achieving high-quality information is a battle that is never really won, in part because what constitutes victory is not clear, as different parties have differing views as to the definition of success [2]. The widely used definition for IQ is "fitness for use." If users of the data feel that quality can be described by such attributes as accuracy, completeness, timeliness, and consistency, is sufficient for their needs, then, from their perspective, the quality of the information available to them is fine [2]. Since the data are increasingly used to support organizational activities, it is obvious that poor data quality may negatively affect organizational effectiveness and efficiency. Such negative impacts are often manifested as the direct and indirect costs of prevention [1]. Errors in data cause a variety of problems and raise the costs in several areas. The cost to recognize and detect errors is not trivial. Detecting data errors in programs often takes as much, or more, of the analysis and programming efforts than the main logic. The earlier an error is detected, the cheaper it is to correct [7]. Despite the potential costs of bad data quality, organizations often ignore data quality altogether. The data quality cannot be improved independently of the processes producing the data, rather than the context in which the data is to be used [1]. As we are aware that the strong push to gain competitive advantage and meet emerging demands of the users has increased the volume, variety and sources of data. Similarly, large data volumes, heterogeneous data types, widely distributed data sources and multiple stakeholders characterise a typical University Examination System. The very existence of any organization

[1]International Centre for Distance Education and Open Learning, Himachal Pradesh University, Summer Hill, Shimla, India

[2]Web Tech Computer Education, Sanjauli, Shimla, India

Email: [1]vikas_doeacc@rediffmail.com, [2]anjana_vashisht@yahoo.com

can be threatened by poor information quality (IQ). The data on which examination results are based and upon which the future of thousands of students depends if inaccurate, out-of-date, incomplete or has other types of problems can put a big question mark on the credibility of the university. The automation of such examination system demands high data quality management system in place to avoid garbage-in-garbage out. Various studies reveal that it is better to curb data errors at source rather than to wait for their reflection in other sub-systems. The Examination Wing of Himachal Pradesh University, Summer Hill, Shimla collects students' detail using traditional examination forms followed by manual data entry for compilation of results. But the data collected using these traditional forms is not 100% fit for compilation of results and generations of various pre examination reports such as cut lists, centre statements, admit cards, signature lists, etc. This data is also not completely healthy for compilation of results and generations of post examination reports such as tabulation sheets, gazettes, result cards, tabulation sheets in addition to various decision making reports for authorities. So a need was felt to design ICR compatible examination form by reengineering old form and development of user-friendly, robust, secure and reliable data entry software for effective management of students' database.

2. OBJECTIVES

1. To evaluate the opinions of the examination branches on data quality dimensions considered important during automation of results.
2. To study the limitations of traditional examination form being used by the Himachal Pradesh University to collect appearing students' detail.
3. To design ICR compatible examination form for automatic conversion of handwritten data into machine readable form and development of software prototype for manual data entry of students' detail to mitigate the risks of data quality.
4. To recommend on how to improve data quality by fetching data at source.

3. RESEARCH DESIGN AND METHODOLOGY

This study is conducted in the Examination Wing of Himachal Pradesh University, Summer Hill, Shimla and specifically on three examination branches - BA, B.Com and B.Sc. Part-I, II & III whose results are complied using computers. A convenient sample survey of 11 dealing persons who assist in compilation of results of undergraduate classes was collected. The above sample comprises one dealing person from each volume (volume contains around 30 examination centres) of undergraduate section of Examination Wing of Himachal Pradesh University. There are total eleven volumes - three each from B.A. Part-I & II, two from B.A. Part-III, one each from B.Sc. and B. Com Part-I, II and III. These 11 dealing persons were further asked to collect random sample of 5 candidates of their respective volumes using convenient sampling technique to make a total of 55 candidates for analysis of quality of data collected using traditional examination forms. The data collection tool used here was self designed questionnaire having two parts: 1) first part was used to rate the importance of data quality parameter in context of University Examination System, and 2) second part was used to observe the quality of data from already received examination forms. A 5-point Likert Scale (5 = highly important and 1 = not important at all) was used to observe the opinion of the dealing persons. To build an initial list of data qualities, the fifteen data quality dimensions defined by Wang, et. al. (1994): 1) access security, 2) accessibility, 3) accuracy, 4) appropriate amount of data, 5) believability, 6) completeness, 7) concise representation, 8) ease of understanding, 9) timeliness, 10) interpretability, 11) objectivity, 12) relevancy, 13) representational consistency, 14) reputation, and 15) timeliness were discussed in detail followed by brainstorming sessions to conclude a raw list relevant items in context of automation of examination system. These items were arranged then in logical order to give a questionnaire format. Using the literature on information quality and by looking carefully for relevance of data qualities in context of examination system, the items in the questionnaire were reduced to a more manageable 12 items.

4. RESULTS AND DISCUSSION

4.1. Data Quality of Existing Traditional Examination Form

Table 1 shows the summary average for weighted and unweighted sets. The importance score (IS) shows the average importance ranking of each question as rated by 11 dealing persons. The data qualities considered most important e.g. above the upper quartile (4.82) are all about accuracy (sr. no. 1) and appropriate amount or quantity of data for compilation of results (sr. no. 2). The data quality dimensions considered least important, e.g. below the lower quartile (4.09) are relevancy (sr. no. 7) and value added services (sr. no. 11). These two data quality dimensions have been discarded for further analysis and discussion in this paper. Other questions are in between and the median for above importance score is 4.36. The standard deviation shows that there is less variance on data quality opinions. The average scores per data item for traditional examination form are displayed in two modes: 1) raw score (RS) as unweighted ratings (with a theoretical Likert Scale range of 1 to 5), and 2) weighted score (WS). The weighted score

(theoretically ranging from 1 to 25) is obtained by multiplying the raw score by the importance score. It becomes very difficult to analyse the data quality of traditional examination forms, so Data Quality Index (DQI) was calculated using weighted score against the weighted importance score (WIS) -Importance Score X 5. The traditional examination form being used by the Himachal Pradesh University to collect data of appearing candidates scored just 0.69 points in context of data quality. In descending order of data quality available through traditional undergraduate examination form, the timeliness dimension scored highest (0.91 points) followed by accuracy of data (0.83 points), reputation of data (0.73 points), ease of understanding and accessibility (0.68 points each), appropriate volume of data (0.67 points), completeness of data (0.65 points), interpretability of data 0.63 points, representational consistency (0.61 points) and securely retrieval of data (0.45 points).

Table 1
Summary Averages for Weighted and Unweighted Data Sets

Sr. No.	Quality Dimensions	Descriptions	IS	SD	WIS	RS	WS	DQI
1	Accuracy	Data must be correct, reliable and certified i.e. free of errors.	5.00	0.00	25.00	4.14	20.70	0.83
2	Appropriate Amount of Data	The quantity or volume of available data must be appropriate for compilation of results	4.91	0.30	24.55	3.36	16.50	0.67
3	Completeness	Data must be sufficient for compilation of results and generation of pre and post examination reports.	4.82	0.40	24.10	3.27	15.76	0.65
4	Ease of Understanding	Data must be clear, without ambiguity and easy to comprehend.	4.82	0.40	24.10	3.38	16.29	0.68
5	Timeliness	The age of the data must be appropriate for the task at hand.	4.64	0.50	23.20	4.56	21.16	0.91
6	Reputation	Data must be trusted or highly regarded in terms of their source or content.	4.55	0.68	22.75	3.65	16.61	0.73
7	Relevancy	Data must be applicable and helpful for the task at hand.	4.00	0.63	-	-	-	-
8	Representational Consistency	Data must always be presented in the same format and compatible with previous data.	4.18	0.87	20.90	3.03	12.67	0.61
9	Accessibility	Data must be available easily and quickly.	4.09	0.83	20.45	3.40	13.91	0.68
10	Interpretability	Data must be in appropriate language, units and the data definitions must be clear.	4.18	0.60	20.90	3.16	13.21	0.63
11	Value-Added	Data must be beneficial and provide advantages for value added services.	3.82	0.83	-	-	-	-
12	Access Security	Access to data must be restricted and hence kept secure.	4.09	0.53	20.45	2.27	9.28	0.45
		Total	45.28	—	226.4	34.22	156.08	0.69

IS-Importance Score, SD- Standard Deviation, WIS- Weighted Importance Score, RS-Raw Score, WS-Weighted Score, DQI-Data Quality Index

The above DQI indicates that it is becoming very difficult in the University Examination System to maintain data quality using traditional examination form. So there is a need to provide some state-of-the-art solutions to minimise the risks of data quality by reengineering the traditional examination form and designing of interactive user interface utility.

4.2. Technological Solutions to Maintain Input Data Quality

4.2.1. ICR Compatible Examination Form

Today we are living in an era of Information and Communication Technology where offices, schools, colleges and universities are becoming paperless offices. But paper forms are still convenient and less expensive data capture tools for many individuals residing in remote areas who don't have computing and communicating facilities.

Even well-connected individuals do not always have computer connected when they need it. A well designed and easy to use formatted document helps to obtain accurate and complete data right from source.

The table 2 shows the list of changes incorporated in traditional examination form to design an ICR compatible formatted document to fetch quality data at source.

Table 2
List of Changes Incorporated in Traditional Examination Form to Design an ICR Compatible Formatted Document to Fetch Quality Data at Source

Sr. No.	Quality Dimensions to Achieve	Changes Incorporated
1	Accuracy	1. Complete instructions for the candidates regarding filling of their personal and examination detail.
		2. Usage of codes rather than writing full text e.g. Sex Codes, Paper Codes, Exam Centre Codes, Discrepancy Codes, Capacity Codes, Category Codes.
		3. Allocation of separate space for the university fee receipt stamp to avoid overlapping with filled text.
		4. Allocation of space to put discrepancies in coded form on the front page to avoid over writing and clear understanding.
2	Appropriate Amount of Data	1. Allocation of separate space on the first page of the form to put discrepancy (if any). The clear-cut discrepancies play vital role during compilation of results.
		2. Allocation of proper space to fill information in the designated boxes only.
		3. Logical arrangement of data in blocks form with eye catching bold headings/titles.
		4. Complete guidelines for the candidates throughout the examination form to promote him/her to add complete detail.
3	Completeness	1. Encouraged students to use data codes rather than writing of full text.
		2. List of exam centres, paper codes along with examination form as enclosure.
		3. Ample space for writing of personal detail, appearing exams detail, communication address, etc.
4	Ease of Understanding	Encouraged candidates to add their detail in capital letters only for easy recognition and understanding.
5	Timeliness	1. Inclusion of last date of receipt of forms with and without late fee on the front page of examination form with the cost of form.
		2. Proper space allocation for examination month, year, class, year/semester.
6	Reputation in terms of source or content.	1. Space for candidate's signature and specimen handwriting for authenticity.
		2. Space to certify regular candidate's detail by the concerned college principal.
		3. Ample space allocation to certify detail of private candidates by the class-I or gazetted officer.
8	Representational Consistency	Single examination form for the regular and private candidates.
9	Accessibility	1. Routing of regular candidates' examination form from their concerned colleges through principal.

Table. 2 Contd...

Table. 2 Contd...

		2. Collection of private candidates' examination forms at university counters/post mail.
10	Interpretability	1. Usage of English language.
		2. Usage of Hindi Language to fill candidate's name, father's name and mother's name in addition to English language.
12	Access Security	1. Usage of bar code and hologram on the front page of form to avoid duplicity.
		2. Usage of special droppable ink to extract data using ICR.

The figure 1 shows the designed ICR compatible examinations form for undergraduate classes on legal size paper with red ink droppable during recognition process.

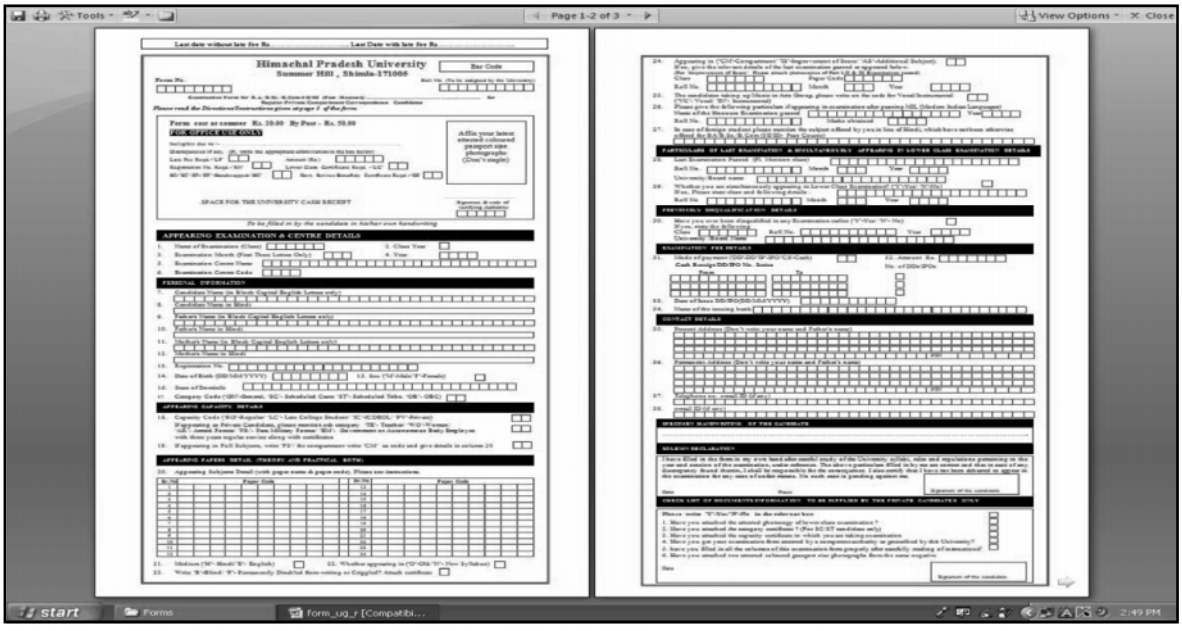

Fig. 1: ICR Compatible Examination Form

4.2.2. Electronic Interface for Manual Data Entry

A small software prototype Students Management Information System (SMIS) was developed to do manual data entry of candidates' records where ICR system is not applicable. The main objective behind the development of this utility was to obtain reliable, authentic, accurate, complete data of regular candidates from their respective colleges. The figure 2 shows the relationship among entities to maintain data integrity and consistency in SMIS prototype.

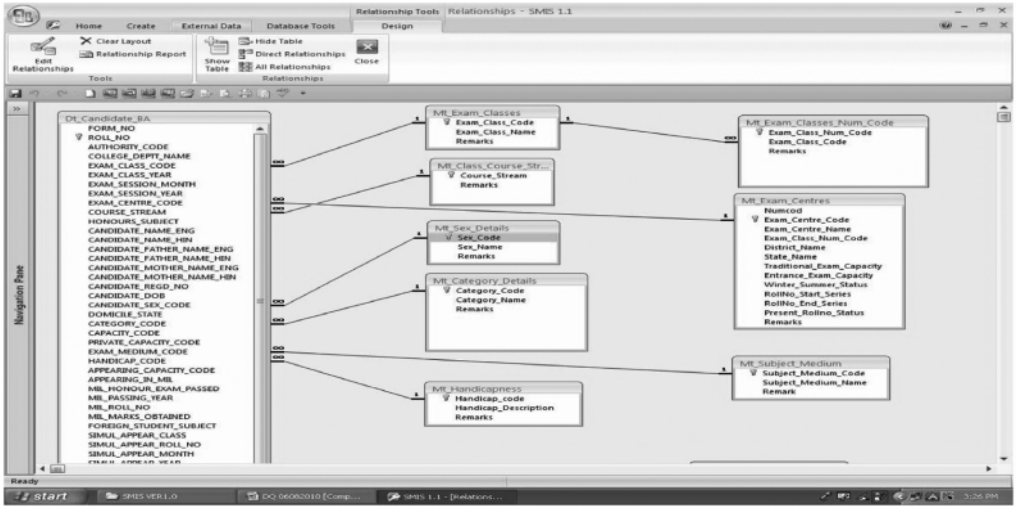

Fig. 2: Relationship of Entities in SMIS

The SMIS prototype was developed using Microsoft Access 2000 because it is easily available within the university/college environment. Further, Microsoft Access is also easy to install, manage and to take backup of data. Inaccurate input data are the most common cause of the errors in data processing. Errors entered by the data entry operator can be controlled by perfect input design. The perfect input design facilitates to convert user-originated inputs to its computer-based format. The main goal of designing input interface is to make entry as easy, logical and free from errors as possible. During input design process, there is a need to know the following tips: 1) allocate space for each field, 2) field sequence must match with the source document, and 3) specify data format [2]. The figure 3 shows the screen shots of Student Management Information System ver. 1.1 used to fetch quality data at source.

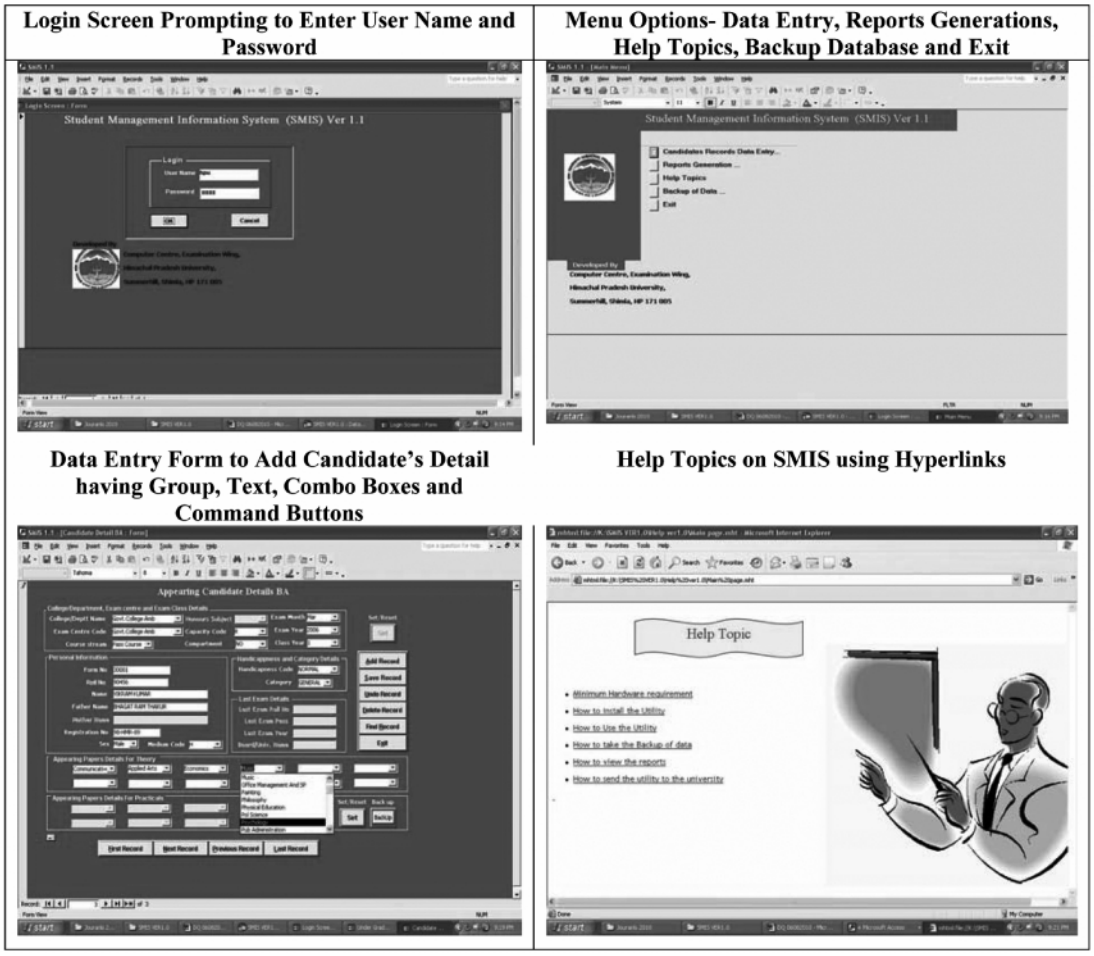

Figure 3: Screen Shots of Student Management Information System Ver 1.1 Utility

The table 4 shows the provisions made in data entry utility to fetch quality data during manual data entry process.

Table 3
Provisions Made in Data Entry Utility- Students Management Information System to Fetch Quality Data During Manual Data Entry Process

Sr. No.	Quality Dimensions to Achieve	List of Actions Taken
1	Accuracy	1. Used entity integrity and referential integrity rules to maintain database consistency and integrity.
		2. Used data validation checks during data entry process.
2	Appropriate Amount of Data	1. Provision of mandatory fields to promote the user to enter complete detail of candidates such as roll nos, examination centre, appearing paper codes, etc.
		2. Usage of combo boxes to add only appropriate amount of data such as examination centre, month, year, appearing capacity, etc.

Table. 3 Contd...

Table. 3 Contd...

3	Completeness	1.	Data fields are arranged logically to encourage user to add complete detail of candidate.
		2.	The candidate information is divided into blocks using group boxes to promote user to fill complete information.
4	Ease of Understanding	1.	Help facility (hyperlink) is available on installation and usage of utility for new users.
		2.	The first form pops up at startup of utility which prompts for user name and password followed by menu of choices for classes, data entry, report generation, help, etc. using switchboard options.
		3.	For easy data entry, the layout of the data fields is designed according to the physical arrangement of the data items on the examination forms.
		4.	The data items are further grouped logical and divided into blocks for clear visibility and understanding.
		5.	Every form and reports have unique titles for easy recognition.
5	Timeliness	1.	The examination month and year options are enabled only for the current going examination sessions.
		2.	Examination type such as regular, compartment or reappear is asked before entering candidate's detail
6	Reputation	1.	The data entry using this utility is done at the colleges' end for the regular candidates whereas for the private candidates, it is done in the Computer Centre of Examination Wing.
		2.	The computer generated pre examination reports such as cultists, centre statements, admit cards, signature lists, etc. are verified by the concerned examination branch officials from the original forms to maintain data quality.
8	Representational Consistency	1.	The structure of table is unique for all types of candidates' data- regular, private, etc.
		2.	The generated reports have also the unique format for types of candidates.
		3.	Proper error handing mechanism to avoid crash of application.
9	Accessibility	1.	User-friendly interfaces to use this utility using linked switchboards, forms and reports.
		2.	Menu based options for better accessibility of application.
		3.	Help facility for easy installation and usage of utility.
10	Interpretability	1.	The English language is used for data entry.
		2.	The reports generated using this utility for pre examination phase has unique format and easy to understand.
12	Access Security	1.	Accessible to only authorized users who have assigned user name and password.
		2.	The other background details of MS-Access are hidden to the users to avoid them to make changes/customize the utility.

5. CONCLUSIONS AND RECOMMENDATIONS

Assuring quality of information is not only important but also difficult to achieve. It has different meanings to different users but in automation of examination system, the quality is determined on accuracy, completeness, timeliness, consistency, secure access, etc. The quality of input data-students' detail plays very important role for generation of various pre and post examination information in addition to reports meant for different categories of user. Implementation of ICR compatible examination forms can help the universities to curb input data quality anomalies at source without indulging extra human efforts. Further, designing of small software prototypes having user-friendly interfaces, proper data validation and integrity checks, help topics, etc. can also help the universities to collect accurate, valid, complete and authentic data by implementing such utilities at colleges' level to fetch regular candidates' data. This will also help the universities to save data entry costs and additional human efforts. But to make these state-of-the-art solutions a success story, there is a need to have a proper monitoring system of technical and managerial persons to assist each other from collection of data to compilation of results. In addition to above, efforts should be made to enhance awareness level among students on usage of ICR compatible forms and proper training to data entry operators on feeding complete, accurate, valid and authentic details of candidates.

REFERENCES

[1] Abate, Marcey L., Kathleen V. Diegert. (1998), "A Hierarchical Approach to Improving Data Quality," Sandia National Laboratories and Heather W. Allen, Heather Allen & Associates, Data Quality, 4, No.1.

[2] Awad, Elias. M. (1990), "Input/Output and Forms Design, System Analysis and Design", New Delhi: Galgotia Publications.

[3] Ballou Donald, Stuart Madnick, and Richard Wang. (2004), "Assuring Information Quality", Journal of Management Information Systems, < www.jmis-web.org/articles/v20_n3_p9/index.html>, 20, Nov. 3.

[4] Lee, Yang W. (2004), "Crafting Rules: Context-Reflective Data Quality Problem Solving", Journal of Management Information System, 20, No. 3.

[5] Wang, Richard, Diane Strong, and Lisa Guarascio. (1994), "An Empirical Investigation of Data Quality Dimensions: A Data Consumer's Perspective", Working Paper TDQM-94-01, MIT TDQM Research Program, E53-320, 50 Memorial Drive, Cambridge.

[6] Wang, Richard, Thomas Allen, Wesley Harris and Stuart Madnick. (2003), "An Information Product Approach for Total Information Awareness" <web.mit.edu/tdqm/www/tdqmpub/IEEEAero03.pdf>

[7] Viking. (2009), "The Importance of Power/Precision Data Entry to Document Imaging, A White Paper", 2005. Viking Software Solutions. <http://www.vikingsoft.com/pdf/importanceofppde.pdf>