Project Report

1 MCQ Exam Data Processing

1.1 Data Pre-processing

After scanning all MCQ exam papers, data that comes sometimes is in Comma Separated Value(CSV) format and sometimes is in normal text file(TXT) format. In various scanning machines several types of formats are used for scanned data.

So main task in data pre-processing step is converting all files in one format, either CSV or TXT. But most of the time CSV is preferred for convenience. Based on the machine used for scanning, structure of the scanned data is changed to a common format. For example, some machines don't use any serial number. In that case serial number is added in the files that scanned with that specific machine.

Generally column names also added during Data Pre-processing step.

1.2 Data Merging

After completing Data Pre-procssing step, all scanned data from various directories are merged into one single file. Further queries are made on this file.

1.3 Data Cleaning

1.3.1 Separating Duplicate Data

The first step of Data Cleaning is separating duplicate data. Duplicated data generates when same file is scanned multiple times. This checking is done by checking the duplication of Litho Code 1, Litho Code 2 and Roll Number fields of Top answer sheet. For Bottom answer sheet Litho Code 1, Litho Code 2 and Answer fields are checked.

1.3.2 Generating Subject Code

For Top answer sheets subject code is generated based on the first digit of a student's Roll and then add it to the main file. As sometimes student made mistake in Subject Code so the subject code filled up by student is ignored and Generated Subject Code field is used for further use.

1.3.3 Separating Invalid Roll

For Top answer sheets Roll field is checked with student's previous database. If a roll is absent in database that roll is marked as invalid roll and separated from the file.

1.3.4 Separating Invalid Set Code

For Top answer sheets Set Code field is checked. If no set code or multiple set code is filled up then that entry is marked as invalid and separated from the file.

1.3.5 Separating Unmatched Litho Code

For both Top and Bottom answer sheets, Litho Code 1 and Litho Code 2 are checked if two are matched or not. If for one entry Litho Code 1 and Litho Code 2 are not matched than the data point is separated from file. After re-scanning the corresponding answer sheets the data is entered the main file again.

1.3.6 Separating Duplicate Roll

For Top answer sheet if two entry contain same Roll number then those entries are separated from the file. After that corresponding answer sheets are carried out. When checking is done that which entry is actual and which entry is wrong, right entry is set back to the main file.

1.3.7 Generating Report Based on Venue

Based on the Venue Code field in Top answer sheet, a report is generated. This report contains venue wise, subject wise total student count.

1.4 Matching Top and Bottom Data

After completing the Data Cleaning step, Top and Bottom data are matched based on the Litho Code 1 and Litho Code 2 fields from both Top answer sheets and Bottom answer sheets. Four time this checking is done. First time, Litho Code 1 field of Top answer sheet is checked with Litho Code 1 field of Bottom answer sheet. The entries that matched during this checking, that are merged together as same entries. Second time Litho Code 1 from Top answer sheets are checked with Litho Code 2 of Bottom answer sheets. Third time Litho Code 2 of Top answer sheets are checked with Litho Code 1 of Bottom answer sheets. And finally Litho Code 2 of Top answer sheets are checked with Litho Code 2 of Bottom answer sheets.

After this four cross checking, the unmatched entries are separated. Further processing is done on file where Top data entries and Bottom data entries are merged.

1.5 Result Generation

The entries that are merged in previous step, Answer field is added there from Bottom answer sheet. Based on the Subject Code and Set Code field the Answer field is checked with the given solution. Based on the checking, mark is calculated and new Mark field is added to the file that is used.

2 Written Exam Data Processing

(As most of the operations for written exam data are same as MCQ exam data so details of the operation is skipped here.)

2.1 Data Pre-Processing

2.2 Data Merging

2.3 Data Cleaning

2.3.1 Separating Duplicate Data

2.3.2 Generating Subject Code

2.3.3 Separating Invalid Roll

2.3.4 Separating Invalid Set Code

2.3.5 Separating Unmatched Litho Code

2.3.6 Separating Duplicate Roll

2.3.7 Adding Written Mark Field

For written exam, two type of written marks are in Bottom sheets. Examiner Mark and Head Examiner Mark. If for any entry Head Examiner mark existed then that mark is treated as Written Mark on the other hand Examiner Mark is treated as Written mark. Based on this two field, Written Mark column is added to the file.

2.3.8 Separating Invalid Written Mark

If any entry contains a mark that is either a non numerical character or grater than 100 then that entry is treated as invalid mark and separated from the file. After checking, the entry is either added or remain separated based on the situation.

2.3.9 Generate Report Based on Venue Code

2.4 Matching Top and Bottom Data

2.5 Result Generation

Result generation for written exam is little different from MCQ exam. As mark is already in Bottom sheets, so based on that mark result is generated.

3 Venue Distribution

For venue distribution, two files are provided. One file comes from Deputy Commissioner's office that contains student capacity for each available cen-

ter. Another file comes from Teletalk, this file contains each student's information like Roll, Subject Code etc.

The main task here is to assign venue to each student so that the capacity of venue doesn't exceed. This process have to done for each center. And one another issue needed to handle that a student from school level should must have the venue that is marked as school level. Similarly for a student from college level should must have the venue that is marked as college level.

4 Grouping for Viva

After getting the list of eligible student for viva and similar subjects, each student is assigned to a specific group for a specific date and time.

5 Viva Taking Interface

In this step, a Graphical User Interface is designed. In this interface, a teacher have to log in according to Group Number, Board Number and viva Date.

Once logged in successfully, for given Group number, any valid Roll with relevant information can be retrieved. After completing the viva obtained marks also can be added to the file using the marks field in the interface.

This interface also provide a utility of showing the report for the given Group. In this report, how many student already attended in viva and how many absent are also shown.