ASIAN JOURNAL
ISSN: 0976-3376

Available Online at http://www.journalajst.com

ASIAN JOURNAL OF SCIENCE AND TECHNOLOGY

Asian Journal of Science and Technology
Vol. 4, Issue, 12, pp. 007-010, December, 2012

RESEARCH ARTICLE

AUTOMATED DATA CAPTURING AND RECOGNITION TECHNOLOGY TO MINIMIZE HUMAN INTERVENTION IN UNIVERSITY EXAMINATION SYSTEM

***Mohini and Dr. Amar Jeet Singh**

Department of Computer Science, Himachal Pradesh University, Shimla, India

Received 29th September, 2012; Received in Revised from; 28th October, 2012; Accepted 11th November, 2012; Published online 17th December, 2012

ABSTRACT

Manual data entry from hand written data is very time consuming, tedious and prone to many errors more so where there is bulk amount of data is involved. This is the area where there is a requirement of automated data capturing and recognition technology. Using a recognition engine to convert text or handwriting from the printed page into computer readable characters saves up to 90 percent of the time it would take to enter the information manually. To fetch the data with minimal human intervention there are three types of recognition engines for this purpose and these are Optical Character Recognition (OCR), Intelligent Character Recognition (ICR) and Optical Mark Reader (OMR). Generally. OCR is best for machine print, or type, ICR is better for converting handwriting data and OMR is best suitable to detect the absence or presence of a mark, but not the shape of the mark. Automated data capturing is rapidly becoming an integral and necessary component in any organization. This not only saves the cost but also increases the speed and accuracy over manually entered data. This paper report on the study of use of these technology in various processes of examination system especially in fetching awards from OMR technology to get almost hundred percent accuracy.

Key words: OCR, OMR, ICR, AutoRec.

INTRODUCTION

Due to technological advancement human has been successful in getting handwritten data converted into machine readable form without wasting much time and cost. Computer can understand only alphanumeric characters as American Standard Code for Information Interchange (ASC II) typed on a keyboard where each character or letter has a unique code. However, computer itself does not recognize characters and numbers from scanned image. This has been overcome by applications of Artificial Intelligence to match images of characters available on scanned documents and convert them into their ASC II equivalents to get readable text. An Intelligent Character Recognition (ICR) System for handwritten forms includes functional components for form registration, character image extraction and character image classification. An effort is needed so that written words get separated into individual characters. Using ICR technology, computer system becomes artificially intelligent to recognise characters based on their shapes. The matched characters are directly converted into machine readable form. For unrecognised or doubtful characters, human intervention is needed to convert them to correct characters. So, ICR technology is seemingly a good machine facility to human operators to minimise their data entry time, decrease human drudgery and increase overall productivity. But in tasks where 100% accuracy is desired such as awards in examination system, this technology does not give desirable results. Himachal Pradesh University, Shimla, India is using ICR technology to fetch students awards. So here, the demand is of that technological which can give 100% accuracy. In this paper we are discussing the practical problems in data fetching from handwritten forms through ICR. Before discussing the recognition results, it is instructive to review various component of ICR technology.

Processing of ICR Forms

Examination system of any university is most important component in University Administration. Manual Examination system is very tedious job and due to human intervention it sometimes leads to errors which tarnish the image of university. Himachal Pradesh University has initiated a process to use the ICR technology, an Automated Data Capturing/Recognizing System to minimize human intervention and for fetching data through ICR, AutoRec software is being used. An *AutoRec* is an intelligent data capture and perfection software that facilitates the conversion of hard copy data into computer readable form /electronic data with minimum human interventions. The various components of "AutoRec" Flow Process are shown in Figure 1. ISCAN1 and ISCAN2 are two scanner stations where we can scan ICR forms. There are different processes to fetch the data from ICR forms, these are briefly discussed below:

Designing of Template: The first step in fetching data from ICR awards is to design a template for the document whose data has to be recognized. This is necessary to design the form according to position of the fields. Type of fields, position of data fields, length of field, type of recognition and other properties associated with any data item which we want to

**Corresponding author:* mohini_bhardwaj@rediffmail.com

recognize. Frame is designed and registration points are set and template is registered.

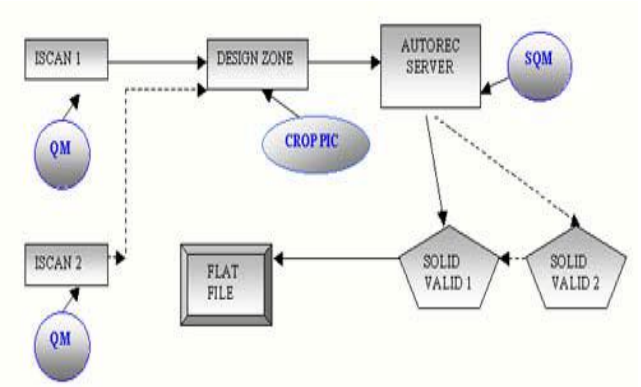


Fig. 1: AutoRec Flow Process (Source : http//www.fil-flan.com)

Image Capture & Scan: Different batches are created and forms are scanned. The forms which are scanned are images and are placed in different batches. We can scan more than 100 forms at a time from a scanner according to the capacity of the scanner.

Image Processing: In this process scanned image is processed. Here different processes are involved viz. image rotation, deskewness, despeckling, border removing, edge noise or entire background, locating data field and their identification and registration of the form.

Batch Processing: The batches are processed according to designed template.

Recognition: After processing the batches next step is to recognize them. The recognition starts processing one form at a time from the batch folder and marks the data fields in different colors depending on the recognition confidence. It is capable of recognizing free hand writing, hand written numbers and alphabets, check boxes, bar codes, OCR characters and magnetic ink characters. The recognized images are then passed to the data validation.

Data Validation: In this process, data is distributed for validation to different users. Operators are deployed for data validation. In data validation, there is a split screen, the actual image and the recognized data are shown separately. The user can see the image with the naked eye, check and correct any unrecognized data elements. After doing the required validation data can be exported, the data can be exported to various data formats e.g. flat ASCII file, PDF, HTML or ODBC compliant data bases.

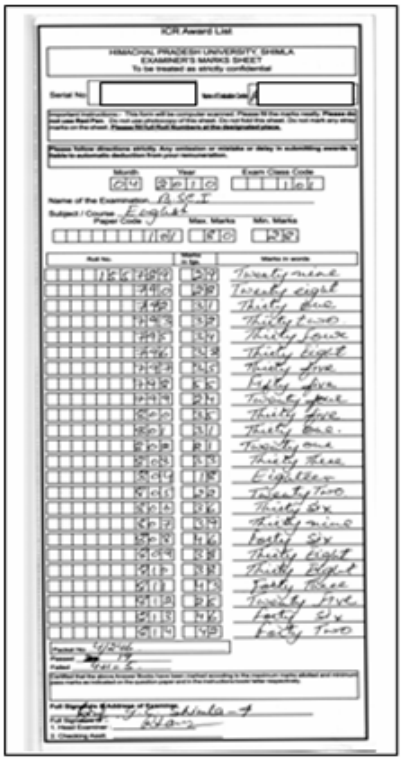
Exported Files: The exported files are ready for data processing

Related work: Presently university is using ICR technology to fetch awards from hand written data to machine readable form. Deodhare, Dipti, N. N. R. Ranga and Suri has suggested two most significant technologies, Intelligent Character Recognition (ICR) for hand written documents and Optical Mark Recognition (OCR) for data capture from printed documents. An Intelligent Character Recognition (ICR) System for handwritten forms includes functional components for form registration, character image extraction and character image classification. Rosen, Richard, Vinod and Patrick has suggested that ICR technology along with FAX system is a viable cost-effective option for processing CES reports presently being received via paper FAX. Overall, the cost of the ICR system would be recovered with in a year. Fax ICR technology is sufficiently accurate to provide high quality data with much less operator time than is presently used for full key entry of the data. ICR mode of data fetching is not 100 percent accurate as it is not able to recognize poor handwriting. Sometime evaluators marks stray marks on the award list and rollnos are also not properly written. Viking has stated that ICR engines can achieve very high recognition rates when the documents are properly designed, printed and controlled. Nonetheless, a 1% error rate on a printed page with 3,000 characters means there is an average of 30 errors on each page. This would be unacceptable for a typist, but may be adequate for information that will only be read occasionally. Kevin has found three practical methods of getting data from a piece of paper into a computer. These three methods are to key the data, to use Optical Mark Reading (OMR) or to use Imaging combined in some way with Optical Character Recognition (OCR). These are reliable and validated transfer of completed form data to a computer system for further processing and has suggested that OMR is a fast, accurate and cheap method for collecting census data. According to Bergeron OMR systems approach one hundred percent accuracy and only take .005 seconds on average to recognize marks.

ICR Form for fetching data from Awards List for university examination : As discussed in related work, ICR is a good technology that can be used to minimize human intervention to get accuracy. But recognition of hand-printed forms can be challenging. Form design is vital to ICR accuracy. ICR reads images of hand-printed characters from paper and converts them into machine-readable characters. The ICR award list which is currently being used in HP University is shown in Fig. 2.

Fig. 2 : ICR Award List

The award list shown above is designed in red colour. To capture the data, red colour is dropped and data written in the boxes is fetched. Accuracy of data depends upon the handwriting. There are other practical problem which the evaluators make while filling the awards in the ICR forms. Although instructions have been provided but still evaluators do make mistakes. Some of the problems are shown below:



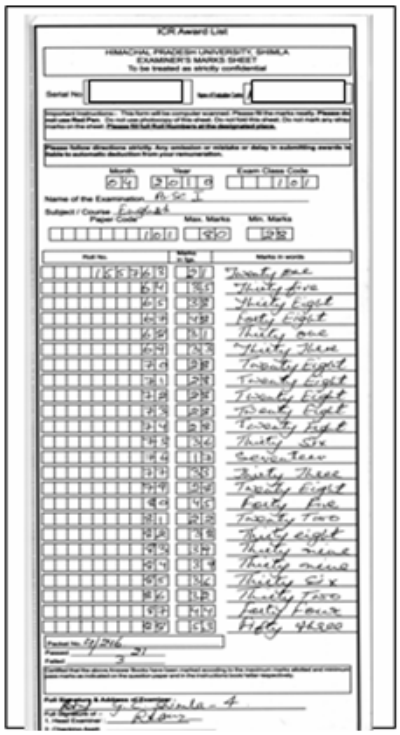


Fig 3-4 : Complete roll nos are not written, only last three/ two digits of roll nos are written. ICR is not intelligent enough to recognize candidate's complete rollno.

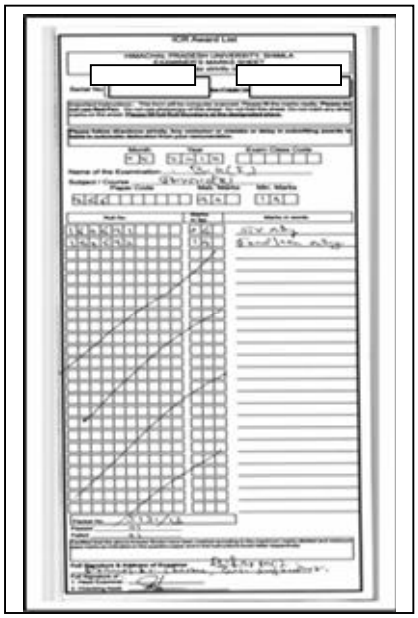


Fig. 5 : Stray marks on ICR marks list

Processing and recognition of handwritten documents is relatively difficult. The cursive scripts, variability and separation of characters, document skew, accommodating variability of strokes, learning the human characteristics are very difficult to model. Most of the handwritten document understanding systems are employed in specific narrow domain. HP University is adopting ICR technology in fetching awards of BSc. And BCom classes only. As stated above, ICR technology is not hundred percent accurate. It sometimes misinterpret the data and even it does not have doubt its recognition. The Figure below explain this fact. In the figure ICR recognize the Figure as 5594 whereas the actual figure is 5534 as shown in the scanned image.

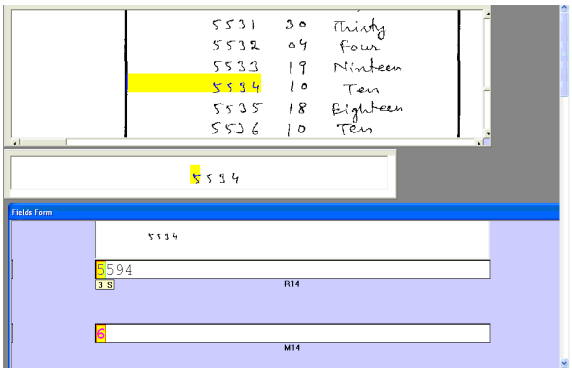


Fig 6 : Misrecognition of figure 5534 as 5594

OMR for accuracy in Awards

ICR is less accurate than OMR and it also requires some editing and verification. But examination system demands 100 percent accuracy in awards fetching. For awards fetching we must use that technology which gives more accuracy than ICR. OMR is the fastest and most accurate of the data collection technologies. It is also relatively user-friendly. OMR technology detects the existence of a mark, not its shape. According to Webster's Online Dictionary, Optical

mark recognition (OMR) is the scanning of paper to detect the presence or absence of a mark in a predetermined position. The accuracy of OMR is a result of precise measurement of the darkness of a mark, and the sophisticated mark discrimination algorithms for determining whether what is detected is an erasure or a mark. Sen, Patel has analysed that OMR scanner can maintain a throughput of 1,500 to 10,000 forms per hour. This activity can be controlled and processed by a single PC workstation, which can handle any volume the scanner can generate. Increasing the throughput simply requires upgrading the scanner. OMR scanning is more accurate than ICR. It consistently provides unmatched accuracy when reading data, exceeding the accuracy of expert key-entry clerks because it eliminates transcription errors. Optical mark recognition (OMR) is the scanning of paper to detect the presence or absence of a mark in a predetermined position. Optical mark recognition has evolved from several other technologies. In the early 19th century and 20th century patents were given for machines that would aid the blind.

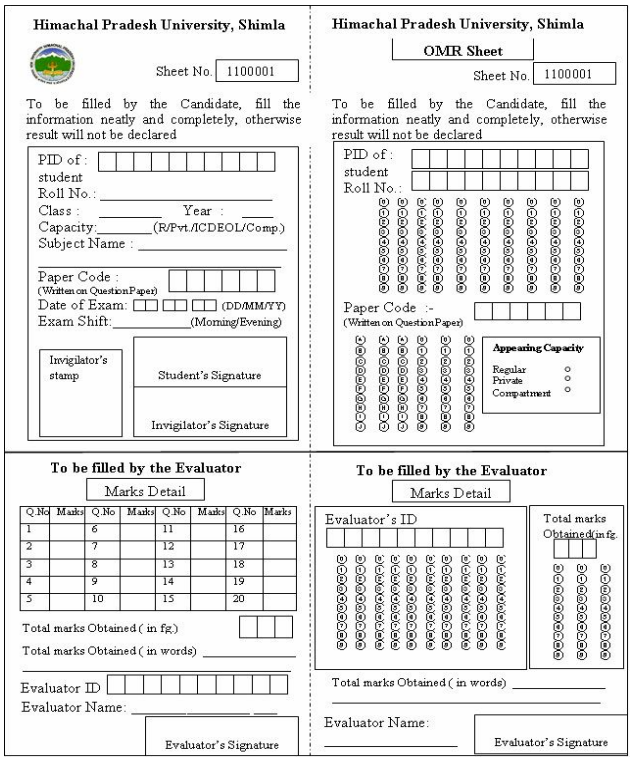


Fig. 7 : Design of OMR sheet(first page)of answer sheets

In a traditional method awards lists are prepared from the answer sheets and in doing so there are possibilities that evaluator may write wrong data. To avoid this OMR sheet may be attached to the answer sheet and evaluator will write marks obtained by student. So instead of writing on the award list, these OMR sheets will be directly scanned. Lay out of OMR sheet is designed in Fig.7.

Conclusion

Technology plays a vital role in fetching data where there is bulk amount of data. It should be kept in mind that there will definitely be mistakes while entering the data manually. OMR, OCR, and ICR technologies all provide a means of data collection from paper forms. When compared to manual data entry, ICR does not have human intervention so there are fewer chances of manual errors, but when compared with OMR technology it is less accurate. Inaccurate data is expensive to correct and university as well as student has to face its consequences. So for fetching up of awards from answer sheets OMR is the suitable technology.

REFERENCES

Bergeron, Bryan P. (1998, August). Optical mark recognition. Postgraduate Medicine online. http://www.postgradmed.com/issues/1998/08_98/dd_aug.htm Bookrags. (n.d.) Optical Character Recognition

Sen, D. J., R. N. Patel, U. Y. Patel (2010). Modern Database Technology Needs Optical Mark Reading, International Journal of Pharmaceutical and Applied Sciences/1 (2)/2010

Deodhare, Dipti, N.N.R. Ranga and Suri R. Amit (2005). "Preprocessing and Image Enhancement Algorithms for a Form-based Intelligent Character Recognition System", Internal Journal of Computer Science and Application, Technomathematics Research Foundation, 2005,

Fil-Flan, "AutoRec" ,2006, Available [online] http://www.fil-flan.com/index.htm

Haag, S., Cummings, M., McCubbrey, D., Pinsonnault, A., Donovan, R. (2006). Management Information Systems for the Information Age (3rd ed.).Canada: McGraw-Hill Ryerson.

http://www.ancsdaap.org/cencon98/papers/drs/drskevin.pdf

http://www.bookrags.com/sciences/computerscience/optical-character-recognition-csci-02.html

http://www.fcsm.gov/03papers/Rosen.pdf

http://www.vikingsoft.com/pdf/importanceofppde.pdf

http://www.websters-online-dictionary.org

Kevin Orchard(1998), 18th Population Census Conference: 26–29 August 1998, Program on Population, East-West Center, Honolulu, Hawaii USA

Rosen, Richard, Vinod Kapani and Patrick Clancy (2003). "Data Capture using FAX and Intelligent Character and Optical character recognition (ICR/OCR) in the Current Employment Statistics Survey (CES)", 2003, Federal Committee on Statistical Methodology.

Viking (2009). "The Importance of Power/Precision Data Entry to Document Imaging, A White Paper". 2005. Viking Software Solutions. 2009. Vol. II(II), pp 131 - 144.

****