

# Intelligent Framework for Auto Filling Web form using Scanned Documents

S. D. Bandari<sup>1</sup> Ankita Jagtap<sup>2</sup> Namrata Mane<sup>3</sup> Swarali Garud<sup>4</sup>

<sup>1,2,3,4</sup>Department of Computer Science & Engineering  
<sup>1,2,3,4</sup>DACOE, India

*Abstract*— Nowadays, people use online services or various platforms to complete various tasks, such as form filling, online shopping and trip planning using web applications. Generally, users are required to enter the information into web forms. Users often have to type same information repetitively into different web applications or forms. It is a tedious job for all the user to fill in web forms with same information. To save users or organizations from typing unnecessary information and increase creative productivity, it is critical to propagate and pre-filling user inputs across different web applications. However, the existing systems approach do not support this need well. User have all the information on the documents so it is easy to fill the forms using scanned documents.

**Keywords:** Form Filling, Scanned Documents, Tedious, Pre-Filling

## I. INTRODUCTION

There are many things we do on the internet and most of them involve filling up forms. It is common that when you want to register as a member of any website, you are asked to provide information. It is not too much trouble if the form is simple one to fill but if form requires lot of information, it may be a bit tedious to fill it up. The most annoying situation is when you have completely filled up the form and for some reason the submission failed forcing you to refill the whole form again. A large amount of information is same for the various services. To save end-users and the any organization member from such tedious process, it is more convenient for them if the information, which required commonly among different services, can be generated and filled using the scanned documents. Recently, researchers in various industry and academics have developed various tools and approaches towards the problem. Web browsers provide web form auto-filling tools, such as Mozilla Firefox Auto fill Forms and Google Chrome Auto fill Forms, to help end-users fill in web forms. In general, the above-mentioned auto filling tools stores the values entered by a user. The recorded or stored information is used for prefilling by identifying the parameters names. The tools also allow user to modify the recorded and filled information manually.

## II. LITERATURE VIEW

A. Noman Islam, Zeeshan Islam, Nazia Noor, A Survey on Optical Character Recognition System, *Journal of Information & Communication Technology*

This system uses Optical Character Recognition (OCR), which is a piece of software that converts printed text and images into digitized form such that it can be manipulated by machine. Unlike human brain, which has the capability to very easily, recognize the text/characters from an image. Machines are not intelligent enough to perceive the information available in the image. Therefore, a large

number of research efforts have been put forward that attempts to transform a document image to format understandable for machines.

B. Derek Edwin Pappas, Palo Alto, A survey on Intelligent Data search Engine in 2009.

This system automatically extracts the information that matches a predetermined criterion from one or more web pages at one or more web sites. The extracted information includes at least one extracted data-field value associated with one of the one or more extracted data-field names.

## III. METHODS

### A. Data Extraction

OCR is known as optical character recognition and it is the technology that allows software to convert the machine printed text or image-based text on scanned documents. Data extraction software uses simple zone ocr can be employed to convert specific region of page to usable data.

### B. Field Matching

The field matching methods and different field matching algorithms are very important part of the matching rules matching criteria. They help to determine how one field in one record is compared with the same field in another record. This will help to recognize the data and fill the accurate data into the field.

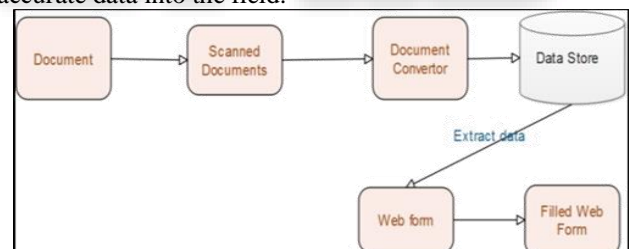


Fig. 1: Architecture

### C. Data Analysis

Data analysis is a process, which is required for inspecting, cleansing, transforming and modelling data to discover useful information, informing conclusion, and supporting decision-making. This also help to improve the accuracy of the system.

### D. Data Requirements

The data is necessary as inputs to the analysis, which is specified based upon the requirement of those directing the analysis or customers who will use the finished product of the analysis. In addition, the data is depending on the requirements of the form.

### E. Data Cleaning

Once the documents are scanned and converted, the data may be incomplete, contain duplicates, or errors. The need

for data cleaning will arise from problem in the way that data is entered or stored. Data cleaning is the process of correcting and preventing these errors. Common tasks include record matching, identifying inaccuracy of data.

#### F. Data Products

A data product is a web application that takes the input from scanned documents and fills the form as output. It may be based on model or algorithm.

#### IV. CONCLUSION

Current system is manual; we propose an intelligent framework to help organization to fill web forms and save their time from repeatedly entering the same data to web forms. Our framework automatically collects the inputs from scanned documents and convert the scanned document into text file. Then we analyze this text file and generate database in structured format. Finally, with the help of this database our framework fills the web forms automatically. The manual work is reduced by using scanned documents.

#### ACKNOWLEDGEMENT

We would like to thank all those who have supported us to complete our work.

#### REFERENCES

- [1] Mozilla Firefox Add-on Autofill Forms, <https://addons.mozilla.org/enUS/firefox/addon/autofill-forms/?src=ss>.
- [2] Google Chrome Autofill Forms, <https://support.google.com/chrome/answer/142893?hl=en>.
- [3] HARTMANN M. and MUHLHAUSER M. (2009)'Context-Aware Form Filling for Web Applications', ICSC' 09. IEEE International Conference on Semantic Computing, 2009, pp. 221- 228.
- [4] <https://www.ufsm.br/orgaossupplymentares.cpd/content>
- [5] WANG S., ZOU Y, UPADHYAYA B. and NG J., 'An Intelligent Framework for Auto-filling Web Forms from Different Web Applications ', In Proceedings of 1st International Workshop on Personalized WebTasking (PWT 2013) on IEEE Services, June 2013 - Santa Clara, California, USA. IEEE Computer Society Press.
- [6] BOWNIK L., GORKA W., PIASECKI A. (2009) ' Assisted Form Filling ', Engineering the Computer Science and IT. InTech, October 2009. ISBN 978-953-307-012-4.
- [7] RUKZIO E., NODA C., LUCA De A, HAMARD J.and COSKUN F. (2008) 'Automatic form filling on mobile devices ', Pervasive Mobile Computing, vol. 4, no. 2, pp. 161-181, 2008.