Use of OCR Technology for Data Extraction Using Amazon Textract

Sumit Muddalkar[1], Kiran Kolte[2], Ayush Batra[3], Ayush Naphade[4], Neha Lokhande[5]
Project Guide, Department of Information Technology[1]
Project Group Leader, Department of Information Technology[2]
Project Group Member, Department of Information Technology[3, 4, 5]
Shri Sant Gajanan Maharaj College of Engineering, Shegaon, Maharashtra, India

Abstract: *In the digital era of twenty first century, everything is becoming automated, and information is stored and transfer in digital forms. But there are many situations where data is not stored in digital form and it is essential to extract text from those hardcopies to store in digitized form. The latest technology such as Text recognition software has completely changed the process of text extraction using Optical Character Recognition. Therefore, this paper introduces the concept of OCR technology, explains the process of extraction using Amazon Textract tool and current research in the area. Detailed information and working methodology of Amazon Textract. Its comparison with other OCR tools and its scope. This paper will help other researchers in the field to get an overview of the technology.*

Keywords: Amazon Textract, Optical Character Recognition, Machine Learning, Google AI

REFERENCES

[1]. Jamshed, Memon Maira, Sami, Rizwan Ahmed Khan and Mueen Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)", School of Computing, Quest International University Perak, Ipoh 30250, Malaysia, Department of Computer Science, Shaheed Zulfiqar Ali Bhutto Institute of Science and Technology, Karachi 75600, Pakistan. Faculty of IT, Barrett Hodgson University, Karachi 74900, Pakistan. Department of Software Engineering, Faculty of Science and Technology, Ilma University, Karachi 75190, Pakistan.

[2]. Rishabh Mittal, "Text extraction using OCR: A systematic review", Department of Computer Science and Engineering, Amity school for Engineering and Technology, Amity university Uttar Pradesh, Noida (UP), India.

[3]. Thomas Hegghammer, "OCR with Tesseract, Amazon Textract, and Google Document AI: a benchmarking experiment"