Extraction and Verification of Information from Semi-Categorized Data

Sangeetha L[1], S M Sushmitha[2], Bindu P[3], Sanjana Y[4], Sunitha S[5]

Computer Science and Engineering[1-5]

Rao Bahadur Y. Mahabaleswarappa Engineering College, Ballari, India

Abstract: *Earlier, organizations mainly handled fully structured data stored in databases and well-organized files. With the growth of digital content, much information now appears in semi-structured forms such as reports, invoices, logs, and web-scraped data. Their inconsistent layouts make manual processing slow, error-prone, and non-scalable. This creates the need to accurately extract relevant information and track how it transforms across different formats in a mixed data environment. To address this, the proposed system provides an automated method for extracting and validating information using intelligent parsing, pattern recognition, rule-based checks, and database-driven verification. It combines web-scraping, preprocessing, structured mapping, and an embedded verification engine that checks extracted data against rules or trusted sources. Experimental results show that the system significantly reduces manual effort, improves accuracy, and reliably converts semi-structured inputs into validated structured data.*

Keywords: Information Extraction, Semi-Categorized Data, Data Validation, Pattern Recognition, Automation, Web Scraping, Data Clean Up

Copyright to IJARSCT
www.ijarsct.co.in

DOI: 10.48175/IJARSCT-30506

ISSN
2581-9429
IJARSCT