Abstract

An effective document processing system must be able to recognize structured and semi structured forms that is written by different persons’ handwriting. In this work we have developed a method and system that can process structured form document layout and recognize its contents. Our approach has been applied here in the context of Indian railway reservation/cancellation requisition system with encouraging results. In reality, handwritten data usually touch or cross the preprinted form frames and texts, creating complex problems for the recognition routines. In this paper, we address these issues and attempted to solve the problem for Indian Railway Reservation system using our custom built form processing software and Tesseract open source character recognition engine.
Reference

- http://code.google.com/p/tesseract-ocr
Index Terms

Computer Science
Text Mining

Key words

Optical Character Recognition
Handwritten Document Analysis
Form Processing
Tesseract OCR