TEXTRACT: ANALYZE SCANNED DOCUMENTS

Team 7
Wen-Ling Ku
Amlendu Kumawat
Maria Moy
Raghuram Sirigiri
Aditya Tomar
Kexin Yang

Our Git Repository

WHAT IS TEXTRACT?

Textract is a machine learning service that automatically extracts text, handwriting, and data from scanned documents. It can quickly automate document processing.

> **"Why trouble your experts when Textract can do it faster?"**

Variety of Documents

PDFs, Images, Tables, Forms & OCR software

Languages Detected

English, German, French, Spanish, Italian & Portuguese

Accurately Extracts

Text, Handwriting, Tables & other data forms

USE CASES

Financial Sector

Interest rates, Application names and Invoice Totals

Public Sector

Government Applications and Federal Tax Forms

Business Operations

NDAs, Lease Agreements, & Terms and Conditions

CURRENT FLOW

PDF or Image

Expert manually inspects files

Sorts files and highlights the mismatches

TEXTRACT FLOW

PDF or Image

Textract automatically extracts files

Sorts files and highlights the mismatches

IMPLEMENT TEXTRACT