

Document Entity Extraction



Overview

JioCloud Document Entity Extraction is built specifically for Indian documents and languages. It captures key information like names, phone numbers, dates, and organisations from scanned, printed, or typed text. The service works across multiple Indian scripts, supports API-based integration, and includes a live testing playground. With consumption-based billing and a self-service dashboard, teams can automate document processing at scale—without depending on manual workflows or external support.

Key Features

- **Support for Indian languages**
Works with documents written in Hindi, Tamil, Bengali, Telugu, Marathi, Kannada, and other widely used Indian scripts.
- **Multiple entity types**
Detects and extracts names, organisations, phone numbers, email addresses, locations, dates, numbers, hashtags, and URLs.
- **Handles unstructured content**
Processes data from paragraphs, scanned forms, email text, and handwritten inputs.
- **Live testing via playground**
Upload samples and instantly check how entities are extracted before going live.
- **Easy API integration**
Designed to connect smoothly with CRMs, document workflows, or e-governance systems.
- **Self-service control panel**
Creates, monitors, and manages your service and credits independently.
- **Usage-based pricing**
Pay per character or document, with clear visibility into how much you consume.

Benefits

- Speed up backend processes by automating document parsing
- Reduce errors and manual effort in form-based workflows
- Capture data from Indian scripts and mixed-language text
- Maintain consistency for audits, KYC, and compliance
- Feed structured data into analytics or downstream systems
- Scale entity extraction without infrastructure costs

Supported Language

Category	Details
Hindi	Full
Bengali	Full
Tamil	Full
Telugu	Full
Kannada	Full
Malayalam	Full
Marathi	Full
Gujarati	Full

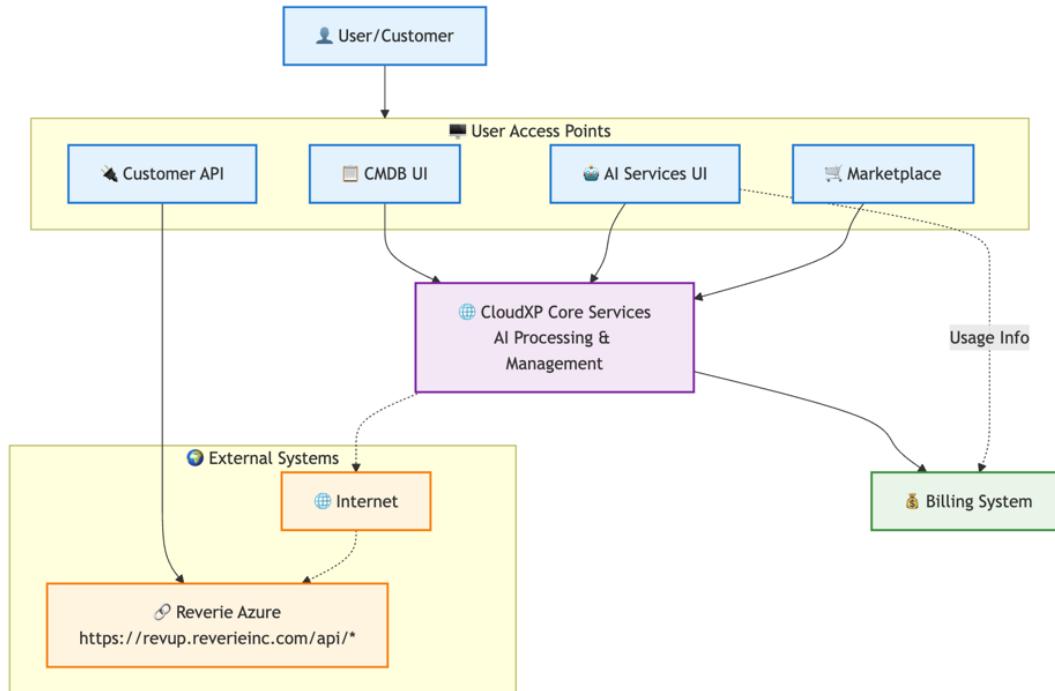
Category	Details
Punjabi	Full
Odia	Full
Assamese	Full
Urdu	Full
Sanskrit	Full
Konkani	Full
Dogri	Full
Manipuri	Full

Category	Details
Bodo	Full
Santali	Full
Kashmiri	Full
Maithili	Full
Sindhi	Full
Nepali	Full
English	Full

Technical Specifications

Category	Details
Entity Types	Person, Organization, Location, Date, Amount, etc.
Input Format	Text (Plain, Structured, PDF, OCR output)
Output Format	JSON with entities and confidence scores
Access Protocol	REST API over HTTPS
Authentication	Bearer Token
Average Latency	-
Rate Limiting	Tier-based; adjustable quotas available
Web Playground	Upload documents and receive entity extraction results
Deployment Model	Fully managed SaaS
Billing Model	Per document or per-character based pricing
Security	HTTPS; no storage of extracted data or document content

Architecture Diagram



Use Cases

- Resume screening**
 Extract candidate details from regional-language resumes and auto-fill recruitment systems.
- KYC form processing**
 Capture names, addresses, and contact info from scanned forms in Indian scripts.
- Government scheme applications**
 Identify Aadhaar, income, and location data from regional application forms.
- Legal document analysis**
 Extract parties, dates, and addresses from contracts and petitions across languages.
- Hospital intake forms**
 Pull emergency contacts and insurance data from handwritten or typed patient forms.