

Three layers between a document and a clean system record.

Most tools solve one of them. DocXtract owns all three — extraction, validation and integration — with no templates, on any document.

Document automation programmes rarely fail at extraction alone. They fail in the gaps between tools: a scanner that captures but does not validate, an OCR engine that returns text but not meaning, and an RPA bot that posts data nobody checked. Each handover adds latency, cost and a place for errors to enter unseen.

The 2026 buying pattern is consolidation — IDC's latest vendor assessment describes the industry's key challenge shifting from handling unstructured documents to building end-to-end workflows fed by reliable data. This brief maps the three layers between an incoming document and a clean record in your system of record, and how DocXtract handles each.

AT A GLANCE

- Any document, any format — no templates
- 98%+ field accuracy, 5,000 pages/hour
- Validation before data reaches the ERP
- Live in days; 1-day integration

The three layers



01 Extract — context, not coordinates

Layout and language are read together, so any document works on first submission — no template to build, no format to configure, no retraining when a vendor changes their layout. Handwritten notes, skewed scans, tables spanning pages, mixed-language content and multi-document PDFs are handled in the same pass, at 98%+ field accuracy and up to 5,000 pages per hour.

03 Integrate — post where work actually happens

Pre-built connectors for SAP, Oracle, Tally and Zoho, plus UiPath and Power Automate for existing automation workflows, and a REST API returning real-time JSON for everything else. Typical integration is a one-day exercise.

02 Validate — the layer most tools skip

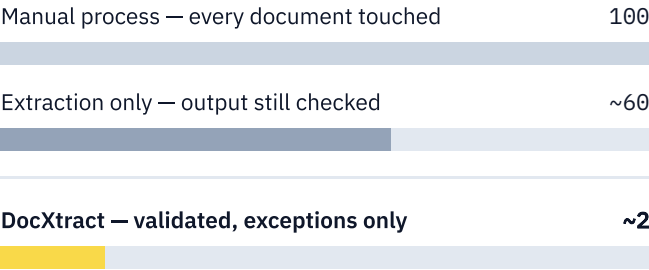
Multi-layered, context-aware validation engine that applies semantic understanding, arithmetic cross-checks and duplicate detection all run before anything is posted. Wrong data is caught inside the platform, not discovered downstream in the ERP.

Each extracted field carries a confidence score and a reference back to its source page, so downstream systems receive data your own controls can be applied to.

Exceptions are a design decision

Exception handling is seamless with DocXtract. As it flags those could not pass straight through due to missing data, irrelevant text/number.

FIG. 1 — WHERE EFFORT GOES, PER 100 DOCUMENTS



Illustrative. Actual touchless rate depends on document mix and configured confidence thresholds; established in a pilot.

What each layer replaces

LAYER	TYPICAL TODAY	WITH DOCXTRACT
Extract	Template OCR maintained per format; mixed PDFs split by hand and new vendors become configuration projects	Template-free extraction with automatic splitting; a new format works on first submission
Validate	Spreadsheet checks and reviewer judgement, applied inconsistently	Configured rule set applied identically to every document, every time
Integrate	Re-keying into the ERP, or brittle bots that break when a screen changes	Native connectors and REST API; structured JSON posted directly

A realistic rollout

1. **Pick one document with real variance.** The long tail, not the cleanest supplier — that is where the business case sits.
2. **Benchmark against a live sample.** Field-level accuracy measured on your own documents, not a demo set.
3. **Connect and scale by document type.** Integrate in a day, run in parallel briefly, then add types once touchless rate holds.

Measure it on Accuracy & STR

Accuracy is necessary but not sufficient. The metrics that moves costs and processing time and **straight-through rate** — the share of documents completing end-to-end in seconds with no human touch. Baseline it per document type before you start.

- 98% Accuracy across document type
- ~2% exceptions
- Seconds in DocXtract vs Hours spent maintaining formats

NEXT STEP

Map your workflow against the three layers.

Send a sample document set — any format, no preparation needed. We report measured field-level accuracy and a projected touchless rate against your current process.

Book a Demo

SOURCES

1. IDC MarketScape: Worldwide Intelligent Document Processing Software 2025–2026 Vendor Assessment — shift from document handling to end-to-end workflows fed by reliable data.
2. DocXtract product documentation and benchmarking, docxtract.io, 2026 — accuracy, throughput, connector list, deployment timelines.
3. Figures in Fig. 1 are illustrative of typical deployment patterns and are not client results; touchless rate is established per customer during a pilot.