



OBT INTELLIGENCE SUITE

AI Products for SharePoint



OBT IA CRAWLER

AI Classifier for SharePoint

Product Sheet

Commercial and executive brief

Version 2.0 · July 2026

Confidential document — Ortega Technologies (OB Technologies)

What OBT IA Crawler is

OBT IA Crawler is the AI document classifier of the OBT Intelligence Suite. It sweeps a data source—a file server or a SharePoint site—extracts the text of every document, classifies it with a language model, and produces structured metadata plus a report of inventory, sensitive-data exposure and proposed information architecture. It is the value layer that turns a SharePoint Online migration into a governed, searchable and compliant repository.

Design principle: in Phase 1 the product only reads and proposes. It does not modify a single document or metadata field until the client reviews the report and approves (ADR-07, non-destructive).

The problem it solves

- **Ungoverned legacy repositories:** millions of files in folders or on-premise SharePoint, with no metadata, full of duplicates, obsolete content and scattered personal data.
- **Migrations that just move the mess:** lifting to SharePoint Online without classifying reproduces the chaos in the cloud.
- **Invisible compliance risk:** no one knows where PII, contracts or regulated documentation live, nor who has undue access.
- **Unfeasible manual classification:** hand-tagging hundreds of thousands of documents does not scale.

How it works

The engine walks the source in a resumable, idempotent way, distributes the work through a queue and classifies every document with a triage funnel that discards the trivial cheaply and escalates only the ambiguous or sensitive to larger models. For each document it produces business and sensitivity metadata:

- **Document type** (contract, invoice, report, procedure...), project, client and date.
- **Area or department** and document language.
- **Sensitivity** (public, internal, confidential, restricted).
- **PII presence** and its type (email, phone, national ID, IBAN, card), validated with checksums (Luhn, mod-97, ID letter).
- **Confidence** of the classification; anything doubtful is flagged needs-review, never fabricated.

Operating phases

Phase	What it does	Output
1 · Read & report	Read-only sweep, inventory and proposed classification	Report (PDF/Excel/JSON) + cost estimate
2 · Client	The client reviews the report and the	Approval and agreed taxonomy

Phase	What it does	Output
validation	metadata architecture	
3 · Write & migrate	Metadata written to SharePoint columns during or after migration	Classified, governed repository

On this base the recurring services are built: *confidential-data assessment*, *monitoring subscription (continuous DLP)* and *vertical sector packs (pharma first)*.

Architecture in one line

A Node/TypeScript engine deployed as Azure Functions v4 (Flex Consumption plan) in the EU, with self-hosted OCR on Azure Container Apps, a Service Bus queue, secrets in Key Vault and classification by Mistral (La Plateforme, EU) by default. LLM and OCR layers are swappable by configuration (Mistral, Claude, OpenAI, Azure OpenAI; RapidOCR, Azure Document Intelligence or Mistral OCR).

Differentiators

- **Non-destructive, approval-gated:** Phase 1 never touches the source; writing is enabled only with client approval.
- **EU data sovereignty:** LLM and OCR in the European Union; the OCR does not store the image (it processes and discards it). Ideal for pharma and regulated sectors (GDPR).
- **Margin under control:** self-hosted OCR ~25× cheaper than Azure Document Intelligence and a triage funnel that discards ~88% cheaply.
- **Proven resilience:** retries with backoff and rate-limit handling across the whole HTTP layer, a queue with duplicate detection and a dead-letter alert.
- **Bilingual and multilingual:** reports in ES/EN/FR/DE.

Plans

Fixed price by size: each plan is billed as an initial price plus a price per analysed TB. Example: Professional, 5 TB = €3,000 + 5 × €1,500 = €10,500.

Plan	Includes	Initial price	Price per TB
Essential	PoC / audit: read-only sweep + classification, reports. Client supplies their own AI key. No metadata writing.	€1,500	€900/TB
Professional	Production governance: OBT manages the LLM (Mistral EU), optional OCR, metadata writing after approval.	€3,000	€1,500/TB
Enterprise	Large volumes / regulated: all AI providers, assisted migration, guaranteed EU residency, priority support.	€6,000	€2,200/TB

Enterprise > 25 TB: custom. Optional project services: self-hosted OCR deployment (from €1,200), assisted migration + mapping (from €2,500), target architecture (from €900) and training (from €600). Prices excl. VAT, July 2026.

License distinction and calculation

All three licenses are calculated with the same formula —initial price + price per TB— but each enables a different scope: Essential does not write metadata (report only), Professional writes after approval with an OBT-managed LLM, and Enterprise adds all AI providers, assisted migration and priority support. Total price by analysed size:

License	Initial price	€/TB	1 TB	5 TB	15 TB
Essential	€1,500	€900	€2,400	€6,000	€15,000
Professional	€3,000	€1,500	€4,500	€10,500	€25,500
Enterprise	€6,000	€2,200	€8,200	€17,000	€39,000

Formula: total = initial price + (TB × price per TB). The license calculator (attached Excel) returns the price of all three licenses for any volume.

In short

The initial report is the entry point: low price, no migration commitment, and with the real document count in hand the classification and migration are quoted. Cost and price scale together: never quoted blind on unknown volume.

Ortega Technologies (OB Technologies) · soporte@obtechnologies.com ·
www.obtechnologies.com