

Aufwandsarme Dokumentenindexierung und Metadatenextraktion mit IDA

Indexierung ist ein Prozess in Dokumentenmanagementsystemen, der dazu dient, Dokumente zu strukturieren und zu kategorisieren, um den Abruf von Informationen zu erleichtern. Er macht Dokumente leicht zugänglich und durchsuchbar, indem er sie in digitale Formate umwandelt und mit Metadaten wie Datum, Autor:innen und anderen relevanten Elementen kennzeichnet.

Erfahren Sie, wie die IDA-Software-Suite den Dokumentenindexierungsprozess vereinfacht und optimiert.

Reduzierung manueller Aufwände

Überragende OCR- und ICR-Genauigkeit für schwierigste Szenarien inklusive Handschrift

Minimierung des Wartungsaufwands

Geringer Trainingsaufwand für sich ändernde Dokumentenlayouts

Sicherstellung der Compliance

On-Premises- oder Private-Cloud-Deployment

TYPISCHE HERAUSFORDERUNGEN

Digitalisierungsinitiativen stoßen während der Indexierungsphase oft auf Verzögerungen. Um **genaue Informationen** aus verzerrten, qualitativ minderwertigen Scans mit maschinen- und handgeschriebenen Texten **abzurufen**, ist oft erheblicher menschlicher Aufwand zur Validierung und Korrektur erforderlich. Wenn die OCR-Ergebnisse nicht ausreichend genau sind, greift der Prozess oft auf komplett manuelle Verarbeitung zurück. Darüber hinaus kann die **Aufrechterhaltung von regelbasierten Dokumentenklassifikations- und Datenextraktionsverfahren** zur Bewältigung neuer oder modifizierter Dokumentenlayouts und detaillierterer Kategorien sowohl komplex als auch kostspielig sein.

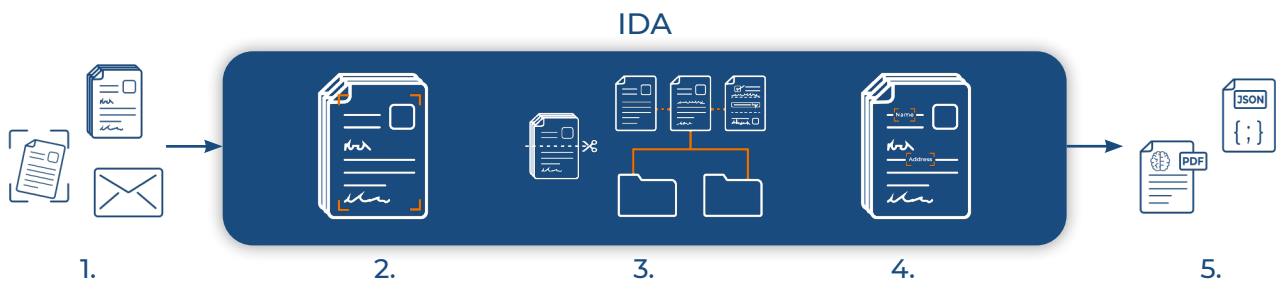
DIE LÖSUNG

PLANET AI's Intelligent Document Analysis ermöglicht eine **mühevolle automatische Indexierung und Metadatenextraktion** aus großen Dokumentenvolumina. Durch die Kombination patentierter Kerntechnologie mit modernsten Machine-Learning-Features bietet IDA **beispiellose Genauigkeit bei OCR und ICR**, wodurch der Bedarf an manuellen Korrekturen selbst in den anspruchsvollsten Szenarien minimiert wird.

Mit dem **regelfreien Ansatz zur Dokumentenklassifikation und Datenextraktion** erfordert IDA nur minimale Trainingsdaten und geringen Wartungsaufwand. IDA kann als Java-Anwendung oder via Containerisierung mit Docker **on-premises (lokal) oder in einer (privaten) Cloud** bereitgestellt werden.

SO FUNKTIONIERT'S

IDA-Workflow für aufwandsarme Dokumentenindexierung und Metadatenextraktion:



1. Eingabe: Physische und elektronische Dokumente über Scanner, Postfach, E-Mail usw.

2. Recognition: OCR- und ICR-Fähigkeiten auf Basis der patentierten PerceptionMatrix

3. Trennung und Klassifikation von Dokumenten: Regelfreie Trennung großer aneinanderhängender Dokumente und Klassifikation basierend auf Few-Shot Learning

4. Metadatenextraktion: Datenerfassung einzelner Felder aus Dokumenten ("zonal data extraction")

5. Ausgabe: PDF or PDF/A A (alle Konformitätsstufen) mit Textebene, die die OCR-Ergebnisse enthält, sowie optional hervorgehobene Metadaten und/oder JSON mit Metadaten, einschließlich Positionsdaten, Confidence Score usw.

CUSTOMER SUCCESS STORY

Dokumentenindexierung findet Anwendung in verschiedenen Szenarien, wie beispielsweise Records Management im Geschäftsprozess-Outsourcing (einschließlich Scan-Dienstleistungen), Dokumenten- und Content-Management sowie digitalen Bibliotheken und Archiven.

Unser **renommierter Kunde** bietet seit über 50 Jahren **Geschäftsprozess-Outsourcing-Dienstleistungen** im Gesundheitswesen, den öffentlichen Sektor und für Unternehmenskunden an. Er hatte mit einer **Automatisierungsrate von nur 50% bei der Dokumentenklassifikation** zu kämpfen. Als Ergebnis war erheblicher manueller Aufwand erforderlich, um Dokumente zu korrigieren und zu validieren. Der Few-Shot-Learning-Ansatz von IDA ermöglichte eine schnelle erste Implementierung und **erhöhte die Automatisierungsrate auf 90%, was zu einer Reduzierung des manuellen Aufwands um 80% führte.**