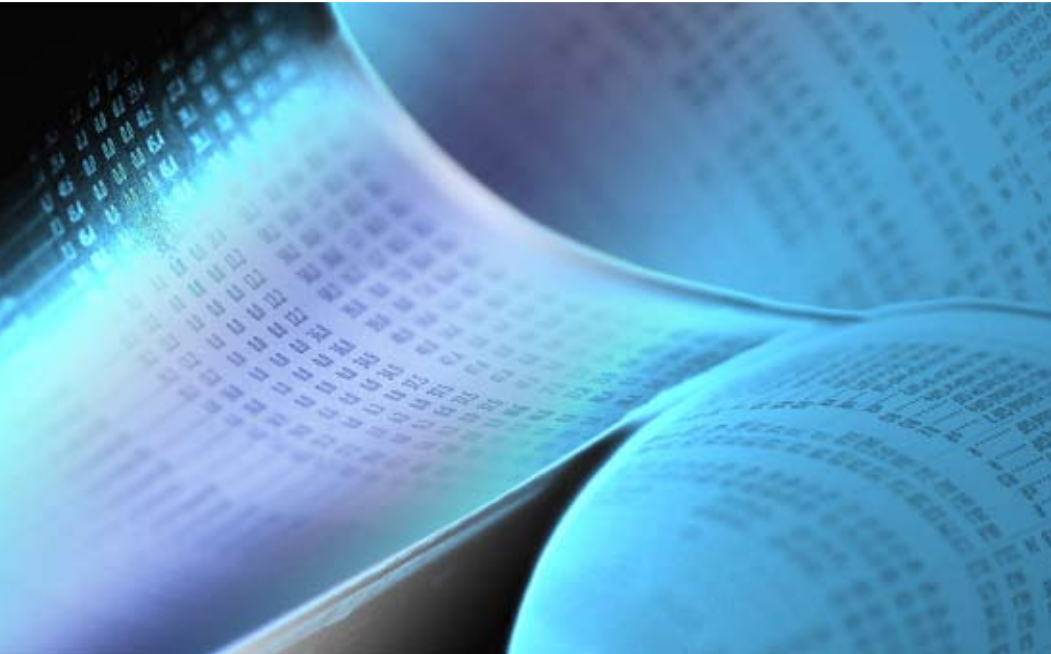
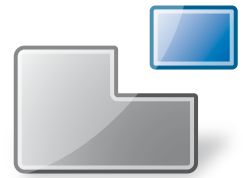


>>

ELO DocXtractor II FORM

Die Komplettlösung für schnelle Formularverarbeitung



Erhöhen Sie Ihre Effizienz mit maximaler Produktivität und einem Minimum an Arbeit

ELO DocXtractor II Form eliminiert die hohen Kosten der Formularerfassung vom ersten Tag an. Das intelligente Modul ist die ideale Lösung bei der Verarbeitung strukturierter Dokumente wie beispielsweise Formularen und bringt alle hierfür erforderlichen Funktionalitäten mit. Eine große Menge an Daten wird so in kürzerer Zeit und bester Qualität erfasst. Dies führt zu flexibleren und vor allem schneller ablaufenden Geschäftsprozessen.



ELO DocXtractor II FORM

Die Komplettlösung für schnelle Formularverarbeitung

>>

Wettbewerbsvorteile durch Zeitgewinn

Effizienz vom ersten Tag an

Formulare sind heute nach wie vor ein großer Bestandteil in der Geschäftswelt. Täglich erreichen Unternehmen Formulare unterschiedlichster Art wie Aufträge, Anfragen, Antworten usw. Diese kommen im Unternehmen per Post, Fax oder Internet in unterschiedlichster Qualität und mit unterschiedlichsten Feldinformationen an. Es spielt keine Rolle, ob die Informationen handgeschrieben oder gedruckt sind; die manuelle Erfassung ist in der Regel sehr aufwendig und fehleranfällig. Mit dem **ELO DocXtractor II Form** kann eine große Menge an Daten in kürzerer Zeit und in besserer Qualität erfasst werden. Der **ELO DocXtractor II Form** fügt sich nahtlos in die **ELO ECM-Suite** ein. Durch das einfache Einrichten von Formularen und die einzigartige Anwenderfreundlichkeit ermöglicht dieses flexibel erweiterbare System einen erfolgreichen Einstieg in die Geschäftsprozessoptimierung.

ELO DocXtractor II Form ermöglicht Ihnen:

- Fokussierung auf eigentliche, gewinnbringende Aufgaben
- Kosteneinsparungen
- Effizientes Arbeiten durch weniger Fehler
- Schnellere Weiterverarbeitung durch schnellere Verfügbarkeit und höhere Datenqualität
- Flexiblere Prozesse
- Einen qualitätsgesicherten Datenexport in die Datenbanken der Folgesysteme (z.B. in der Auftragsbearbeitung oder bei Behörden zur schnelleren Weiterverarbeitung)



Funktionsweise ELO DocXtractor II Form

Bild-Vorverarbeitung: Einwandfreie Ergebnisse

Nachdem die Dokumente eingescannt worden sind, folgt die konfigurierbare Aufbereitung der gescannten Dokumente für die OCR. Folgende Verarbeitungsschritte können hier genutzt werden:

- Vordruck- und Linienentfernung (z.B. bei Überweisungsträgern)
- Rotationstoleranz und Upside-Down-Korrektur für 180° gedrehte Images
- Entrauschen für bessere Bildqualität
- Automatische Wiederherstellung von entfernten Punkten für die OCR (z.B. Punkte in der Datumsangabe oder bei Umlauten)
- Drehwinkelkorrektur für schief gescannte Images
- Entfernung von Stanzlöchern

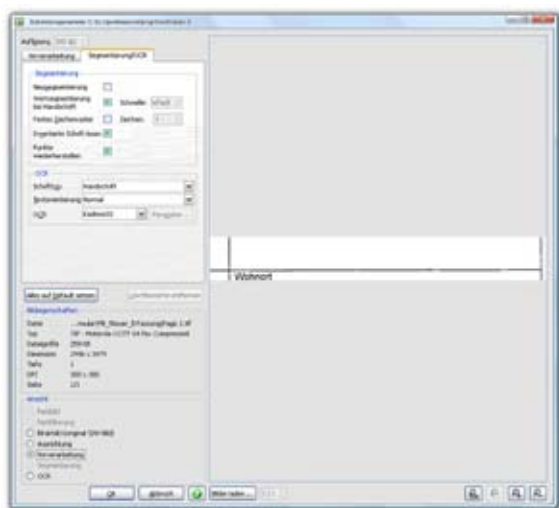


Abb. 1: Imaging Vorverarbeitung auf Dokumenten- und Feldebene

Hat der Kunde eine Scansoftware, die eine Bild-Vorverarbeitung anbietet, kann diese Funktion innerhalb des **ELO DocXtractor II Form** einfach deaktiviert werden.

Intelligente Klassifikation

Die Klassifizierung der Dokumente kann in **ELO DocXtractor II Form** auf Basis des Layouts sowie auf Basis einfacher inhaltlicher Verfahren erfolgen. Bei der layoutbezogenen Klassifizierung wird ein Dokument aufgrund seiner Form und Gestalt analysiert. Die inhaltsbezogene Methode interpretiert den gesamten Inhalt des Dokuments bzw. Teile des Inhalts.

Mögliche Form-Klassifikationen:

- Klassifizierung aufgrund des Layouts
- Klassifizierung aufgrund von Suchmustern (Text an einer bestimmten Position)
- Klassifizierung anhand von Barcodes
- Klassifizierung anhand der Seitengröße

Bei der layoutbasierten Klassifizierung benötigt der **ELO DocXtractor II Form** zum Training lediglich ein Beispieldokument, um alle erforderlichen Merkmale zu analysieren. Dokumente mit einem nahezu identischen Layout werden daraufhin automatisch erkannt. Durch die Angabe eines Suchmusters, sprich einer bestimmten Position im Text, kann eine 100% Klassifikation sichergestellt werden. Weitere Klassifikationsmethoden sind im **ELO DocXtractor II Mailroom** verfügbar.



>>

Intelligent und strukturiert

Zielsichere Verarbeitung

Ein Formular besteht aus verschiedenen Feldtypen. Nachdem der **ELO DocXtractor II FORM** das Dokument klassifiziert hat, erfolgt die Extraktion der klassenspezifischen Felder.

Folgende Feldtypen werden unterstützt:

- Adresse (zur Erkennung von Adressfeldern)
- Anker (zur Ortung einer genauen Position im Dokument)
- Ankreuzfeld (zur Erkennung von Ankreuzfeldern)
- Barcode (zur Barcode Dekodierung innerhalb eines Dokumentes)
- Suchmuster (Definition von Such- und Ergebnismustern)
- Tabelle (zur Extraktion kompletter Tabellen)
- Text (zur Extraktion von Informationen an bestimmten Positionen)
- TopDown (Auslesefunktion von Informationen aus der Datenbank)
- Diverse Sonderfelder wie z.B. Stapel-, Vorgangs- und Dokument-ID

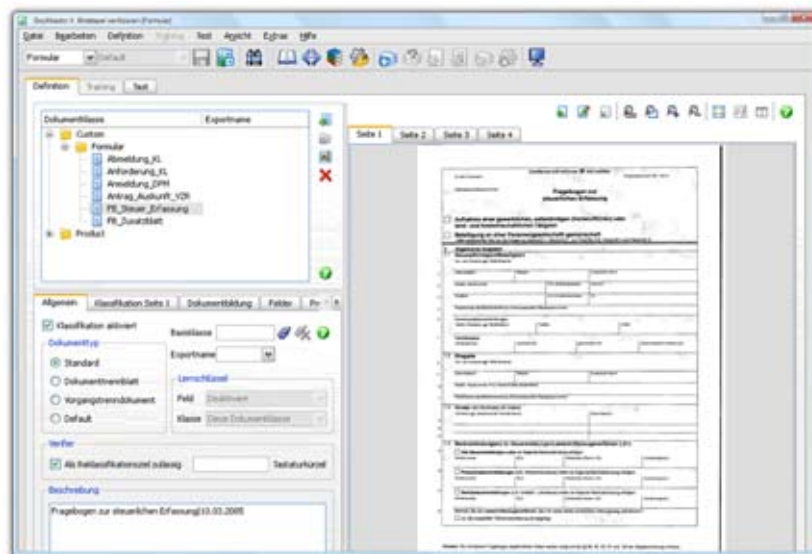
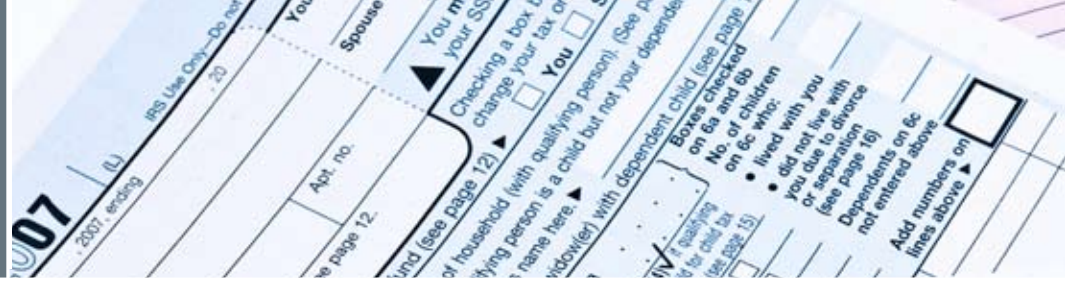


Abb. 2: Dokumentendefinition und Formularverifier im ELO DocXtractor II FORM



Der **Feldtyp „Anker“** bezeichnet die Verankerung einer Position innerhalb des Dokumentes zur genauen Ortung eines bestimmten Informationsfeldes. Dies ist notwendig, wenn das Dokument z.B. beim Ausdrucken nur noch einen bestimmten Prozentsatz seiner Ursprungsgröße besitzt.

Der **Feldtyp „Ankreuzfeld“** wird zur Erkennung von Ankreuzkästchen in einem Dokument genutzt. Der ELO DocXtractor II Form erkennt automatisch, ob

das Kästchen angekreuzt ist oder nicht und wertet diese Informationen aus.

Über den **Feldtyp „Barcode“** können mit einem Barcodeleser, Barcodes auf einem Dokument decodiert werden. Das System findet den Barcode unabhängig von dessen Position auf dem Dokument.

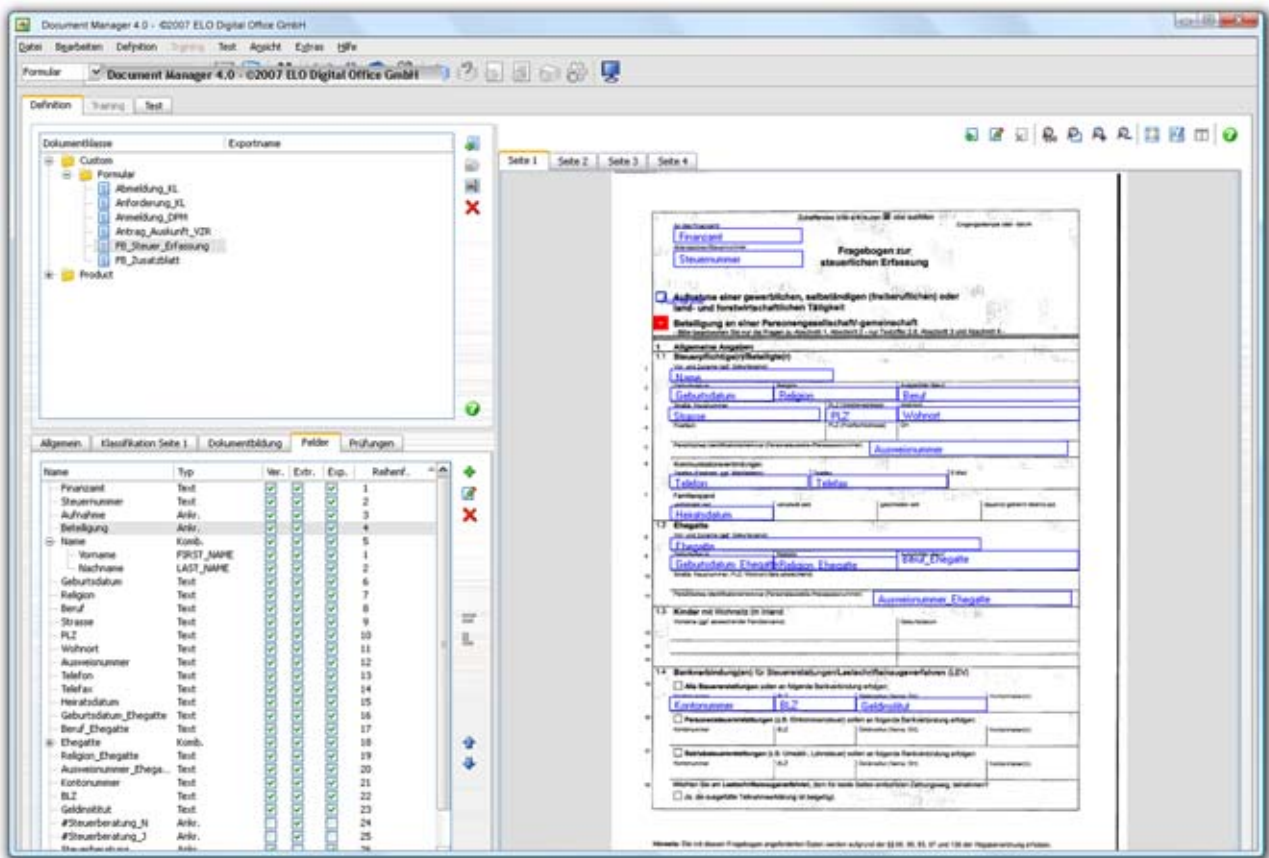


Abb. 3: Felder für die Extraktion



>>

Flexible Anpassung

Mit Hilfe des **Feldtyps „Suchmuster“** gibt es die Möglichkeit aneinander gekoppelte Such- und Ergebnismuster zu definieren. So kann in einem freien Brief z.B. das Wort „Datum“ das Suchmuster darstellen und der darauffolgende Wert wie z.B. „21.10.2008“ als Ergebnismuster übernommen werden.

Der **Feldtyp „Tabelle“** wird zur Extraktion kompletter Tabellen verwendet. Es können Tabellen fester Position (z.B. auf Formularen) und freie Tabellen (z.B. auf Rechnungen) definiert werden. Da die Struktur von Tabellen sehr unterschiedlich sein kann, werden diverse Extraktionsstrategien angeboten.

Der **Feldtyp „Text“** dient der Extraktion von Informationen mit fester Position in Formularen. Mit Hilfe von regulären Ausdrücken, wie z.B. „tt.mm.jj“ für Datum oder dem Aufbau des Wertes für Rechnungs- oder Steuernummern, ist es möglich die Struktur des Inhalts vorzugeben und anhand dieser Kriterien auch exakt zu verifizieren.

Der **Feldtyp „TopDown“** (Abgleich von Daten) wird immer genutzt, um Informationen auszulesen, die in einer Datenbank abgelegt sind. Der Abgleich ist hochperformant und tolerant gegenüber OCR-Fehlern oder unterschiedlichen Schreibweisen. Auch stark verfälschte OCR-Ergebnisse führen dann noch zu guten Abgleichwerten.

Kosten sparen in der Nachbearbeitung

Sind die Feldtypen analysiert worden, bietet der **ELO DocXtractor II Form** eine einzigartige Möglichkeit der Prüfung und Korrektur. Der **ELO DocXtractor II Form Improver** ermittelt auf der Basis aller ausgelesenen Informationen und ihrer Alternativen das Optimum aller Werte. Dadurch steigt die Erkennungsqualität und der Nachbearbeitungsaufwand wird stark reduziert.



>>

ELO DocXtractor II Form

Die Komplettlösung für schnelle Formularverarbeitung



ELO® erhalten Sie bei:



ELO Digital Office GmbH · D - Stuttgart · www.elo.com · info@elo.com
ELO Digital Office CH AG · CH - Zürich · www.elo.ch · info@elo.ch
ELO Digital Office AT GmbH · A - Linz · www.elo.at · info@elo.at

