

Datenerfassung für grosse Volumen von Formularen und unstrukturierten Dokumenten

The screenshot displays the PAPERIN software interface, which is divided into several functional areas:

- Manager [deutsch]:** The main application window with a menu bar (Engine, Eingabewarteschlange, Ausgabewarteschlange, Optionen, Werkzeuge, Hilfe) and a toolbar. It shows a list of input files (arena-001.tif to arena-010.tif) and a list of output files (InterSport_Average, Intersport_err, Intersport_ok, Intersport_Export).
- Form Recognition Client (v 4.47):** A sub-window for processing documents. It includes:
 - Buttons for Start, Stop, Optionen, Grösser/Kleiner, and Hilfe.
 - OCR options: Kadmos, ScanSoft.
 - File information: Aktuell Dateiname: arena-002.tif, Formular: Arena.
 - Page information: Dateien Gut: 2, Mittel: 0.
 - A status bar showing "00000002 ? Bereit".
 - A preview of the scanned document with red bounding boxes around key fields.
- Form Recognition Results:** A detailed view of the extracted data from the document:
 - Rechnungs Datum:** 22.02.2001
 - Rechnungs Nummer:** 562407
 - Blatt Nr:** 1
 - Lieferschein:** (empty field)
 - Betrag Lieferung L:** 19.61
 - Betrag Porto L:** 0.00
 - Zwischen Summe L:** 19.61
 - Betrag MWST L:** 3.14
 - Betrag Summe L:** 22.75
 - Document Text:**
 - Rechnung
 - Kunde: 71402 Re-Nr.: 562407
 - Re-Def.: 22.02.01
 - Verb.: ISPO 507140 Rab.: 3.00%
 - Auftrag vom: 05.09.00 12 894
 - DER, BA: MOBI Seite 1
 - prozessieren
 - Stck. Preis Wert
 - ng vom: 22.02.01 Auftrag: 234364
 - LS: 144868
 - helmstraße 16
 - 1638 Ludwigsburg
 - 140
- Log:** A bottom-left pane showing system logs, including file recognition events:
 - 11:10:14 C:\PaperIn4\Queues\Input
 - 11:10:15 C:\PaperIn4\Queues\Input
 - 11:10:15 C:\PaperIn4\Queues\Input
 - 11:10:15 C:\PaperIn4\Queues\Input
 - 11:10:32 arena-001.tif recognized , =
 - 11:10:36 arena-002.tif recognized , =

PaperIn...

- ersetzt manuelles Erfassen von Dokumenten für weitere Prozesse
- ermöglicht Auswertungen formularmässig festgelegter Felder
- ermöglicht Auswertungen von unstrukturierten Dokumenten sowie deren Zuweisung zu einem Workflow
- erlaubt die Auswertung von Faxen und Überführung der Inhalte an definierte Prozesse.
- erkennt lateinische und kyrillische Druckschriften, u.a. nach dem Kurzweil ICR-Verfahren (Intelligent Character Recognition), ohne Vorgabe der Schriftfamilie.
- ausrichten schräg gescannter oder gedruckter Seiten; präzise Positionierung definierter Felder anhand graphischer Merkmale auf dem Dokument; unterdrücken störender Linien, unerwünschter Zeichen und Schmutzpartikel.
- separiert definierte Prozesssteuerungen in Verzeichnisse
- basiert auf einem modularen Konzept. Die einzelnen Funktionen und Programmteile können auf individuelle Bedingungen verschiedenste Aufgabenstellungen erfüllen. Alle Module greifen nahtlos ineinander über und können beliebig leistungsbezogen erweitert und ergänzt werden.
- multilinguale Benutzeroberfläche

PaperIn...

Server Module:

- **DocRec Manager:** Verwalten der Datenbank für Konfigurationen, Benutzer, Projekte
- **FormrecWS:** Dokumentenerkennung

Client Module:

- **DocRec Designer:** Definition von Formularen und Dokumenten
- **DocRec PostEdit:** Korrektureditor mit Imageanzeige
- **DocRec Admin:** Definition der Benutzer, Projekte und Zugriffsrechten

zusätzliche Module:

- **Dokumentenanalyse:** Inhaltliche Erschliessung

Technische Spezifikationen:

Schrifterkennung/Schriftarten: Druckschrift (Omnifont¹), Glyph³, Prüfbox, Handblockschrift² und Barcode

Schriftgröße: 8-72 Punkt bei 200 dpi

Lesebereich: Unbeschränkte Anzahl Felder, frei definierbar

Erkennungsklassen: Alphanumerisch, numerisch, Währung, Datum

Prüfroutinen: Eingebaute und benutzerdefinierbare Prüfroutinen

Benutzerschnittstelle: Graphische menugesteuerte Oberfläche

Formulardefinition: Interaktiv, anhand des gescannten Dokumentes

Vorverarbeitung: Entfernen von Linien, Kästchen und Schmutzpartikel, Ausrichten des Image

Nachbearbeitung: Interaktiv, imagebezogen

Stapelbetrieb: Multiqueue Management

Bildverarbeitung: optional können ausgerichtete Images abgespeichert werden

Bildspeicherung: in verschiedenen Auflösungen und Formaten, z.B. CCITT G3/G4

Datenausgabeformat Text: ASCII, ISO, Xdoc², PDF³, XML

Für Datenbank: Im Text ausserhalb liegendes Bild referenziert oder in Datei eingebettet.

Betriebssystem: NT 4.SP6⁴, Windows 2000⁴, Windows XP⁴

PC Voraussetzungen: Intel Pentium III, min. 1GHz, min. 256 MB RAM, minimum SuperVGA; 19" Bildschirm empfohlen

Hersteller:

ARPA Data GmbH
Albisstrasse 36

CH 8134 Adliswil

Tel: +41 1 709 09 79

¹Registriertes Warenzeichen von Scansoft™

²Warenzeichen von RE Recognition Technology GmbH

³Warenzeichen von Xerox Corporation

⁴Registriertes Warenzeichen von Microsoft™

PAPERIN

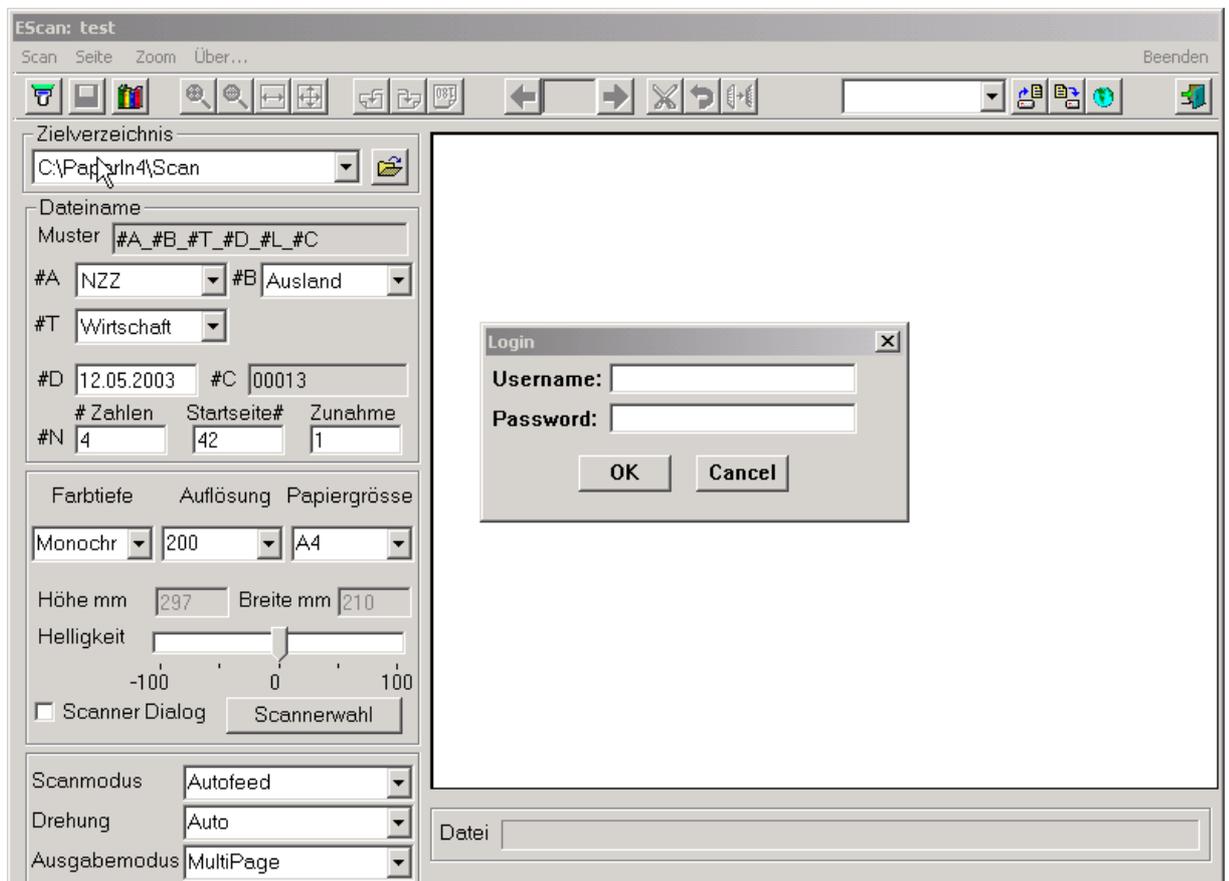
Multifunktionales Datenerfassungssystem

Update / Neue Eigenschaften

Auslieferung Mitte Februar 2003

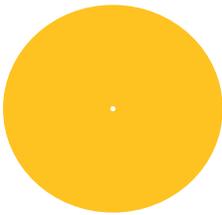
PaperIn Scan

- Freie, einzigartige Strukturierung zur Vergabe von Dateinamen.
- Beliebiges Zusammenfügen von Namensteilen (z.B. #A, #B, usw.) und Einfügen des aktuellen Benutzers (#L) in den Dateinamen.
- Passwort gesteuerte Zutrittsberechtigung mit Verlinkung zu Scan Profilen.
- Ersetzen und Ergänzen einzelnen Seiten in einem Multipage-Dokument.
- Dokumentidentifikationsnummer hält mehrere Dokumente zum gleichen Betreffnis zusammen (DokumentID #C).
- Felder mit unterlegten Listen zur Auswahl von Namenselementen.(#A, #B, #T).
- Unterstützte Scanner: Twain kompatible.
- Ausgabe als PDF auf Anfrage.



PaperIn Manager / FormRecWS

- Automatischer Start beim Booten.
- Läuft im Hintergrund als Service.
- Für unstrukturierte Dokumente Inhaltsanalyse mit firmenspezifischen Begriffen wie Personen, Materialien, usw. zur Klassifizierung für die elektronische Aufarbeitung von Eingangspost, Projektdokumentationen, allgemeine Firmendokumente.



PaperIn PostEdit

- Passwort Verwaltung: Der Benutzer erhält Zugriff auf definierte Verzeichnisse oder auf Dateinamen, welche seine UserID enthalten.
- Feldweises Generieren von Listboxen oder RadioButtons.
- Definition verschiedener Validierungsaufgaben.
- Umbenennung des ursprünglichen Dateinamens im Ausgabeprozess.
- Zuschaltbare Fachwörterlexikas zur automatischen Korrektur von Wörtern sowie deren xml Auszeichnung.
- Notizfenster mit Editierfunktionen.
- Korrektureditor für mehrsprachige Texte.

PaperIn DocClip

- Neu: gothische Schrifterkennung (Eigenentwicklung).
- Ausgabe im xml Format.
- Zuschaltbare Fachwörterlexikas zur automatischen Korrektur von Wörtern sowie deren xml Auszeichnung.

Hersteller:

ARPA Data GmbH
Albisstrasse 36

CH 8134 Adliswil
Tel: +41 1 709 09 79