



Papyrus FreeForm®

KWH (Usage)	Bill Demand	KVAR Hours	Code	Amount
1	637		ACTL	
1			ACTL	24.00
				\$5.44
600 KWHs @ \$.061376 Per KWH				\$36.83
37 KWHs @ \$.076197 Per KWH				\$2.82
24 KWHs @ \$.00 Per KWH				\$.00
637 KWHs @ \$.000328 Per KWH				\$.21
637 KWHs @ \$.010850CR Per KWH				\$6.91CR
\$38.39 @ \$.057 Per Dollar				\$2.19
Current Electricity Charges				\$40.58

Häufig unter einer Gruppe von Objekten des Typs "Betrag"

Typischerweise rechts von

Wahrscheinlich im unteren Drittel der Seite

FreeForm® Ansatz

Selbstlernendes Modul speziell zur Erkennung und Extraktion unsortierter und unstrukturierter Dokumente

Produktbeschreibung

FreeForm® ist ein Softwaremodul für die automatisierte Erkennung von unsortiert eingehenden, gescannten Geschäftsdokumenten mit unbekanntem Strukturen und Layout. Das System erkennt auch Geschäftsdokumente wie Eingangsrechnungen, Lieferscheine, Bestellungen, Anträge, Bewerbungsschreiben, etc. **FreeForm®** basiert auf modernen Methoden der Mustererkennung und verwendet modernste Techniken in den Bereichen Schrifterkennung (OCR, ICR, Voting), assoziative Datenbanken, Fuzzy Logic und Neural Networks.

Funktionsweise

Die Option **FreeForm®** ermittelt aus einem Dokumentenimage den Namen der Dokumentenklasse sowie die für diese Klasse definierten Felddaten. Das Konzept erlaubt eine N:M Beziehung zwischen Images und Dokumenten: ein Dokument kann sich über eine oder mehrere Seiten erstrecken, genauso kann ein Image mehrere Dokumente beinhalten wie z.B. eine Zeitungsseite mehrere Artikel umfasst.

Kontakt

Europa

T: +43-2236-27551
F: +43-2236-21081

Amerika

T: 817-416-2345
F: 817-416-1223

Asien/Pazifik

T: +65-6339-8719
F: +65-6336-6933

E-Mail

info@isis-papyrus.com

Internet

www.isis-papyrus.com

© 2012, ISIS Papyrus, ersetzt alle bisherige Dokumentation.

1. Jänner 2012

Arbeitsweise

FreeForm® verarbeitet gescannte Dokumentseiten in drei Schritten:

Image Vorverarbeitung

FreeForm® bereitet die Images für optimale Erkennungsergebnisse vor:

- Erkennung der Dokumentenausrichtung und Autorotation
- Schräglagenkorrektur
- Schmutz- bzw. Hintergrundbereinigung.

Dokument Klassifikation

FreeForm® bewertet das Image anhand von Klassenmerkmalen und Verarbeitungsregeln:

- Schlüsselwörter an vorgegebenen Positionen oder frei im Text
- Textphrasen an vorgegebenen Positionen oder frei im Text
- Grafiken, Linienelemente und Hintergrundfarbe an vorgegebenen Positionen, sowie Seitenformat

Datenextraktion

FreeForm® extrahiert und liest alle benötigten Felddaten zur erkannten Dokumentenklasse. Dies geschieht durch Auswertung folgender Merkmale und Regeln:

- Absolute Positionsangaben
- Relative Positionsangaben durch Hinweisobjekte
- Mustervergleich (Auftreten von definierten Synonym- oder Datentyp-Objekten)

Leistungsmerkmale

- Verarbeitung von Dokumenten mit unbekanntem Format
- Klassifizierung und Sortierung von gemischten Dokumentenstapeln
- Kombination von Methoden des klassischen Formularlesens mit Volltextanalyse von freien Textbereichen
- Einbinden in eine effiziente und ergonomische Produktionsumgebung mit integrierter Nachbearbeitung
- Integrierte Definitions- und Trainingsumgebung
- Interaktive Festlegung von Regeln, Parametern und Zonen mit sofortiger Verifizierung in einem Testlauf
- Statistische Analysen des verarbeiteten Dokumentengutes mit der Möglichkeit die Ergebnisse für das Tuning zu verwenden

Definition und Training

Ergonomie - FreeForm® erspart die komplexe und zeitaufwendige Erfassung von kryptischen Definitionen und Regelwerken. Trainiert wird intuitiv und anhand konkreter Beispieldokumente - learning-by-example.

Die Auswahl der relevanten Klassifikationsmerkmale erfolgt während der Design- und Trainingsphase alternativ:

- Automatisch - im Auto-Learn-Modus ermittelt **FreeForm®** die relevanten Merkmale zur bestmöglichen Unterscheidung der Dokumentenklassen anhand von vorgegebenen Imagebeständen für jede Klasse selbsttätig
- durch den qualifizierten Benutzer (Superuser) - grafisch und einfach durch Auswahl relevanter Objekte des Images per Mausklick.

Analog erfolgt das Lernen von Extraktionsregeln direkt am konkreten Beispieldokument.

Kontrolle - „What gets measured gets done!“ Daher ist in die **FreeForm®** Entwicklungsumgebung ein Werkzeug zur laufenden, statistischen Überwachung und Analyse des aktuellen Trainingsstatus integriert. Der Benutzer wird über die Auswirkungen seiner Definitionsschritte unmittelbar informiert. Problemfälle und ihre Ursachen werden direkt aufgezeigt.

Repository - FreeForm® verfügt über ein Object Repository. Einmal erstellte Definitionen wie

- Synonyme, Worte oder Phrasen
- Feldtypen - einfach oder zusammengesetzt (z.B. Adressblocks, Tabellen)
- Dokumentklassen und Subklassen (Typen)

werden übersichtlich verwaltet und können bequem wiederverwendet werden; d.h. neu zu erstellende Definitionen werden von bewährten, bestehenden abgeleitet.

Das bedeutet:

- rasche Anwendungserstellung
- einfache und effiziente Wartung und Weiterentwicklung der Anwendung.

Integration/Ergänzende Produkte

Die Option **FreeForm®** ist für den Einsatz mit Papyrus Capture optimiert. Durch die ActiveX Schnittstelle kann diese Option auch von anderen Anwendungen und Erfassungs-Plattformen, welche die Integration von ActiveX Servern unterstützen, genutzt werden. Das ISIS Capture Rohformat ist dokumentiert (ASCII/UNICODE mit Angaben zu Image, Zonen und Einzelzeichen). Der **FreeForm®** Applikationsdesigner ist eine eigenständige Anwendung, die im laufenden Betrieb des Document Capture Systems nicht benötigt wird.

Technische Voraussetzungen

- Hardware
PIII 800+ Mhz, 512 MB RAM, 2GB HD.
Pro Klassifikationsmodul sollte zur besseren Performance je ein Reco-Server vorgesehen werden
- Software
Papyrus Capture Basissystem

Dateiformate

FreeForm® unterstützt eine große Anzahl unterschiedlicher Dateiformate, Bspw. Text (DOS, ANSI, Unicode, UUEncode, MIME), Microsoft Word (ab Version 6.0), Microsoft Excel (ab Version 2.x), HTML, RTF, WordPerfect (ab Version 6.0), Wordstar, Microsoft Works, PowerPoint, LotusWordPro, Microsoft Outlook Mail Format, AmiPro.

Schulungen

ISIS bietet seinen Kunden ein spezielles Inhouse Schulungsservice in Form des **Papyrus Capture** Workshop.