



# Papyrus Classify

*Selbstlernendes Programm zur Klassifizierung von Dokumenten aller Art zur automatischen Sortierung und Verteilung.*

**Dokumentenklassen**

DocClass01  
DocClass02  
DocClass03  
DocClass04  
DocClass05  
DocClass06  
DocClass07  
...

**Klassenname:**  
Stromrechnung

**Felder:**  
Kundenname  
Adresse  
Rechnungsdatum  
Rechnungsbetrag  
...

**Prozess:**  
WFL043

- Felder
- Geometrische Merkmale
- Schlüsselwörter
- Satzerkennung

## Produktbeschreibung:

**Papyrus Classify** ist ein selbstlernendes Modul zur Klassifizierung von Dokumenten, das vielfältige Anwendungsmöglichkeiten in den Bereichen automatische Sortierung und Verteilung von elektronischem Posteingang, Fax und Papierpost eröffnet.

Intention von **Papyrus Classify** ist es, den ungefilterten Fluss einlangender Daten nach gewünschten Kriterien zu beurteilen, in frei definierbare Kategorien einzuordnen und damit die Informationen für Unternehmen zu erschließen und gezielt an den zuständigen Stellen zur weiteren Verarbeitung vorzulegen.

Eine konkrete Anwendung von vielen ist das automatische Vorsortieren von elektronischer Eingangspost wie bspw. das Aufteilen von E-Mails an unterschiedliche Abteilungen – Rechnungswesen, Bestellung, Support, etc. – oder an einzelne Sachbearbeiter bzw. Kundenbetreuer. Auch für die automatische Klassifizierung von Texten, die in verschiedenen Sprachen vorliegen, nach der jeweiligen Sprache ist **Papyrus Classify** hervorragend geeignet.

## 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

## Funktionen/Highlights

- **Selbstlernendes System**  
Das bedeutet, dass die Regeln, die für das automatische Zuordnen der Dokumente nötig sind, nicht von einem Administrator spezifiziert, gewartet bzw. den sich ändernden Anforderungen angepasst werden müssen, sondern dem System für jede gewünschte Dokumentenklasse eine bestimmte Anzahl an Dokumenten (typischerweise zwei bis drei Dutzend) zum Trainieren zur Verfügung gestellt werden. *Papyrus Classify* lernt aus diesen Trainingsdaten selbständig Regeln, nach denen in weiterer Folge die Klassifizierung vorgenommen wird.
- **Einfaches Nachtrainieren des laufenden Systems**  
Ein *Papyrus Classify*, das im Echtbetrieb eingesetzt wird, kann laufend nachtrainiert werden; Grundlage dafür sind jene Dokumente, die nicht richtig zugeordnet werden konnten. Damit ist eine kontinuierliche, langfristige Optimierung des Systems gewährleistet, bzw. können die Regeln flexibel an Änderungen der Dokumentklassen angepasst werden.
- **Ausnahmebehandlung**  
Dokumente, die nicht eindeutig einer Klasse zugeordnet werden konnten, werden ausgesondert und dem Administrator zur weiteren Beurteilung vorgelegt. Diese Dokumente werden dann manuell einsortiert, dabei können sie auf Wunsch für ein Nachtrainieren des Systems verwendet werden.
- **Zuordnung von einem Dokument zu mehreren Klassen**
- **Mehrstufige Klassenhierarchie**  
Dies ermöglicht ein mehrstufiges Vorsortieren der Dokumente; beispielsweise wird ein Dokument einer allgemeinen Klasse „Rechnung“ einer darunterliegenden Klasse „Lieferant A“ zugeordnet.
- ***Papyrus Classify* erledigt** den Großteil seiner Arbeit im Hintergrund. Sollte ein Eingriff des Administrators, oder eines Anwenders erforderlich sein, steht eine sehr klare, einfach und intuitiv zu bedienende Oberfläche zur Verfügung.

## Arbeitsweise

Die beim Training selbständig erlernten Regeln wendet *Papyrus Classify* zur Klassifizierung unbekannter Texte an. Dabei berechnet das Programm für jedes Dokument die Wahrscheinlichkeiten, mit der dieser Text jeder Klasse zugeordnet werden kann. Die Zuordnung erfolgt in jene Kategorie mit der höchsten Wahrscheinlichkeit, sofern die bei der Definition der Klassen festgelegte Mindestwahrscheinlichkeit erreicht wird. Die solcherart vorsortierten Dokumente werden dann an die zur Bearbeitung bestimmte Stelle (Niederlassung, Abteilung, Person, etc.) weitergeleitet. Dokumente mit einer zu geringen Wahrscheinlichkeit für eine bestimmte Klasse werden in ein sogenanntes Reject Directory gestellt, aus dem sie der Administrator mit einem Mausklick der richtigen Kategorie zuweisen kann. Auf Wunsch werden diese Texte zum weiteren Training des Systems herangezogen. So kann *Papyrus Classify* im laufenden Betrieb ständig dazulernen und seine Wissensbasis erweitern. Möglicherweise falsch klassifizierte Texte gibt der Anwender an den Administrator weiter, der mit diesen wie mit den nicht klassifizierbaren Dokumenten verfährt.

## Voraussetzungen organisatorisch

Bevor *Papyrus Classify* zum Einsatz kommt, gilt es, die gewünschten Kategorien, in welche die Dokumente unterteilt werden sollen, festzulegen. Aus jeder dieser Kategorien wird eine gewisse Anzahl von Texten (üblicherweise ca. 20 bis 30) benötigt, die man *Papyrus Classify* als Trainingsmaterial zur Verfügung stellt. Anhand dieser Information lernt das Programm selbständig, wodurch sich die einzelnen Klassen voneinander unterscheiden. Natürlich können auch während des laufenden Betriebs auf einfache Weise neue Kategorien hinzugefügt werden.

## Voraussetzungen technisch

- 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

Unterstützt eine große Anzahl unterschiedlicher Dateiformate beispielsweise 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.