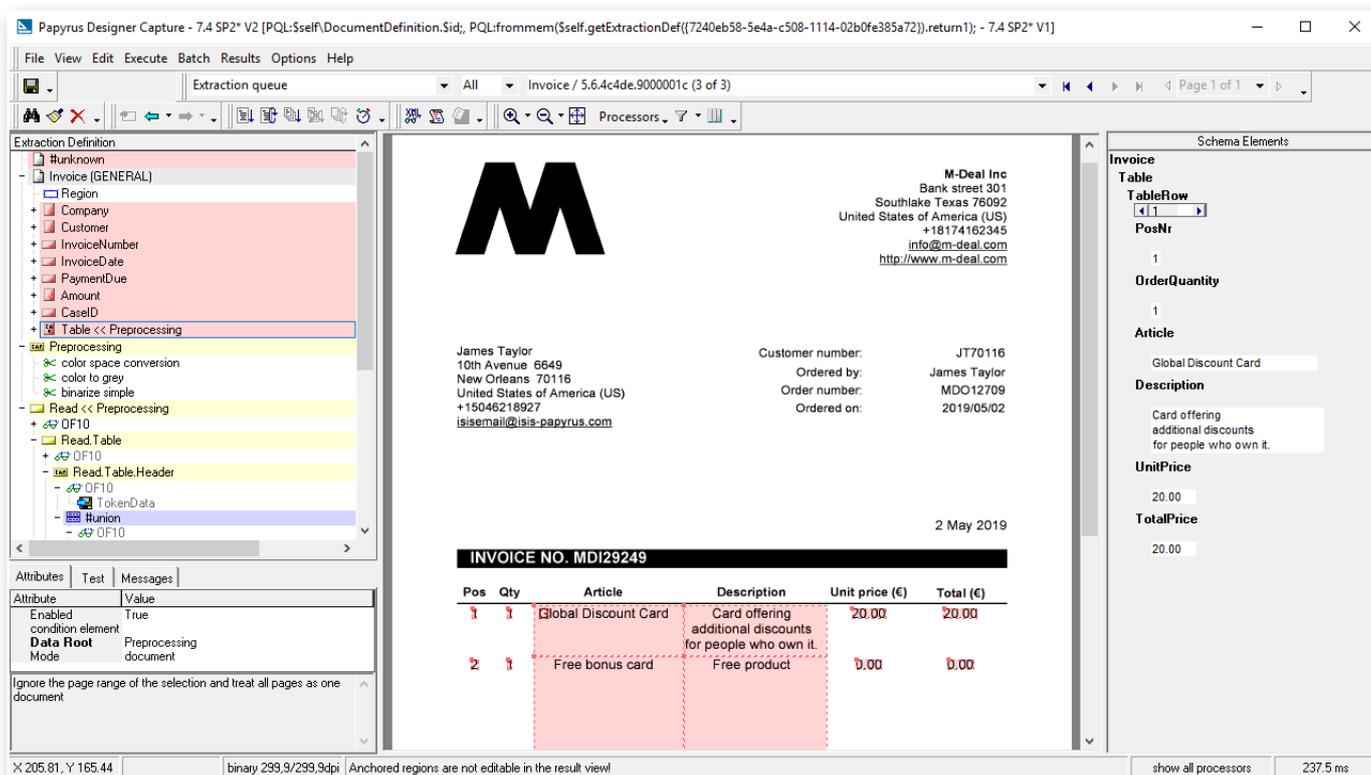


## Papyrus Designer/Capture



### Descrizione

Il Papyrus Designer/Capture è un toolkit molto potente per la definizione del set-up di classificazione e di estrazione dei dati per il riconoscimento automatizzato dei dati di svariati tipi diversi di documenti.

La componente Papyrus Designer/Extract semplifica la creazione delle definizioni di documenti di lavoro scansionati o già in formato immagine, di cui non si conosce la struttura ed il modello, come fatture, liste e tabelle complesse. Fornisce gli strumenti necessari all'inserimento dei parametri di estrazione dati per gestire ogni tipo di documento dalla struttura nota.

Papyrus Designer/Classify si compone di un modulo di auto-apprendimento per la classificazione di documenti ed offre una vasta gamma di possibilità applicative per automatizzare la sistemazione delle pagine e la distribuzione dei documenti attraverso posta elettronica, cartacea e fax.

Nella sfida dettata dalle esigenze moderne di salvare solo le informazioni ed i contenuti di interesse al business da altissimi volumi di corrispondenza sia cartacea che elettronica, il Papyrus Designer/Capture è molto di più di un semplice strumento di impostazione delle regole di riconoscimento dei testi. Di fatto, ha anche un'eccellente capacità di identificare i contenuti attraverso il riconoscimento di pattern, ovvero di schemi e contesti, parole chiave, posizionamenti relativi, regole, ecc. e loro combinazioni.

*Potente strumento senza eguali per la definizione delle regole di classificazione e di estrazione dei dati da ogni tipo di documento.*

### Contatti

**Italia**  
P: +39 0125 645500  
F: +39 0125 6455150

**Europe**  
P: +43-2236-27551  
F: +43-2236-21081

**America**  
P: 817-416-2345  
F: 817-416-1223

**Email & Web**  
info@isis-papyrus.com  
www.isis-papyrus.com

© 2020, ISIS Papyrus,  
sostituisce la documentazione precedente.

## Vantaggi principali

- Digitalizzazione dei contenuti: definizione delle regole di estrazione dei contenuti di interesse su alti volumi di dati in ingresso
- Definizione di regole di classificazione ed estrazione sia per documenti FixForm che FreeForm®
- Supporto di documenti provenienti da tutti i comuni canali di inbound e formati di file
- Ampia selezione di funzioni disponibili per la definizione delle regole di classificazione ed estrazione
- Supporto di vari motori di riconoscimento per vari tipi di contenuti
- Insieme al Papyrus Recognition Server fornisce una configurazione di alto livello per automatizzare la digitalizzazione dei contenuti
- Rapida definizione e test di nuovi tipi di documenti

## Caratteristiche

### Componenti del Package

- Papyrus Designer/Extract
- Papyrus Designer/Classify

### Pre-processing dell'immagine

Papyrus Designer Capture si occupa della preparazione dei dati dell'immagine per ottimizzare i risultati di riconoscimento, sia automatici che manuali. L'applicabilità avviene soprattutto in un contesto di immagini scansionate o fotografate.

- Rimozione di sporchie ed ombreggiature
- Rimozione dei segni di spinzatura
- Rotazione automatica
- Binarizzare
- Conversione da colore a toni di grigio
- Rimozione di box, linee, griglie e combinazioni
- Alleggerimento di colori
- Ampliamento
- Ridimensionamento
- Raddrizzamento in base ai bordi
- Raddrizzamento in base al contenuto

### Motori di riconoscimento

Il contenuto di immagini pre-processate è riconosciuto grazie ad una serie di potenti motori di riconoscimento classificabili nei gruppi seguenti:

- Optical character recognition (OCR) – testo scritto a macchina
- Intelligent Character Recognition

- (ICR) – testo scritto a mano
- Optical Mark Recognition (OMR) – selezione di caselle
- Codici a barre, codici QR, DataMatrix
- Logotipi

## Papyrus Designer/Classify

The Papyrus Designer/Classify è utilizzato per definire le categorie da classificare, i set di estrazione e la strategia di classificazione.

Il Designer/Classify supporta l'automazione di molteplici passaggi. Permette anche una semplice gestione di dati di esempio. Il Designer/Classify è pienamente integrato in Papyrus Objects.

Gli step di classificazione supportati possono essere riassunti come segue:

- Parole chiave e loro combinazioni
- Layout
- Testo avanzato
- Logotipi

Prima di iniziare a lavorare con Papyrus Designer/Classify, devono essere già state definite le categorie di documento richieste. Con il supporto di un certo numero di documenti di esempio (una dozzina) per ciascuna categoria, il sistema viene formato in modo che possa apprendere le specifiche proprietà di ciascuna classe di documento. In modo molto semplice è poi possibile aggiungere nuove categorie all'occorrenza.

Il Papyrus Designer/Extract permette di classificare aggiungendo della logica anche in fasi successive – durante l'estrazione

## Papyrus Designer/Extract

Durante la fase di definizione dell'estrazione viene trovato il contenuto rilevante attraverso le possibilità principali riportate qui di seguito e le combinazioni di esse.

- Pattern (contesto)
- Parole chiave
- Regione
- Posizione
- Punti di ancoraggio
- Righe
- Layout
- Logotipi

## Struttura del processo di definizione base

- Definizione di tutti i tipi di documento che possono arrivare
- Definizione degli elementi di interesse in ogni tipo di documento
- Definizione per ogni elemento di dati:
  - Attributi generali
  - Ancoraggi
  - Schemi ricorrenti per quell'elemento
  - Condizioni (regole) che devono essere rispettate
- Test delle definizioni con vari documenti
- Comparazione dei risultati con risultati precedenti nella fase di estrazione dei dati.

Vari strumenti di supporto visivo vi permettono di definire nuovi documenti rapidamente:

- Mostra i candidati per l'estrazione
- Mostra regioni e connessioni
- Strumenti di etichettatura

Il modulo Papyrus Capture/Extract permette una definizione di riclassificazione o di classificazione basata su un'analisi molto approfondita sulla struttura del documento. Questo fornisce una grossa flessibilità per il disegno del documento e per le definizioni e la lavorazione di documenti complessi.

## Formati di dati

- TIFF 6.0 o superiori (anche multi-pagina)
- JPEG, JPEG 2000, BMP, GIF, PNG
- HTML, CSV, TXT (plain text)
- PDF 1.4 o superiori, PDF/A (ISO 19005)
- AFP
- Office formats:
  - DOC/DOCX/DOCM/ODT
  - PPT/PPTX/PPTM/ODP
  - XLS/XLSX/XLSM/ODS

## Prerequisiti

- Windows 7 o successive; Windows Server 2012 o successive
- Office Standard per PC
- Papyrus WebRepository
- Papyrus Capture /Office-In per supporto di formati Office
- Papyrus PDF-in per supporto di PDF nativi