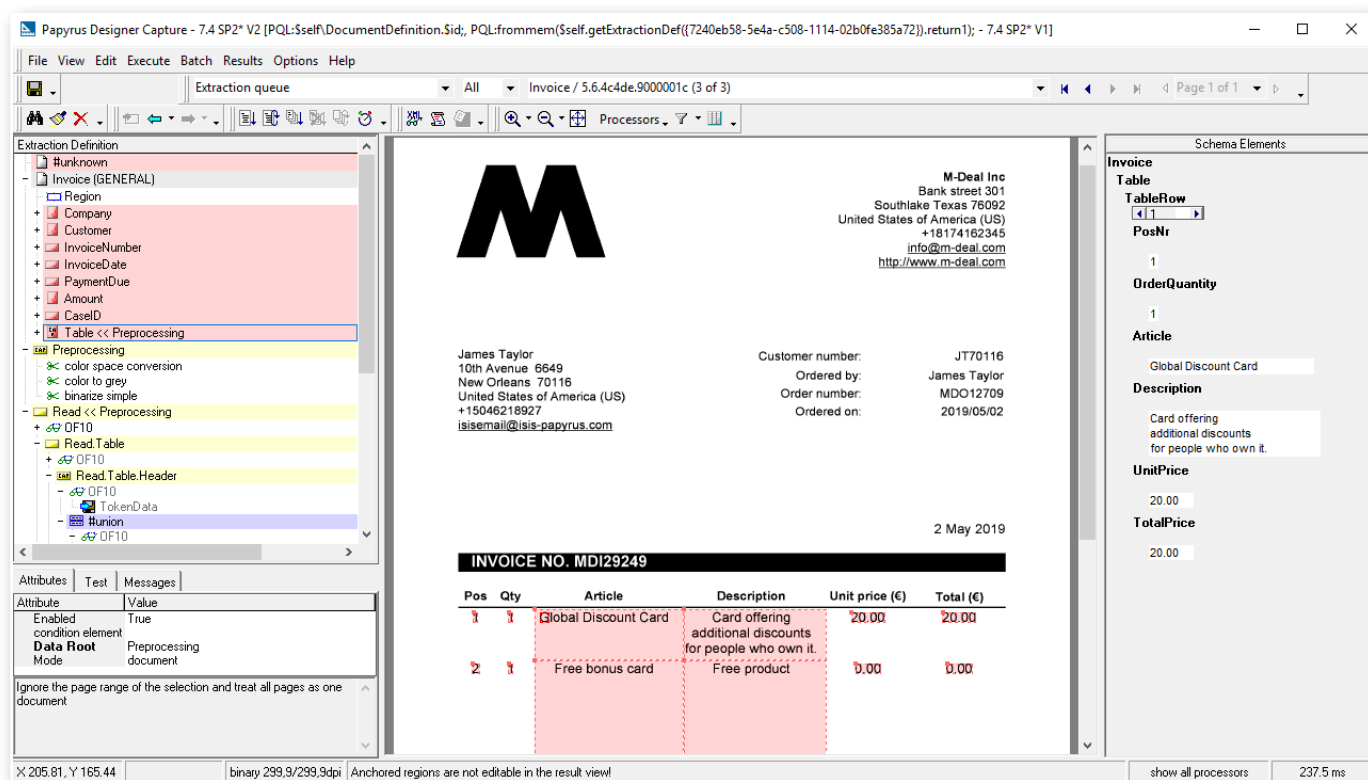


# Papyrus Designer/Capture



## Product Description

The **Papyrus Designer/Capture** is a very powerful toolkit to define classification and data extraction setups for automated data recognition of different kinds of documents.

**Papyrus Designer/Extract** component helps to easily create definitions of scanned or electronic unsorted business documents with unknown structure and layouts such as invoices, lists and complex tables. It also provides all necessary tools for a prompt definition of data extraction parameters to process all kinds of structured form layouts.

**Papyrus Designer/Classify** component supports a self-learning module for classification of documents that offers a broad range of possibilities for application in the field of automated sorting and distribution of electronic documents, fax and paper mail.

Accepting a challenge of the modern world to extract only business relevant content out of high volume inbound paper and digital correspondence, the **Papyrus Designer/Capture** is much more than just a text recognition design tool. It has an excellent capability to identify required business content by means of pattern recognition, keywords, relative positions, rules etc. and a powerful mixture of these means.

*Unmatched power to  
define classification and  
data extraction on all  
types of documents.*

## Contact

### Europe

T: +43-2236-27551

F: +43-2236-21081

### America

T: 817-416-2345

F: 817-416-1223

### Asia Pacific

T: +65-6339-8719

F: +65-6336-6933

### Email & Web

info@isis-papyrus.com

www.isis-papyrus.com

© 2020, ISIS Papyrus,

replaces all previous  
documentation.

## Key Benefits

- Content digitizing: extraction definition of business relevant content out of high volumes of inbound data
- Classification and extraction definition for both FixForm and FreeForm® documents
- Support of documents from all common inbound channels and file formats
- Powerful set of classification and extraction definition features
- Support of diverse recognition engines for different kinds of content
- In collaboration with Papyrus Recognition Server, provides a high-level automated setup for content digitizing
- Fast definition and testing of new document types

## Features

### Package Components

- Papyrus Designer/Extract
- Papyrus Designer/Classify

### Image Preprocessing

Papyrus Designer Capture allows automatic or adjusted preparation of the image data for optimal recognition results. This is especially important for scanned and photo images.

- Dirt removal, Despeckle
- Punch hole removal
- Automatic rotation
- Binarize
- Convert color to gray
- Box, line, grid and combs remover
- Drop out color
- Dilate
- Erode
- Deskew on border
- Deskew on content

### Recognition Engines

Content of preprocessed images is recognized by a powerful set of recognition engines belonging to the following main groups:

- Optical character recognition (OCR) – machine written text
- Intelligent Character Recognition (ICR) – hand written text
- Optical Mark Recognition (OMR) – check boxes
- Bar code, QR code, DataMatrix
- Logo

## Papyrus Designer/Classify

The Papyrus Designer/Classify is used to define categories to be classified, the extraction of feature sets and the classification strategy.

The Designer/Classify supports automation of several definition steps. It also enables the easy handling of sample data. The Designer/Classify is fully integrated into Papyrus Objects.

Supported classification steps based on:

- Keywords and its combination
- Layouts
- Advanced text
- Logos

Before you start working with the Designer/Classify, the required document categories have to be defined. With a number of sample documents (a few dozen) for each category the system is trained in order to learn the specific properties of each document class. New categories can also be added easily and conveniently.

The Papyrus Designer/Extract component empowers classification capabilities by letting to define some classification logic even on a later stage – during extraction.

## Papyrus Designer/Extract

To discover relevant content the following main possibilities and their mixture are usually used in the extraction definition

- Patterns
- Keywords
- Regions
- Positions
- Anchor points
- Rules
- Layout
- Logo

### Basic Definition Process Structure

- Define all document types that occur
- Define the elements of interest on each document type
- Define for each data element:
  - General attributes
  - Related anchors
  - Patterns that apply to the element
  - Condition (rules) that have to be fulfilled
- Test the definitions with various documents

- Compare the results with previous generated “golden files” of intended extract data.

Various display support tools enable you to define new documents fast:

- Display candidates for extraction
- Show regions and coherences
- Convenient tag tools

The Papyrus Capture/Extract module provides a possibility for document reclassification definition or classification definition based on the very deep document structure analysis. This gives a big flexibility for the document design and processing definitions of complex documents.

## Data Formats

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

## Prerequisites

- Windows 7 or later; Windows Server 2012 or later
- Standard office PC
- Papyrus WebRepository
- Papyrus Capture /Office-In for office formats support
- Papyrus PDF-in for native PDF support

## Order Information

*Papyrus Designer/Capture*  
for Windows

## Training

ISIS Papyrus offers in-house workshops and standard courses for user training.

## Maintenance

The Terms and Conditions for ISIS Papyrus Software Products apply. Free service period is 6 months after installation. A maintenance agreement is offered optionally for updates and product support.