



# Papyrus Designer Package/Capture

**Document & field structure**

**5 OCR/ICR engines**

**Definition of parameters**

**Powerful tools for teach-by-example**

**Display of evaluation images**

**Immediate preview of extraction results**

The screenshot shows the PapyrusDesigner / FreeForm Designer - [Workshop] interface. It includes a menu bar (File, Edit, View, Sample, Recognition, Help), a toolbar, and several panels. On the left, a 'Class' tree shows a hierarchy for 'Invoice' with sub-items like 'Invoicedate', 'Invoicenummer', 'A\_Invoice', 'Customernumber', 'Amounts', 'Region', 'VAT', 'Brutto', 'C\_VAT', and 'Customer'. Below this is a 'Parameters' table with columns for 'U.', 'Express.', 'Ignore', and 'Prior'. The main workspace displays a scanned invoice with a grid overlay. On the right, a 'Training' panel shows 'Result' data for the 'Invoice' class, including fields like 'Invoicedate', 'Invoicenummer', 'Customername', 'Brutto', and 'Netto'. A 'Positions' table at the bottom right shows extracted data for 'Article', 'Count', and 'PricePerStk'.

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

## Product Description

The *Papyrus Designer Package/Capture* is the convenient ISIS Designer toolkit family to define classification and data extraction set-ups for every type of documents. Essential features to speed up the customizing process include re-use of predefined parameter settings, teach-by-example wherever possible and self-explaining rules are as a matter of course.

*Papyrus Designer/FreeForm*® is an extremely powerful software module to easily create definitions for automated data recognition of scanned or electronic but unsorted business documents with unknown structure and layout like invoices, lists etc.

*Papyrus Designer/FixForm* provides all necessary tools for speedy definition of data extraction parameters to process all kind of structured forms layouts.

*Papyrus Designer/Classify* 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.

## Contact

### Europe

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

### America

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

### Asia Pacific

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

### E-mail

info@isis-papyrus.com

### Internet

www.isis-papyrus.com

© 2012, ISIS Papyrus, replaces all previous documentation.

January 1st, 2012

## Benefits

- Faster definition and testing of new document types than any other product
- A special design-tool for each of the three basic document understanding requirements (Classification, Data extraction FixForm and FreeForm®)
- Ergonomic, self-explaining user shell – no programmer's skills required
- Universal function Repository offers re-use of generated definitions wherever possible
- Analyze and statistic functionality (regression tests) included

## Package Components

- Papyrus Designer/FreeForm®
- Papyrus Designer/FixForm
- Papyrus Designer/Classify

## Common Features

Image Pre-processing

All designers allow automatic or adjusted preparation of the image data for optimal recognition results:

- Despeckle
- Dilate
- Erode
- Line removal, Dirt removal
- Punch hole removal
- Automatic rotation
- Binarize
- Convert Color to gray
- Deskew on border
- Deskew on content

Image Recognition of anchor, bitmaps, lines, barcode, OMR, text: 5 OCR/ICR engines cover every recognition challenge, Character types easy selectable

Flexible Post OCR Processing with filter patterns, space compression etc.

## Designer/FreeForm®

The "star" among the Capture Designers allows both document type based generation and generic definitions and layout-based region extraction. The combination of both results in shortest time definition effort with best results possible.

- Generic Definition
  - Rule based extraction
  - Maintainable rule system
  - Advantage: works on document layouts never seen before

- Layout-based Definition
  - Learn-by-example
  - Template based, region orientated data extraction
  - Optimization to specific attributes of document types

## 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 in minutes, not hours:

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

## Designer/FixForm

Persistently designed on XML structure principles, the Papyrus IDEX designer pools more than a decade of experience with forms extraction.

The workspace contains all necessary parts

- Image frame to display the example document to be read out
- Extraction tree with all defined data elements and their accompanying parameters like zone or recognition type
- Attribute frame with detailed attributes of each selected element
- Processing result frame for immediate visualization of results.

## Designer/Classify

The Designer/Classify is used to define the 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 and supports test runs and result analysis. The Designer/Classify is fully integrated into Papyrus Objects.

## Supported classification steps based on:

- Image attributes of the whole page or in certain regions
- Keywords at predefined positions or anywhere in the text
- Text phrases at predefined positions or anywhere in the text
- Rules

The Designer/Classify visualizes the automatically generated categorizing attributes and allows to

- Modify parameters like the relevance of keywords, classification threshold
- Analyze the recognition, reject and substitution rates and at which documents they occur
- Adapt classification method for special document types.

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 while the system is in full operation. Continuous re-training as a seamless by-product to manual classification is possible.

## Data formats

Supports a large number of file formats, Image (TIFF, JPG, PDF, AFP) and Text (DOS, ANSI, Unicode), Microsoft Office files.

## Technical Prerequisites

- Pentium IV or higher
- 512 MB RAM (1 GB recommended)

## Order Information

*Papyrus Designer Package/Capture 7*  
for Windows 2K/XP/2K3/VISTA/2K8/7

## Training

ISIS offers a special in-house training service for its clients, the Papyrus Designer Capture Workshops. Please contact us for more details.

## Maintenance

The Terms and Conditions for ISIS Software Products apply. Free service-period is 6 months after delivery. A maintenance agreement is offered optionally for updates and hotline.