

docas[extract]

XML-Data capturing out of output data

Rule-based XML-Raw data extraction out of spool files

docas[extract] technology puts you in the comfortable position of being able to participate in ebusiness without making any modifications whatsoever to existing applications. One of many applications addressed by docas[extract] is the generation of electronic invoices from available print data and their publication in various media (B2B/B2C e-billing platforms, Web services, email, WML). No changes are required to your existing infrastructure. When using docas[extract] it's possible to extract document information from available proprietary mass print data including AFP, PCL, PDF etc. and transform them into semantic XML data. XML offers the ideal portability technology for information via different platforms, applications and organisations.

Overview

- Visual Business Rule Designer
- Creation of structured XML documents for the Web from print data
- Print data index extraction
- Automatic document classification
- Data Mining
- High volume data conversion
- Extensive support for AFP, PCL and numerous other input and output formats
- Comprehensively supports integration into existing customer applications

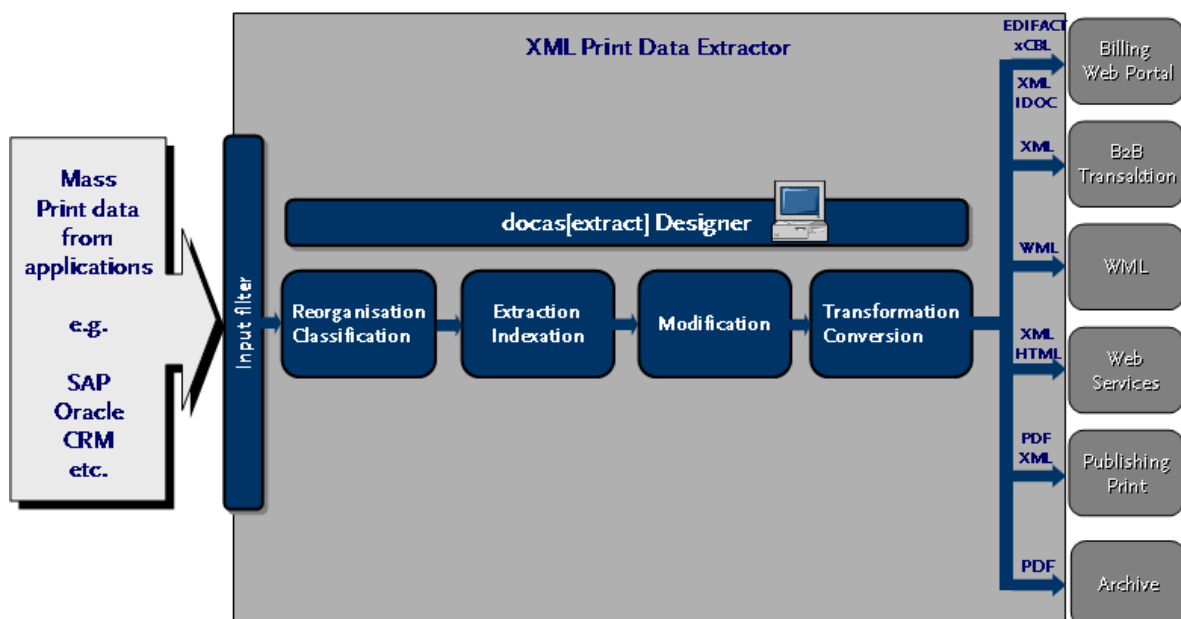
Why convert documents into XML?

The main reason behind XML's popularity is the underlying technology. It offers a high level of portability for disseminating

business information to different platforms and applications. Its strength lies in the possibility of exchanging structured data among various organisations.

But XML technology is just as important in creating, archiving, indexing and publishing business documents. Nowadays unstructured (and semi-structured) data within companies are stored at various locations and typically comprise 80% of all corporate data/information.

Probably the most decisive reason for converting documents into XML is when they have to be published in various different media (multi-channel publishing).



Rule-based data extraction

A new technology enables XML data to be extracted from proprietary print data streams. Rule based data extraction lets you export print data into XML without changing or reprogramming existing systems. With the help of a simple visualisation tool, you define how document content should be displayed in XML.

The Business Rule Designer

docas[extract] contains a visualization tool that lets you define how document content should be displayed in XML. The rules enable fully automatic conversion of print data into XML format. The Designer can also handle complex document structures.

The docas[extract] Designer is used to analyse documents and formulate assignment rules. Document content can be tagged with the WYSIWYG editor and linked to the corresponding data fields. The items are administered in lists for rule formulation. During the definition process the content

is extracted and displayed based on the model document. Declaring complex conditions including line breaks, page breaks, optional elements, dynamic text and multicolumn tables is a breeze.

Integration functionality

docas[extract] can be run in batch mode and integrated into existing process chains or operated "on the fly". The Software Development Kit (SDK) features the option of embedding the functions into analogous development environments like C, C++ or Java.

Document format transformation

docas[extract] not only converts documents and/or print batches into semantic XML data, but into other output formats for multi-channel publishing as well. For example, electronic invoices from a print data stream can be output in XML and other standard marketplace for-mats.

Supported operating systems

docas[extract] runs on Windows NT, 2000 and XP, Linux, zLinux, AIX and Sun Solaris.

The screenshot shows the Business Rule Designer interface. On the left, there is a list of existing rules. In the center, there is a table of extracted data. On the right, there is a detailed view of a selected extraction rule, showing its template, rule ID, and description. Below the rule view, there is a current XML element and a document image.

docas[extract] enables the mining and extraction of XML data out of legacy spool files. No change of the legacy application is necessary because docas[extract] fits seamlessly into any existing document creation infrastructure.

If it is mandatory to enable the corporate data for ebusiness or other repurposing scenarios, then...

docas[extract] is the right answer!

For further information, please contact:

Docas AG
Sumpfstrasse 26
PO Box 947
6301 Zug
Switzerland

Tel: +41 (41) 747 43 90

Fax: +41 (41) 747 01 41

Email: info@docas.com