



Automating PDF Document Creation: PDF-on-the-Fly and More

Seybold Seminars, Amsterdam 2004

**Thomas Merz
President**

**PDFlib GmbH
München, Germany**

www.pdflib.com



Why PDF-on-the-Fly?

- ▶ PDF preferred over HTML for various types of documents
- ▶ Content is not available in advance
- ▶ Number of documents or pages is very high
- ▶ Content must be customized according to user input, database, etc.
- ▶ Conventional approach for creating PDF doesn't work in these cases:
 - because of performance problems
 - because of real-time requirements
 - because of the required manual effort
 - because of platform (operating system) availability



Trends

- ▶ Everything needs a Web frontend (except your kitchen sink)
- ▶ Automation: create documents in unattended mode
- ▶ Customized and personalized documents (marketing, printing cost, storage)
- ▶ Digital printing and short runs
- ▶ One-to-one marketing



Generic PDF Creation

- ▶ Generate page content using a low-level programming interface
- ▶ Many low-level products available (cheap, but limited functionality)
- ▶ Pros:
 - very fast
 - simple to use
 - complete freedom over content creation
 - no knowledge of PDF internals required
 - attach to existing software such as formatting engines, document systems
- ▶ Cons:
 - must manually code the page layout and formatting
- ▶ Applications:
 - invoices, statements, reports with simple layouts
- ▶ Products:
 - PDFlib, iText, TallPDF.NET, Big Faceless Java PDF library



Generic PDF Manipulation

- ▶ **Manipulate PDF documents on an object level**
- ▶ **Pros:**
 - access to all PDF internals
 - modifications to existing PDFs possible
- ▶ **Cons:**
 - slow
 - very complex API
 - requires detailed knowledge of PDF internals
 - very expensive, difficult to license
- ▶ **Applications:**
 - general PDF creation and editing
- ▶ **Products:**
 - Adobe PDF Library (distributed by Datalogics), Appligent SPDF, GlobalGraphics



Report Generators and Database Drivers

- ▶ Create PDF output directly from a database or XML input
- ▶ Pros:
 - use existing report generators and languages, no new API
 - comes for free with some databases
- ▶ Cons:
 - some create bad PDF output since database vendors are not PDF experts...
 - limited functionality
- ▶ Applications:
 - PDF-enable existing database applications
- ▶ Products:
 - database: Oracle PDF driver, Crystal Reports
 - XML: Big Faceless Report Generator, TallPDF.NET, ReportLab



Recombining Existing PDF Pages

- ▶ Work with existing PDFs to create new ones
- ▶ Pros:
 - input PDF pages can be designed manually with any application
 - easy programming on a per-page basis
 - can be combined with generic content creation (e.g. add page numbers)
- ▶ Cons:
 - limited to page granularity (no text reflow across page boundaries)
- ▶ Applications:
 - all documents which can be assembled from existing parts, such as customized catalogs
 - document assembly
 - simple imposition
- ▶ Products:
 - PDFlib+PDI, ActivePDF Toolkit, Appligent StampPDF, AppendPDF



Forms Merging

- ▶ Use PDF form fields and FDF/XFDF to customize documents
- ▶ Client-side merging doesn't require server software, but suffers restrictions with Acrobat Reader
- ▶ Pros:
 - existing forms can be implemented quickly
- ▶ Cons:
 - no prepress features, high-level formatting, high-quality typography
 - resulting field contents are editable unless the fields are flattened
 - preparing large forms with Acrobat's form tool may be cumbersome
- ▶ Applications:
 - existing forms with simple layout requirements (especially government)
- ▶ Products:
 - Appligent FDFMerge, ActivePDF Toolkit



Document Converters

- ▶ Convert existing documents in any format (e.g. MS Word) to PDF
- ▶ Pros:
 - use existing (manual or automated) processes for document creation
- ▶ Cons:
 - not suited for dynamic content creation (only dynamic PDF creation)
- ▶ Applications:
 - publish enterprise paperwork in PDF
- ▶ Products:
 - ActivePDF DocConverter, AdLib eXpress Server



Standard Desktop Software

- ▶ Use ready-made desktop applications to create PDF output
- ▶ Caution: not recommended for server deployment!
- ▶ Pros:
 - no custom programming required
- ▶ Cons:
 - heavy-weight and slow
 - stability and performance problems
- ▶ Applications:
 - environments where these applications are heavily used anyway
- ▶ Products:
 - Microsoft Word, TeX, custom PostScript output plus Distiller/Ghostscript



XSL/FO Formatters

- ▶ **Sophisticated formatting based on XSL/FO standards**
- ▶ **Requires intimate knowledge of XML, XSL, FO and related tools**
- ▶ **Pros:**
 - **powerful formatting based on standards**
 - **existing XML/XSL workflows can be modified to create PDF output**
- ▶ **Cons:**
 - **steep learning curve and high development effort for XML and XSL creation**
 - **FO is comparatively simplistic, e.g. prepress features are missing**
 - **requires DTD or schema**
- ▶ **Applications:**
 - **large-scale automated document production: books, catalogs**
 - **structured documents**
- ▶ **Products:**
 - **(Apache FOP), RenderX XEP, AntennaHouse XSL Formatter, Advent 3B2-FO, Adobe Document Server**



PDF Personalization

- ▶ Personalize PDF templates with custom text or image content
- ▶ Add template information to PDF
- ▶ Pros:
 - highly flexible
 - easily prepare PDF templates with common DTP tools and Acrobat
- ▶ Cons
 - limited to page granularity (no text reflow across page boundaries)
- ▶ Applications:
 - one-to-one marketing
 - direct mailings
 - automated production of personalized business cards
- ▶ Products:
 - PDFlib Personalization Server, Lexigraph (formerly Think121), Verity LiquidOffice (formerly Cardiff)



Evaluation Criteria

- ▶ **Suitability for server deployment: robust, thread-safe, no memory leaks**
- ▶ **Scalability**
- ▶ **Speed**
- ▶ **Availability on operating system platforms**
- ▶ **Availability for programming languages/development environments**
- ▶ **Implementation and integration effort: complexity of API**
- ▶ **PDF feature set and support for auxiliary file formats (fonts, images, ICC, ...)**
- ▶ **Cost**



Case Study

- ▶ **Speaker: Guy Spriggs of Rosetta Internet Database Solutions, Amsterdam**
- ▶ **Project: online ad booking system based on PDFlib**



Resources

- ▶ Hundreds of PDF tools: www.pdfstore.com
- ▶ Seybold Seminars Web site: www.seybold365.com
- ▶ Moderator: Dr David Penfold, penfold@eps-edge.demon.co.uk
- ▶ Speaker: Thomas Merz, tm@pdflib.com
- ▶ Seybold Director of Content: Cynthia Wood, Cynthia.Wood@mlii.com
- ▶ Seybold VP & General Manager: James Smith, James.Smith@mlii.com