

Content-Centric Applications and the xDoc Converter

Financial services firms across the globe are beginning to develop *content-centric applications*, or applications that leverage existing and on-going content to provide their businesses with a valuable competitive edge.

The Problem

For financial service firms, *information is their business*. Be it generating custom financial reports or matching titles to mortgage brokers, financial services firms traffic in information and are defined by the data that passes through their systems. Because information is their business, financial service firms want to leverage their previous investments, and extract more value from their existing content by offering enhanced content services for their customers. In a word, they are trying to *repurpose* their content. Dictionary.com defines the word “repurpose” as follows:

re·pur·pose (rĕ-pûr'pōs)
To use or convert for use in another format or product.

This describes perfectly, at a high-level, what these firms are trying to accomplish. They are typically looking to convert their legacy documents and ongoing content-streams into XML format so that they can be uploaded to a SQL or XML database for quicker searching and client access. They are typically looking to break their content down into granular chunks, so that it can be reassembled dynamically for custom reports or contracts. And they are typically looking to publish their content across different formats, be it HTML, Word, or PDF.

But these firms also have several challenges to overcome. First, *they process information on an ongoing basis*. Financial services firms have thousands of documents to process each month, and because of that volume, they find it cost-essential to handle the content in a “lights-out” fashion – i.e. at night, or without any human intervention whatsoever. Second, financial services firms *have to handle content that comes from many different sources*, including governmental and regulatory bodies, licensing partners, employees and customers. Finally, *they have to handle content that comes in many different forms, and follows many different styles*, be it a PDF table of asset data, a Microsoft

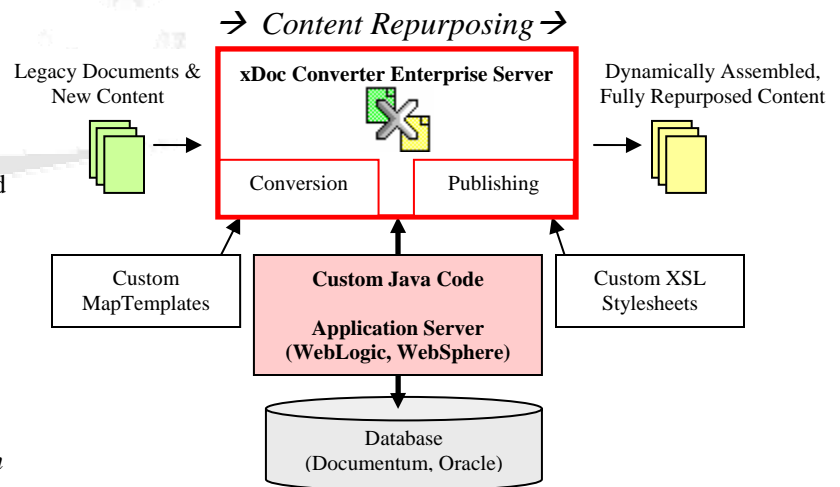
Word internal report, or an HTML file of regulatory information.

The Solution

To create enhanced service offerings, financial services firms have developed in-house, content-centric applications. These applications are web-based, distributed, multi-tier, Windows / Linux / Solaris server-based applications that receive content either via hot-directories, web-based input, or dedicated application plug-ins.

And they have used the CambridgeDocs xDoc Converter Enterprise Server to repurpose their content. They have used it to convert legacy information into XML format, and they have used it to publish dynamically assembled PDF, RTF, and HTML content.

Because their characteristics and repurposing goals are often similar, the content-centric applications that these xDoc customers and others like them are developing typically have a similar, high-level architecture, as shown below.



As this diagram illustrates, these xDoc customers developed their content-centric applications using Java and J2EE technologies, in order to a) avoid being locked into a Windows / Office environment, b) leverage high-powered Solaris systems, or c) leverage cheaper Linux systems. They typically used either WebSphere or WebLogic Application Servers as the core means of accessing their database systems.

In addition to the custom Java code, IT engineers at the financial services firms also created a set

of custom rules, called MapTemplates, which direct the xDoc Converter Enterprise Server to convert their legacy Word, PDF, or HTML content into an XML format that follows a proprietary schema or DTD. These MapTemplates are then fed into the xDoc programmatically, along with the files that need to be converted, and the resulting XML output is uploaded via the Application Server into some sort of database, typically Documentum, Oracle, or Software AG.

Finally, in addition, these content-centric applications typically use custom XSL stylesheets to take output from these databases and dynamically assemble HTML web pages, PDF contracts, or RTF reports, using chunks of content from the legacy documents plus business data.

Conclusion

This document describes, at a high-level, the goals of some of CambridgeDocs financial service firm clients, the challenges they faced, and the solution they adopted. While financial services firms have typically been the most high-profile CambridgeDocs customers, a number of other firms have also developed or are starting to develop content-centric, content-repurposing applications for their businesses. For example, Sun Country Airlines has recently developed a content-centric application to streamline and enhance maintenance procedures, and Southwire, a \$1.7B wire manufacturer, developed a content-centric application for the conversion of HTML web catalog pages to PDF format for printing. Content-centric applications represent a growing market that CambridgeDocs will continue to address.



To Contact CambridgeDocs and find out more about Content Centric Applications and the xDoc Converter:

*CambridgeDocs, LLC
186 Alewife Brook Pkwy, Suite 300
Cambridge, MA 02138*

Phone: 617-349-3606

Email: Sales@cambridgedocs.com

Web: <http://www.cambridgedocs.com>