

ZAPTHINK ZAPNOTE™

XMLCITIES AUTOMATING XML DOCUMENT CREATION

Briefing Date: September 2001

Analyst: Ronald Schmelzer

Abstract

In order for any of the advantages of XML technology to be applied, XML documents must first themselves be created – either from scratch or converted from “legacy” document formats. The XMLCities’ solution provides automated and “non-intrusive” conversion of paper and legacy documents requiring minimal manual operation.

All Contents Copyright © 2001-2002 ZapThink, LLC. All rights reserved. Reproduction of this publication in any form without prior written permission is forbidden. The information contained herein has been obtained from sources believed to be reliable. ZapThink disclaims all warranties as to the accuracy, completeness or adequacy of such information. ZapThink shall have no liability for errors, omissions or inadequacies in the information contained herein or for interpretations thereof. The reader assumes sole responsibility for the selection of these materials to achieve its intended results. The opinions expressed herein are subject to change without notice. All trademarks, service marks, and trade names are trademarked by their respective owners and ZapThink makes no claims to these names.



Converting Legacy Content into XML Documents

In order for any of the advantages of XML technology to be applied, XML documents must first themselves be created – either from scratch or converted from “legacy” document formats. While the former case is one of using advanced integrated development environments (IDEs) or embedded tools, the latter case is more complicated. How can documents that are unstructured be converted in an automated manner into structured documents?

Typically, conversion of large quantities of documents such as Word, PDF, Quark, or even paper legacy documents to XML requires intensive, high-cost manual operations. In addition, those that are performing the conversions rarely have programmer or developer-level skills requiring conversion tools to be user friendly, seamless and efficient. In addition, XML content creation requires extensive knowledge of XML syntax and a user can only create limited XML documents based on pre-defined template. Outsourcing document conversion brings with it language problems and large effort coordination costs.

It is this challenge that XMLCities and other vendors in the category of document conversion and automated XML document creation address. The XMLCities’ solution provides automated and “non-intrusive” legacy to XML content creation and conversion requiring minimal manual operation.

Originally known as Paravision, the company built PDF, HTML and RTF conversion tools for large service bureaus. Two years ago they saw the emergence of XML technology and consequentially the company decided to make a shift. In 2000, they began a focus on XML and subsequently changed their name to XMLCities. XMLCities’ product suite is composed of XML Capture and XMLMaker, each of which leverage a number of other technologies developed by the company.

XML Capture Suite

XMLCapture Suite is aimed at automating XML document creation through automatic and intuitive document tagging. The technology is designed for non-development users leveraging GUI and WYSIWYG environments to manipulate a large volume of XML documents. In effect, the system allows a content creator to use familiar word processor or other document creation software and then utilize the XMLCapture Suite to turn the documents into XML in the background and then can output to PDF, RTF, HTML, Images and XML formats.

XMLCapture performs this task through an innovative auto-tagging feature that uses pattern recognition to assign data to appropriate DTD tags based on comprehensive rules set-up by

TAKE CREDIT FOR READING ZAPTHINK RESEARCH!



Thank you for reading ZapThink research! ZapThink is an IT market intelligence firm that provides trusted advice and critical insight into XML, Web Services, and Service Orientation. We provide our target audience of IT vendors, service providers and end-users a clear roadmap for standards-based, loosely coupled distributed computing – a vision of IT meeting the needs of the agile business.

Earn rewards for reading ZapThink research! Visit www.zapthink.com/credit and enter the code CITYDOC. We'll reward you with ZapCredits that you can use to obtain free research, ZapGear, and more!

For more information about ZapThink products and services, please call us at +1-781-207-0203, or drop us an email at info@zapthink.com.

content administrators. Rules can be established with "IF-THEN-AND-OR" type strings in order to assign tags based on key words, phrases, document location, data type, or other identifiable patterns or identifiers. XMLCities also is developing a pattern-scripting language for client customization.

The XML Capture Suite was designed with service bureaus, corporate IT and content management departments in mind and has seven different modules supporting large volumes of seamless valid and well-formed XML document creation and conversion. This suite includes PageGenie Enterprise for document OCR and capture from paper, GenieWriter for converting PDF files, PageGenie Controller for production batch file dispatching and batch load balancing, XDTD Editor for DTD creation and editing, XMLPattern for defining pattern rules for auto-tagging, PageGenie XML for intuitive XML tagging, both auto and manual tagging, and XMLProofer for advanced XML quality assurance.

PageGenie Enterprise turns scanned paper-based documents into a proprietary intermediate file format called PageGenie Meta Format (PML). This is then passed for tagging to PageGenie XML. GenieWriter converts PDF files into PMF for use by other modules in the suite during the XML conversion processes. XDTD Editor creates, imports, and transfers DTDs in preparation for XML tagging. XMLPattern allows content managers to define intelligent rules for PageGenieXML's auto-tagging process. PageGenieXML is used for automatic and/or manual XML tagging based on pre-configured DTDs. PageGenie Controller provides enterprise-scale batch processing with multiple workstation support. It manages the XML conversion or creation work process, dispatching work from initial input, auto-tagging, manual tagging, proofing, and posting. Finally, XMLProofer verifies and validates created files XML against the DTD tree. This module links the DTD tree, XML document, and original scanned image and highlights errors for operator intervention and modification..

The amount of automated tagging that can be accomplished depends upon the type of document to be converted and the resolution of the DTD structure assigned. With fairly structured documents, over 90% of the XML tagging can be accomplished with the Auto-tagging feature. For other unstructured data, 50-80% of the Auto-tagging may be accomplished. Non-technical operators use the system by associating content to a DTD tree, and the tagging is done automatically. Users can then quickly browse through documents to determine what has not been tagged and quickly identify other data that has been previously tagged.

XMLMaker Suite

XMLMaker Suite is very similar to XMLCapture, except that it is designed for smaller volumes of document conversion. In addition to the above features of XMLCapture, XMLMaker has an additional module called XSLT Maker that allows created documents to be transformed into other documents or formats as well as accepting RTF formatted files as input documents.

The XMLMaker Suite includes six components and has two optional pieces. This includes GenieWriter Lite for converting PDF to RTF, XDTD Editor for DTD preparation for XML tagging, XMLMaker for automatic and manual visual tagging, XMLEditor for quality assurance and validation of XML, XSLT Maker for open XSLT editing, and optionally XMLPattern for defining pattern recognition rules for auto-tagging, and PageGeniePro for OCR document capture.

The included XSLT Maker module is a next generation XSLT design tool facilitating advanced net publishing for efficient transformation from XML-to-XML or XML to any output format such as PDF, HTML, WAP and cHTML.

Competition

XMLCities faces competition from two fronts: document conversion utilities and XML document creation tools. On the former front, a number of vendors including Progressive Information Technologies (PIT) and a number of conversion services offer challenges, while on the latter front SoftQuad, Arbortext, and others provide a certain amount of document conversion capability. Hypervision and I4I also offer compelling tools to convert Word documents to XML. However, these tools and services require document creators to use their development environments to create XML documents, while XMLCities allows content users to use whatever development environment they are already comfortable with.

Customers & Release History

XMLCities is now in evaluation mode and has released beta early last month with about eight evaluations in progress. They will release product in early 2002 incorporating feedback from these beta users. Some industries currently testing the product include governmental and military organizations and human resources seeking to convert resumes..

There are three kinds of users mainly: large service bureaus, system integrators and content management OEMs, and end users who have lots of content that needs to be converted to XML. In all of these cases, the company is seeking organizations with large volumes of documents to convert.

XMLCities prices its products and services based on overall capacity, customization, and/or an unlimited usage license.

Key Conclusions & Recommendations

- XMLCities' suites provide a good option for companies with large volumes of paper or electronic documents that seek a good automated method for XML conversion.
- However, these documents need to be fairly regular in nature for successful automation to occur. Regardless, the company also has a manual approach for conversion.
- Competitive pressures exist and XMLCities should careful position itself as a less intrusive option versus Content Creation utilities and as a more automated approach versus other content conversion tools.

Profile: XMLCities (September 2001)

Date Founded: 1997 (Company renamed in 2000 from Paravision)

Funding: Privately-held, Joint Venture with Mitsubishi Corporation

CEO / President: Dr. Evan Huang

Products:

- XMLCapture Suite
- XMLMaker Suite

Address:

XMLCities, Inc.
1619 South Main Street
Milpitas, California 95035

URL: www.xmlcities.com

Main Phone: 408-934-7878

Contacts:

George Nishiura gnishiura@xmlcities.com
Wes Lagrone, wlagrone@xmlcities.com

Related Research

- *XML in the Content Lifecycle* Report (ZTR-CL100)
- *Corel ZapNote* (ZTZN-0118)
- *HyperVision ZapNote* (ZTZN-0249)

About ZapThink, LLC

ZapThink is an IT market intelligence firm that provides trusted advice and critical insight into XML, Web Services, and Service Orientation. We provide our target audience of IT vendors, service providers and end-users a clear roadmap for standards-based, loosely coupled distributed computing – a vision of IT meeting the needs of the agile business.

ZapThink's role is to help companies understand these IT products and services in the context of SOAs and the vision of Service Orientation. ZapThink provides market intelligence to IT vendors who offer XML and Web Services-based products to help them understand their competitive landscape and how to communicate their value proposition to their customers within the context of Service Orientation, and lay out their product roadmaps for the coming wave of Service Orientation. ZapThink also provides implementation intelligence to IT users who are seeking guidance and clarity into how to assemble the available products and services into a coherent roadmap to Service Orientation. Finally, ZapThink provides demand intelligence to IT vendors and service providers who must understand the needs of IT users as they follow the roadmap to Service Orientation.

ZapThink's senior analysts are widely regarded as the "go to analysts" for XML, Web Services, and SOAs by vendors, end-users, and the press. They are in great demand as speakers, and have presented at conferences and industry events around the world. They are among the most quoted industry analysts in the IT industry.

ZapThink was founded in October 2000 and is headquartered in Waltham, Massachusetts. Its customers include Global 1000 firms, public sector organizations around the world, and many emerging businesses. ZapThink Analysts have years of experience in IT as well as research and analysis. Its analysts have previously been with such firms as IDC and ChannelWave, and have sat on the working group committees for standards bodies such as RosettaNet, UDDI, CPExchange, ebXML, EIDX, and CompTIA.

Call, email, or visit the ZapThink Web site to learn more about how ZapThink can help you to better understand how XML and Web Services impact your business or organization.

ZAPTHINK CONTACT:

ZapThink, LLC
11 Willow Street
Suite 200
Waltham, MA 02453
Phone: +1 (781) 207 0203
Fax: +1 (786) 524 3186
info@zapthink.com