

ZAPTHINK ZAPNOTE™

COMTEX IMPLEMENTING NEWSML

September, 2001

Analyst: Ronald Schmelzer

Abstract

COMTEX, an organization that serves as an infomediary between news providers such as Associated Press and the various news distribution and display sites is making use of NewsML, a format for the structure and delivery of news content stories. This note details their use of the format and some detail on NewsML.

All Contents Copyright © 2001 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.



Distributing News Content

XML is making waves everywhere where information and content is present, and the news industry is no exception. Responsible for the creation and distribution of thousands of new items of content on a daily basis, the various businesses and organizations in the news industry have come together to develop and promote an XML-based format for news interchange. One of these organizations leading the way is COMTEX, an organization that serves as an infomediary between 75 news providers such as Associated Press and the various news distribution and display sites.

The challenge in aggregating and distributing news content is that each of the different news feeds provides news and content information in their own format. Without XML, aggregation of news information is chaotic and overly complex. However, with the development of NewsML by the International Press Telecommunications Council (IPTC), a new means for open, non-proprietary exchange of news story information is possible.

Prior to the development of the XML format, COMTEX utilized a modified version of the ANPA 1312 standard as the envelope around its news distribution formats. However, recent demands by customers to a more flexible and targeted content offerings required the development of a new format based upon XML. In March 2001, COMTEX News Network announced its support for NewsML as a means for packing news information for delivery to their subscriber base.

NewsML

NewsML created by IPTC as a continuation of efforts in standardizing news story content. The basic problem that the format is trying to solve is that news organizations are providing news content in different, sometimes proprietary formats. Many news sources publish their stories in "broadcast format", which is a text-delimited or fixed-length text format. The first goal was trying to produce a single format using XML as the means for defining the language. A number of major news organizations including AP and Reuters are part of this effort. With NewsML, they publish in one, industry-wide format. NewsML has a tag vocabulary that describes the body of content as well as a means for extending the body with custom tags. On top of NewsML, COMTEX has developed a means for categorizing news stories and adding key words. In this manner, NewsML can be used as a basis for value-added news offerings and services.

NewsML also provides rich metadata capabilities that allow it to be further enhanced and targeted for delivery. The NewsML format does not designate strict rules governing which vocabularies are to be used to describe content, thus allowing for the association of metadata

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 COMNEWS. 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.

from both proprietary and standardized vocabularies. COMTEX takes advantage of this flexibility utilizing several IPTC vocabularies as well as its own proprietary tags that encode such information as company profile and stock ticker data. Another feature of the language allows users to track the evolution of news items over time. This allows providers such as COMTEX to pass along updates and revisions to content to customers who are dependent on time-sensitive delivery of information, such as breaking news stories, sports scores, or stock information. Furthermore, the NewsML format allows for individual news components to be packaged together as a "story", allowing for coherent transmission of news content to subscribers.

Other XML News Formats

Besides NewsML, there are a number of other XML-based formats for news and content exchange. However, these differ in some substantial ways. The Information & Content Exchange (ICE) specification is a protocol that rides communications between parties. While NewsML is a one-way standard for marking up news documents, ICE is a full protocol that includes handshaking and other basic messaging features. While AP and Reuters are also backing ICE, it is unclear how these organizations plan to make use of the format or how committed they are to its long-term support. Another similar format is the PRISM specification that aims to simplify content exchange as well. At some point, there will need to be clear differentiation and value propositions attached to each of these formats.

COMTEX's use of NewsML

COMTEX uses XML not only for packaging of output but also as an internal representation of news item during processing. Each of the approximately 20,000 unique content units that are delivered to COMTEX each day from its network of information providers is parsed into a valid XML document based on a proprietary, internal-only DTD. This XML document is combined with metadata gathered as part of COMTEX's value-added services, and then transformed using XSLT into valid NewsML documents. The NewsML standard provides the metadata envelope describing the news content contained within. The textual content is marked up using a subset of the NITF DTD.

While COMTEX is just an observer of the NewsML working group and doesn't take an active role, the organization sees XML as an evolving standard that many of their customers are adopting with significant advantages over a proprietary file format. They would like to see more publishers adopt XML standards so that time to implementation for news content is dramatically lowered. Currently, it takes from one to two months to implement a new content source in the COMTEX system.

Key Conclusions & Recommendations

- NewsML as a specification is gaining in usage and adoption across many different kinds of users. As a format for encapsulation of news "stories", NewsML is a compelling format. Companies should evaluate NewsML as per their more specific needs.

-- continued --

Profile: COMTEX	(September, 2001)
Date Founded: 1981	
Funding: Publicly listed (OTC BB: CMTX)	
CEO: Charles W. Terry	
Specifications:	
• NewsML	
Address:	
4900 Seminary Road Suite 600	
Alexandria, VA 22311-1811 USA	
URL: www.COMTEXnews.net	
Main Phone: 1 (703) 820-2000	
Contacts:	
Heather Henderson hhenderson@comtexnews.net	
Slawek Ligier sligier@COMTEXnews.net	

Related Research

- *XML in the Content Lifecycle* Report (ZTR-CL100)
- *Corel ZapNote* (ZTZN-0118)
- *HyperVision ZapNote* (ZTZN-0249)
- *NewsEdge ZapNote* (ZTZN-0239)

-- continued --

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