

ZAPTHINK ZAPNOTE™

NEWSEGE *CATEGORIZING AND EXCHANGING NEWS CONTENT WITH XML*

Briefing Date: December 20, 2001

Analyst: Ronald Schmelzer

Abstract

Since there is an almost infinite, endless stream of news and information to be tapped, the primary issue is not aggregating this content but accurately classifying and sorting it so that a single user can pinpoint and receive the specific information they are interested in without being inundated by a tidal wave of news and information. NewsEdge has developed an innovative and effective means for aggregating, categorizing, and applying taxonomy to news and story information. The system provides real-time Java-based "push" technology that sends out a formatted XML stream as the news or data information is generated. This "feed" delivery option can be tailored as to its frequency, level of specificity, and method of transformation.

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.



Classification and Taxonomy of News Information

Since there is an almost infinite, endless stream of news and information to be tapped, the primary issue is not aggregating this content but accurately classifying and sorting it so that a single user can pinpoint and receive the specific information they are interested in without being inundated by a tidal wave of news and information. This information content can consist of news feeds, web site content, and specific topical information such as medical images, frequent flyer programs, etc.

While the Internet and news feeds serve as the primary source for external information, there is an even greater wealth of information and data stored within the confines of any enterprise or organization. While some clusters of information are neatly categorized in corporate databases, data that is even more important may be buried in corporate email or in complex text and image documents scattered across hundreds of desktops. There needs to be a way to unlock that information and distribute it to the right sources so that it can be taken advantage of. After all, what is the point of storing information if it is never going to be used?

However, classification is not an easy art. Much information exists at the intersection point of multiple different hierarchies and classifications. For example, a document about unemployment figures is, to some degree, about economic news generally, as well as about employment, legal, and regulatory issues. Another document concerning personnel practices in Japan relates both to personnel management generally and to Japanese business. Any document belongs to several hierarchies simultaneously, at one level or another, to one degree of relevancy or another.

NewsEdge: XML in and XML out

Recently acquired by Thomson Corporation, and formed as a merger between DesktopData and Individual, Inc., NewsEdge has developed an innovative and effective means for aggregating, categorizing, and applying taxonomy to news and story information. The company innovated the concept of the "Digital Awareness Marketplace", which is the ability for a company to get a single source of information that is aggregated from multiple media outlets in a personalized, customized manner.

The information that NewsEdge collects is stored within a "content refinery", where it is aggregated and categorized. The Content Refinery receives over 200,000 text-based stories per day from multiple sources around world through FTP, leased line, satellite, and other methods. The system then mixes and matches this information using special tags and a patent-pending process that combines technology with human processing to enable

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 NEWXML. 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.

placement into a global taxonomy. The system is real-time and automated, but also a manual process. This tagging operation forms a classification system that is consistent across different sources, and enables automated classification of thousands of stores and delivery of stories that are relevant to a specific user's needs.

In essence, NewsEdge is in the business of structuring data. Their cutting-edge technology is really their methodology for classifying data rather than the data itself. NewsEdge calls their ability to provide that information in a logically organized format, prioritized according to relevance, with a visible connection to related material "Structured Content, in Context." Since the classification scheme is of the utmost importance, rather than borrowing an existing taxonomy and structure from other industries, NewsEdge built their taxonomies from over 10 years of practical experience, global acceptance, and use of industry experts. The system also "concordes" tags from other classification systems and incorporates them in the final classification system. The end result is not only a representation of information but also what belongs at each node in an information tree. Thus, the system addresses the major question, "What types of information belongs in a classification, and what types do not?"

Originally based on proprietary technology, the content refinery was rebuilt in 1999 on top of an XML DTD based on the News Industry Text Format (NITF) DTD. NITF was developed by the International Press Telecommunications Council (IPTC), an independent international consortium of the world's leading news agencies and publishers. It is a standard that is open, public, proven, well used, well documented, and well supported. NITF supports the identification and description of a tremendous number of news characteristics. Highlights include who owns the copyright to the item, who may republish it, and who it's about, what subjects, organizations, and events it covers, when it was reported, issued, and revised, where it was written, where the action took place, and where it may be released, and why it is newsworthy, based on the editor's analysis of the metadata. NITF formatted documents are produced by The Associated Press, Dow Jones, The New York Times, Deutsche Presse-Agentur and others. It is also generated by such news aggregators as ScreamingMedia and Wavo.

Once NewsEdge aggregates and classifies its informational content, it is transmitted via electronic feeds in the form of XML and software components to customer sites, which are typically the information provider or library part of a company. The system distributes its information through a subscription-based service. Of course, the company is extremely flexible in how it publishes its data, using FTP, file exchange, Java, and other sources to get its classified, aggregated data out to users.

The system provides real-time Java-based "push" technology that sends out a formatted XML stream as the news or data information is generated. This "feed" delivery option can be tailored as to its frequency, level of specificity, and method of transformation. A user can choose between receiving all the information, a headline feed only, or interact directly with information database in real time to retrieve the full-text of the required articles as needed.

The NewsEdge "XML On-Demand" system provides a simple request/response delivery platform with direct access to the content repository. NewsEdge hosts all the informational content, eliminating any human or capita resource requirements for storage and processing of information. Through XML and Web Services, the system provides an API that supplies tools to search for the content users require and enabling integration with applications. In addition, NewsEdge can actually host the application in an ASP-like environment providing an XML news feed that is dynamically transformed to a customizable HTML interface.

Customers & Release History

NewsEdge serves about 1500 clients today and claims the majority of the top Fortune 100 companies as customers. The company works with most major pharmaceutical, financial services, energy, and automotive industry vendors. Within these companies, their primary

users are the knowledge workers that need to be aware of market changing events. One of their news and information products is aimed at Public Relations professionals who maintain queries and are immediately notified when new stories break that interest them.

Competition & Alternatives

NewsEdge's primary competition comes from other news aggregation and categorization vendors including Lexis/Nexis and Factiva. However, there are clearly differentiations between these companies. Factiva is in essence "co-opetition" since they also sell their content to NewsEdge, but compete for certain types of business. However, NewsEdge has better taxonomy and notification methods. In addition to this primary competition, a lot of smaller players have emerged to address different aspects of news aggregation, classification, and distribution. In fact, most Content Management vendors have launched some form of distributed XML application. One of these emerging companies is eNow, which has a specialized desktop application for news collection and display.

In addition, emergence of new data and content formats, such as NewsML may result in new entrants into the news aggregation, classification, and distribution space. As these formats emerge, it is quite possible that NewsEdge will support them and add to their capabilities.

Key Conclusions & Recommendations

- NewsEdge has a compelling solution for news and information aggregation and classification that uses XML for distribution and storage.
- As NewsEdge expands its use of XML, they should look at specifications such as ICE, PRISM, RSS, and NewsML to see if they are appropriate for use to distribute aggregated news content to their customers
- Companies looking for a robust news feed that supports XML exchange should investigate the NewsEdge offering.

Profile: NewsEdge	(December 2001)
Date Founded: 1989	
Funding: Publicly-traded Company	
Acquired by Thomson Corporation [TSE: TOC] in Sept. 2001	
CEO / President: Clifford M. Pollan	
Profiled Products:	
<ul style="list-style-type: none">• NewsEdge service• XML on Demand	
Address:	
80 Blanchard Road	
Burlington, MA 01803	
URL: www.newsedge.com	
Main Phone: 781-229-3000	
Contacts:	
Charlie White charlie.white@newsedge.com	
David Scott david.scott@newsedge.com	

Related Research

- *XML in the Content Lifecycle* Report (ZTR-CL100)
- *PRISM* ZapNote (ZTZN-0122)

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