

NEDINE: AN INTELLIGENT GLOBAL NEWS PUBLISHING AND DISTRIBUTION SERVICE BASED ON WEB SERVICES AND PEER-TO-PEER TECHNOLOGIES

Markus Schranz

Research & Development, presstext Nachrichtenagentur GmbH, Josefstaedter Str. 44, Vienna, Austria

Keywords: Web Services, global news distribution, Peer to peer architecture.

Abstract: Modern Internet technologies have been opening innovative aspects in research for more than two decades and entrepreneurs and pioneers have been transferring latest technical developments to successful businesses. Business oriented application domains like media get increasingly interested in applying such modern technologies. The specific application domain news publishing and news distribution is facing significant performance requirements and high loads of information handling while transferring hundreds of thousands of messages daily to the research and commercial auditorium. Although communication technologies are well employed by modern organizations to generate, publish and distribute their information, specific application domains like online news publishing and distribution involve multinational and multilingual requirements and enormous amounts of data to be transported. This paper discusses an approach that integrates news agency services from existing European organizations supported by University research in the area of global information services, distributed information management and AI in order to form an intelligent multinational business news publishing and distribution network, based on Web Services and a peer-to-peer inter-agency communication network.

1 INTRODUCTION

By using the Internet to reach hundreds of millions of people all over the world mass distribution of information has come to a new level. Electronic publishing has introduced a wide application range within the information management field, including the digitisation of ancient archives, the manageability and availability of large amounts of data or the visualization and provision of contents to the open public (Paepen, 2002; Schranz 2004). Modern intelligent approaches have enriched the field of electronic publishing with the research results of information retrieval methodologies, eLearning, theories of infocracy, security, privacy, semantic relations and metadata management towards the Semantic Web (Berners-Lee, 1998). Recent research work within the EU-funded project NEDINE (Nedine, 2004) focuses on the creation and the conception of an intelligent news publishing and distribution network consisting of existing local news agencies that use modern Internet and distributed computing technologies to build up a new kind of multinational and multilingual news

distribution service. Research fields such as network communication (Yu, 2004) have been involved to create a scaleable peer-to-peer architecture, artificial intelligence is utilized to identify the most relevant related news articles within the multinational Nedine service. The vector space model (Salton, 1993; Wong, 1985) has been utilized to provide easy access to related and most relevant business news.

This paper is focused on the integration of modern Internet technologies such as Web Services (Haas, 2002) and peer to peer architectures (Birman, 2003) to create a scalable and high traffic information exchange and distribution network. A brief introduction to the investigated application domain is presented in section 2. Requirements analysis and the systems architecture is described in section 3. As proof of concept the prototype implementation and innovative features are presented in section 4. A brief summary and an outlook to further development and practical use conclude the paper.

2 BUILDING A NETWORK OF NETWORKS

The application domain we are focusing on has built a business on top of the technological features and basic services available on the Internet: Several PR-companies and news agencies in Europe have utilized modern synchronous and asynchronous Internet technologies such as web-based information access frontends for business news archives or email-based mass news distribution or content provision for next generation user devices. A media industry has emerged that uses complex and technologically challenging Internet services to create, aggregate, exchange, publish and distribute current business news.

The discussed project has been initiated to build a network to integrate multiple European national information sources consisting of participating news agencies, PR agencies and independent journalists into an international information service for news professionals and decision makers. Existing services are based on modern Internet technologies like Web application servers holding the business logic in the services middleware (Cvitkovich 2005) and maintaining a local network of content providing editors and commercial customers as well as thousands of subscribers. Each business news service provides a set of features to these specific user groups that need to be normalized and shared within the integrated network of media partners.

Technologically, recent implemented services lack homogenous implementation models, data structures and communication protocols. With modern Internet technologies like Web Services for the information exchange and P2P architectures to manage a scalable integration of several local service providers a meta-network has been designed within the scope of the project NEDINE.

The implemented network discussed within this paper is targeting at the integration of such business cases, thus allowing news agencies of different European countries to share contents and exchange business news towards an integrated network for news aggregation, creation and dissemination. Aside from the obvious business benefits of such a service integration, there are necessary steps to be taken to technically and organizationally bring the services and the systems of the existing news agencies to an integrated network. Beyond this, further development shall attract additional agency partners throughout Europe to join the network in order to

- (1) have access to relevant business news at an international level
- (2) to offer a distribution and dissemination interface for their customers that provide news to the network

3 TECHNICAL ARCHITECTURE

Internet technologies in the area of business news distribution involves technical features to manage scalability and performance in mass information provision (millions of page impressions per month) and mass distribution (millions of electronic mails sent per day).

The goal is to set up a sophisticated news platform and a high performing distribution network based on convenient digital news exchange technologies. Experiences from local providers give figures of hundreds of thousands of electronic mails per day and about the same amount of visits to the online business news archives (Web presentations). NEDINE has also been preparing for different submission channels by using standardized data formats like XML news applications (NewsML(IPTC 2003)).

3.1 Requirements Analysis

In order to build a pan-European network of business news brokers, we have been concentrating on common features and identified technological concepts shared by the participating news agencies within the NEDINE project consortium. The news agencies manage their content platforms by providing news of different areas of interest via typical Internet protocols and interfaces. All partners utilize Web servers as part of their service. External resources like national and international content providers are used by the agencies to enrich the contents for the national platforms.

The integration work was started in the requirements analysis by identifying and describing the existing platforms and services in order to come up with a feasibly information exchange interface and an integrated network service. The integrated news distribution network includes several services that originates from different sources: some are provided by the existing local services and others are provided by the integrated network capabilities and features. In order to fulfil the multilingual and multinational requirements of the integrated news distribution network, existing solutions needed to be adapted and new software had to be designed and implemented.

Technologically, the legacy applications at the previously existing systems provided only a subset of the required functionality of the designed multinational and multilingual news distribution service and therefore needed to be extended with the some unified features: a unified access to local articles, the standardized distribution of foreign

articles and a unified presentation of network articles.

Table 2: Requirements analysis – Services available locally and in the integrated network

Local system

Services	Data Design and Processing
News creation and management	Handled in relational database schemata Manipulated by web-based content management engines
News aggregation	Aggregation of third party information (b2b news exchange); Upload of business news by customers; Handled by local protocols, mainly web-based applications
News distribution	Handled by email services to individually managed profiles of subscribers

Integrated network

News presentation	XML-based, mainly web-oriented interfaces to provide free access to business news via Web browsers
Multinational distribution	Aggregate business news from other partners and use local distribution
International enrichment	Provide most relevant semantically related news for each article; focus on local strength and partner enrichments
Multilingual news distribution	Provide business news for all subscribers and opinion leaders in targeted business languages
Multinational and -lingual news service	Offering multilingual and international business news tailored to the individual needs of the subscriber for media access

3.2 Architectural details

The requirements analysis has shown the need for a scaleable integration of exiting systems with little intervention and adaptation to the heterogeneous software components. The final decision taken in the project is based on a peer to peer network which is a unique way to create a network for news exchange between European news agencies. New technologies like NewsML, Web Services via SOAP or XML in general, have been used to create this decentralized system and connect it to all participants.

The architectural concepts handle data management, publishing interface (main focus on web technologies using web application servers) and the interface connections. The interfaces between the communicating peers and the interfaces between a peer and an existing local system are defined in detail by Web Service descriptions. The WSDL (W3C,2001) details are out of scope of this paper. Every news agency is only aware of the SOAP interface of the peer that is directly assigned to it.

Existing services use modern web services to retrieve remote business news information to be published by Internet technologies for the local service and/or distributed business news via local channels and the network for remote distribution.

The peer to peer concept was chosen since technically all current and future network partners can be equipped with the same piece of software. The network communication can be handled within the peer logic, only the interface between the peer and the existing service needs to be implemented by the joining partner. All activities within the multinational and multilingual network are triggered by the actions of one of the participating partners. The communication types within the system include Service registration, information upload, multinational news distribution requests and news enrichment requests.

3.3 Use Cases

Based on the requirements, the technical architecture and the communication types the created NEDINE network architecture can be explained in different use cases and possible scenarios during service utilization. To demonstrate Web services and peer to peer communication in action we chose a specific scenario on delivering a business news item multinationally in multiple languages.

The originating network partner (in Figure 1 named CIA) sends several messages to its peer, one per language (Czech and English, in this specific case), each with a request for enrichment.

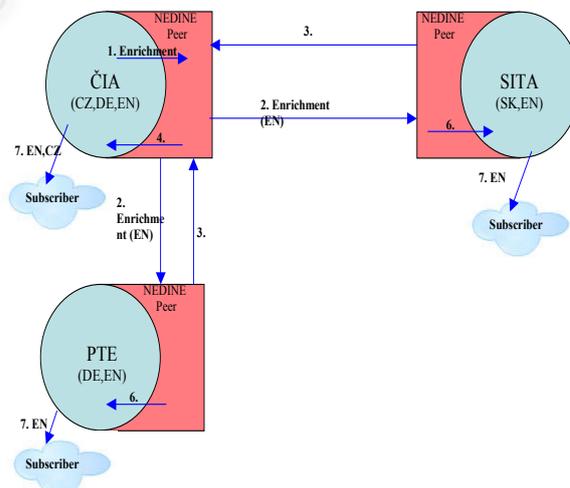


Figure 1: Use Case “article distribution”

The answer to this message will contain links to a set of articles, those related to the one in the message both locally and in the countries supporting the

language of the article, in the same language. After several enrichment processes (per language and country), the result set has to be merged to one final result set.

The originating network partner sends a message to its peer with a push request containing a push set with all the articles in the different languages. The answer to this message indicates whether the article has been effectively pushed into the countries supporting at least one language of the article (but in all the languages supported that match with the languages of the article).

The receiving network partners (in Figure 1: PTE and SITA) send a message to their peer with a polling request and receive pairs of an article (the article in the initial push request, in the use case) and a set of articles, those related to the other article in the pair either locally or in the originating country (in the use case, CZ), in one language.

4 CONCLUSIONS

The main goal in our research has been the creation of a business-oriented intelligent business news publishing and distribution network, engaging modern Internet technologies such as Web Services and P2P networking. Beside the technological challenges of using state-of-the-art Internet technologies the approach is highly driven by industrial and commercial ideas. The research experiences from project partners have supported the creation of a well-defined service architecture, a communication and interaction model and specific network features that lead to a service result that is more than just the connection of existing business news distribution services.

XML expertise and system architecture experts for high performance distributed systems cooperate in this project to provide a platform for multinational business information exchange.

Technically, the network integration has been defined and its functionality and availability is proven in running prototypes, demonstrating the business opportunities at several running sites yet. On the business side NEDINE produces a digital news distribution service, connecting existing news providers with new markets. The focus is on the efficiency of content aggregation and distribution, in order to provide an affordable solution for interested partners in participating countries.

ACKNOWLEDGEMENTS

The Nedine research documentation has been evaluated by the Technical University of Vienna, lead by Prof. Schahram Dustdar and Christian Platzer. This work was partially funded by the EU eContent project NEDINE (News Distribution Network, EDC-22225).

REFERENCES

- Paepen, B. et.al. 2002. OmniPaper: Modern Approaches for an Intelligent European News Archive. In *Proceedings of the IADIS WWW/Internet2002 Conference*, Lisbon, Portugal, November 13-15, 2002
- Schranz, M. & Paepen, B., 2004. Architecture design and application for an Intelligent Distributed News Archive. In *Proceedings of the ICC3 8th International Conference on Electronic Publishing – ELPUB2004*, Brasilia, Brazil, June 23-26, 2004.
- Berners-Lee, T. , 1998. Semantic Web Road map, <http://www.w3.org/DesignIssues/Semantic.html>
- NEDINE, 2004. – Intelligent News Distribution Network, EC eContent EDC 22225, <http://www.nedine.org/>
- Yu, S., Liu, J. & Le, J. 2004. Decentralized Web Service Organization Combining Semantic Web and Peer to Peer Computing. *ECOWS 2004*: 116-127
- P. Birman, K., 2003 *IEEE Internet Computing: Peer To Peer - The League of SuperNets*. In *IEEE Distributed Systems Online 4(10)*, 2003
- Salton, G., 1993. *Introduction to Modern Information Retrieval, volume 1*. McGraw-Hill, Inc.
- Wong, W.Z. & Wong, P. and Patrick Wong., 1985. *Generalized vector space model in information retrieval*. ACM
- Haas H., 2002. World Wide Web Consortium - Web Services, <http://www.w3.org/2002/ws/>
- Cvitkovich, A. & Schranz, M. (2005) Applied Middleware: Object-Oriented Content Management Components with Mason. In *Proceedings of the IADIS Applied Computing Conference AC2005*, Algarve, Portugal, February 21-24 2005
- NewsML,2003. IPTC-NewsML Standard, <http://www.newsml.org>
- World Wide Web Consortium (W3C), 2001. Web Services Description Language (WSDL) 1.1, WSDL<http://www.w3.org/TR/wsdl/>.