

2004

Scalability, Interoperability and Movement of Data on the Web Using XML Technology

Emmanuel Opara

Prairie View A & M University

Omprakash Gupta

Prairie View A&M University

Follow this and additional works at: <http://scholarworks.lib.csusb.edu/ciima>

 Part of the [Management Information Systems Commons](#)

Recommended Citation

Opara, Emmanuel and Gupta, Omprakash (2004) "Scalability, Interoperability and Movement of Data on the Web Using XML Technology," *Communications of the IIMA*: Vol. 4: Iss. 4, Article 7.

Available at: <http://scholarworks.lib.csusb.edu/ciima/vol4/iss4/7>

This Article is brought to you for free and open access by CSUSB ScholarWorks. It has been accepted for inclusion in Communications of the IIMA by an authorized administrator of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.

Scalability, Interoperability and Movement of Data on the Web Using XML Technology

Emmanuel Opara

Prairie View A&M University, Prairie View, TX 77446
(936) 857-4011, Fax (936) 857-2105; Emmanuel_Opara@pvamu.edu

Omprakash Gupta

Prairie View A&M University, Prairie View, TX 77446
(936) 857-4010, Fax (936) 857-2243; Omprakash_Gupta@pvamu.edu

ABSTRACT

Today's economy pulses with electronic vibrancy. eXtensible Markup Language (XML) has prompted individuals and organizations to discover new and creative ways of enhancing existing technologies and creating new ones. This study investigates emerging XML-related technologies that cover a broad range of industries and their organizational systems. XML provides access to a plethora of technologies for manipulating, structuring, transforming and querying data.

INTRODUCTION

With the explosive growth of e-commerce, more and more companies are using the Internet to exchange data with their business partners in what is called Business to Business (B2B) transactions. Enterprise systems have created an engine of nonstop wealth generation by gaining momentum from the flow of cheap, easy, instantaneous transactions on the World Wide Web. Electronic marketplaces are the backbone of our new economy. Since various organizations use different platforms, applications and data specifications, exchanging data become difficult. Organizations and various business partners are forced to rely on establishing protocols and data formats so as to be able to engage in electronic commerce. Since electronic commerce allows more efficient and effective data exchange among business entities, standards for data formats and protocols need to be implemented widely in the industry for an effective B2B, peer to peer, node to node, machine to machine communication.

Agital 2001, Boumphrey and Tittel 2000, Dodds 2000, Dick 2000, among others found evidence that standards for data formats and protocol in question is eXtensible Markup Language {XML}.

XML stands for eXtensible Markup Language. This means that it does not have any fixed set of tags. XML is not a programming language such as C++, Visual Basic etc, but is a Meta language. This further means that syntax or rules that are defines by XML can be used to create other markup languages (Dornan 2003). XML offers interoperability using a flexible, open standard giving B2B client's new ways to access and deliver data. It ensures that structured data will be uniform and independent of applications or vendors (Goldfarb 2000). In industrial automation, XML is the driving force in data exchange, manipulation, and transmission. XML, in collaboration with HTML and Simple Object Access Protocol {SOAP}, will provide the foundation for Web-Based Human Machine interface (Kobielu 2003).

XML uses markup (angled brackets, element and attributes) to add meaning to text document. It is a subset of the Standardize Generalized Markup Language (SGML). Arbortext 2001, Coyle 2002, Dornan 2003 among others noted that programmers, systems end-users and Web developers might identify some similarities between XML and HTML. This assumption could be based on the fact that both XML and HTML originated from the family of SGML. Dyen 2003, concludes that both XML and SGML are text-based formats that provide the methodology for describing document structures using markup tags such as words surrounded by '<' and '>'. XML is designed to describe data using a Document Type Definition (DTD) or Schema. XML has enormous potential for designing and enabling enterprise applications. As noted by the W3C-endorsed standard for document markups, XML defines a generic syntax to describe data with simple yet flexible tags included in the documents. Markup Languages (such as XML or SGML) are designed to add structure and convey information about document and data. In" markup" languages, the main mechanism for supplying structural and semantic information is by decorating the document with "elements" comprising of a "start tag", occasionally some content, and an "end tag". Example is <tag> data</tag>. XML has been standardized by the World Wide World Consortium {W3C} as the international markup standard.

Since the application of XML technology is becoming universal among businesses, it is generally accepted that this technology is not only useful for describing new document formats for the Web, but is also appropriate for describing structured data (Tibco 2002). An illustration of structured data include the information that is contained in network protocols, spreadsheet and program configuration files.

According to Dodds, 2000, XML is becoming the enterprise interoperable data formats because the technology can easily represent both tabular data and semi-structured data. Examples of tabular data format include relationship database or a spreadsheet while semi-structured data include a Web page or a business document. Dick 2000, Didie, 2000, Holland 2000, among others noted that XML is now classified as the lingua franca of information interchange.

In addition to being able to represent structures and semi-structured data, this technology has distinctiveness that has caused it to be extensively adopted as a data representation format {Travis, 2000}. The distinctiveness include XML platform-independence, support of internationalization by being fully Unicode compliant, extensible, etc.

Dornan 2003, and Hollander 2000, found evidence that although XML is in a text-based format, yet one can read and edit an XML document using standard text editing tools. XML is not handicapped to any programming language, operating system or vendor software, because of its platform independence, XML technology is useful as a means for achieving interoperability between different programming platforms and operating systems.

Dyers (2003) and Coyle (2002) noted that the exposure of data on the internet using XML technologies have stimulated the proliferation of XML data sources. Examples of information data sources include business documents, database and inter-business communication, Microsoft.NET Framework, Microsoft SQL ServerA.ATM and Microsoft Office®. These technologies have made it possible for developers and end users to create network messages, and develop XML documents (Kobielus 2003).

XML was developed because of the need to represent and manage data on the Web. HTML, GIF and JscriptTM were the standard for visual display and user interface on the Web. These standards allow a page to be created once and be displayed at different times by the receivers. Representation and management of data between different platforms was problematic. HTML, for example, does not support a common way of representing data so that software can search, move, display and manipulate data. Secondly, XML and XML extensions have now become the standard for Web-based data transactions such as managing e-commerce business processes that connect B2B businesses with its clients.

More recent studies have shown that B2B businesses have begun developing XML-based voice files to deliver information through wireless phones to its clients (Holland 2002, TIBCO 2002, Didie 2000, Arciniegas 2000). For example, Detroit-based General Motors Corporation, a B2B enterprise system, has claimed that they will be providing OnStar XML-based wireless voice files to deliver weather forecasts, news sports scores and stock updates to wireless phones built into 30 models of their premier 2002 cars (GMC Annual Report 2001).

According to Chen 2000, if a business strategy is to deliver Web content to devices, then they need to be using XML and XSL (Extensible Style sheet Language) because 86% to 94 % of firms are committed to using this strategy to attain competitive advantage.

PURPOSE OF THIS STUDY

This study will examine how XML facilitates universal data access and mobility. Also, to better understand the relationship between XML, HTML and other enabling technologies.

XML TECHNOLOGY

XML as a widespread service platform, provide a technical specification and a critical stepping stone in Web design and development. XML version 1.0 is a published, restricted, open standard that is accessible to developers to build. All XML documents have conforming attributes of SGML documents (Travis 2000). Similar to all conforming SGML profiles, XML supports flexible definition of new document structures and extension of existing structures. But, XML

defines the logical interrelationships among data fields within the document, while SGML's traditional orientation has been towards the definition of relationships among document elements such as titles, headers, paragraph and footnotes (Tibco 2002).

An XML document's application context resides within the document itself and in one or more external documents and data sources such as directories to which it points (Walsh 2002). Inside the document are markup constructs such as tags, declarations, an attributes, that frame the content elements and provide metadata that describes, in human readable plaintext, what that content means and how it hangs together logically.

Outside the content-bearing XML document are related entities. These are XML schemas, namespaces, style sheet, transformation maps, and work flow definitions that collectively instruct XML-enabled applications how to interpret, render, and process its content. The external entities are schema and XML 1.0 Document Type Definitions (DTDs). Schema is a logical template that defines the permissible vocabulary of elements and fields in a particular type of XML document (Sterling Commerce 2001). Figure 1 shows the typical sequence of events for receiving, parsing, and processing an XML document.

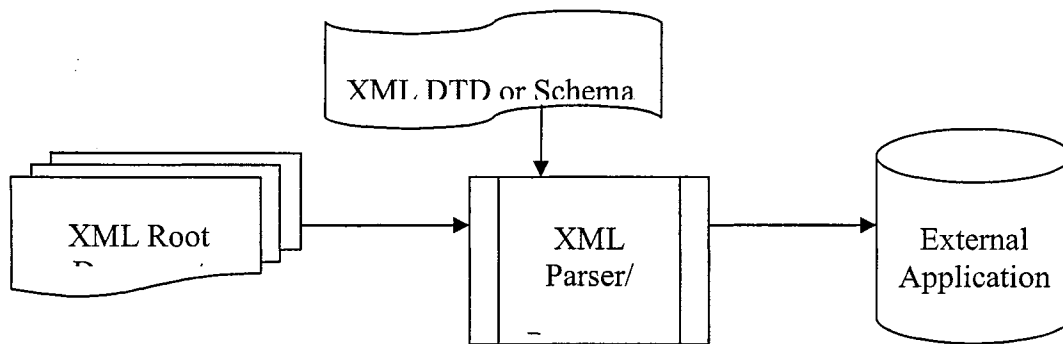


Figure 1: XML Processor/Parser in Action

In the above figure, an XML processor/parser receives an inbound document, parses it, and checks to see if it contains well-formed XML, validates it against a DTD or schema, and exposes its structure and content to external applications via Document Object Model (DOM) APLs.

XML and HTML: Is there a Relationship

Study has indicated that HTML and XML documents consists of elements, each of which have a “start tag” (such as <Opara>), an “end tags” (such as </Opara>), and the information between the tags (referred to as the contents of the element). Elements can be annotated with attributes that contain metadata about the element and its contents (Dick 2002).

HTML is about user interface while XML is about data. Dynamic HTML describes display and user interaction. Didie 2000, noted that XML describes information. This can be supported in the argument, that XML can add information to an HTML document and that HTML can display information expressed in XML format. See Table 1 for XML and HTML analysis.

XML	HTML
Excellent for data Structure	Used for display
Presentation independent	No knowledge of data
Open Language	Closed Language Standard
Case sensitive {because of Unicode}	Not Case sensitive
Tags without content and those which are empty elements must use “/” notation	Empty tag like requires nothing
White space, including line breaks, is significant within content	Special white space is ignored

Table 1: Relationship between XML and HTML

XML is case sensitive while HTML is not. This means that in XML, the start tags <Gupta> and <gupta> are different, while in HTML, they are the same. XML introduces the concept of well-formedness while HTML does not. The rule of well-formedness of XML removes some of ambiguity imbedded in processing markup languages. Other differences show that HTML has predefined elements and attributes whose behavior is well specified while XML does not. In XML, document authors can create individualized vocabularies that are specific to their application or organizational needs. Current research indicates that XML vocabularies exist for a large number of industries and applications. According to Dornan 2003, Web documents (XHTML), network protocols (SOAP), financial services (FPML) and financial filings (XBRL) are examples of XML vocabularies that exist in industries and system applications. The inability of laying emphasis on predefined elements and attributes that specify the methodology on how an XML document is delivered or displayed enables document programmers or developers to concentrate on generating documents that has only relevant semantic information for their exclusive problematic arena or domain.

Complaint about HTML is that is not extensible. The set of tags is fixed and fairly display-centered, which makes it difficult to add information such as revision histories or to mark-up displayed text. Dodds 2000 noted that it is not easy to add semantic information to HTML pages. HTML is not easily extended for data representation, due to its nature as a display language and partly because it was not for open extensibility. To resolve this problem, Microsoft is coordinating with the W3C to define a format for putting XML data inside HTML pages. By extending HTML to allow arbitrary XML data elements, a wide range of applications can use HTML as the primary document or display format, and also use XML embedded within these documents to hold application-specific data.

A Simple Anatomy of an XML Document

The following illustration is an XML document that shows a client’s order from Barnes and Noble book store. It is interesting to see how the document represents the rigidly structured data that describes information about books as well as the semi-structured data containing special instructions and comments about a special client.

```
<?xml version="1.0 encoding="isbn-768932-1"?>
<?xmlstylesheet href="orders.xml"?>

<order id="45698587">
  <client id="client0881">
    <first-name>Emmanuel</first-name>
    <last-name>Opara</last-name>
    <address>
      <street>College of Business</street>
      <city>Prairie View</city>
      <state>TX</state>
      <zip>77446</zip>
    </address>
  </client>
  <items>
    <book>
      <price>25.99</price>
      <author>Gupta</author>
      <title>Web Interface</title>
    </book>
    <book>
      <price>65.99</price>
      <author>Gupta</author>
      <title>Oracle SQL</title>
    </book>
  </items>

  <!-- Always go the extra mile for the client to make them happy -- >
  <special-instructions xmlns:html="http://www.w3.org/2004/xhtml">
    <html:p>if client is not at home address then attempt to leave package
      at one of the following available addresses listed in order of which
      should be attempted first
    <html:ol>
      <html:li>Secretary Office Room 2C229</html:li>
      <html:li>Deans Office Room 2D111</html:li>
      <html:li>On doorstep</html:li>
    </html:ol>
    <html:b>Note</html:b> Remember to leave a note for Dr Opara detailing where
      to pick up the package when he returns to his office.
    </html:p>
  </special-instructions>
</order>
```

It is interesting to note that this document begins with the optional XML declaration that specifies what version of XML is being used and character encoding used by the document. Next is the introduction of an XML style sheet processing instruction, which is used to bind a style sheet containing formatting instructions to the XML document for use in a more imaginative and descriptive manner in user applications such as Netscape and the Internet Explorer (IE) browsers. Further, it is interesting to note that processing instructions are used to embed application-specific information in an XML document. It is also important to note that majority of applications that process the contents of the XML document cited above would ignore the xml-style sheet processing instructions. However, applications used for displaying XML files such as Netscape or IE browsers would use the information in the processing instruction to identify the location of the style sheet that has the special instructions for displaying the document.

Valid and Well-Formed XML

Reiterating the findings above, XML has strict rules. These are illustrated in Table 2.

- | |
|--|
| <ol style="list-style-type: none">1. Every start tag should have an end tag [<code><name>Gupta</name></code> or comments <code></></code>]2. Attribute values must be in single or double quotes [<code><client id="1" /></code> or <code><Client id='1' /></code>]3. Tags cannot overlap [<code><c><d></c></d></code> is not allowed]. They should be properly nested accordingly [<code><c><d></d></c></code>].4. One top/root element is allowed at all times5. Strict rules for element names such as <code><234 /></code> is not allowed6. Elements may not have two attributes with the same name |
|--|

Table 2: XML Rules

Any document that conforms to XML 1.0 rules that stipulates that one root element, matching start and end tags, attributes in quotation marks, is known to be a Well-Formed XML document.

Further XML documents follow the rules that the programmer or developer defined. These formats are structural XML rules that include the following: the presence of hierarchy of elements, attributes, values data types, child-element occurrences etc. Any XML document that follows the above mentioned rules can be called a valid XML document. Programmers can write DTDs/XML schemas (XSD) to validate XML documents. The rationale is that DTDs/XML Schemas help in making sure that XML structures resemble the result the programmer expected.

XML supports special characters. In a well formed XML document, entity references are used for five special characters, while character references are used for other special characters. Table 3 illustrates these characteristics.

EXAMPLE	FORMAT	IS USED FOR
Entity References are used for five special characters	1. & 2. < 3. > 4. ' 5. "	1. & 2. < 3. > 4. ' 5. "
Character references that are for other special characters	1. #0174 or ® 2. Ω or Ω	1. ® 2. Ω
Characters between 0 to 31 (except CR, LF, and tab) are not allowed in XML documents.		

Table 3: XML Special Characters

XML AND THE WEB

This technology preserves, and describes information for easy processing and publishing to facilitate multiple intelligent uses. This process makes searching, indexing and locating client and system's information easier in a systematic manner. XML is used as a data definition and manipulation language. For systems integration of applications in enterprise intranet, XML technology will allow for more intelligent use of data and timely processing to facilitate critical decision-making within the enterprise. Since data is being exchanged in the XML format, it can be easily integrated from different platforms. An example (see Figure 2), using Microsoft's BizTalk specification that consist of a framework description, a repository for integrated-industry schemas, and a process for posting and validating schemas. The figure shows how XML communicate with other software packages.

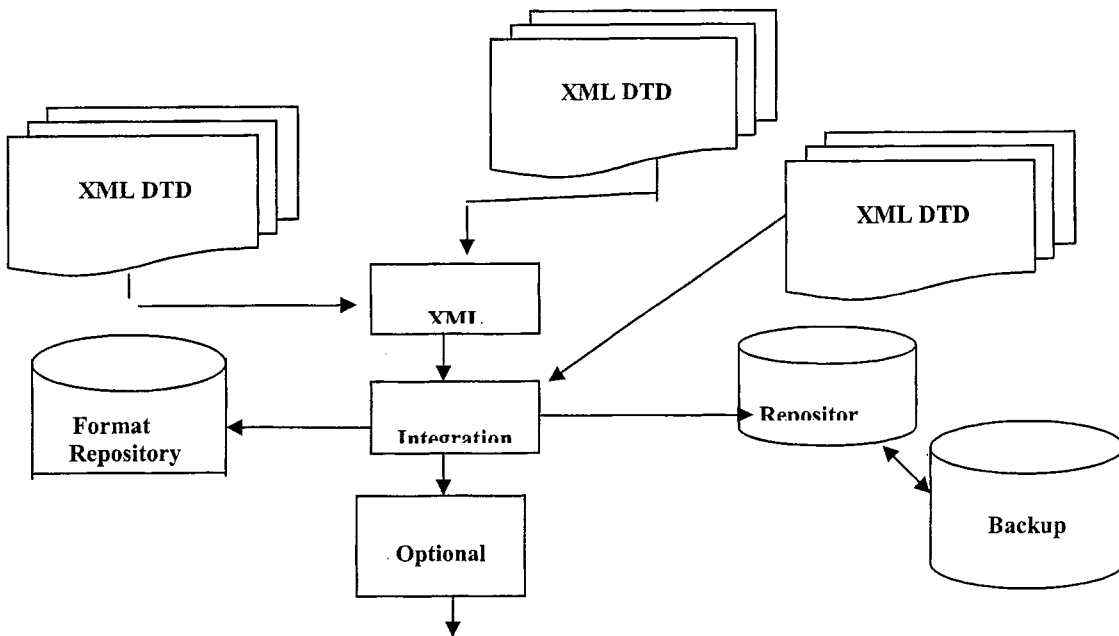


Figure 2: Microsoft's BizTalk Specification Model

The BizTalk server (see Figure 3) provides B2B document interchange and provides a server for routing information between clients, suppliers and customers based on the specific value. Enterprise Resource Planning (ERP) leading technology companies such as SAP, BAAN, Peoplesoft, JD Edward, are developing application integrated Component (AICs) as shown in Figure 3. These AICs provide the capability for software to integrate with Microsoft's BizTalk. For example, an alarm occurs that generate an XML-based BizTalk's representation of a report. This is sent to a BizTalk server, and the BizTalk Server then shares the information with other application or the Web (see Figure 3). The BizTalk Server is a scalable product that will facilitate the interchange of BizTalk Framework-encoded information.

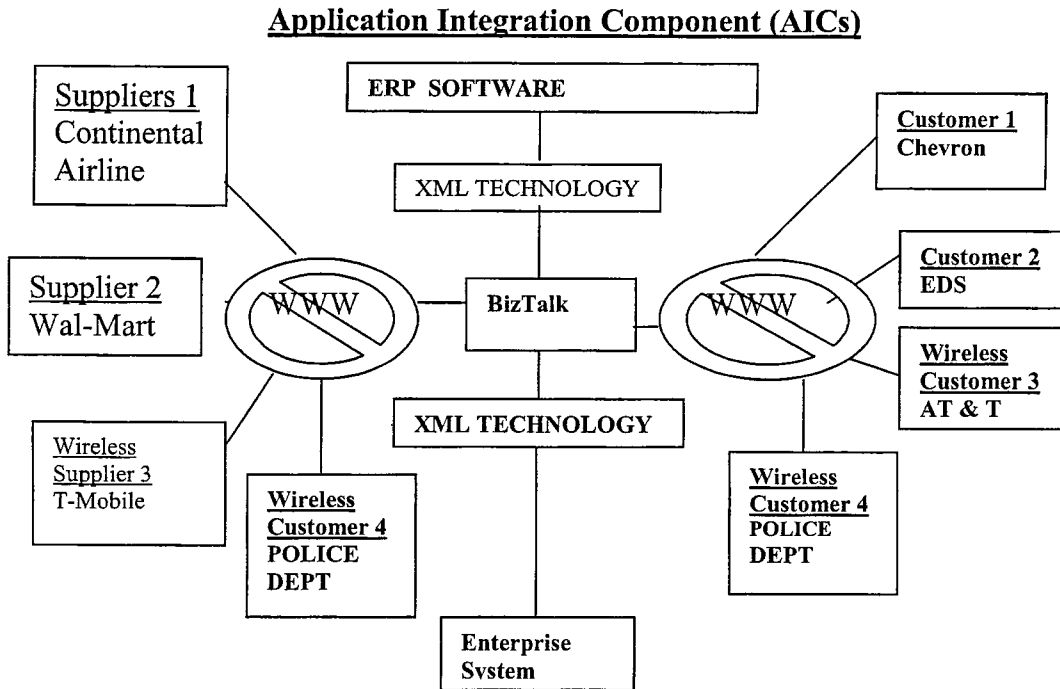


Figure 3: The BizTalk Server in Action

xml and the future

This study found that by using text-based XML syntax, platform interoperability and extensibility problems were resolved when programmers use XML as a data representation format. The notation that XML 1.0 syntax is text based and easy to parse has prompted the emergence of XML as the premiere data interchange format when cross-platform interoperability becomes problematic. The abundance of XML parser for various operating systems made it possible for disparate entities on different platforms to standardize XML as the interchange format when sharing of information is necessary.

Angle Brackets + Unicode = Interoperability

Table 4: Availability of Unicode

The equation on Table 4 indicate that the availability of Unicode make XML suitable for sharing information across global networks such as the World Wide Web. A Unicode is the international standard whose goal is to specify a code matching every character needed by every written human language to a single unique integer number called code point. This coding system

supports the worldwide interchange, processing, and display of the written texts of the diverse languages and technical disciplines of the modern world (Goldfarb 2000).

For software developers building Web applications and line-of-business Internet software, XML provide a powerful, flexible format for expressing data. This format could be a wire format for sending data between client and server, a transfer format for sharing data between applications, or a persistent storage format on disk.

XML offers a mechanism for adding meta-data or meta-content to HTML documents on the Web. For end-users, it promises to provide a much richer set of Web applications for browsing, communication, and collaboration (Boumphrey, 2000).

The technology development cycle for XML is now the “excitement phase”. What does the future hold? Additional research will be done to answer that question. However, the next few years will show the impact of XML on e-commerce as new information retrieval capabilities will be possible from an XML enabled Web base infrastructure. Also, with the increased capabilities of businesses to store and process data, the role of the browser will change from one presentation to that of being an application tool. This change will result from the combination of XML based data and HTML embedded scripting or Java to customize the presentation to the user’s needs. Further research will confirm all the mentioned possibilities.

Potential Limitations of XML Technologies

Although XML is associated with many advantages as high-lighted in this paper, yet high cost is involved in the application phase. Implementing XML as a data format for enterprise development promises extensibility and flexibility. Associated cost includes but not limited to degraded system performance and the lack of scalability of enterprise-class environment. Managing information as XML raises three issues related to enterprise scalability and robustness. The first is that XML is not designed for fast information retrieval. The second is that XML is a verbose method for exchanging data over a network. As a result, it becomes problematic. The third issue is that XML elements are not defined as native data types. Instead an enterprise application must parse the XML tags and interpret the values before beginning to process the application. The added steps required for converting XML-tagged data into predefined data types degrade the performance of an enterprise application system. These are a few of XML limitations.

CONCLUSION

We have provided an overview on how XML facilitates universal data access. It has demonstrated XML as a plaintext, Unicode-based meta-language that is designed with the mechanism for describing structured and semi-structured data, which provides access to a rich family of technologies for processing variety of data. XML is growing in importance in the industrial automation world. XML, in collaboration with HTML and SOAP, will provide the foundation for Web-Based Human Machine interface. The future of B2B as well as XML is

through the use of web services because XML Web Services exposes useful functionality to Web users through a standard Web protocol use in SOAP.

Extensible Markup Language (XML) offers interoperability using a flexible, open standard giving clients and business entities new ways to access and deliver data. The growing use of XML will improve Web-browsing applications for viewing, filtering, and manipulating information on the Internet. Web services which are built on the Simple Object Access Protocol (SOAP) will be discussed in the next paper. This paper will discuss issues such as the secured HTTP (HTTPS), Secured Socket Layer (SSL) 3.0, Transport Layer Security (TLS) 1.0 and X.509 digital certificates.

REFERENCES

- Agital. (2001, August). *XML e-business solution available for free 30-day test*. (retrieved November 25, 2001), <http://www.agital.net/company/press>.
- Arbortext. (2001). XML for managers. Arbor text Think Tank. (retrieved September 10, 2001), http://www.arbortext.com/Think_Tank/XML_for_Managers/xml_for_managers.html.
- Boumphrey, F., & Tittel, E. *XML for Dummies*. Foster City, CA: IDG Books Worldwide, 2000.
- Boumphrey, F.; Greer, C.; and et. Al., *Beginning XHTML*. USA: Wrox Press, 2000.
- Byous Jon, Web Services and J2EEvTechnology: One Platform," Java.sun.com, July 8, 2002.
- Chen, A., "Getting to XML: The new B2B lingua Franca will need to live with EDI. Here's how to mix the two." eWeek (retrieved 7 August 2000), <http://www.zdnet.com/eweek/stories/general/0,11011,2610043,00.html> 2000.
- Coyle, Frank, P. "XML, Web Services and the Data Revolution." Boston. Addison-Wesley Information Technology Series, 2002.
- ick, K., *XML A Manager's Guide*. Reading, MA: Addison-Wesley, 2000.
- Didie, M. "Getting into i-Mode." (retrieved 20 September 2000), <http://www.xml.com/pub/2000/09/20/wireless/imode.HTML> 2000.
- Dodds, L., "Gentrifying the Web: Taking a look at XHTML, the XML-Deviant finds that although the W3C HTML Activity is moving forward, the rest of the web is still lagging behind." (retrieved 13 September, 2000), <http://www.xml.com/pub/2000/09/13/xhtml/index.html> 2000.
- Dodds, L., "Schema Round-up." (retrieved 6 September, 2000), <http://www.xml.com/pub/200/09/06/deviant.html> .
- Dornan, Andy. "XML: The End of Security Through Obscurity?" Network Magazine, April 2003 Issue.
- Dyer, Russell. "The XML.com Interview: Liam Quin." O'Reilly XML.com, April 06, 2003 Goldfarb, C.F., & Prescod, P. *The XML Handbook*. Upper Saddle River, NJ: Prentice Hall, 2000.
- Holland, R., "More and more, XML diving into the 'stream.'" eWeek (27 July 2000) <http://www.zdnet.com/eweek/stories/general/0,11011,2608541,00.html>, 2000.
- Holland, R., "XML bridges e-apps." eWeek (7 August 2000), 2000. <http://www.zdnet.com/eweek/stories/general/0,11011,2585447,00.html>.
- Hollander, J., "Netscape Sees XML in Its E-Commerce Future." E-Commerce Times (retrieval September 2000) <http://www.ecommercetimes.com/news/articles/990309-2.shtml>
- <http://www.xml.com/pub/2000/09/xlink/index.html> 18, 1999.
- Kobielu, James. "Web Services Work, But Will They Scale?" Business Communications Review, April 2003.

Morgenthal, J. P. and La Forge, B. *Enterprise Application Integration with XML and Java*. Upper Saddle River, NJ: Prentice Hall, 2001.

Oracle. *Using XML in Oracle database applications*. (retrieved November 25, 2001)
<http://technet.oracle.com/tech/xml/info/htdocs/otnwp>, 2001.

Sterling Commerce. *XML and e-Business*. (retrieved September 25, 2001)
<http://www.sterlingcommerce.com/solutions/products/ebi/wp/xmlbiz%5Fwp.html>, 2001.

Sum Microsystems. *Sun Microsystems leads industry effort to simplify XML development*. (retrieved November 25, 2001), <http://www.sun.com/smi/Press/sunflash/1999-12/sunflash.991207.2.html>. 1998.

TIBCO. *TIBCO Extensibility*. TIBCO Software, Inc. (retrieved September 2, 2002),
2002. <http://www.tibco.com/solutions/products/extensibility>.

Travis, B., *XML and SOAP Programming for BizTalk Servers*, Redmond, Washington, Microsoft Press, 2000.

Walsh Kathy and Shin Sang, "Discover and Publish Web Services with JAXR." *Java World*, June 14, 2002.

