

Enhancing Knowledge Management With XML

Mirza B. Murtaza, (E-mail: mmurtaza@mtsu.edu), Middle Tennessee State University

Timothy H. Greer, (E-mail: tgreer@mtsu.edu), Middle Tennessee State University

Abstract

Knowledge Management involves gathering, categorizing, storing and sharing of knowledge. There are several available tools that can be used to build a knowledge management infrastructure to achieve its goals. This paper discusses issues involved in designing an information Portal using XML-based tools. As compared to traditional HTML-based portals, use of XML offers several benefits - it provides a great way of efficiently aggregating, classifying, and presenting both structured and unstructured content over the Internet or similar networks.

Introduction

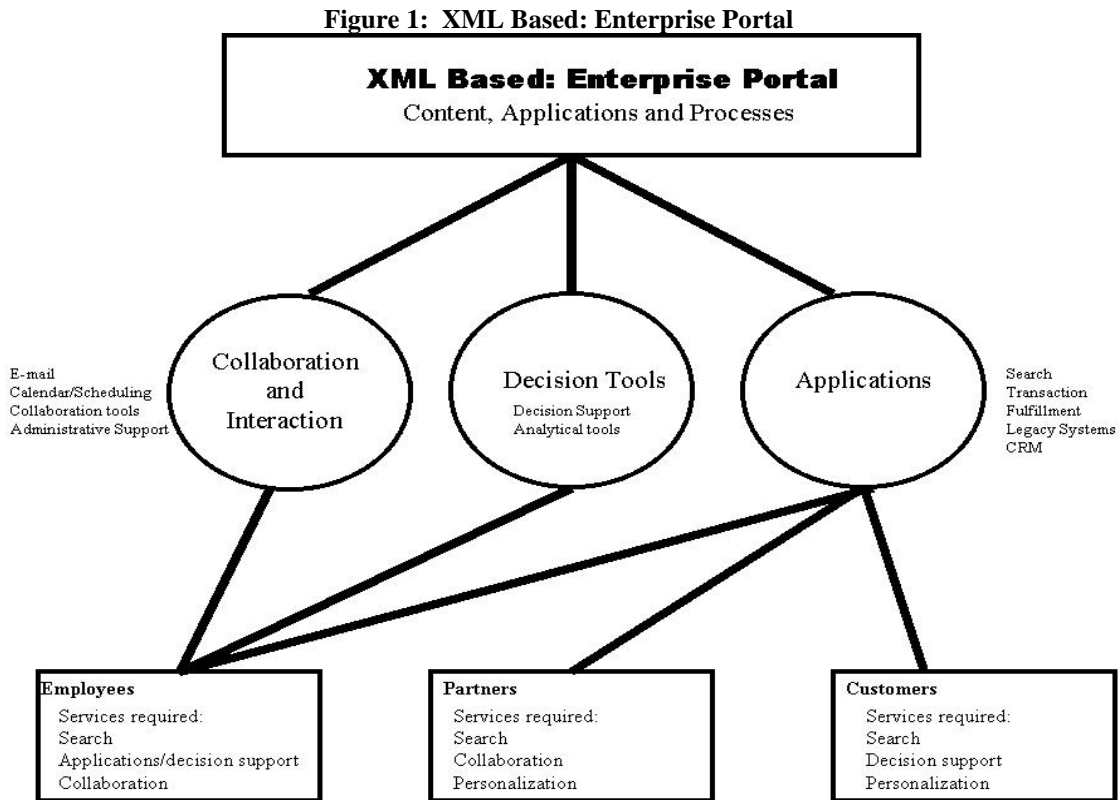
*K*nowledge is one of the most valued resources within an organization. Knowledge may provide the ultimate competitive advantage. Knowledge is a resource that is difficult to imitate and has to be accrued over a period of time. According to Hijri (2001) "Knowledge is the only factor of production that is not subject to diminishing returns". Knowledge Management focuses on ways of gathering, categorizing, storing, and sharing of knowledge. It strives to improve efficiency, effectiveness, and competency of individuals within an organization. Consequently, the field of knowledge management is highly inter-disciplinary and encompasses both technological tools and organizational processes in order to achieve its goals. Knowledge is different from information – what are the differences is a hotly debated issue. In general, information is defined as 'useful data' or 'value-added data' while knowledge goes beyond that. According to Duffy (2000), knowledge is information that has been enriched through interpretation, analysis, and context. The increased pace of change in the business environment along with the information technology revolution has led to the recent interest in knowledge management. Cole (1997) has discovered that employee attrition, decentralization, globalization, and an increase in networked organizations are also among the contributing factors for this increased interest in knowledge management. All these reasons are in some way due to recent technological advances. Although it is true, as suggested by Malhotra (2000), that too much attention to technology for knowledge management may be risky, any knowledge management initiative should at least consider the latest advances in technology. Any technology or process that would help sustain growth and distribution of knowledge within an organization would be a tremendous asset. This paper is concerned with information technologies that can be used to build a knowledge management infrastructure. More specifically, this paper discusses issues related to developing an XML-based Intranet Portal for knowledge management. Intranet portals allow users to have an aggregated transparent view of the organization's information systems. These information systems comprise the main components of a knowledge management system. The Intranet portal will use XML-based tools to present, store, and export documents. XML was chosen over HTML because of XML along with its related tools possess the ability to structure as well as present documents.

A knowledge management initiative needs to document all organizational activities that take place within an organization. This includes all activities, transactions, and correspondence between employees, customers, and partners. These activities, transactions, and correspondence may include things such as email, scheduling, customer relationship management, or the use of a decision support tool. According to Davenport, Long and Beers (1998) the four main objectives of knowledge management include:

Readers with comments or questions are encouraged to contact the authors via email.

- Improved access to knowledge through easy knowledge transfer among individuals and among organizations.
- Enhancement of the knowledge environment by facilitating and rewarding knowledge creation, transfer and use by knowledge workers.
- Knowledge capture through the creation of knowledge repositories.
- Management of knowledge as an asset to generate new income and/or to reduce costs.

The framework presented below focuses on these objectives. An information portal would create an easy to use interface among individuals, customers, and partners. XML-based tools would facilitate knowledge representation and exchange, due to their data handling capabilities. An Intranet portal would also provide a way to reward and facilitate knowledge creation and management. An example would be posting solutions and advice on potential problems. Each solution would be credited and rewards would be possible based on the value of the posting. This framework would facilitate knowledge creation and use throughout the organization. Increase revenue should follow from an increase in the knowledge of each employee, customer, and partner. When productivity of its employees is higher, an organization will have more satisfied customers and a better relationship with existing partners. Figure 1 illustrates the basic elements of an Enterprise Portal.



Intranets and Portals

The idea of a networked corporation is not new. All through the later part of the last century, there has been incredible investment in computer networking infrastructure. However, those networks were built using proprietary software from various vendors. The development of these networks involved a large financial investment. In addition to high costs, another disadvantage of those networks was that they were not able to provide global accessibility nor could they interconnect all the computers and database resources within an enterprise. With the popularity of the

Internet in 90s, along with its open standards and protocols (Transmission Control Protocol/Internet Protocol - TCP/IP, Hypertext Transport Protocols – HTTP), it became apparent that there was a way to build cost-effective and efficient networks within a company called intranets. An intranet is essentially an intraorganizational network that uses Internet technology and it can be accessed by authorized users via a standard web-browser. All the standard Internet features like web pages, newsgroups, discussion groups, and electronic mail are available for use within the organization through the intranet.

Most intranets today use firewalls for protection. These firewalls are hardware/software units that check incoming and outgoing data traffic to ensure that an unauthorized person could not access or illegally manipulate corporate data. Additionally, it can restrict internal users from accessing some unwanted external websites. Nowadays, most intranets are used to support corporation-wide activities like publishing company information, E-mail, groupware-based work activities, back-end database access, online workflow management, employee training, etc.

Large, as well as smaller, organizations have adopted intranets, at least initially, for the sole purpose of supporting the intraorganizational communication activities. In a study performed by Lai (2001) the major reasons cited for intranet adoption by corporations include use of open standards and protocols, ease of use, multiplatform compatibility, cost effectiveness and the ability to bring data and documents together.

The purpose of a portal, whether on the Internet or on an intranet, is to aggregate information or knowledge from diverse sources and present it in a uniform interface. Most enterprise information portals can be categorized as either collaborative information portals or decision information portals (Techguide.com 2001). The purpose of collaborative information portals is to facilitate communication and interaction among employees, customers, partners, and others who need to share information about any projects or issues. As the available bandwidth increases, these portals could be extended to include audio and video conferencing as ways for organizations to reach a more diverse audience.

Decision information portals are a step-up from the collaborative portals. The purpose of these portals is to provide internal enterprise information along with selected external data to executives, knowledge workers and business partners around the world. In essence, a user of a Decision portal gets access to an enterprise's business intelligence and data warehouses at the same time. For instance, if a financial institution, as suggested by Saaticioglu, Stallaert, and Whinston (2001), would use market research to develop proprietary indices that focus on the needs of individuals or small businesses, it will be called a Decision portal. The customers would be able to access the proprietary financial instruments on these indices and can select customized optimal portfolios, and use trading mechanisms to help establish and rebalance portfolios.

The major goal of a portal is to provide each user (employees, partners and customers) a personalized and integrated view of corporate information, business intelligence, and applications as illustrated by Figure 1. Therefore, in addition to providing an easy to use web interface, a portal must also have a rich set of features and tools for collecting, categorizing, integrating, and personalizing the information content. Traditionally, HTML was used to deliver content to a user's computer screen. However, increasingly users will access information through Internet devices (hand-held computers, mobile devices) and therefore, XML-based tools as the delivery mode would be preferred. Table 1 summarizes the major functions of an enterprise portal.

Markup Languages, Metadata, and Taxonomy

Originally, portal sites were designed using Hypertext Markup Language (HTML), the language of the World Wide Web. However, it is neither efficient nor feasible to convert every data source to HTML. Additionally, as contended by Seligman and Rosenthal (2001), HTML has some major shortcomings that limit its use in representing information from different sources. HTML focuses on presentation only, it does not “understand” the content. Additionally, it does not offer any way of going beyond its fixed set of pre-defined tags and thus, it is not suitable for content that uses a special format like mathematical or chemical formulas. Therefore, to overcome the limitations of HTML, in 1998 the World Wide Web Consortium (W3C) proposed another standard – XML.

Table 1: Major Functions of an Enterprise Portal

Information presentation	XML is preferred due to its data handling capabilities. XML provides structure and meaning to a document.
Information access	A portal provides its user specific information related to the task.
Personalization	Allow users to customize the content that they want to access every time they log on.
Search	Help the user find and access business content, this search could be directory-based and/or keyword-based.
Collaboration	A portal would also include tools that allow its users to communicate and collaborate with each other. Examples include calendar and scheduling, chat, news and discussion boards, and real-time data/video/audio conferencing.
Workflow management	Allow management of a set of interrelated business tasks. Once a task in the workflow sequence is completed, the following task can start.
Information subscription	A portal user can request a specific type of information delivered on a regular basis. Alternatively a user can choose to be notified when new content is added or existing content is updated.
Information security	Enforce the security policies and standards of an organization so as to ensure that users can only access the content they are authorized to view.
Content management, content adapters and application integration	Every portal would have a way managing the content through portal directory that maintains the metadata. Content adapter and application integration terms refer to mechanisms for importing databases and other files, applications (e-business, back-end), decision support and other tools into the portal.
Caching the content	An additional feature that can improve efficiency is caching. It will improve portal performance if frequently used information is cached.

Extensible Markup Language (XML) is a standard devised to bring about the separation of content and presentation of a document. XML is concerned with content, whereas HTML is used to present the documents. XML and its associated tools provide structure as well meaning to the data. An additional problem associated with HTML that XML is able to solve is that of data validation. A Document Type Definition (DTD) can be associated with an XML document and this document must adhere to the data format described in the associated DTD. For handling complex documents, W3C proposed XML schema specifications that even include data relationships and constraints. The structure of a document must conform to the relationships and constraints defined in the schema.

XML is a meta-language that can be used to define other markup languages. By itself, it neither defines tags nor does it prescribe the syntax, therefore the users are able to define any tags they prefer. In order for XML to be useful, it requires additional tools and technologies that are used for presenting and communication of XML data. In most instances, the term XML includes not only the meta-language itself, but also all the additional tools used for manipulating and using XML data.

One of the major advantages of XML is that it allows a single data source with multiple views of that data. The view is dependent on the medium or channel of distribution and the audience. For example, a news story stored in an XML data file can be delivered either through a web browser on a computer (using HTML), a hand-held computer, or a cell phone (using WML- wireless markup language). An additional advantage of XML is that it offers an extensive infrastructure for exchange and validation of structured data.

The term Metadata refers to data about data, it is a way of documenting information about datasets or knowledge repositories. When exchanging information between two data sources, knowledge of the metadata is essential. Metadata is used to define the structure of an XML document or file, which is published in a separate Document Type Definition (DTD) file for reference by other systems. According to Finkelstein and Aiken (1999), a DTD file defines the structure of an XML file or document in a similar fashion as the Database Definition Language (DDL) file is used to define the structure of a database, although with a different syntax. XML can be used to document the metadata of one system so that it can be linked and integrated with the metadata used by other systems. Even legacy files and other databases can now be integrated relatively easily. Traditionally, this task was achievable only through Electronic Data Interchange (EDI) – a relatively complex and expensive way of achieving the desired result. In comparison to EDI, XML is simple to use and relatively inexpensive to implement for both small and large

sized organizations. XML is in an ideal position to become a major part of the application development mainstream. It provides a bridge between structured and unstructured data. Both types of data can be delivered via XML and it can be converted to HTML for display in web browser.

Together with metadata, XML is a key component in the design, development, and deployment of intranet Portals. In Internet or intranet portals, XML provides a great way of efficiently aggregating, classifying, and presenting content. Within the enterprise, XML can be used to export data and knowledge content from the portal. This is an important step towards Enterprise Application Integration of internal and external business systems.

Designing Portals for Knowledge Management

A knowledge portal is essentially an information portal with an additional feature that it focused on knowledge rather than information only. As suggested by Firestone (2000), it provides information along with meta-information about what information can be relied upon for decision-making purposes. According to Baker and Baker (1999), in most corporations, only a small fraction of data resides in a structured database, a larger portion of information is unstructured – letters, reports, etc. So far, there has been no easy way of managing and allowing access to this type of data. Among the existing tools, XML-based content management is the only one that can bring-in both structured and unstructured data.

A major initial task in developing a portal using XML is to define the metadata, this can sometimes be extremely difficult, and it essentially depends on the quality and quantity of existing documents. Gregory (2000) has suggested that the process of describing or categorizing content, i.e., taxonomy, takes a great amount of time in portal development. Every industry and every business has its own “language”; it is not easy to agree on the terms or even nail down the meaning of those terms. However, it is essential for proper categorization of knowledge to come up with a standard taxonomy. The content of data determines the hierarchies that are essential for an effective search and smooth navigation.

A portal is traversed with the use of directory entries that are typically organized by subject area, which, in turn, is related to a specific business topic or concept. The standard taxonomy that defines the subject area is an essential element for portal development and for its ongoing maintenance and update. Some portal servers use a categorization engine or manager that follows some business rules for the taxonomy in order to determine the appropriate subject area with which to associate metadata about business content. To apply the rules, either the information extracted from the business content or from the metadata about the business content is used.

Another important task in developing a portal is to determine what specific group or groups are the audience. Watson and Fenner (2000) recommend that the focus of the portal must match with established business practices and employee behavior in order for a portal development project to be successful. This would help determine what are the required features or service for the portal. It is essential that a portal developer match the requirements of its audience with the features it offers. Otherwise, either the audience would not find requisite information or they will be inundated with unnecessary information. Additionally, the content appropriate for each viewer group must also be ascertained. This content requirement could be used to determine what types of application integration would be required, in terms of database access, legacy application data, decision support and information tools.

Most of the existing application programs do not save data in the XML format directly. Therefore, some indirect means of achieving this are required. There are some software solutions that can help convert legacy application data to XML format. Another alternative is to utilize XML messaging for content transportation from non-XML application to XML-based portal server. In the latter case, the XML message can extract metadata along with the content. From the metadata, a portal server can index the content, and give users the ability to retrieve and view the original application data through the portal.

The Simple Object Access Protocol (SOAP) is an evolving W3C standard for the exchange of information over a network. It provides a set of tags to indicate the roles of different pieces of information being sent over the web using the HTTP (web) and SMTP (mail) transport protocol. With SOAP, a program sends a "request", which is

a short XML document that describes a method to invoke on a remote machine and any parameters to pass on to it, to a SOAP server. The server tries to execute that method with the given parameters and sends a response back to the requesting program. This response is either the result of the execution of the request or the appropriate error message. Thus, SOAP allows sending a request to execute a task from an application on one computer to another computer that may have a completely different platform. Although, there are some other means of implementing this kind of information exchange, XML along with HTTP is better suited for the Internet/intranet. The SOAP based tools are expected to facilitate business-to-business integration, according to Seligman and Rosenthal (2001) among all business partners and not just among the big players as the case was with traditional EDI.


Development and implementation of the portal is also a complex and time-consuming task. There are a large number of portal server solutions available in the market, but unfortunately, there is no single best solution. None of the solutions is specifically focused on the task of knowledge management. Of course, another alternative is to build a portal from the ground up, which will extend the design and development phase further and it will require a tremendous amount of financial, time and human resources.

A completed initial portal is not the end in itself but just the beginning. As the computing professionals learned the hard way, systems sometimes outlast all expectations. Design of a portal should include 'learning' features. The portal should perform traffic analysis, learn from viewer's usage of services and modify its presentation to match viewer needs. As mentioned earlier, technology by itself may not be sufficient for knowledge management; it requires support from an organization's culture. According to Duffy (2000), there is evidence indicating that people with shared objectives and goals recognize the benefits of knowledge sharing. However, a properly designed and well-thought out knowledge portal will go a long way in supporting and building a community of shared interests, ideas and pursuits.

Conclusions

Knowledge management is an area where an organization can achieve a competitive advantage. The impact of knowledge management on the organization is through an increase in efficiency and effectiveness of the individuals within the organization. There are new technologies that would enable the development, implementation and utilization of knowledge management within an organization. Enterprise portals facilitate, create, and allow an easier flow of knowledge within the organization. The enterprise portal utilizes XML-based tools to present, disseminate, and store documents. XML is preferred over HTML due to its data handling capabilities. XML structures the data as well as conveys meaning about the data, whereas HTML is only concerned with the presentation of the data.

Suggestions For Future Research

There are various areas of research that can be pursued in order to better understand the use of portals for knowledge management in organizations. For instance, a survey of those organizations that have developed these portals would assist in discovering what the critical success factors are for such a project. Additionally, an analysis of successful projects will help in determining how structured and unstructured data has been integrated in these portals and what role markup languages like XML have played in it. 

References

1. Baker, S., and Baker, K. (1999) "Portals to Paperless Prosperity," *The Journal of Business Strategy*, 20(6), November/December 1999.
2. Cole, R. E., Ed., (1997) "Knowledge and the Firm," Special Issue of *California Management Review*, spring 1997, 40(3).
3. Davenport, T. H., De Long, D.W., and Beers, M.C. (1998) "Successful KM Projects," *Sloan Management Review*, winter 1998, 39(2).
4. Duffy, J. (2000) "Knowledge Management: What Every Information Professional Should Know," *The Information Management Journal*, 34(3), July 2000.

5. Finkelstein, C., and Aiken, P. (1999) *Building Corporate Portals with XML*, McGraw-Hill 1999.
6. Firestone, J. D. (2000) *The Metaprise, the AKMS and the Enterprise Knowledge Portal*, Working paper No. 3, Executive Information Systems, Inc., <http://www.dkms.com/MetapriseEKPrev32000.PDF>.
7. Gregory, J. (2000) "Portal Development: Roadblocks on the Information Superhighway," *Intranet Corner*, http://www.intranetjournal.com/articles/200010/pic_10_12_00a.html.
8. Hijri, K.K. (2001) "Exploring Data Mining Implementation," *Communications of the ACM*, 44(7), July 2001.
9. Lai, V. S. (2001) "Intraorganizational Communication with Intranets," *Communications of the ACM*, 44(7), July 2001.
10. Malhotra, Y. (2000) "Knowledge Management for E-Business Performance: Advancing Information Strategy to Internet Time," *Information Strategy: Executive's Journal*, Summer 2000, 16(4).
11. Saatcioglu, K., Stallaert, J., and Whinston, A. B. (2001) "Design of a Financial Portal," *Communications of the ACM*, 44(7), July 2001.
12. Seligman, L., and Rosenthal, A. (2001) "XML's Impact on Databases and Data Sharing," *IEEE Computer*, 34(6) June 2001.
13. Techguide.com (2001) "E-Intelligence for Informational Portals." <http://www.techguide.com/titles/portals.shtml>.
14. Watson, J. and Fenner, J. (2000) "Understanding Portals," *The Information Management Journal*, 34(3), July 2000.

Notes