Untangle Your Information… Four Steps to Integrating Knowledge with Topic Maps

Executive Summary

For years, organizations have sought to improve the way they share information and knowledge with employees, partners and customers. But the platforms that exist to deliver such content – Web sites and corporate databases – are not designed to support or facilitate an integrated knowledge management strategy. As a result, this goal has remained somewhat elusive for many organizations.

Moreover, many companies have yet to fully resolve the challenge of integrating data and knowledge from disparate software systems, although the adoption of standards has made it easier to combine content from different platforms and software applications. For example:

- *Network standards* have solved the physical issues of connecting different computing systems.
- *Internet protocol standards* have made the integration of disparate platforms transparent and seamless.
- *Web standards* have allowed information to be shared among people using a wide variety of tools. XML, for example, has made large amounts of data accessible to a variety of systems.

But when large volumes of data become more accessible, the process invariably runs into the problem of how to convey knowledge about the content and the interrelationships between its individual elements.

New *knowledge-sharing standards*, such as topic maps, address this challenge and are now emerging as a viable approach for organizations seeking to find a better way to share, interpret and respond to the wealth of information now available. By allowing enterprises to represent the knowledge stored in a wide variety of applications, topic maps not only enable different applications to share information, but also the intrinsic metadata describing that information. In this paper, we will examine these challenges, detail four steps for implementing topic maps and examine several real-world applications.

While organizations have been aware of topic maps for several years, some are just beginning to consider them a key business tool. As a result, the momentum toward topic maps is gathering steam. *Business 2.0* reported that the software for a related knowledge model, concept mapping, is predicted to reach $2.4 billion by 2007.

Clearly, enterprise knowledge managers will sharpen their focus on topic maps and other related knowledge sharing technologies as they seek to make content assets easier to find, manage and retrieve. Organizations that learn how to do this better than anyone else will be poised to improve how they respond to new threats and emerging opportunities.

Background

Prior to the Web, information management was relatively simple. Customer records, buying patterns and competitive information were stored in databases and access was strictly controlled by database managers, who doled out information on an as-needed basis. The Internet and proliferation of information technology to all workers generated growing demands for access to real-time data and information. Across many organizations, increasing numbers of employees sought better access to data that would help them do their jobs.

Information Management (1994 – 2004)

To put more information closer to key knowledge workers, enterprises implemented new ways to improve the creation, management, and dissemination of data. Of the new products that emerged over the last decade, three major technologies gained increasing prominence:

- Content management systems
- Digital asset management systems
- Search engines

Content Management Systems

Many organizations turned to content management systems as a unified way to create and store content. While the definitions of content management systems vary depending upon tool vendors, most products share a core set of features: version control, access control, automated workflow, integration of authoring and publishing environments, and metadata support for content.

While these are ideal systems for creating content, they don't solve the problem of integrating *knowledge* because the systems are by nature centered on documents – not information. The authors, editors and certain key players know where to find information, but until it's made available to the public through a website or manual, it's not accessible to others. Even once it is made available, the users are limited to those systems or applications in which the information is stored.

Digital Asset Management Systems

Digital asset management systems create a centralized repository for digital files that allow the content to be archived, searched and retrieved. Administrators then grant other users permission to review and store data in these systems.

This makes sense for small teams or groups that understand the limits of the data sets and need to collaborate on projects. This becomes a problem, however, when other groups need to view or incorporate that data into broader systems. To make it available to broader audiences, someone needs to convert this data from a legacy format into XML and then tag and index the content with metadata, a time-consuming and potentially expensive process.

Search Engines

Many companies have created extensive corporate intranets and Web sites to deliver content to a wide range of audiences. These platforms offer excellent ways to deliver targeted information to specific groups. The problem arises when you need to search for information or data on other Web sites.

Almost all Web sites contain numerous resources – Web pages, images, sounds, and other data. The application that drives those resources, namely HTML, is format oriented mark-up, which means it's not structured to communicate anything about the information or knowledge contained within its pages. This

makes it difficult to tap that embedded knowledge without understanding the technology (i.e. HTML), as well as the designer's approach to building those pages.

To find information or knowledge in a Web page, complex algorithms must be developed to create integrated data models. Search engines such as Google and Yahoo! then crawl through the resources on a Web site to infer its intent and add it to their overall knowledge model. Users might be able to find what they're after, but only after sifting through dozens, if not hundreds, of hits for unrelated information.

While these technologies work well, they all suffer from a common weakness: they don't present information in the manner people want to retrieve information – by subject matter. These technologies are non-subject-centric, which means that they do not really deal with the knowledge embedded within clumps of content. Even metadata is typically defined by who wrote it, when it was written and when it was published. Metadata provides nothing more than a string of possible keywords, but there is no real understanding of context, relationships, etc. of those subjects.

Information Challenges: 2005 and Beyond

To build the next generation of knowledge sharing technologies, organizations need to find new ways to overcome the limitations imposed by current content repositories and delivery systems. In an era when companies track the daily and weekly results of product launches so that they can modify promotional and marketing campaigns on the fly, the ability to deliver information to the right people at the right time becomes an imperative for success. To achieve this goal, organizations need to address three key challenges:

Overcome information overload

- On a typical day, workers may sift through 80-100 e-mails – from SPAM to critical corporate directives – and visit dozens of Web sites in search of information about clients, competitors or new industry trends. The problem of information overload only gets worse inside an enterprise, where applications are not designed to facilitate access to knowledge. A recent article on ZDNet reported that many executives are growing increasingly frustrated that while their corporate IT systems deliver huge amounts of information, they are often unable to translate that data into actionable intelligence.

- In addition, enterprise applications are often driven by huge combinations of technologies: Web sites, content management systems, desktop applications, enterprise applications, and a variety of databases, some running on different platforms and servers. Providing a seamless view of information in this environment becomes extremely difficult, if not impossible, given the existing technical and informational challenges

Integrate disparate information sources and multiple data formats into one cohesive knowledge base

- In most corporate environments, knowledge exists in many places – from proprietary corporate databases to the computer files of key subject matter experts. The challenge facing the corporate world today is to collect or connect the data from those varied sources and make them available to the people who need that information to do their jobs.

- Until now, most organizations have had to replicate knowledge as new technologies are updated and old ones replaced. This is an inevitable result of making knowledge subordinate to technology platforms. Sharing knowledge across platforms will continue to be an unattainable goal until organizations resolve the wide disparities in internal structure according to technology, platform and developers. Some organizations try to solve this by using search engines, but as we noted before, this approach hardly delivers true knowledge integration.

Contextualize search results to suit the needs of particular audiences

- These two issues or challenges create another problem for organizations. Employees, particularly in areas like sales or consulting, are demanding that knowledge managers customize information – from search results to e-mail messages – to fit their needs. In theory, this makes perfect sense. Why should a sales rep get the same information about a product launch as a software engineer? Both need to understand the product, but in varying degrees of detail. The sales person needs more information about offerings and pricing, while the engineer needs more facts about the underlying technology platform.

- In response, some enterprises are beginning to deliver more targeted information through Web portals or targeted e-mail blasts. This approach is effective, but it still has several shortcomings. For one, the employees need to classify their content based on keywords or by filling out profiles. Not everyone takes the time to do so. Even more, they still cannot search through the corporate Intranet or a data warehouse with powerful search engines that enable them to find the precise nuggets of information they are seeking. They may find what they are seeking, but usually not without an exhaustive effort.

Topic maps represent a potential and compelling solution to help enterprises overcome these challenges and improve knowledge sharing throughout their organizations.

What Are Topic Maps?

First introduced in 1993, topic maps were formally adopted as an ISO Standard in 2000. Since then, various organizations have looked for ways to unleash their potential for improving the way individuals find and manage information. Topic maps serve as a *knowledge interface layer* between storage devices, such as Web sites or databases, and user interfaces like Web browsers. They enable guided navigation through enormous volumes of content, and just as importantly, they let users share knowledge independently of proprietary systems.

Moreover, topic maps are subject-centric, one of the key factors that supports navigation and contextual search results.

Topic maps also create *semantic networks,* which provide a meaningful description of the relationship between subjects. They enable information or business objects to be connected semantically via the following methods:

- Defining the interaction between a person and specific domain of knowledge
- Ensuring that relevant subjects populate the interface
- Specifying important relationships between subjects
- Indicating relevant qualifications of subjects and relationships
- Citing all relevant occurrences of subjects

Like a subject index in a book, topic maps list all the subjects of knowledge that are appropriate for a given domain. In that sense, they tell users how to:

- Find common occurrences of one subject
- Distinguish subjects of the same name
- Indicate alternative names for subjects

One major advantage of topic maps is that they are *based on open standards*. An organization that adopts topic maps will be able to share the data and knowledge within those maps with other groups, whether a business partner or a merging company. In fact, one topic map can be merged with one, or multiple maps of related information. The possibilities for expansion are quite significant.

This means that topic maps also support group-level projects that require extensive collaboration by multiple parties. These projects can also be aggregated across the enterprise when needed, which lessens the risk of an all-or-nothing system implementation.

Topic maps also exist independently of any underlying storage mechanism – the data can be stored in a database, XML, spreadsheets, word processing documents, Web pages, or anything else. Topic maps can also point to other types of electronically stored content via:

- Search engines
- Information navigation
- Intelligent software agents

Topic maps are thus a logical extension of standards-based platforms like XML (Extensible Markup Language). XML and many of its related standards represented the first wave at bringing standard information encoding to the masses of software developers. Sharing a common coding format allows data to be made accessible to a wide range of systems, regardless of the platform or application in which the information originated.

Business Issues Addressed by Topic Maps

Although they should not be viewed as a solution to all knowledge integration problems, topic maps can help organizations address the information challenges we described earlier.

Information overload

- Because topic maps create a *subject-centric overlay of information on the Internet*, topic maps enable knowledge seekers to cut through the clutter and find matches to their search queries without having to sift through dozens of unrelated topics. Topic maps and related taxonomies also enable organizations to deliver more targeted information to users, giving them an opportunity to populate the fields with the terms and phrases that they want people to read and understand – such as new strategic positioning or customer issues.

- One European tax agency, for example, maintained a Web site that served up 9 million pages to 700,000 unique visitors each month. To help citizens navigate more efficiently through this wealth of information, they built a topic map that enabled citizens to find the instructions they were seeking as well as download the applicable tax software.

Finding information in disparate data sources and multiple formats

- Because they are based on open standards, topic maps can be integrated with any storage system to present the information that resides within, even if those databases are running on different applications. Users with a basic understanding of the topic they are seeking are able to find the right information without having to log in to diverse systems or know the various identifiers other systems might be using.

- A leading European car manufacturer, for example, created a topic map that linked together knowledge from different content management systems (databases and Intranets) to help engineers searching for the right parts to use in designing a vehicle. Over the years, the various design teams had created dozens of knowledge bases filled with information about different standards and quality requirements for each vehicle model. But it would have taken months to create a unified database, so building a topic map overlay proved to be the most logical choice, one that ultimately resulted in a 40% improvement in the design, development and reusability of the auto manufacturer's knowledge bases.

Contextualizing search results to the needs of different audiences

- One major advantage of topic maps is that they allow end users to personalize their search and to extend it into different systems. While personalization may be of interest most of the time, there may be times when you want to search outside of your usual search parameters. When those parameters are turned, however, contextual search returns hits according to various contexts. Instead of 200,000 hits, your search queries are grouped into a number of categories (or contexts), that enables you to more easily

distinguish what you are seeking. The effort to track down information is no longer so exhaustive.

- Take the scenario of an organization that created a topic map overlay that merged data from an accounting database and customer call center records. As knowledge-driven applications, topic maps enable different sets of data to be combined. This means that users don't have to know the specific nuances of the accounting system to find out information about a customer. The invoice may have both a customer name and billing number, while the customer call center record would include recent contacts. By entering a customer name and account number, someone from accounting could pull the information they are seeking about billing information, while a customer service analyst could do the same, both from their own system. This would also give another employee who, for example, was researching patterns in queries from customers who'd just received billing statements, the ability to extract and combine data from both systems. Information entered in a knowledge-driven application is less dependent on which system the employees use, as long as it's available to the application.

- Contextualized search also holds significant promise for public sector Web sites. A European government agency created a national portal to bring together cultural information held by both the agency hosting the portal as well as external institutions. The agency had learned that they needed to organize information along multiple pathways, not related to the institution involved, but according to the cultural theme. Creating this type of themed organizational structure allows visitors to find the information they want, regardless of its physical organization or location. The use of contextual search results (Ibsen [museum] vs Ibsen [festival]) reduces visitor frustration following unwanted links. The themes are further used to support a faceted search interface in which each choice narrows the search results. The national portal also broke new ground by not only setting up collaborative relationships with the external sites to which they link, but by using a method of unique identification of subjects (Published Subject Indicators or PSIs). In this way, they are able to easily automate a large amount of the link maintenance.

How to Implement Topic Maps: 4 Steps to Integrating Knowledge

Organizations that plan to integrate knowledge via topic maps will achieve better success by following these four key steps:

- Locate the sources of knowledge within your organization
- Capture critical knowledge, then evolve and refine the model
- Integrate knowledge-driven applications with topic maps
- Migrate legacy data into the knowledge base

Step 1: Locate the sources of knowledge within your organization

- The first step is to identify the experts who own the knowledge. They can help create meaningful taxonomies, and specify associations between pieces of information. Finding out where knowledge is generated is an important starting point. In the process, you may even find ways to streamline and consolidate disparate knowledge sources. Other knowledge may be less tangible. For example, customers may want access to account information locked away in your call center's database that is not widely available to the general public.

- Once you've located your knowledge sources or experts, you also need to move to an independent knowledge model that relies upon their input. The ability to specify associations between pieces of information, as well as metadata about the associations, enables knowledge management professionals to control and shape the new information model by creating meaningful taxonomies that overlay the raw information.

- This is a dramatic departure from traditional application development, which views knowledge only within the confines of the application in which it is created or stored. Those applications won't disappear in a knowledge-centered computing environment. Once organizations start building knowledge-centered applications, they will stop treating information as an afterthought, but rather as the centerpiece of application development. In this environment, organizations and developers will begin to look at their content objectively and try to identify the following elements:
  - The core subject matter content at a component level (e.g., the "noun")
  - Metadata that describes the content (e.g., the "adjective")
  - The associations and relationships between different components of information (e.g., the "verbs")

- XML provides an ideal, standards-based framework for representing these elements. And many organizations have worked with content analysis and data modeling experts to facilitate this task.

Step 2: Capture critical knowledge, then evolve and refine your knowledge model

- As a first step, encode information and knowledge into standard formats, starting with the basic, most critical items and building from there. Reviewing search engine results, for example, can help you fill in the gaps in your knowledge model. Evaluate server logs to determine what users are searching for, where they are looking, and what they will ultimately use. Another key consideration is to remember that knowledge is rarely static and needs to adapt over time. That means organizations need to set up plans for continually enhancing, refining and improving their knowledge models.

- There are two ways in which you can refine applications to ensure that your organization continues to surface and exploit the knowledge within its ranks.

Provide doorways to other applications using basic links

- If an end user is reading about a subject on a Web site and finds that other applications or Web sites have related information, then *related to* links could be provided. While this opens a door to other applications, each one retains its own integrity. This is a useful model for basic knowledge integration, allowing each application to maintain control over its information. User experiences can be customized, enabling each site to retain its own branding.

Integrate information and knowledge models between multiple applications

- Actual taxonomies and information could be read directly out of other systems to provide a seamless user experience, which would provide a unified view of the information, no matter what the source.

- In either design model, information and knowledge models become the centerpiece of the information technology infrastructure. Applications become subservient to these repositories, enabling users to locate and use information flexibly, making static navigation bars obsolete.

- In the place of the knowledge model, system architects and information professionals create systems that directly respond to a user's needs by providing flexible presentations of knowledge and information.

Step 3: Integrate Knowledge-Driven Applications using Topic Maps

- Once an enterprise has made the commitment to build knowledge-centered applications, the next step is to build the topic maps that will overlay those applications. This starts by creating an interface – or authoring tool – that can be used to build the taxonomies and ontologies that support the topic map. Next, developers need to create a system for managing topic map content. Once that is complete, the company is ready to integrate the topic map with other knowledge tools.

- A critical part of this process is to reach out to subject matter experts and involve them at the ground level, encoding information into the models to create the taxonomies, define the information categories and provide an identity mechanism for those components.

- This would also be a good opportunity for the organization to bring in knowledge management professionals who have a formal understanding of categorization, taxonomies, and related subjects. Hiring a professional consulting firm is one option, as is finding people who have strong backgrounds in library sciences or related fields.

Step 4: Migrate Legacy Data into the Knowledge Base

- In our experience, we've found that starting with a phased implementation is the best way to launch any data migration project. This will enable your organization to determine how it should structure its taxonomies, which information models to use as content repositories – customer databases, product manuals, etc. – and to find the best ways to make information accessible to employees, as well as to customers.

- Once the topic map has been built, organizations can easily merge data from one topic map to another, setting the stage for dramatic benefits. This can be accomplished quite easily by writing an application that reads one database field and then compares the information to one or more additional databases. Organizations can make this process even more precise by enlisting experts to use authoring tools that allow them to create associations manually.

- In some respects, building topic maps will always be a labor-intensive process. There are only so many shortcuts to properly categorizing, managing and merging knowledge. In some situations, it may make more sense to outsource some or all of the data conversion to an XML services company.

Defense Intelligence Agency

Challenge… A government agency, a primary supplier of national level intelligence, needed to find a way to organize and classify mountains of disparate data from multiple information sources into actionable intelligence. The agency had recently expanded their reporting and analysis to cover non-traditional intelligence missions, providing information to other national agencies, planners, and decision makers that have intelligence requirements.

At the same time, the agency recognized that it needed to find a better way to provide other agencies and external customers with more relevant, timely, concise and complete intelligence answers; presented in a digital environment that was easily accessible, consistent, flexible and transparent.

Solution… The agency realized that one way to make sense of all the information was to create topic maps that would compile data in ways that eased the retrieval of more accurate, time-sensitive information. To that end, they built a topic map component that enables the creation, revision and management of topic maps. Due to the sensitive nature of the content, consultants built an authoring tool that enabled the defense agency analysts to create their own taxonomies, so that they could continue to populate the topic maps with the correct terms, phrases and associations.

Benefits… The topic map enables agency analysts to read the summary of a document, and then decide whether they want to read more or move on to another document, saving them the step of actually opening the file. This more precise search feature also allows them to exclude an entire range of subjects and focus on only on the exact associations that they are looking for in a particular area. This will help them analyze more information at a faster rate, deliver better intelligence and help other defense agencies respond to emerging threats.

Conclusion

Not every company, government agency, or non-profit organization will come to the conclusion that it makes sense for them to embrace topic maps. On the other hand, we do believe that it's time for organizations and IT professionals to begin considering the benefits of knowledge-centered application architectures as a cornerstone of enterprise and organizational knowledge management.

To that end, there are three steps any organization can take to prepare for the day when topic maps and other new knowledge sharing technologies become more widely accepted as knowledge-sharing standards:

1) *Familiarize your employees with the concept of knowledge-driven methodologies*. This will help them get ready for an environment where information can be pulled from multiple platforms. And it will also help them begin to comprehend how faster information retrieval may change the way they do their jobs.
2) *Consider how such a shift would impact your current resource structure*. Knowledge-driven applications may shift staffing requirements and skill sets needed for projects that require huge amounts of data to be processed and analyzed.
3) *When possible, encourage investment in knowledge-driven applications*. Any organization that does so will be in a much stronger position to provide the best information to employees and focus on making content assets easier to find, manage and retrieve.

Regardless of the next step you take, organizations that learn how to integrate knowledge will be positioning themselves to take advantage of new information-sharing technologies – and improve how they respond to new threats, emerging opportunities and other future developments.