

Classification, Taxonomies and You

Verity White Paper



Classification, Taxonomies and You

July 2004

Table of Contents:

Taxonomies: One-size-fits-all <i>and</i> Made-to-measure	3
Putting Classification & Taxonomies to Use	4
Basic Browsing	4
Multiple-taxonomy Browsing	4
Enhanced Search	4
Tagging Applications	4
Monitoring/Alerting Applications	5
Content Evaluation Applications	5
Different Ways to Build a Taxonomy and Classify Documents	5
Verity's Approach.....	5
Business Rules	5
Training	5
Diagram 1: Business Rules	6
Diagram 2: Training.....	6
Diagram 3: Thematic Mapping	6
Thematic Mapping.....	7
Combinations of Business Rules, Training and Thematic Mapping	7
Where to Get Started	7
1. Define the Taxonomy Team	7
2. Analyze Dependencies: Users/Content.....	7
3. Deploy the Taxonomy	7
a. Taxonomy Creation.....	7
b. Taxonomy QA	8
c. Taxonomy Publication	8
d. Taxonomy Maintenance	8
Verity and Intellectual Capital Management	9
Glossary	10

Publisher's Note: Information contained in this document is intended for guideline purposes only. Verity product documentation supersedes information contained in this document. The situations described in this document are offered as examples; actual configurations and results will vary from system to system.

Classification, Taxonomies and You

Classification. It's something so basic that we do it every day without realizing it. Classification is the act of grouping like objects together. It helps us to make sense of the world around us. Whenever there is too much information, the mind organizes the information to process it and make quick decisions. We're constantly organizing information in our day-to-day tasks. For example, think about how you organize information in Microsoft Outlook or in Windows Explorer. Do you create folders, or categories, to organize your information? Is there a hierarchy to these categories?

Why do you organize your email that way? The answer is obvious—so you can quickly and consistently get information you need. Imagine how long it would take to find information if you didn't have any categories and each time you needed an email you had to sort through hundreds or thousands of randomly organized files.

If you organize your own information to make it easier to find, why expect less from your corporate intranet or website where the amount of content is growing at exponential rates? The issue of too much information is a huge problem in today's "wired" enterprises. The growth of electronic content in the last few years has been staggering, and companies of all sizes are spending significant amounts of money to store it all. Unfortunately, many companies haven't made similar investments in the management and retrieval tools that are needed to make all this information the source of competitive advantage that it could and should be. How many times have you experienced a problem finding information? Waited for the results of a search only to be presented with a seemingly endless number of documents that you have to sift through for the one you need to complete the task at hand? Or even worse, faced a blank results list that said, "Zero Results Found"?

A waste of time, right? IDC "conservatively" estimates that the average knowledge worker spends 15% of his time, or six hours a week, searching for information, and that 50% of all searches fail to locate the desired information.¹ That means your knowledge workers are wasting 7.5% (half of the time they spend searching) of their workday because they can't find the information they need. Now take 7.5% of your company's payroll, and that's how much you're losing every year to ineffective search.

As high as that cost may seem, it's the least of your problems. The real cost of employees not being able to effectively view, locate and retrieve intellectual capital across your business comes from:

- Duplicating pre-existing work because somebody couldn't find it with a full-text search. IDC estimates that 15% of the average knowledge worker's time is spent recreating existing information.¹

- Re-purchasing information products and services that you've already bought and paid for, but can't find on your network.
- Inconsistent decisions being made across your company because people in different departments are basing them on different sets of documents.
- Opportunities that are missed because you didn't identify them in time, or because the inability to locate information dragged out R & D, manufacturing and/or sales cycles for too long.

All of these costs are avoidable. The underlying cause is that many companies have outgrown the ability of a search box on an intranet page to solve all of their users' information retrieval problems. The majority of users type ambiguous, single-word queries into search boxes. They haven't been trained how to use advanced search features and strategies, and they don't know how to construct complex Boolean queries to narrow down their search results. And they shouldn't have to. A good information retrieval solution should be flexible enough to work just as well for novice users as for experts.

Taxonomies: One-size-fits-all and Made-to-measure

How do you create a "one-size-fits-all" content retrieval system in an information environment as complex as yours? One of the easiest and most effective ways is to create a corporate taxonomy, or several "made-to-measure" taxonomies that organize the same content in different way to meet the unique needs of different groups of users or achieve specific business objectives. Researchers have found that deploying taxonomies can reduce the amount of time it takes to find information by 50%²—the same amount of time IDC believes is lost to ineffective search.

So, what exactly is a taxonomy? A taxonomy is a hierarchical structuring of categories from generic to specific that break down a given domain. They aren't new and cutting edge. In fact, they've been around for centuries. From ancient Greek philosophers to librarians of the present day, taxonomies have been used to organize all kinds of information. Why? Because they work. Think of your own user experience. Would you rather get 2,000 documents returned to a search query, or be presented with categories that help you narrow down your search and discover new and related information?

The same taxonomies you deploy to make information quicker and easier for your employees and customers to find can be used as the foundation for real-time business applications such as:

- HIPAA and other regulatory compliance initiatives
- Tax credit analysis
- Email mining
- Other high-value, content-centric business applications

Whether you want to find information, take action and produce results—instead of sorting through search results—or need to evaluate content in order to achieve specific business objectives, taxonomies and classification are the solution.

¹ *The High Cost of Not Finding Information*, IDC, April 2003.

² Chen, Hao and Dumais, Susan, *Bringing Order to the Web: Automatically Categorizing Search Results*, 2000, Proceedings of CHI'00, Human Factors in Computing Systems, pp. 145-152.

There aren't any alternatives, and the benefits are simple: taxonomies and classification provide the structure and organization that makes your enterprise information more manageable, usable and therefore more valuable.

Putting Classification & Taxonomies to Use

When most people think about classification and taxonomies, they think about browsing through categories and sub-categories to find the information they need. While this is probably the most common use, classification and taxonomies can deliver a lot more value to your company as the foundation of HIPAA compliance, tax credit analysis, email mining and other high-value content evaluation applications.

Using classification and taxonomy tools such as Verity K2 Enterprise and Verity Profiler you can deploy:

- Basic browsing
- Multiple-taxonomy browsing
- Enhanced search
- Tagging applications
- Monitoring/alerting applications
- Content evaluation applications

Basic Browsing

The most straightforward use of taxonomies is integrating browse navigation with search on an intranet or website. End-users can either enter a search query or browse the taxonomy to find what they need. Or they can perform a scoped search while navigating through the taxonomy: once they find a category of interest, they simply search within it to focus their results. Alternatively, searching the entire taxonomy can lead to related categories that the user didn't think of. The best practice for intranet and Internet sites is to tightly integrate search and taxonomy browse at all levels.

Take the example of an employee at a door and window manufacturing company who needs information on troubleshooting the Microsoft Windows 2000 operating system on his computer. If he searched the company's entire intranet for "Windows 2000 troubleshooting," he would be deluged with instruction manuals for the hundreds of types of glass windows the company manufactured in the year 2000. With a taxonomy, the employee could either browse directly through the IT category to the Microsoft Windows 2000 subcategory, or he could dramatically reduce the number of results returned by limiting his search for "Windows 2000 troubleshooting" to the entire IT category.

Multiple-taxonomy Browsing

In addition to browsing through linear taxonomies, Verity also gives users the ability to view two or more taxonomies relationally. Verity Relational Taxonomies, known in the industry as faceted navigation, lets users navigate through more than one taxonomy to discover relationships and locate specific information. In effect, users browse two or more different taxonomies that organize the same documents in dif-

ferent ways, and view the intersections where they meet. So when a user navigates through one taxonomy, the results in the other taxonomies dynamically change.

Let's take the simplified example of an engineer looking for left-handed pipeline widgets. In this case, there are two taxonomies: one for parts and another for industries. First, the engineer navigates through the "parts" taxonomy to the category labeled "widgets," then down to the subcategory "left-handed," which has datasheets describing hundreds of left-handed widgets. This dynamically changes the results in the "industry" taxonomy, so the user is presented with only the industries that have specialized left-handed widgets: agriculture, energy, manufacturing and so on. By clicking "energy," and then drilling down through the subcategory "oil & gas" to "pipelines," the user can narrow the results to view only information where the "left-handed" and "pipelines" subcategories intersect: datasheets about left-handed pipeline widgets.

Enhanced Search

Remember that taxonomy building involves two tasks: the creation of the hierarchy, and the definition of the business rules that route documents to the appropriate category. The rules for categories are typically synonyms, acronyms and quasi-synonyms. These rules can be used as a "synonym ring" to enhance the results of search queries. (In Verity terminology, synonym rings are a special case of the "thesaurus control file.")

For example, the query "cell phone" would return documents that contain the term "cell phone," as well as documents that contain the synonyms and quasi-synonyms listed for it in the synonym ring such as "cellular phone," "cell," "mobile phone" and "wireless phone."

Tagging Applications

In addition to using the taxonomy for information retrieval, it is also possible to leverage the taxonomy's business rules to tag documents. Traditionally, manual tagging has been expensive and time-consuming. Accuracy and consistency are also called into question as authors may spend hours writing a document, but will spend little or no time actually assigning subject metatags to the document. The two most common problems with manual tagging are that:

- Authors may not assign the correct tag
- If the document is about multiple subjects, authors may not tag the document with multiple subject tags.

This is where Verity Profiler comes in. Verity Profiler can be integrated with content management systems and actually tag documents in real time using the business rules that define the taxonomy's categories. These tags can be presented to authors when they're writing documents to ensure consistent tagging, they can be used to enable advanced search applications such as Verity's Parametric Selection and to classify new documents as soon as they enter the system.

Monitoring/Alerting Applications

Even better than giving people the right search tools and up-to-date taxonomies is sending them relevant information as it is created in real time. This is absolutely critical in environments in which the value of content depreciates with time. As new documents are created, Verity Profiler can evaluate them against the categories in your taxonomy. When a document is created that matches the rules for, say, the “Left-handed widget” category, Profiler triggers an alert to the people in your purchasing, engineering and manufacturing departments who want to be kept abreast of the fast-moving left-handed widget market.

Content Evaluation Applications

Regulatory compliance. R & D tax credit analysis. Email mining. This is intellectual capital management at its best, and the work you do creating and deploying taxonomies sets the groundwork for it. In the case of a HIPAA compliance application, Verity K2 Enterprise can be used to create a category or categories with business rules that define protected health information (PHI). Documents that contain PHI are then assigned to these categories so they can be reviewed and removed from publicly accessible repositories if necessary. Verity Profiler can then use the same PHI category(ies) to evaluate new documents as they are created, identify those with PHI, and alert the appropriate compliance officer.

For more information about content evaluation applications, download the Verity white paper, *Evaluating Enterprise Content to Increase Business Value: Proven Examples of How Verity Makes Information Actionable*.

Different Ways to Build a Taxonomy and Classify Documents

There are several different approaches and technologies that one can use to build a taxonomy and classify information. The two main approaches in the industry are the use of business rules and training. Training is also referred to in the industry as automatic classification, but it is important to remember that there is still work involved with training. You have to identify the documents that are representative of each category so the machine can learn from them. The learning is automatic on the machine’s part, but the training takes varying degrees of effort on your part.

The second thing to realize is that training is an essential tool in the taxonomist’s toolbox, but it shouldn’t be the only tool. Verity’s training technology produces easily viewed and understood business rules that can be readily modified if necessary. At the heart of Verity’s approach is the understanding that any classification solution must provide both the ability for domain experts to control the end result and automation to jump-start classification efforts.

Lower costs and increase value with distributed, collaborative taxonomy creation and management

Verity Collaborative Classifier lets you distribute the management functions of your taxonomy to subject matter experts across your enterprise. This lets the people who best know the subject matter of specific categories collaborate with the knowledge engineers who specialize in classifying information. Individuals can be assigned editor of specific categories or entire “branches” of a taxonomy, and their recommendations automatically routed to “publishers” who approve or reject them. Once approved, the changes are published in real time to the taxonomies that users see. Similarly, when documents enter the system, they can be automatically published to categories that they match with a specified level of confidence. Or, documents that fall below that level but above a “propose” level can be routed to publishers who decide whether or not to include them in the proposed categories.

The benefits to your business of distributing taxonomy management this way are simple:

- Taxonomies that capture the domain expertise of your greatest intellectual assets—your people—in collaboration with the best practices developed by your knowledge engineers.
- Taxonomies that organize information the exact way your employees use it.
- Lower costs because fewer dedicated resources are required to manage your taxonomies.

Verity’s Approach

Verity lets you combine a number of taxonomy and classification methods to produce specific results. This flexibility is essential to creating taxonomies that are relevant to your users. The three main methods Verity provides are:

- Business rules
- Training
- Thematic mapping

Business Rules

Domain experts (Diagram 1, #1) create the business rules (#2) for each category. These rules define which documents belong in the category based on a number of criteria, such as the occurrence of words or phrases, the weighting specified for those words, their location and so on (#3).

Training

Administrators select prototypical documents (also called exemplary documents) that are representative of a category (Diagram 2, #1). Once selected, training documents are passed into the software (#2), the technology “learns” what they’re about and creates a rule for the category (#3). This rule can be read and modified by your librarian or domain experts to ensure the highest accuracy. Once all the categories are trained, documents are automatically classified into the correct one(s) (#4).

Diagram 1: Business Rules

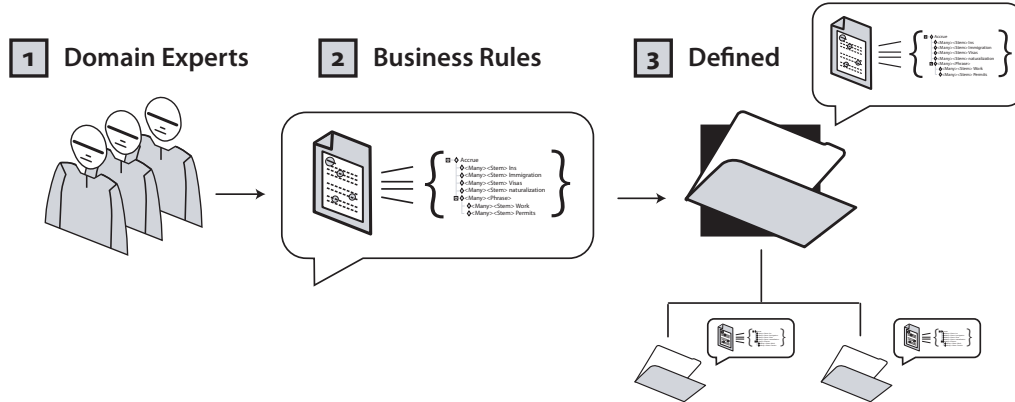


Diagram 2: Training

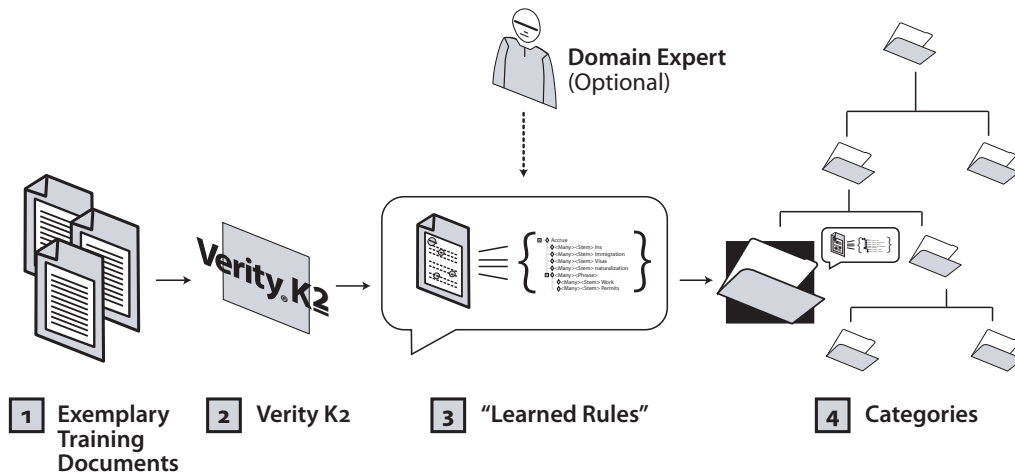
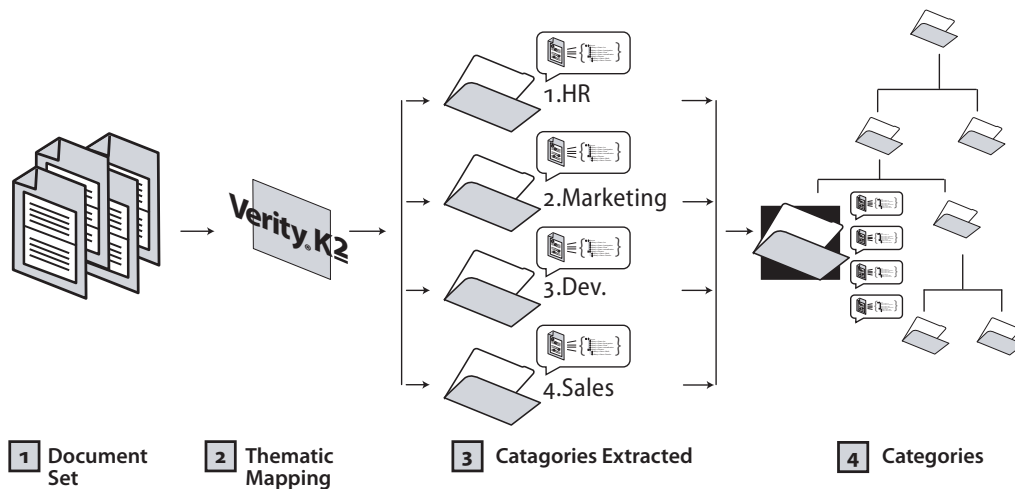


Diagram 3: Thematic Mapping



Thematic Mapping

Another technology that is unique to Verity is Thematic Mapping. Using Thematic Mapping, the software automatically extracts key concepts (Diagram 3, #2) or themes from a document set (#1), arranges the themes in a hierarchical fashion, defines each theme with business rules, and assigns a name to each category that makes sense to end-users (#3). Documents are then assigned to the automatically created categories (#4). This can be extremely useful when you have to build taxonomies from scratch.

Combinations of Business Rules, Training and Thematic Mapping

The ability to combine methods as necessary is critical to ensure the accuracy of documents assigned to categories, and to create categories and taxonomies that are relevant to end-users in the first place. For example, with Verity your taxonomist can manually re-rank the results in a taxonomy to designate “best bets” in one, a few or all categories. In the case of a human resources taxonomy with a category specifically for information on 401K benefits, the most used documents are probably the 401K enrollment and 401K rollover forms. Verity lets you designate these forms as best bets that will always be the first and second documents in results lists for the query “401K” and when the category is viewed. Verity’s approach also lets you add documents to, or delete documents from, a category if necessary.

Where to Get Started

Embarking upon a taxonomy project can be intimidating, whether you’re beginning from scratch or you have existing taxonomies to use as a starting point. Verity has found that although there are as many “right” taxonomies as there are companies, there are three basic steps common to all successful taxonomy projects.

1. Define the taxonomy team
2. Analyze dependencies
3. Deploy the taxonomy

1. Define the Taxonomy Team

One of the first questions that most companies ask is, “Who should build these taxonomies?” Corporate librarians are usually best suited for the task. Librarians are first and foremost trained to help people find information. This is their job day in and day out, and they understand the problems people encounter better than anyone else.

Don’t have a corporate librarian? Don’t worry. You can utilize anyone in your organization who has in-depth understanding of your enterprise content and who can act as subject matter experts. Taxonomy building involves technology and is an iterative process, so a good relationship with your IT staff is also helpful. The better they understand what you’re doing, the better they’ll be able to support you. Finally, involve key business stakeholders to ensure that the taxonomy meets the needs of the various organizations in your company.

2. Analyze Dependencies: Users/Content

Taxonomy projects have several components and dependencies. Before actually getting to the task of building the taxonomy, it is necessary to sort these out. You’ve already identified the members of the taxonomy team, but who will actually use the taxonomy? How will they use it? What high-value content needs to be classified? Is there content that shouldn’t be classified? Do your documents have metadata? What is the quality of that metadata?

The most important thing to realize is that taxonomy projects seldom start from scratch. There are usually a number of internal or external resources that can assist in taxonomy design or the creation of business rules, including existing thesauri (lists of terms that are related), product lists, site maps, search logs, and so on. But if you don’t take the time to analyze and plan your project, you’ll spend even more time recreating what you’ve already got—which is one of the things taxonomies are supposed to prevent.

3. Deploy the Taxonomy

You’ve brought together a team, analyzed your dependencies and put together a plan. Now, it’s time to build the actual taxonomy and the rules that define its categories. This can be broken down into four steps:

- a. Taxonomy creation
- b. Taxonomy QA
- c. Taxonomy publication
- d. Taxonomy maintenance

a. Taxonomy Creation

The two major tasks required to create taxonomies are:

- i. Creating the categories and hierarchy.
- ii. Creating the business rules that define the categories and the documents that will be assigned to them.

The categories and hierarchy captures the browse functionality that lets people drill down through subcategories or navigate across categories. It is what your end-users actually see. However, creating business rules at each category level is the more labor-intensive of the two tasks. These rules generally consist of synonymous terms and acronyms, with the preferred term becoming the category label. Take the example of a category labeled “Symmetric DSL.” The category name is “Symmetric DSL,” but the rules created for this category include “symmetric dsl,” “symmetric digital subscriber line,” “single-line dsl,” “single-line digital subscriber line,” “sdsl” and other terms commonly used to refer to “Symmetric DSL.” Carefully labeling each category allows you to enforce your corporate terminology. This ensures that everybody in an organization is using the same vocabulary, which minimizes miscommunication internally and with external stakeholders such as customers. Carefully choosing the business rules for each category ensures you can find all of the relevant documents in each category even if others don’t use your corporate terminology.

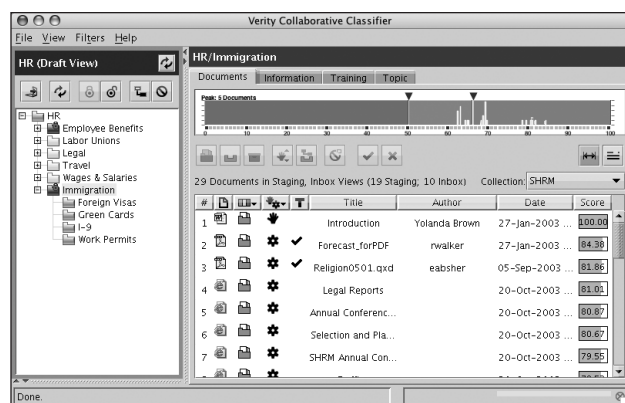
Depending on your starting point and resources, there are a number of ways to create both the categories and the business rules that define them. A good bottom-up approach is to run Verity Thematic Mapping on all of your content to automatically discover the key themes and concepts represented in it. These themes can be used to develop the categories and/or business rules. During this phase it is also important to leverage existing resources: search logs, existing thesauri, site maps, synonym lists, products lists, acronym lists.

Conversely, a good top-down approach is to jumpstart the taxonomy creation process with existing structures, such as Verity Taxonomies. Built by experienced knowledge engineers using best practices learned over hundreds of consulting engagements, Verity Taxonomies let you rapidly deploy industry-specific taxonomies that can be combined with existing corporate taxonomies or easily customized to meet company and industry requirements. These pre-built taxonomies are updated on a regular basis to ensure the categories and the business rules that define them are up-to-date. Vendors such as Factiva and LexisNexis also provide pre-built structures in a number of industries that can be used to help you create taxonomies.

Another approach to creating taxonomies is to take advantage of your company's subject matter experts. Verity's best practice is to interview subject matter experts—the people who know both what your company knows and how the taxonomy will be used—to help create categories and/or business rules. By using a combination of these approaches, your taxonomy team can create a draft taxonomy. Once consensus is reached on the draft, the final taxonomy and its business rules can be built.

b. Taxonomy QA

Between building the final taxonomy and publishing it for use is one of the most important steps: quality assurance (QA, a.k.a. testing). During this phase of taxonomy deployment, subject matter experts should review the populated taxonomy. This means clicking on each category, reviewing the docu-



Verity Collaborative Classifier lets you distribute taxonomy maintenance to your subject matter experts and knowledge engineers. For more information, see “Lower costs with distributed, collaborative taxonomy creation and management” on page 5.

ments in them and asking questions like: Are the right documents in the right categories? Are documents that should be in a category missing? Are there too many documents in a category? Too few?

If categories are overpopulated, either refine the rules that define them or create additional subcategories. You can do this automatically by applying Thematic Mapping to the documents in these categories to help break them down into sub-categories. If categories are underpopulated, consider removing them or collapsing them with sibling or parent categories. Are there any unclassified documents? Review them using Thematic Mapping to determine if categories need to be added or refined. Also, analyze the taxonomy hierarchy to make sure that users do not have to drill down too many levels before finding relevant information. And remember to document what you find; the changes you make and why you made them should be captured in a Taxonomy Change History.

c. Taxonomy Publication

Once the QA phase of the taxonomy is finished, it's time to publish the taxonomy to end-users. Here, it's a good idea to provide a basic tutorial on your intranet or Internet site that shows how to leverage the taxonomy for information retrieval. It should be obvious how to use the taxonomy, but taxonomies are still new on many sites so it never hurts to provide useful tips to the end-user. Most users are familiar with browsing taxonomies, but they don't understand how to leverage integrated search and browse to refine their result sets by limiting searches to specific categories or subcategories (i.e. scoped searching). If the taxonomy is for an intranet application, schedule education webinars for your various user communities. Finally, provide a feedback mechanism so users can suggest additional categories or other improvements.

d. Taxonomy Maintenance

The job is not over yet! Now that the taxonomy has been published, it's time to get into maintenance mode. Why? Because over time your employees will create documents about new technologies, products, customers, markets and so on. Similarly, the way your employees use your existing content will change on a continuous yet unpredictable basis.

Let's go back to our Microsoft Outlook example. Today you might have several folders that you use to capture sales and marketing information on different products, competitors and customers. What if you suddenly were not able to add any more folders and you had new email coming in about a new competitor, your latest product or from a prospective customer? Those emails would just sit unorganized in the Inbox. The same is true for your corporate intranet or Internet site. You need to invest in maintenance so the taxonomy continues to be relevant and usable.

How much time needs to be dedicated to maintenance? This depends on several factors, including:

- The size of the taxonomy
- Content volatility (how often concepts within an organization change)
- Dedication of resources
- How frequently content is added

Ensure that mechanisms are in place to update and revise the taxonomy through regular taxonomy audits. Plan for changes such as new product releases or mergers and acquisitions that would result in additional categories.

In addition, the taxonomy team should review search logs regularly. Cross-reference queries in the logs against the taxonomy to evaluate new categories and/or business rules that should be added. Remember the feedback mechanism we mentioned above? Review what your users are telling you and make changes as necessary. Also, schedule regular testing with the site's most prolific users to determine additional categories or refinements to business rules that will make the taxonomy more relevant to them.

All this may sound daunting, but Verity has found that once the benefits of regular taxonomy maintenance are understood, most taxonomy teams are more than willing to absorb maintenance tasks into their day-to-day routines. Distributing management of specific categories and subcategories to the subject matter experts who know the content best can greatly simplify the maintenance process and lower costs. Distribution also enables collaboration between your taxonomists and the people who actually use the taxonomies. (For more information on distributing taxonomy management, see “Lower costs with distributed, collaborative taxonomy creation and management” on page 5.)

Verity and Intellectual Capital Management

Headquartered in Sunnyvale, California, Verity provides software that enables organizations to maximize the return on their intellectual capital investment. The company's intellectual capital management (ICM) solutions provide integrated search, classification, recommendation, monitoring and analytics across the real-time flow of enterprise information, along with question and answer interfaces for effective online self-service. Other Verity ICM solutions capture content and drive automated business processes. Verity technology also serves as a core component of more than 260 applications from leading independent software vendors.

Verity software solutions are used by more than 11,500 customers in the private and public sectors. Customers include American Express, AT&T, Bristol-Myers Squibb, Cap Gemini Ernst & Young, Cardinal Health, Cisco Systems, Documentum, Dow Jones, Financial Times, Hewlett-Packard, Home Depot, Lotus, Kaiser Permanente, META Group, SAP, Siemens, the State of California, Stellent, Sybase, the U.S. Departments of Energy and Justice, and the U.S. Army.

Glossary

Acronym: A word formed from the initial letter of each word in a compound term. For example, ICM is the acronym for intellectual capital management.

Authority file: An authority file is similar to a synonym ring, but goes one step further. In addition to listing synonyms, it also specifies which terms are preferred and which terms are variants. Authority files are used to enforce corporate terminology while controlling vocabulary.

Automatic classification: The automatic creation of the rules that define which documents belong in a category, based on a set of documents that represent the category's subject. See also Training.

Boolean queries: Boolean logic, developed by a mathematician named George Boole in the mid-1800s, lets you specify relationships between concepts using the terms AND, OR and NOT. Boolean queries use these relationships to help you narrow down search results. For example, the query “dog AND cat” would return only documents that contain both of the words “dog” and “cat”. The query “dog OR cat” would return documents that contain either one of the words “dog” or “cat”, in addition to those that contain both words. The query “dog NOT cat” would return documents that contain the word “dog,” but none that contain the word “cat;” documents that contain both “dog” and “cat” would not be returned. one of the words “dog” and “cat”, in addition to those that contain both words.

Business rules: The rules that define which documents belong in a category. These can be automatically generated, created by domain experts, imported from existing taxonomies or a combination of the three.

Content volatility: How often concepts within an organization or taxonomy change. Not to be confused with how often the actual documents in a taxonomy change.

Controlled vocabulary: A system that controls or normalizes the different words used in an industry or company. Taxonomies, synonym rings, thesauri and authority files are examples of controlled vocabularies.

Domain expertise: A person's knowledge about a domain (subject).

Exemplary documents: Documents that are representative of the type of documents that should be assigned to a category. These are used to train categories.

Faceted navigation: See Relational Taxonomies.

Intellectual Capital Management: The process that combines human knowledge and experience (both implicit and explicit) with the information and data in an enterprise for the purpose of exploiting greater value.

Parent category: The category directly above a subcategory.

Relational taxonomies: Two or more taxonomies that organize the same set of documents in different ways, and that can be independently browsed to locate information at their intersections. For example, used cars could be organized by geographic location in one taxonomy, and by manufacturer and year in a second taxonomy. Someone in Miami could browse through the geographic taxonomy from “North America” to “United States” to “Florida” to “Miami,” and then through the manufacturer/year taxonomy from “European Manufacturers” to “BMW” to “2001.” What the person would see are the cars where the two taxonomies intersect: 2001 BMWs for sale in Miami, Florida.

Scoped search: Full-text queries in which the searches are limited to a specific category and its subcategories.

Sibling categories: Categories at the same level within a taxonomy. For example, in the geographic taxonomy described in relational taxonomies above, the subcategories “United States,” “Canada” and “Mexico” directly under the parent category “North America” would all be sibling categories.

Synonym: A word or term that has the same or similar meaning as another word or term.

Synonym ring: Synonym rings are controlled vocabularies that are lists of synonyms. They capture all of the synonymous ways to refer to a concept. For example, the word PTO is synonymous with the phrases “personal time off” and “paid time off.”

Taxonomy: A taxonomy is essentially a controlled vocabulary that provides a way to control or normalize the different words used in an industry or company for the purpose of information retrieval.

Thematic Mapping: Verity's Thematic Mapping technology evaluates documents across your enterprise in various applications and repositories, identifies themes or concepts, arranges the themes in a hierarchical fashion, defines each theme with business rules and assigns a name to each category that makes sense to end-users.

Thesaurus: Thesauri are similar to taxonomies, but also contain terms that are related to the search term. Thesauri were created not only for information retrieval, but information indexing, as well.

Training: When most vendors talk about “automatic classification,” they really mean training. This is where administrators select prototypical documents (also called exemplary documents) that are representative of a category. The classification software then “learns” what the documents are about and automatically creates a rule for the category. This rule should be readable and modifiable by your librarian or knowledge workers to ensure the highest accuracy. Once all the categories are trained, new documents are automatically assigned to the correct one as soon as they enter they system.

Verity, Inc.

894 Ross Drive
Sunnyvale, CA 94089
t. 408.541.1500
f. 408.541.1600

Verity San Diego

3220 Executive Ridge Drive
Vista, CA 92083
t. 800.659.8755
f. 760.936.4800

Verity New York

230 W. 41st Street
12th Floor
New York, NY 10036
t. 646.366.9500
f. 646.366.9540

Verity Washington, DC

2941 Fairview Park Drive
8th Floor
Falls Church, VA 22042
t. 703.289.8800
f. 703.289.8840

Verity Chicago

200 South Wacker Drive
31st Floor
Chicago, IL 60606
t. 312.674.4525
f. 312.674.4571

Verity Boston

275 Grove Street, 4th Floor
Newton, MA 02466
t. 617.663.5716
f. 617.663.5382

Verity Canada

400 – 4th Avenue SW
Suite 2500
Calgary, AB T2P 0J4
Canada
t. 403.750.4000
f. 403.750.4100

Verity Australia

115 North Oatlands Road
Yarrambat, VIC 3091
Australia
t. 61.3.9436.2024
f. 61.3.9436.2078

Verity Mexico

Camelia 253-6
Colonia Florida
Alvaro Obergon
Mexico D.F.
01030 Mexico
t. 52.5.661.7126
f. 51.5.598.9526

Verity Benelux

Coltbaan 31
3439 NG Nieuwegein
The Netherlands
t. 31.30.669.2120
f. 31.30.662.2094

Verity United Kingdom

The Pavilions
Kiln Park Business Ctr.
Kiln Lane
Epsom KT17 1JG
United Kingdom
t. 44.1372.747076
f. 44.1372.747071

Verity Germany

Robert-Bosch-Strasse 9 „Europahaus“
D-64293 Darmstadt
Germany
t. 49.6151.86082.0
f. 49.6151.86082.20

Verity France

14, Place Marie Jeanne Bassot
92593 Levallois Perret Cedex
France
t. 33.1.41.49.0450
f. 33.1.40.89.0981

Verity Japan

15th floor, Cerulean Tower,
26-1 Sakuragaoka-cho, Shibuya-ku, Tokyo,
150-8512, Japan
t: 81.3.5456.5792
f: 81.3.5456.5733

www.verity.com

© 2004 Verity, Inc. All rights reserved. Verity®, the Verity logo, Ultraseek®, TOPIC®, KeyView®, TeleForm®, and Knowledge Organizer® are registered trademarks. Verity® Portal One™, Verity® LiquidOffice™, Verity® LiquidCapture™, Verity Profiler™ are trademarks of Verity, Inc., in the United States and numerous other countries. All other trademarks or symbols are those of their respective owners.

**Sales and Product Information**

software-sales@verity.com or 888.328.7335

Partnerships

partners@verity.com

Technical Support for Existing Customers

tech-support@verity.com

Verity, Inc.

894 Ross Drive, Sunnyvale, CA 94089 | t. 408.541.1500 | f. 408.541.1600 | www.verity.com