



# GlobalWisdom™

## A People-Smart Approach to Categorization The Convergence of Manual & Automated Methodologies

### Overview

We are all familiar with the statistics on information glut – how the World Wide Web is expected to grow by 7.3 million pages a day. How 60 to 70% of corporate information remains unstructured. And how the Fortune 500 are expected to lose \$31.5B over the next two years due to information inconsistencies.

Enterprises are turning to integrated solutions coming from a number of market orientations, including document management, content management and enterprise information portals. However, a consensus is forming about one thing – a high-quality taxonomy, that is, a categorization scheme for business content, is an essential component for effectively managing information, whether for search, retrieval, content delivery, collaborative workflow or customer resource management.

### What makes a good taxonomy?

While there are many approaches to creating taxonomies, experts from a range of backgrounds agree that all effective taxonomies share certain characteristics:

- Taxonomies must be created and maintained by people with expertise in the subject matter;
- The systems themselves must be both easy to use and to maintain;
- A taxonomy must reflect the needs and organizational logic of its users and their business;
- The categories must make sense to humans;
- Access to information must be concept-based.

Traditionally, there have been two approaches to categorizing information, automatic and manual, each with its own advantages and drawbacks (see sidebar on page 2). At GlobalWisdom we believe there is a better way. Our technology exploits the strength of both traditional approaches, and eliminates the weaknesses. How? We believe people have the inherent smarts, not the software. This core belief is woven throughout our technology, and offers two benefits far beyond that of our competitors – one, a “**network effect**” that exponentially captures individual and team actions to continually improve the system, and two, the facilitation of true **critical thought** and analysis.



### Manual Categorization

Manual categorization isn't just the Yahoo model, where editors put each Web page into one and only one category. At its best, a manual approach is highly effective because it reflects the subtlety of human judgment and expert experience.

But there are drawbacks: These systems take a lot of expert attention to configure, populate and maintain. Subject matter experts are expensive to train and represent a commitment to an ongoing cost. Also, they have different biases than the information users – such systems can never fully reflect the diverse information requirements of their users.

To illustrate: it's like trying to set up your weekly grocery list. You can define a routine that decides you need more milk if the corn flakes box lost more than a quarter of its weight this past week. In some sense, that's automated (if the corn flakes box is sitting on an electronic scale all week). But what if you switch to a different brand, or type of cereal? What if somebody picked up a spare quart of milk in the middle of the week? All the power of the system comes from the labor-intensive manually devised set-up.

### Automated Categorization

The strength of an automated approach is that it is much more cost effective, and easily scales for large document sets. But it lacks the manual strength of expertise, and automated categories that make sense to the system are often meaningless to human beings.

Automated categorization tends not to track and capture, or aggregate and share. It tends to cluster and dictate, and then be stubborn when you want to fix it or when the content changes.

## I. All about Categorization

Software companies in the categorization space like to claim that their functionality encompasses both automated and manual methodologies. Unfortunately, many of these claims fall short of meeting business needs. Why? Essentially, they are extending their core technologies, rooted in one or the other approach, to move toward this convergence of automated and manual methods. There are limitations to this approach, though, as core algorithms are best at delivering on their original intent.

There are currently three "hybrid" approaches to the semi-automated creation and maintenance of taxonomies:

1. **Classification by example** creates categories by analyzing a training set. Editors name a topic, then provide a set of documents that represent the topic (the larger the set the better). The software then captures all similar documents to automatically populate the topic. This allows an automated system to take over after manually defining the example sets.

**Advantages:** by using examples as guides, topic definitions come closer to capturing how people have framed them in the past. Automated population of topics can be very efficient.

**Disadvantages:** updating or changing the taxonomy involves adding new training sets. Best results come from large document sets, which increases effort.

2. **Rule-based** employs a set of rules that identify relationships among words and metadata that are predictive of topics to determine how documents are classified. Again, these must be manually defined, turning to the software to populate the topics.

**Advantages:** does not require a document training set.

**Disadvantages:** changing or updating the topics involves interruption of workflow to go back to the rule set and redefine a rule. Rules can become quite complex, requiring high level of expertise to update. Using words means that it can be difficult to capture the subtlety of concepts that can be expressed in many different words and phrases.

3. **Statistical/linguistic** extracts patterns statistically from text, and maps them into a manually created lexical database or thesaurus. The mappings are used to detect semantic connections between terms, e.g., synonyms, and to guide the classification of documents.

**Advantages:** no training set. Creates categories as they are discovered in text. Automated results can be edited by hand.

**Disadvantages:** improving the quality of automated categorization for your business requires refining the thesaurus with the services of a knowledge engineer.



### About Topic Maps

Topic Maps are an ISO standard (ISO 13250) and XML standard (XTM v.1.0) for the organization and representation of knowledge. Bryan Thompson, President, CTO and co-founder of GlobalWisdom, is a founding member and an author for the XTM 1.0 specification.

Topic maps are rooted in the recognition that the familiar structure of indexes, such as the ones that appear in books, imply a common, powerful set of assumptions about categorizing information. There are three main elements: **Topics**, **Occurrences** and **Associations**.

**Topics** are anything that we need to talk about. They could be as concrete as a person, or abstract as an idea or an event – just like a topic of conversation. **Topics** can be further identified by types - for example, authors, books or reviews.

An **occurrence** identifies a specific content resource that is an instance of some topic. **Occurrences** can also be broken up into types – i.e., digital photographs, articles, or commentary.

The potency of Topic Maps comes from representing **associations** between topics. For example, *Oliver Twist is a book written by Charles Dickens*, or, *Charles Dickens was influenced by serialization of novels in newspapers*. Further, associations themselves are also topics, meaning that they can be assigned types, e.g., “news-source”, “author-of”, etc., and organized as part of the taxonomy.

In codifying these elements, **topics**, **occurrences** and **associations**, categorization schemas can begin to reflect how people actually use information and how they make sense of their conceptual landscape.

See [www.topicmaps.org](http://www.topicmaps.org).

## The People-Smart Approach

As Figure 1 illustrates, GlobalWisdom alone fully integrates the benefits of both the manual and automatic approaches, not just philosophically but functionally as well.

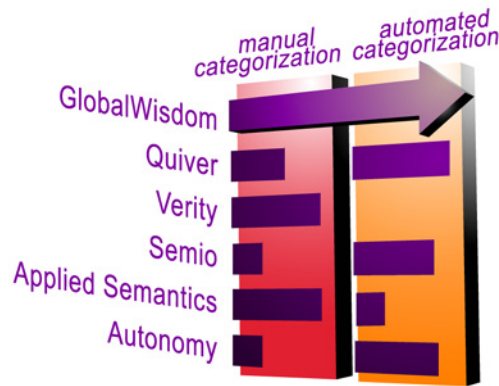


Figure 1. Competitive Comparison: Categorization

GlobalWisdom’s software elevates existing cutting edge efforts to create a true manual/automated synthesis. How?

**We make everybody an expert.** Editors with expertise in a subject matter can be invaluable when setting up a structure or taxonomy. They may be experts in, for example, high tech market trends, semiconductors, hardware manufacturing or PC buying trends, but they are not necessarily the experts in the specific business uses to which people apply that information. Furthermore, the ways in which information is used evolves over time. Ultimately, the end users are the experts on how they best use information, the experts in their own specific needs. The taxonomy must reflect their expertise as well.

**Easy to use, easy to maintain.** The bulk of our explicit feedback occurs within *the workflow*. Users are not required to leave their current tasks to access tools for improving or editing the taxonomy. Rather, GlobalWisdom gathers feedback from how people *already* work with information.

**We enhance and extend existing taxonomies.** GlobalWisdom technology doesn’t replace existing taxonomies and hierarchies – it works with them, extending them to meet the highly personal information requirements of users. Our technology can relate existing taxonomies to non-hierarchical structures as well using Topic Maps (see sidebar) – an international standard for out-of-line metadata. In fact, our technology is fully compliant with the ISO and XTM Topic Map standards – standards that protect the investment in your content assets.

In traditional online publishing models editors “own” the core taxonomy. GlobalWisdom technology allows subscribing businesses to extend the core taxonomy to create unique and customized categorization schema to satisfy their specific business needs.



### Network Effect Defined

Metcalfe's Law is taking over the business world. The value of a network increases with the square of the number of connections. And an Internet company's business model is assured if the value of their services increases with the square of the number of users, or applications, or languages or databases.

A company has the opportunity to replicate this model, the only sure-fire road to success on the Net, if its own users progressively add value to its core capability. As the effect of those users accumulates, it requires less effort for subsequent users to receive the same value. The end result is a steadily improving value proposition – each new customer gets more valuable service, and each new customer costs less to serve.

A network effect also builds a sustainable competitive advantage. Regardless of whether the underlying technology is now or remains superior or even proprietary, the accumulated data becomes harder and harder to replicate as it grows. No other company could build a comparable foundation – since even that company didn't build the foundations, its users did. User-contributed data becomes the most effective barrier to entry that a first mover can create.

**Taxonomies evolve.** No taxonomy is going to be perfect for all customers. Intel, Compaq, and Office Depot all have an interest in what is going on in the semiconductor market. For each company, though, various topics and subtopics will carry more or less importance. Office Depot will likely put more weight on rebate program effectiveness, while Intel would put more weight on changes in chip manufacturing techniques. Compaq might have a general category for chip manufacturing while Intel will need a more fine-grained representation, such as cost of machinery, plant construction, as well as various manufacturing methods. A basic taxonomy for high tech would serve as a starting point for each company, but with GlobalWisdom the taxonomy can evolve to meet each company specific needs.

**We make sense to human beings.** Automated systems can be random and arbitrary. By incorporating user feedback, the topics generated by our software can reflect the subtlety and complexity of human thought.

**Concept based.** Our software does not categorize based on keywords, or by gathering identical text, but by concepts expressed by the indexed documents.

## II. Expert People, Not Just Systems

GlobalWisdom's "people-smart" categorization approach gives our technology the ability to create a **network effect** (see sidebar). GlobalWisdom's **network effect** captures, organizes, aggregates and shares individual actions and decisions (Figures 2 and 3), both implicit and explicit, so everybody benefits from everybody else's expertise, and the system gets progressively smarter – a people-smart taxonomy.

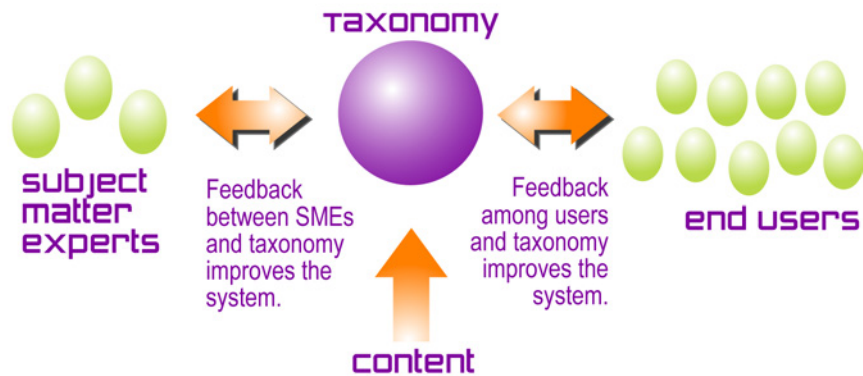


Figure 2. Cumulative Feedback of Network Effect

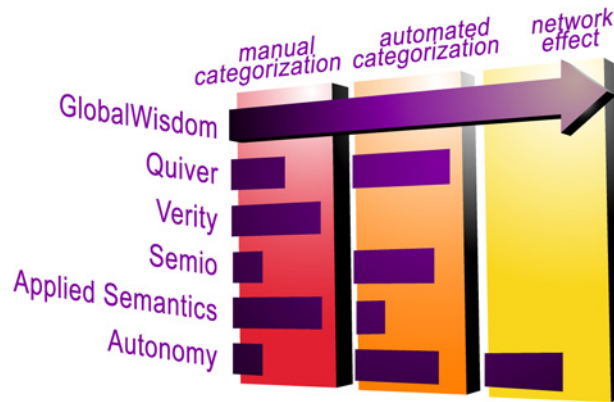


Figure 3. Competitive Comparison: Network

**Multiplying the multiplier.** There are two elements to this approach that extend the reach of the network effect. Consider how different groups, perhaps departments in an enterprise, would use the software (see Figure 4). Clearly, users within the engineering department are going to benefit from aggregation of interactions between engineers, just as users within the marketing department will benefit from aggregation of interactions between market researchers, advertisers, and copy editors. As strategic decisions are made and represented by the software, areas of overlap, disagreement between user groups are highlighted, interaction is fostered and resulting actions captured as well.

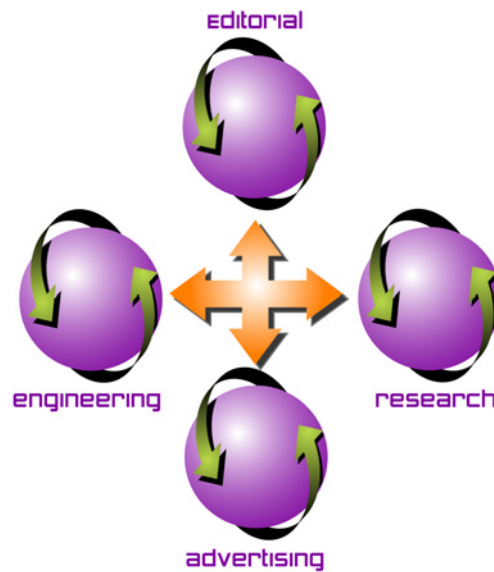


Figure 4. Cumulate feedback of Network Effect within and among discreet groups

There is a two-way cyclical interaction between the **network effect** and **critical thinking**, another key GlobalWisdom differentiator (see part III for more). The **network effect** collects the right information and decisions so that the user can assess, evaluate (**critical thinking**), and make a more informed decision, which is then itself captured and shared. The taxonomy is continually improved and refined as better and better decisions are made.



## Network Effect Case Study

Consider the example of a company that provides quality access to well over 50 databases for scientific and technical research. A team of nearly 5 dozen editors work full time to classify, tag and develop 10-15 keywords per document, as well as editing or re-writing abstracts. These databases average over 10,000 entries per year, and are updated monthly. Speed as well as accuracy is of the utmost importance.

**ROI.** Forrester and other industry analysts estimate a cost of \$100 to fully index a single document using current best practice methods, and \$10-20 to add keywords for a single page abstract. With 500,000 new abstract entries a year, costs for this publisher would run approximately \$10 million annually for abstracts, \$50 million for full documents. If we reduce these costs by just 25%, the ROI for abstracts alone comes to \$2.5 million—the cost of implementing a high-end content management system.

GlobalWisdom's software adds value to three links in this value chain:

- **Reduces the cost of adding keywords to new content:** While the software cannot replace an expert's ability to work with subtlety and complexity, much of the development of keywords and metadata tagging is more mundane. Our software works in conjunction with the editors, nominating the basic classifications, keywords, and metadata tags, allowing the editors to focus on the more complex aspects of organizing documents, as well as refining the abstracts. The feedback loop between this process and the software's learning capabilities continuously improves the quality, accuracy, and the granularity, or level of detail, in the classification system.
- **Increase usage and improve satisfaction for online searches:** With databases containing over 100,000 documents in a single discipline, most searches return too many results – users are unable to sift those results to find the specific information they need. The only current alternative is to restrict search terms, keywords, or the databases themselves, running the risk of leaving out relevant information. Our software lets users restrict a search by topics defined by other users or themselves – and then navigate among the topics and content relevant to those search results. Users will find what they need more easily and over a broader range of databases.

Further, the definitions of those topics become more accurate over time as our technology learns about user needs based on the search results they deem relevant. Users effectively vote with their feet, indicating which content is relevant to the existing topics in the light of their specific business needs. With each search the set of topics becomes richer and more accurate, further improving performance.

- **Improve purchase rates for content included in information updates and email alerts:** With GlobalWisdom's technology, a subscriber can construct their own profile to suit their particular specialty or interest, using the full combined power of topics and saved searches. This profile can then be made available to others



subscribers. This facility for customization and automated knowledge re-use can support increased subscription use and purchase rates. It also provides the content provider with the ability, through their own users, to rapidly identify and adapt to new areas of interest, further raising customer satisfaction.

### III. The Thinking Corporation

At GlobalWisdom, **critical thinking** isn't just about automated learning algorithms. It's about the interaction between the algorithms and all the individual users, and the aggregation of the results. The software can steer users toward issues, nominate choices, and then learn from the resolution. As Figure 5 shows, there is no other competitor whose technology extends to this functionality, either philosophically or in practice.

#### Critical Thinking Defined

Dr. Marvin Cohen – co-founder

Critical thinking is a dialogue with yourself or with others in which your beliefs, plans, inventions, practices and other creations are challenged, defended, replaced and/or improved in order to achieve some practical objective. We think critically when we ask ourselves or others such questions as, How do I know this? Why did I decide to do this? Does this design achieve its intended effect? Are my reasons adequate? Are there other considerations that go against this conclusion or course of action? Is there a better way? What is my real purpose, and am I addressing the issue that really concerns me?

The critical thinking dialogue challenges habits and settled beliefs, exposes hidden assumptions, helps identify and fill gaps in knowledge, brings out alternative approaches that might otherwise never have been considered, speeds up learning, and keeps you on track toward achieving your goals. Critical thinking skill is creative as well as critical: It requires the ability to fashion new solutions, to know when to follow a "gut feeling" and when to question it, to find relevant analogies from other domains that shed light on your own problem and in general to use your knowledge effectively while continually enlarging it.

Critical thinking is not a substitute for knowledge and experience in a particular field, such as medicine, law, or business. But it is a powerful magnifier of such knowledge and a crucial component of learning.

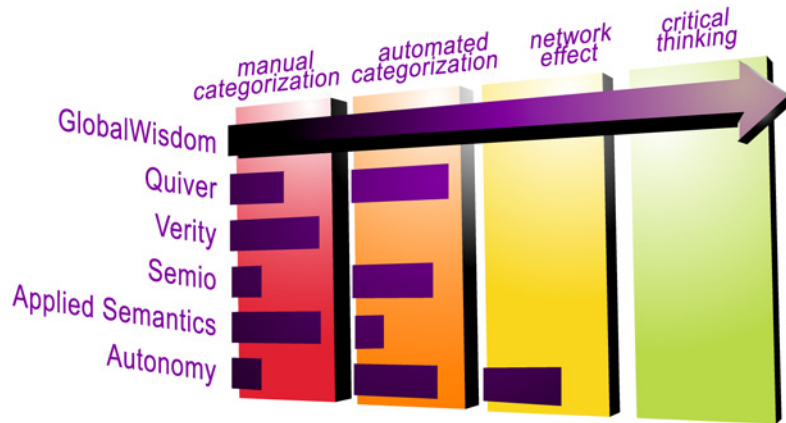


Figure 5. Competitive Comparison: Critical Thinking

A key benefit of critical thinking is the resolution of “disagreements” – how users are alerted when there are different ideas among different users about how a document should be classified. GlobalWisdom technology allows multiple classification schemas to accommodate different viewpoints, permits classification schemas to be merged when they are complementary, uses one schema to correct another when appropriate, helps users break down larger topics into more refined categories based on their interests and those of other users and resolves ambiguities in translating from one scheme to another.

GlobalWisdom’s critical thinking features operate by monitoring the diversity of opinions and usage patterns within communities of practice. When divergent opinions are identified, critical thinking features alert users with editorial responsibility for the relevant subject matter, helping them to uncover the source of the divergent opinions and improve the taxonomy so that it captures the variety of perspectives and needs of its users. Editorial responsibility can either be localized or distributed, depending on the policy and needs of the company.



Here are some specific examples of diverging classifications, each of which receives a different treatment:

- Among groups who care about different things. Engineers and marketing personnel need to use different information or see the same information from a different perspective;
- Among groups who care about the same things but approach them at different levels of expertise or experience (e.g., laypersons, medical students, physicians, and medical researchers);
- Among peers who are addressing the same issues at approximately the same level of expertise, possibly from different theoretical or practical perspectives, but who are willing and able to benefit by sharing ideas and by using one another as sources of error correction;
- Among groups that use different terms to describe the same topic or the same terms to describe a different topic.

GlobalWisdom technology supports collaboration among people who work in the same or related space, and, further, tracks how the relationships among different groups evolve for the benefit of all users.

#### **IV. Summary**

GlobalWisdom technology increases the ROI of enterprise information management systems by capturing, organizing and sharing individual and collective knowledge. Our software enhances publishing, content and knowledge management systems by offering superior and unparalleled functionality for creating, maintaining and evolving powerful taxonomies and delivering business services powered by those taxonomies.

By incorporating a people-smart strategy with proven artificial intelligence learning algorithms, and combining the strengths of automated and manual approaches (while avoiding their weaknesses) we've not only harnessed the immense power of the **network effect** for our customers, but given them the ability to understand and resolve new, unique and individual issues with **critical thinking**. And, we have positioned ourselves and our technology as the industry leader – in thought, innovation and functionality.

Learn more at <http://www.GlobalWisdom.Org>.