NAVIGATE YOUR INTELLIGENCE WORLD BY USING TAXONOMIES

How Applying Structure Will Help You Steer in the Right Direction

Jesper Martell & Gabriel Anderbjörk

Taxonomies - how can such a powerful concept and tool for information management, both generally and in the world of Competitive Intelligence, attract such a variation of opinions? The answer is actually very simple: there are different types of taxonomies and different ways to deploy them. With almost 20 years of practical experience of taxonomies in business, both in CI and Knowledge Management (KM), we are confident to say that it is one of the most powerful tools (if not the only tool) currently on hand to manage, make use, and make sense of the ever growing "big data of unstructured information".

"Finding a piece of information within a large collection of data without a taxonomy is like driving in unknown territory without the benefit of a map or road signs."
- Penny Crossman, Intelligent Enterprise

Taxonomies help people and organizations to quickly bring structure to vast amounts of otherwise unstructured information. They support organizations that need to make quick and well-informed business decisions to overcome one of the largest problems such organizations are facing today, namely being lost in "information fog".

OCEANS OF INFORMATION

Let's start from the beginning by identifying what it is that makes it so hard to find what you are looking for; we call it the **"three stormy oceans of information:"**

1. The Ocean of Overload
2. The Ocean of Storage
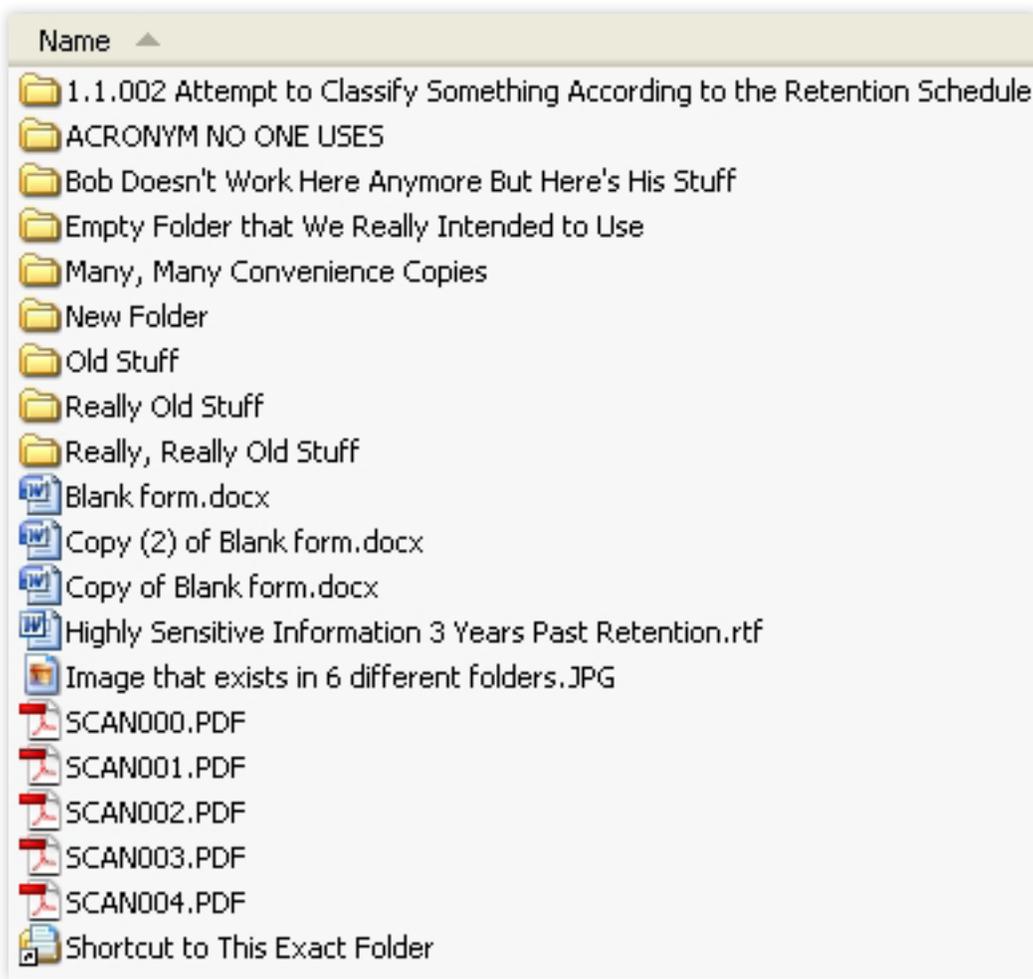3. The Ocean of Retrieval

1

Overload

We all know this already, but the volume of digital data will grow by 40% to 50% per year over the foreseeable future (according to IDC, an IT analyst company). As a consequence, the world's information is doubling every second years and by the year 2020, there will be 50 times the amount of information there is today.

For organizations this means that making quick and well-grounded business decisions become increasingly difficult. It is no longer possible for an individual to manually sift through the amount of information available, a situation that in turn also renders it impossible to manually ensure that all relevant information is at hand to make the best decision.

Storage

The approach of storing information/documents/files etc. in, folder structures predates computers by at least 2000 years. It is claimed that the library in Alexandria is the first known institution to use a structured and labeled system for storing its documents. Despite its long history though, it is a practice with great limitations; a document can only be stored in one place.

But how many documents written today are only about one single topic? Where should a document containing information about a particular company, a particular product, a particular technology, AND a particular market be stored? In the folder for company A or B, the folder for product 1,2,3 or 4, or….? No doubt, every reader of this article recognizes the situation. Most would confess to have a hard time making such decisions even on their own device. Taking that challenge to the "shared network drives" that many organizations use often results in an unmanageable mess of information stored according to different individuals different perspectives of how to best choose the "one best place" for their documents. Everyone has their own way of categorizing things, which makes it hard for others to find.

Name ▲
📁 1.1.002 Attempt to Classify Something According to the Retention Schedule
📁 ACRONYM NO ONE USES
📁 Bob Doesn't Work Here Anymore But Here's His Stuff
📁 Empty Folder that We Really Intended to Use
📁 Many, Many Convenience Copies
📁 New Folder
📁 Old Stuff
📁 Really Old Stuff
📁 Really, Really Old Stuff
📄 Blank form.docx
📄 Copy (2) of Blank form.docx
📄 Copy of Blank form.docx
📄 Highly Sensitive Information 3 Years Past Retention.rtf
🖼 Image that exists in 6 different folders.JPG
📄 SCAN000.PDF
📄 SCAN001.PDF
📄 SCAN002.PDF
📄 SCAN003.PDF
📄 SCAN004.PDF
📄 Shortcut to This Exact Folder

**Image:** *Messy information storage is not uncommon*

3

Retrieval

The two challenges described above results in the third and most costly challenge of them all, the inability for individual decision makers to retrieve the information they need. Due to the volume of information, it is impossible to assess the relevance and accuracy of all incoming information. In combination with a failed model for storing information the result is often *"decisions in the dark"*.

If this challenge is not addressed, the costs associated with not having relevant information for decision making is enormous. Highly competent individuals are expected to utilize the company's proprietary information base and use their knowledge in their job roles. It is unfortunate if they cannot access and make use of the existing information base due to lacking structures and support systems. The consequence is that information often has to be reinvented time and time again.

One of the best tools to overcome the 3 *"oceans"* described above is to use taxonomies. A structured approach to information management makes collection, analysis, and dissemination of information both faster and more efficient!

WHAT IS A TAXONOMY ANYWAY?
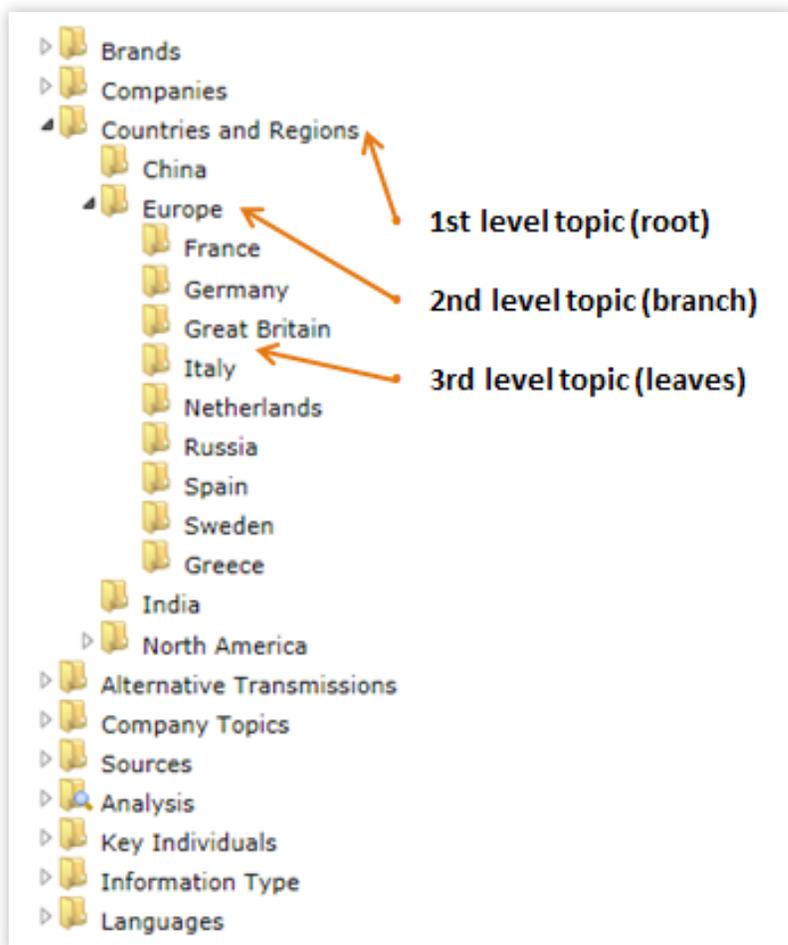
Fundamentally, a taxonomy is a **classification system - a standardized and controlled set of topics, terms, and phrases.** Taxonomies have historically had a large role in biology, e.g. for plant and animal classification. They have also been widely used in library sciences, where books need to be classified according to a structure to make them easy to find. Taxonomies in a business context serve a similar purpose but in business it is used to categorize information to facilitate storing as well as retrieval. As information is primarily used to support decision making, the taxonomy should enable procedures to minimize time spent looking for the information needed and ensuring that all relevant information has been taken into consideration.

Image: *A taxonomy has roots, branches and leaves*

It is fair to say that a folder structure is one possible representation of a taxonomy but in the context of today's business needs, we have already discarded that as a non-sufficient approach. In a modern business context taxonomies are designed for "tagging."

Before we proceed, it is essential to clarify the difference between content-driven and needs-driven taxonomies. Many wrongly claim that taxonomies are not helpful in addressing the challenges at hand, but that is only because they refer to the wrong type of taxonomy.
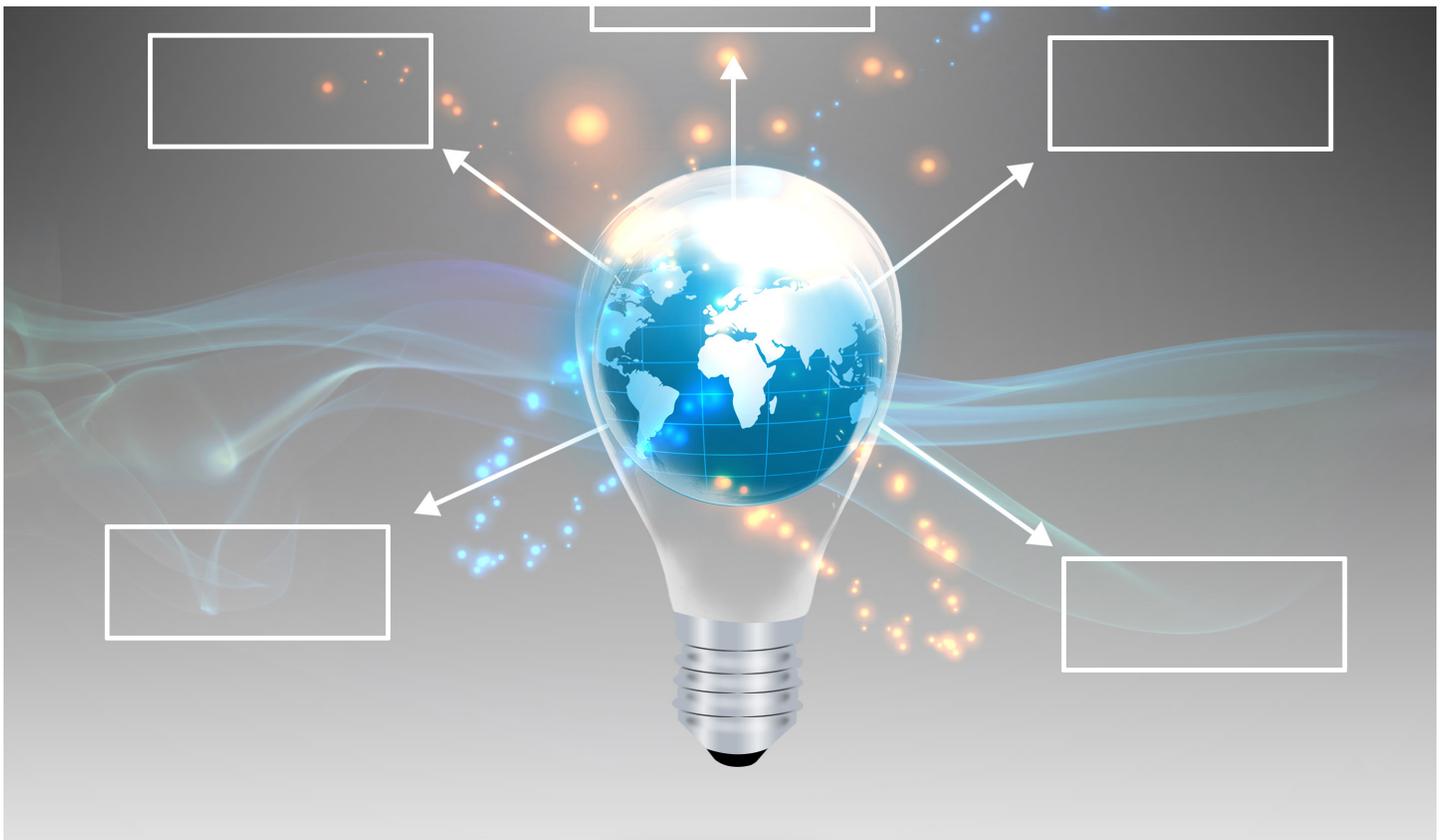
Content driven taxonomies:

Traditionally, taxonomies have been used to classify and codify information that is already at hand. Again, think about a traditional library. There are elaborate systems for knowing exactly on what shelf to put a book about British countryside cooking during the Victorian era. However, this can only be applied on books that you already have, and no one would bother doing this if such a book did not exist.

Unfortunately, the same sometimes applies to corporate taxonomies that are only developed to answer the question **"what information do we have?"** If corporate taxonomies are developed without a single resemblance of the concept of a customer, competitor, partner or other "external factors," they become very inward looking, This misses the point of having a taxonomy, especially seen from a competitive intelligence perspective.

Needs-driven taxonomies:

A needs-driven taxonomy, on the other hand, should answer the question, "**what information should we have?**" We believe that a taxonomy used for CI purposes should reflect the industry and highlight intelligence that is important to the organization. This means that if a topic for some reason is empty, it is not necessarily a negative thing. It could for example be that there is a gap or blind-spot in the understanding of the business environment that has become identified.

Consequently, the procedures for developing a needs-driven taxonomy differ significantly from those of developing a content driven taxonomy. Needs driven taxonomies are key to both Competitive Intelligence and Knowledge Management since both are fundamentally driven by market changes and the need to adapt to the business environment.

An Intelligence Taxonomy should NOT be based on what current information is about, but rather on what the decision makers need.

A NEW APPROACH TO INFORMATION MANAGEMENT

From an individual information user's perspective, the predominant situation is that there is far too much information to digest, it is badly stored and structured, the tagging system, if at all existing, is not meeting my needs and worse, it is a manual task to maintain. A new approach to information is needed!

Organizations need to have a good structure to index and access relevant business information which means a classification system that allows all pieces of information to be categorized with **multiple topics**. Returning to the example above, the choice should not be between Company A, Product 3, or Market 2, the choice should include all of them. A very familiar analogy would be that of tagging pictures on social media sites such as Facebook. No one these days is surprised that a picture can be tagged with multiple "topics" such as grandma, grandpa, sister, best friend etc. The only difference in a business information context is that content is tagged with business relevant topics, instead of relatives and friends.

Information can then be found through multiple entry points, and as a result, organizing information as well as finding and navigating for information becomes structured and efficient. As such, it is driven by future users' needs rather than the storage decision taken by the producer at the time of storage.

This is in stark contrast to "traditional" taxonomies that "classify" information with one label only by the sheer storing of it in one folder (or book shelf). However simple it sounds, classifying information with multiple topics opens up vast opportunities for organizations that have to manage large quantities of information. The way a corporate taxonomy should be created is by assessing the need for different topics in relation to decision mandates throughout the company. Most individual knowledge needs in a company will be related to one or more of the following tags or topics:

- A company (e.g. customer, competitor or supplier)

- A company topic (e.g. finance, strategy, organization, pricing)

- A product/service (e.g. any product from the product catalogue or competing products)

- A product character (e.g. technical spec, pricing, market penetration)

- A country

- A country topic (e.g. economics, security, industry regulations, etc.)

Image: *Example of common tags or topics in a corporate taxonomy*

In addition to the topic areas listed above there are, of course, industry specific topics on regulations, government matters, etc. But as a basic taxonomy, the above can be found in almost every company.

Once the taxonomy is in place, the properties and attributes that inform elements such as documents, etc. will be matched with the taxonomy topics, and each document will consequently be tagged with a descriptive set of topics.

Using this type of thinking, all relevant topics can be assigned to a particular piece of information. As a result, the information has multiple entry points, allowing different users to find the same information, independent of other individuals' perspective of choice.

THE BATTLE BETWEEN MANUAL AND AUTOMATIC CLASSIFICATION

The classification into taxonomy topics can be done either automatically or manually.

Manual classification is done by editors or administrators who have the right to publish and edit documents/ articles, as well as manually add topics based on their individual judgment.

Automatic classification is tuned by creating a set of search rules for each of the taxonomy topics. This type of classification is triggered by identifying key words and search phrases in the content retrieved, which is followed by tagging that content into the right topics. Each topic is associated with an unlimited set of search algorithms or rules (using key words together with operands AND, OR, Phrase, Exact Match).

The main difference between automatic and manual classification is that automatic classification is objective, and manual categorization is subjective. Two people seldom have the same opinion on metadata or topics.

The major benefit of automatic classification is that it allows organizations to go through large amounts of information each day without spending man-days on the filtration, thus allowing more time to be spent on the usage and analysis of the information. Consider yourself reading an article with the purpose to assess whether it has a bearing on your business and, if so, to tag it according to the company taxonomy. As an individual you can read one article at a time and in the best of worlds know the entire taxonomy well enough to tag it properly (although even that is unusual). With automated classification you can instead make sure that the topics themselves "read" all incoming articles and documents simultaneously and with almost exactly the same "assessments" you use yourself decide whether the article should be tagged with a particular topic or not. The difference? The automated topic can read thousands of articles in a fraction of the time it takes you to read one. With an automated taxonomy your document libraries will, despite its size and pace at which it is building up, always have the labelling needed for your decision makers' needs to retrieve relevant information.

The benefit of manual classification, on the other hand, is that the tagging, in itself, is an analysis (specially suitable for more abstract topics such as Opportunity/Threat, Strategy, Positive/Negative), but people usually do not know what topics may be of interest to a department other than their own, plus they do not have the time.
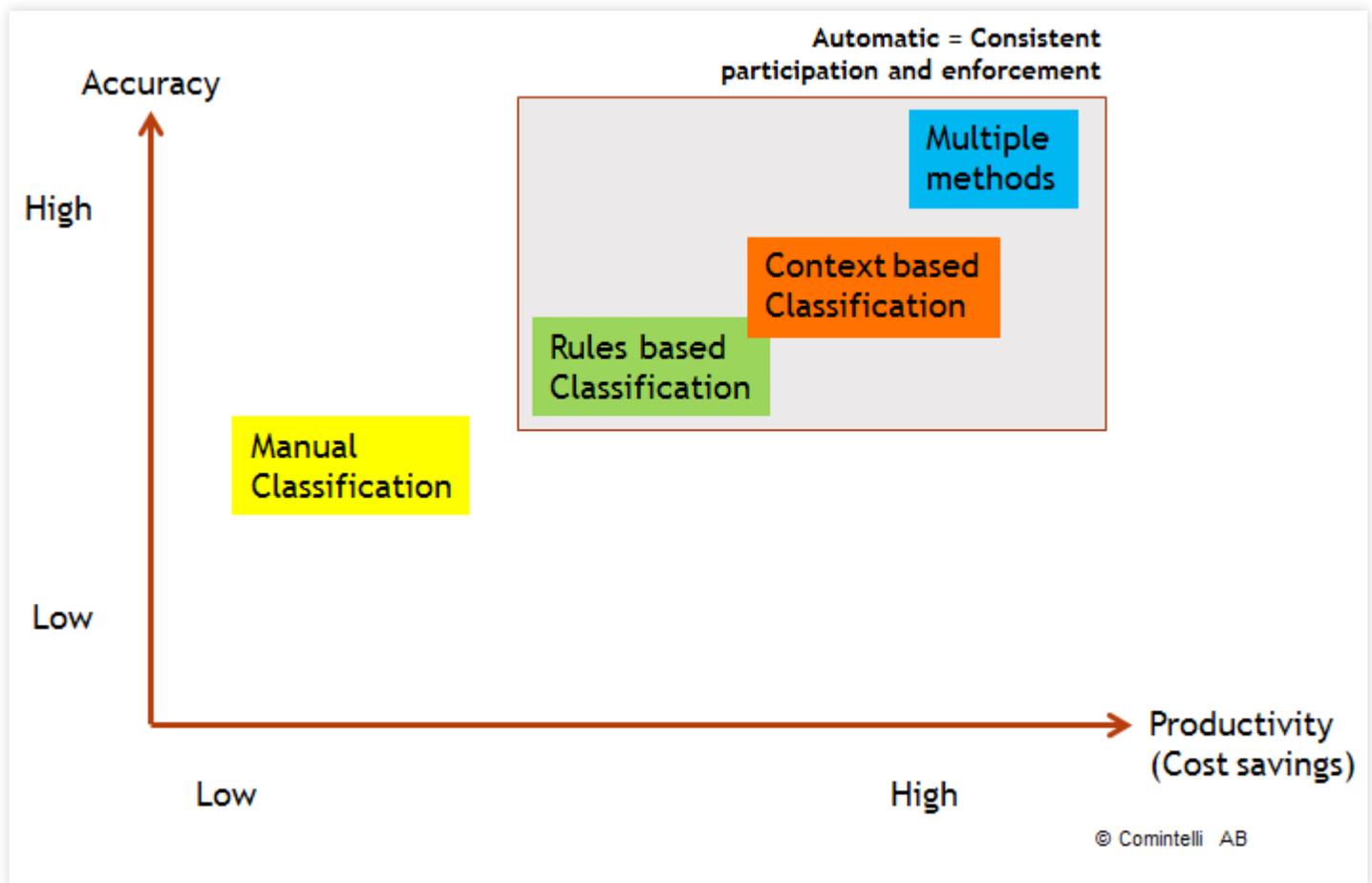

Image: *Pros and cons of automatic vs manual classification*

ADJUSTING YOUR TAXONOMY TO YOUR WORK PROCESS

Great! Now all of your information is being tagged with relevant topics; enabling fast filtering and decision support in perspective of defined decision mandates, that's wonderful! But why limit taxonomy design to that dimension, when you can also apply and customize a taxonomy to support your work process?

 As a CI analyst, a significant portion of the work is also to package content into special reports, create weekly targeted newsletters and/or assess information from perspectives of different analysis models such as PEST, SWOT, Porter Five Forces, Porter Four Corners, etc.

Since a taxonomy should be set up in such a way that you should be able to add and manage the already tagged topics on any document, a very time-saving exercise is to add a dimension to the taxonomy that reflects your work and analysis models. With that in place, the CI analyst can use the business dimension topics to filter out relevant content and then swiftly add topics that reflect the analysis perspective as well as the intended dissemination mode of the information. Such a taxonomy dimension is typically designed as follows but can of course cater for any other work models your particular company may favor.

- Analysis (e.g. PEST, SWOT, Porter 5F) (with subtopics for each; e.g. Strength)

- Output format (e.g. Brief, Alert, Weekly report, Monthly report, Quarterly report, Company report, Product report, Market report)

- Process (Potential, Draft, Selected, Final)

SOME PRACTICAL EXAMPLES

So, in the "real world", what does a taxonomy look like? It can of course be graphically represented in many ways, but the following illustration is an example:
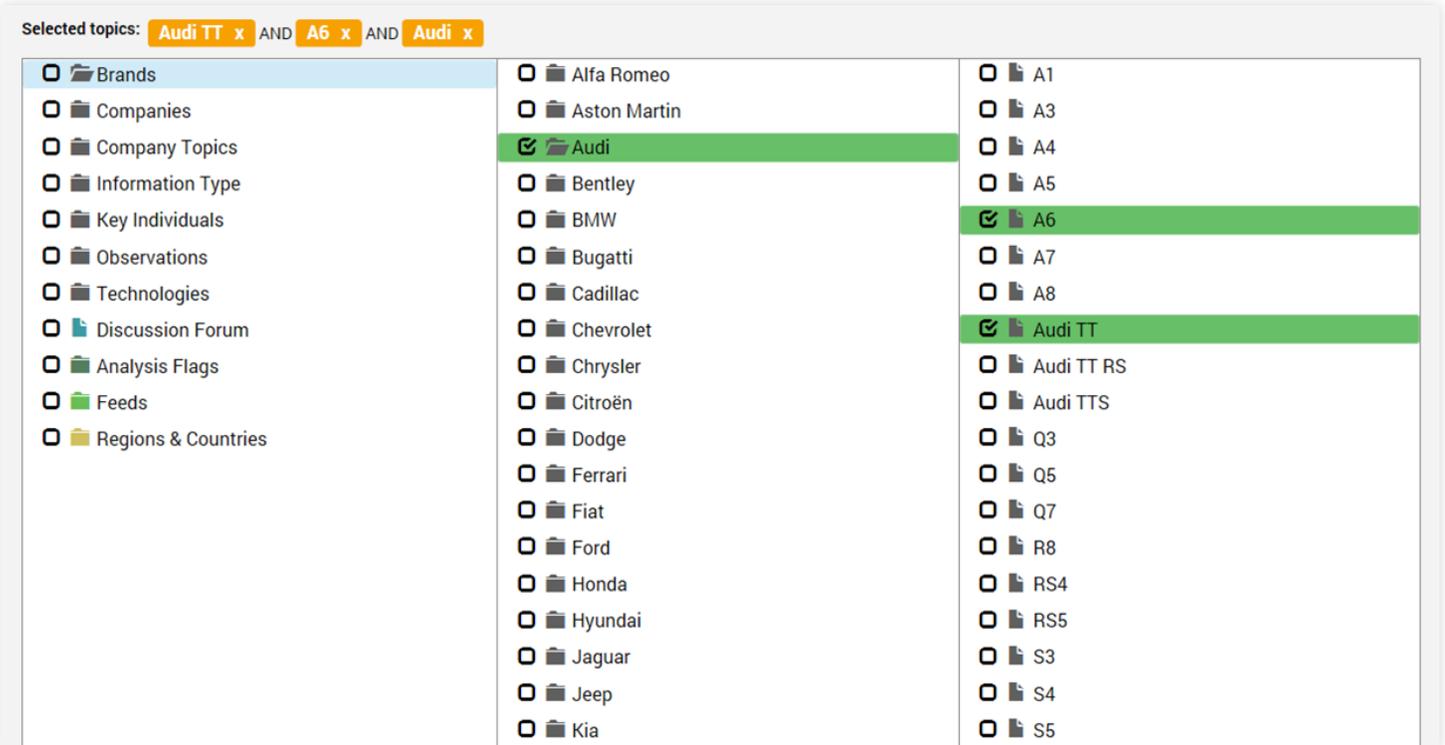
Image: *Example of horizontal taxonomy where the first column contains root topics, followed by 2nd and 3rd layer subtopics.*

By combining topics from the different roots and subtopics, a highly efficient filter is applied. In a given example, out of thousands of documents, the following filter reduces the number to only those that match these selected topics in combination, producing relevance in a matter of seconds!

Image: *Filtering by Company, Technology and Country*

If, as an analyst, you are particularly interested in information related to BMW's Hybrid activities in Germany, set your taxonomy to alert you at any time such a match appears and the taxonomy will keep doing the reading job for you, tirelessly day and night.

When a taxonomy has been created, each of the topics are assigned keywords and synonyms to each of the desired topics, often using Boolean logic or Regular Expression pattern matching. This is what will enable the taxonomy to assign the right topics to each piece of information by scanning the content and allow the taxonomy to do the bulk filtration of the information for you. There are certain types of topics that are more suitable for the automatic classification than others; for example topics with specific names such as Company, Product, and Services names. Other types of topics, for example Context topics, requires more often than not manual classification because of its nature, but the benefit of this is that users can quickly understand a document context by just seeing what topics it has been

tagged with. One example is if you classify a document with a certain company and also "company strategy." By seeing this topic tag, the user will automatically know that this piece of information is about that company's strategy.

By browsing all the information that has been tagged with a particular topic, you instantaneously narrow the scope of your search. When you add topics to your query, the search results become more and more relevant, and you save time by only having to search through a smaller amount of information rather than a long list of articles and documents.

WHAT ARE THE BENEFITS OF USING TAXONOMIES

Taxonomies help your organization deal with the ever increasing amount of information in two main ways;

- Speeding up and improving findability of information;

- Facilitating both manual and automatic classification of information.

Having a taxonomy that automatically helps you tag your information allows your organization to have a better overview of what is going on in your business landscape and free up time to do the analysis of that information. They also allow a better way of sharing information and knowledge within an organization, ensuring that no resources are wasted and that the existing capabilities of the organization are leveraged.

In addition to these benefits, there are also less obvious benefits that are gained from the actual process of creating a taxonomy optimized for business. For example:

- Unifying corporate language and terminology;

- Creating a shared map of the intelligence needs for the organization.

If the taxonomy reflects the intelligence needs of the organization, it will clearly visualize what is important to know for the whole organization and will allow people to start using the same topic names. An often overlooked benefit of a good taxonomy is that it encourages more people to become "intelligence explorers" by creating a shared and updated map of your business landscape, thus improving strategic communication within your organization.

To Navigate: "It is not the ship so much as the skillful sailing that assures the prosperous voyage"

-George William Curtis

Finally, we summarize in 5 quick points:

- **CI Taxonomies** are based on Key Intelligence Topics and decision areas, not the content;

- **CI Taxonomies** are dynamic and adaptable to changes;

- **CI Taxonomies** allow each piece of information to be classified into multiple topics, not just one place;

- **CI Taxonomies** avoid duplicate topics (think in terms of "tags");

- **CI Taxonomies** are simple and easy to use – for everyone!

About the authors:

Jesper Martell is Chief Executive Officer (CEO) and co-founder of Comintelli (www.comintelli.com). Prior to founding Comintelli, Mr. Martell had a career at Ericsson between 1994-1998. Mr. Martell holds a Master of Business Administration (MBA) from the Stockholm School of Economics.

Gabriel Anderbjörk is Chief Financial Officer (CFO) and co-founder of Comintelli (www.comintelli.com). He was Corporate Director of Competitive Intelligence at Ericsson during the years 1996-1998. Mr. Anderbjörk holds a B.Sc. degree in Physics and Mathematics from University of Lund and an MBA from London Business School.

The authors welcome any comments and questions on the content of this article. They can be reached at jesper.martell@comintelli.com and gabriel.anderbjork@comintelli.com.