People try to quantify the value of assets or potential assets as part of the decision process on whether to acquire or maintain those assets. This quantification may be irrelevant in the case of taxonomies for two basic reasons. First of all, as has been pointed out by others, quantitative evaluation may be impossible, or at the very least, impractical. Moreover, in many situations taxonomies are clearly indispensable, or at least invaluable, which pretty much settles the matter of acquiring and maintaining them.

The better thing to evaluate may be whether a taxonomy is important, or indispensable, for a particular situation. It’s somewhat like a piece of machinery that’s needed to do something that no other machinery can do to the necessary level of precision and accuracy.

One type of setting in which taxonomies have proven their worth is the standard search/research platforms of content aggregators and of professional and scholarly associations. In these platforms, interestingly enough, there are two basic knowledge assets: the taxonomy (or thesaurus) and the literature database or set of databases. Only with a knowledge organization system (KOS), such as a taxonomy or thesaurus, can highly accurate and thorough categorization (or indexing) of a corpus of literature within a discipline (or set of disciplines) be achieved, and only with such categorization can excellent recall and precision of search results be achieved.

This thought experiment will focus on research platforms that combine a taxonomy with a literature database. Generally speaking, a searcher using such a platform could be said to be “mining” information. The objective of this thought experiment is to see just how far we can extend the mining analogy suggested by such expressions as “data mining” and “text mining”
(aka “text data mining”), which are often used in other types of semantic analysis. Seeing where parallels can be drawn, where they break down and where things get confusing or just plain silly may help us understand some of the economic characteristics.

The Database or Repository as a Mine

Gold is where you find it, as the saying goes. The database (or, alternatively, a repository to which a bibliographic database is linked) is where the nuggets of information reside, or potentially reside, within the articles and similar resources. Unlike mineral resource mines, this mine does not get depleted when people take its resources. This preservation is partly the nature of knowledge assets, as well described by Thomas A. Stewart, and partly due to the fact that a copy is as good as an original, which is probably a copy, anyway. (Actually, the whole matter of copies versus originals may be irrelevant in a digital context.) The resources, including the nuggets, stay within the repository or database. In another departure from mineral resource mining practices, the organization that maintains the database is likely to find itself in the position of adding resources to keep the material up to date.

Prospecting for Raw Materials

In a mineral resource mine, the target of the mining is rock that contains nuggets and veins; the rock is the ore. This point is where things could get turned around a bit. In accord with the usual use of the metaphor, the searcher is the one doing the mining for information, in the form of relevant articles and similar resources. At the same time, it is tempting to regard the searcher as a prospector looking for a specific kind of information. However, prospecting in the modern world has usually involved finding likely locations to set up industrial-style mining. The prospecting comes first, and then the mine. Our best bet for maintaining the prospecting parallel might be to consider the prospectors of the 19th century gold rushes, or even the individual and small teams profiled in the National Geographic show “Prospectors,” who look for those nuggets and veins and also extract them from the earth without setting up a major mining operation.

In terms of general action, prospecting might be our closest parallel to search, although not a very smooth parallel.

Extraction

In the mining industry, in the extraction of mineral resources from the mine, heavy machinery is often involved to perform the drilling and excavation. On the other hand, there are those individuals and small teams mentioned above who might use small, portable hand tools for extraction. In the context of a digital search platform using a taxonomy, we might repeat the comment of one emailer who boasted that no trees were harmed in the sending of his information, although a large number of electrons were greatly inconvenienced.

Mining also involves another kind of extraction: the extraction of mineral resources from the ore once the rock that may include the desired ore has been taken to the processing location. For the information searcher, this extraction may include an inline tagging function based on a taxonomy, with relevant text passages highlighted or otherwise indicated (see Figure 1).

The information searcher has chosen a particular database according to the fields it covers, knowing that there is a good chance that somewhere in the database there is likely to be information that will be relevant to his or her research. The database makes a large amount of varied information accessible by containing the information equivalent of a huge pile of rocks of the same geologic profile. From the perspective of the mining industry, these rocks would have already been mined.
Sifting, Smelting, Filtering

Now the challenge for the mining company or the Gold Rush ‘49er type of prospector is to smelt or sift, matching the final small, manageable and valuable pile to a predetermined set of characteristics. The information searcher does the same kind of thing, using a platform that matches individual articles to the concepts that the searcher has specified using the taxonomy. With an associated indexing rule base and/or a well-developed set of synonyms, the text in the articles does not need to be an exact match to the terms in the taxonomy. The matches can be to a wide variety of criteria, all used as indicators that the text in the matched articles has the desired “aboutness.”

Refining

When a search platform is associated with a taxonomy, the searcher can take advantage of the hierarchical structure to refine a search using narrower terms (Figure 2). (On the other hand, you can also expand your search using related terms. Would that be like alchemy?)

The Taxonomy as Processor

The taxonomy performs a dual role in the typical search/research platform. It can be used in connection with search, largely because it can be used in indexing. It treats the ore of articles to make them accessible, findable and identifiable.

The Taxonomy as Guide to Additional Resources

With a well-constructed hierarchical structure and well-chosen associative relationships, the taxonomy can lead the searcher or miner to other categories of interest, just as the presence of a particular mineral can suggest where a mining company might look to find another valuable mineral with which it is often found.

Customers

Taking the analogy a bit further, the searcher is the customer, mining for the goods that he or she wants or needs. This direct action is in contrast to the economics of the mining industry, in which the company mines for raw materials. Ultimately, who are the customers? The readers of the report? The beneficiaries of research? In the case of medical research, is it the doctors who acquire information that they can use for diagnosis or treatment? Or is it the patients?

The Big Picture: World Economics

Mining provides the world with valuable resources, otherwise unobtainable, for energy and manufacturing. The taxonomy and database tandem in the research platforms of numerous institutions, associations and other scholarly enterprise is the best means we have for researchers to build on the work of other researchers (and to make their own research available) to achieve advances in technology, medicine and other areas of human endeavor.

For Further Reading