



Artificial Intelligence

As an instrument for organizing large quantities of information or performing extremely complex symbolic operations beyond human capabilities within a normal life span, the computer is an invaluable adjunct to the brain, though not a substitute for it. —Lewis Mumford

There is a new technology emerging in the litigation support industry, referred to as “artificial intelligence.” Not artificial intelligence in the true sense of the word, but artificial in the sense that this new technology purports to “read” and “understand” documents much as we humans do. Essentially, what this technology provides is the ability to perform contextual searching. Search results are provided in ranked order based on the probability that a document contains the query term or is related to the query. This technology does not replace humans by *any* means; it is a new tool that can be used quite effectively when trying to get a handle on a large amount of data very quickly. It has its pros and cons just like any other technology.

Pros

The best use of this technology is in managing large amounts of data. With that in mind, here are the reasons you might consider using it:

Quick access. Electronic capture of data is faster and more efficient than manual coding, allowing for quicker access to relevant information.

No variance. Since documents are processed electronically, there is no variation in the way documents are interpreted.

Synonymous searches. There are multiple ways to say the same thing. The ability to conduct context searches helps because words with similar meaning are likely to be used in similar contexts. This is especially important where there may be misspelled words in documents.

Saves money. This is a broad generalization, but processing data electronically can be, in most cases, less costly than manual labor.

Cons

OCR inaccuracies. The most significant issue in dealing with this technology is the fact that the documents have to be OCR'd (unless you are dealing with electronic data). OCR, at best, is anywhere from 80% – 95% accurate on the documents that it *can* actually read. For argument's sake, let's say OCR is 95% accurate. What this means is that for every 100 words, five of them are wrong. One of those five could have been a key term. In addition, there are many types of documents that cannot be read by OCR, such as handwritten notes, graphics, and photographs. You may have to incur additional costs for manipulating this data.

Excess information. Because of the technology involved, you get a lot of extra information in your database and in your search results. From a searching standpoint, those in the industry will talk about recall vs. precision. What this means is that these services will recall all data about a particular query term, whether it's relevant or not. It will rank the results from most relevant to not relevant, but you still have to look at the documents that are not relevant because they came up in the query for some reason. The precision aspect of searching requires a human to filter and refine the data in order to identify the truly relevant documents.

Too many “others.” Again, based on the technology involved, the documents that these services cannot read or classify (through the OCR process) end up going into a category of “text” or “other.” It does identify for you what it cannot process, but you still have to manually go through those documents to classify or identify them further. This can take a lot of man-hours to accomplish, depending on the size and complexity of the case.

Which date? There is no way a computer can make a judgment on what date may or may not be the important date to pick up on a document with multiple dates. A computer cannot determine whether there was a typographical error in typing the date on a document, nor can it assign a date to a document that has no date.

Syngence and DolphinSearch™

There are other companies that offer artificial intelligence services, but this article focuses on the services offered by the two most predominant names today: **Syngence** and **DolphinSearch**. These services have several things in common, as illustrated in the chart below.

Feature	Syngence	DolphinSearch
Handles Electronic Data	Y	Y
Provides OCR Services	Y	N
Contextual Searching	Y	Y
Boolean Searching	Y	Y
Ranks Search Results	Y	Y
Web Access	Y	Y
Hardware/Software	No special requirements	No special requirements
Works with mainstream document management software	Y	Y
Organize documents in folders	Y	Y
Issue code or "tag"	Yes, but only through iConnect	Y

But there are also some distinct differences between these two services:

Syngence

Technology Overview. Syngence technology is based, generally and in a nutshell, on Bayesian mathematical theories of probability (as is DolphinSearch to some degree). Syngence offers two services: **SynDex™** and **Synthetix™**. Their SynDex technology "reads" documents and creates a bibliographic index with the most commonly requested fields (such as author, recipient, bates number, document date, and document type.) In addition, SynDex recognizes and indexes other information that appears in the text of the document, such as other names, dates, subject and pre-determined key words. Synthetix is a search engine that locates related documents and identifies duplicates and drafts. Synthetix also offers the ability to create a "synthetic" document to use in a search to find others like it. Check them out at www.syngence.com

DolphinSearch™

Technology Overview. DolphinSearch is a text-reading robot powered by a neural network computer model of the pattern recognition capabilities of a dolphin's biological sonar system. The same kind of relational pattern recognition mechanisms that model how a dolphin recognizes objects is used to recognize the meaning of text. Just as a sound wave changes its meaning depending on the context of other waves around it, words change their meaning depending on the context of other words around them. This technology was developed with electronic discovery in mind, but it also works on hard copy documents that have been OCR'd. Check them out at www.dolphinsearch.com.

Summary

The new technology of artificial intelligence is pretty amazing, and can be a valuable tool. Use it wisely. Know its best uses and limitations. Ask your vendors to provide you with a sample of their services; that way, you know exactly what you're getting. And remember....

...however extraordinary computers may be, we are still ahead of them and man is still the most extraordinary computer of all.

—John F. Kennedy

Litigation Information Management

Your support professionals.

Please call us for more information
about our coding, imaging, repository,
and consulting services.

