



## Back to the Future With Hypertext

It's true what they say: If you stick around long enough, everything comes back full circle. I remember the early days of hypertext and the emergence of the World Wide Web in the late 1980s and early 1990s. Then, links were the big breakthrough. The content of the linked information was irrelevant; it could be a document, an archive, a library's catalog, or a webcam displaying the office coffee machine ([en.wikipedia.org/wiki/Trojan\\_Room\\_coffee\\_pot](http://en.wikipedia.org/wiki/Trojan_Room_coffee_pot)). What was important was creating links to enable others to find relevant content. During that same time, value-added online databases still consisted primarily of bibliographic citations and abstracts. What we were searching was not the content of the underlying material but links or citations to that material.

Fast-forward a decade, and we had search engines designed to index all the content within webpages and online database that were chock-full of full-text articles. All of a sudden, metadata and links had become passé. "Who needs indexing when everything's full text?" became a valid question.

Today, with the emergence of text and data mining (TDM), we are seeing that full text isn't necessarily the richest form of content anymore. The questions we can ask and what the answers look like have changed radically. Now, what's important isn't just the information within individual records, but the relationship among those pieces of information.

TDM involves extracting facts from published information and formatting the resulting linked relationships consistently across datasets. An article might have links indicating that Article X describes Disease A; research in Article X was funded by Grant 123; Grant 123 was from Institution Z; and Institution Z is located in ABC. A Wikipedia article on Disease A may have links indicating that Disease A is also referred to as Disease A-1; Conference X covers Disease A; and Treatment Procedure Z is recommended for Disease A. Combine enough of these pieces of information from enough datasets, and you can discover trends about the research landscape that would never be discernable by simply reviewing results from a web or online search. The value of TDM, however, depends on knowing what sources to include and what kinds of connections to monitor.

In a sense, TDM is like a really smart, well-connected researcher. She regularly monitors professional journals, conference proceedings, books, videos, reports, patents, and other material in her field for unexpected findings. She participates in professional conferences, where she meets other researchers in related fields and learns about their current projects. She

monitors organizations and funding sources in her field and analyzes grant patterns. Because she is familiar with information from a wide range of sources, she can see trends and relationships among concepts that would not be obvious to the casual observer. She knows, for example, to watch for new developments from South Korea, based on a conference presentation she heard and a recent uptick she noticed in grants to universities there.

Imagine exponentially expanding that researcher's perspective to include all the information available in her field. And imagine a similar superhuman researcher for every imaginable field of inquiry. That is what TDM initiatives offer to an organization, and, as with any information and knowledge management project, information professionals can play a key role. Info pros know which resources to use, weighing the limitations, restrictions, and cost of each source. We understand how researchers use information—how they approach a problem, how they search for information, what sources they use, and what they do with the information next. Our perspective can be invaluable.

However, one of our biggest challenges is the reconception of what "search results" look like in the TDM world. We usually approach a search with the expectation that we will be iteratively narrowing or expanding our query, reviewing lists of retrieved results, and ultimately providing our clients with either summaries or the full text of articles and other content. However, TDM tools and platforms such as JSTOR's Data for Research ([jstor.org/dfr](http://jstor.org/dfr)), DBpedia's dataset of Wikipedia knowledge ([wiki.dbpedia.org](http://wiki.dbpedia.org)), or Springer Nature's SciGraph ([springernature.com/gp/researchers/scigraph](http://springernature.com/gp/researchers/scigraph)) require info pros to see deliverables in a new light. The ideal output from a TDM project might turn out to be a list of article citations, but it might also be a graphic that shows relationships among the authors citing an organization's patents, or a map of the hotbeds of research in a new field. Whereas we often have a fairly good idea of what we are looking for or expecting to find when conducting a traditional bibliographic search, all bets are off when we bring TDM into the mix. A spirit of exploration and adventure will help as we adjust to a new mindset in which "merely" finding the best article on a topic is no longer enough.

---

*Mary Ellen Bates* ([mbates@BatesInfo.com](mailto:mbates@BatesInfo.com), *Reluctant-Entrepreneur.com*) is still waiting for the search interface Tom Cruise used in *Minority Report*.

Comments? Email the editor-in-chief ([marydee@xmission.com](mailto:marydee@xmission.com)).