by Lynn Dirk

Mapping Knowledge Domains

> The term 'mapping knowledge domains' [describes] a newly evolving interdisciplinary area of science aimed at the process of charting, mining, analyzing, sorting, enabling navigation of, and displaying knowledge. . . . Although thousands of years old, this area has undergone a sea change in the last 15 years, a change fostered by an explosion of the amount of information available, the accessibility of that information due to electronic storage, and the new techniques of analysis, retrieval, and visualization that are made possible by vast increases in computational storage capacity and processing speed and power.

So begins a fascinating look at a new field of science explored in a 127-page supplement of the *Proceedings of the National Academy of Sciences of the United States of America* (*PNAS* 6 April 2004; 101, suppl 1; pp 5183-310, available at at *www.pnas.org*) presenting information from the Arthur M Sackler Colloquium on Mapping Knowledge Domains, which was held in Irvine, California, on 9-11 May 2003. Not surprisingly, the "father" of this field is the originator of citation analysis, Eugene Garfield, who worked closely with the meeting organizers and gave the keynote address, which, unfortunately, is not included in the supplement.

The 19 articles in this *PNAS* supplement illustrate how far methods like citation analysis have evolved beyond identifying highly cited papers and journal impact. The introduction to the colloquium papers is inspirational in its depiction of the potential to mine huge, noisy databases, including "the body of scientific literature", for meaningful and useful information now lost to us because of its volume and density. With new techniques and new goals, the many journal articles that are never cited may at last find use, for example, providing data points on a cartographic presentation—a literal map—of a knowledge domain. Below is a sample of article titles from the *PNAS* supplement and a taste of what the articles contain.

- Coauthorship Networks and Patterns of Scientific Collaboration: "In biology . . . it is far less likely than in mathematics that two of one's coauthors will also be coauthors of one another, a result that has yet to receive a clear explanation."
- Finding Scientific Topics: "A scientific paper can deal with multiple topics, and the words that appear in that paper reflect the particular set of topics it addresses. In statistical natural language processing, one common way of modeling the contributions of different topics to a document is to treat each topic as a probability distribution over words, viewing a document as a probabilistic mixture of these topics."
- Evolution of Document Networks: "Although text analysis has been used for a long time to analyze documents, extract their meaning, retrieve information, and map knowledge domains, link analysis is increasingly used by search engines and digital libraries to estimate the importance or reputation of documents and to map documents into topical clusters."
- The World of Geography: Visualizing a Knowledge Domain with Cartographic Means: "[c]artographers have begun to attempt [to combine] centuries of accumulated cartographic knowledge with modern computational approaches and cognitive insights, toward the visualization of nongeographic information . . . [such as] text document corpi [*sic*] held in digital libraries, user interaction logs created by Web applications, or biological data associated with genome mapping."
- Searching for Intellectual Turning Points: Progressive Knowledge Domain Visualization: "A good example of visualizing thematic changes in a collection of text documents [is] the metaphor of a thematic river to depict temporal changes of word frequencies. An intensified theme can be identified if one can detect increasingly widened word frequency streams."