

Position Paper by Kevin W. Boyack (SNL) and Richard Klavans (SST)

For a session titled *Mapping Humanity's Knowledge and Expertise in the Digital Domain*, to be held at the Annual Meeting of the Association of American Geographers (AAG), Denver, CO, April 5-9, 2005

Organized by Katy Börner & André Skupin

Part 1 – This session

1.1. Describe your main interest in this session.

Our main interest in this session is to learn the current state of the art in 1) large scale knowledge mapping, and 2) the interface between the fields of “knowledge domain mapping” and “visualization, geographic visualization, cartography” with a view towards interface design.

1.2. Which major technical challenges do you see for *Mapping Humanity's Knowledge and Expertise in the Digital Domain*, as laid out in the session description?

We see the following as near-term technical challenges related to large-scale mapping:

- *Cleaning of dirty data (most real data is dirty),*
- *Effective linking of textual data,*
- *Efficient algorithms for calculation of similarity,*
- *Algorithms to extract and summarize text for purposes of labeling maps.*

We no longer see graph layout or clustering for large maps (i.e. >1 million nodes) as major technical challenges.

1.3. Which major non-technical challenges do you foresee?

We foresee the decision about what types of data to use, and the acquisition of appropriate data for large scale maps of science or knowledge as key non-technical challenges.

1.4. Which major opportunities do you envision?

We see the long-term opportunity to understand the innovation process, which will allow innovation to be more effectively managed in the future.

Part 2 – Your research

2.1. Project Name and Web Address

Our current project is titled “A Method of Evaluating Research Using New Innovation, Risk, and Impact Indicators.” In brief, this project aims to create maps of current science with indicators of innovation, risk, and impact at the single paper level. This will enable researchers to locate and view indicators associated with their current or proposed work, and enable institutional comparisons and identification of opportunity and vulnerability. To date, we have created maps of science at the journal level (~7000 journals) for the purpose of selecting, testing, refining, and validating the algorithms that we will use to generate paper-level maps of science. We have also created our first paper-level map of science that locates 833k papers from the 2002 Science (SCI) and Social Science Citation Indexes (SSCI). The papers cluster into nearly 100k separate

These two maps are both from manuscripts that are in preparation for publication. Please do not distribute further.

2.8. Pros and Cons

Pro: Our maps are based on 'all of science' and thus do not suffer effects normally associated with thresholding, or differences in citation practices between disciplines.

Con: As with all multidimensional data, some records will not be accurately placed during reduction to 2-D maps. Some papers end up in clusters where they do not belong.

2.9. Planned Work

By the time of the Spring meeting we will have added ISI's Conference Proceedings Database, and possibly US patents to our maps. This will add a greater technology component to the maps, which is very important to institutions like Sandia. We also plan to calculate maps for additional years, link maps from multiple years, and develop an interface to allow staff at Sandia to use the data to improve their own research.

2.10. Major Publications

[1] Klavans, R., & Boyack, K. W. (2004). Identifying a better measure of relatedness for mapping the backbone of science. Submitted to Journal of the American Society for Information Science and Technology 08/04.

[2] Boyack, K. W., Klavans, R. & Börner, K. (in preparation). Mapping the backbone of science.

[3] Boyack, K. W. (2004). Mapping knowledge domains: Characterizing PNAS. Proceedings of the National Academy of Sciences, 101, 5192-5199.

[4] Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. Annual Review of Information Science and Technology 37, 179-255.

[5] Boyack, K. W., & Börner, K. (2003). Indicator-assisted evaluation and funding of research: Visualizing the influence of grants on the number and quality of research papers. Journal of the American Society for Information Science and Technology, 54(5), 447-461.

[6] Boyack, K. W., Wylie, B. N., & Davidson, G. S. (2002). Domain visualization using VxInsight for science and technology management. Journal of the American Society for Information Science and Technology, 53(9), 764-774.