Metrics for Accessing Heterogeneous Data: Is There Any Hope?

Len Seligman
The MITRE Corporation

Panelists:
Nicholas Belkin, Rutgers University
Erich Neuhold, GMD
Michael Stonebraker, University of California
Gio Wiederhold, Stanford University

Government and industry are investing substantial resources in new technologies for accessing heterogeneous information sources, including text-based corpora, structured data, imagery, geo-spatial data, audio, video, and more.

Managers of programs that fund relevant research face a difficult problem: they are required to justify investment in certain technologies and approaches versus alternate ones. These program managers recognize a need for good evaluation criteria, but there is little consensus on which criteria to use. Metrics are required to help evaluate the contribution of alternate techniques to satisfying the ultimate goal: putting useful information into the hands of users.

A number of metrics have been proposed, including the following:

- Relevance-based measures, including *precision*, the percentage of retrieved information nodes that are deemed relevant, and *recall*, the percentage of all relevant nodes in the entire information space that are actually retrieved.
- Utility-based measures, which measure the "value-added" provided by information returned by a search. Variants include *informativeness* (Tague-Sutcliffe), which measures the extent to which an actual search trail corresponds to some ideal answer trail, and the *overall value* of the information retrieved.
- Cost-based measures, including *user interaction time* and *monetary* cost.
- User satisfaction

Proceedings of the 21st VLDB Conference
Zurich, Switzerland, 1995

The above metrics all measure (either directly or indirectly) the performance of a system in supporting information retrieval. Other dimensions for evaluation include *breadth of coverage* (i.e., applicability across heterogeneous data types and diverse applications), and *life-cycle costs* (e.g., how well suited is the approach to long-term system evolution?).

This panel will address the issue of appropriate evaluation criteria for technologies supporting heterogeneous data access. Panelists will describe specific systems to which they have contributed that perform information retrieval across heterogeneous data sources. They will then describe how they have measured the quality of those systems in terms of the metrics described above. Panelists will then discuss these issues:

- To what extent can we measure the contribution of a technology or approach to the overall process of satisfying a user's information need, from query formulation and refinement through information merging and presentation to the user?
- Which (if any) of the metrics described above are most useful?
- Are there others which are more meaningful or which can be applied more cost-effectively?
- What is the role of empirical evaluation? Can meaningful experiments be conducted at reasonable cost? What are the challenges in doing this?
- How should funding agencies evaluate work in this area?

An important goal of this panel is to encourage discussion of evaluation criteria within the research community. We as a community must consider how we would like to be evaluated and then communicate our ideas to funding sources. It is hoped that this panel will be a first step in making this happen.