Information Cartography:
A Proposed Model for Access to Heterogeneous End-User Databases

Stephen Paling
School of Information Studies
Syracuse University
4-206 Center for Science and Technology
Syracuse, NY 13244-4100
phone: 315-469-2858
fax: 315-443-5806
e-mail: swpaling@mailbox.syr.edu

Maps are among our best information systems. They require little documentation and are commonly used and understood. In contrast, many systems of classification seem to lack this acceptance and ease of use. Anecdotal evidence suggests that this is particularly true of the way collections of databases are classified for online browsing on library Web sites. This paper argues that some of the characteristics that make maps easily usable can be applied to collections of databases. Those characteristics include logical grouping of information, the ability to move smoothly between levels of data, and consistent amounts of data at different levels of representation.

This paper will take a conceptual approach in discussing factors that make maps easy-to-use and readily acceptable, and will sketch some of the implications those factors might have for the classification of online databases. It will start with a description of several cartographic phenomena and their utility. After that it will discuss how these phenomena might be applied to information systems, and will finish with a discussion of what such a cartographically inspired classification system might look like and how it might be built.

Maps create well-formed expectations. We are familiar enough with terms like "road map" and "floor plan" to know what their contents are even before we see them. Bibliographic descriptions carry similar expectations. Anyone familiar with libraries can quickly decipher a bibliographic description for an article, book, etc. Our classificatory descriptions of databases, though, do not achieve this. They tend to feature free-text descriptions with neither consistent elements nor formatting. The elements that users are likely to refer to in distinguishing between databases (topic, scope, features, etc.) are generally described in non-standard language. This kind of classificatory standardization could help reproduce some of the well-formed expectations that maps and bibliographic descriptions engender.

Maps feature pan and zoom, two traits that contribute to their usability. Pan represents the ability to scan across the surface of a map (paper or electronic) to see what features lie next to each other. Zoom allows the user to view selected parts of the map in greater detail. These traits are reproduced inconsistently in collections of online databases. When a user searches in a database, or views its description, the system often does not give the user an effective way to pan to a logically adjacent database or description. Collections of databases do provide some zoom capabilities by allowing the user to move from a list of database descriptions to a single database or small group, and then zoom in further to particular documents. Looking at the way maps (particularly electronic maps) provide this functionality, though, points up both problems and

possible solutions in the way that we classify databases, and with what those classifications afford to users.

"Constant information density" is another key trait of maps that can help elucidate the organization of online databases. As a user zooms from one level of a map to another, they should see a relatively constant amount of data. For example, as they zoom in on a city map, the streets may become less densely packed on the map, but more of the street names and other features will appear. Database descriptions, though, typically demonstrate a relatively poor level of detail compared to the rich descriptions of documents occurring within the databases. A user who has not found what they need in one database sees a precipitous drop in the level of detail when they zoom back out to a list of database descriptions.

The paper proposes Frames of Reference (FoR), a hypothetical system that would address the issues raised here. FoR would use frames in conventional Web browsers to help users maintain the original context of their search. An important part of this context involves the way databases are classified to reveal their relationships with each other. We know from both research and anecdotal evidence that user needs often center around orientation within a search process rather than in specific elements of the search itself. FoR proposes keeping links to logically adjacent, i.e., similarly classified, databases available to the user at all times. This would be similar to seeing the rest of a map on the periphery even while concentrating on a specific part of it.

The final part of the paper discusses how emerging tools like the Extensible Markup Language (XML) can be used to implement systems like FoR that could reproduce pan, zoom, and constant information density in our classification of online databases. In contrast to HTML, which is a page-layout language, XML allows semantic information to be embedded into Web pages. Such information would allow the addition of standard classificatory descriptions of databases. Logical groupings of databases could be built around these descriptions, and even generated dynamically to suit the context of a search.

Questions:

1. What would be the best way to operationalize and test constant information density?

2. If the paper were to be split into multiple pieces, what parts would most merit development?

3. Given that databases can have very different structures, what common features suggest themselves for a classification scheme?