

INTERFACING THE SAND SPATIAL BROWSER WITH FEDSTATS

Hanan Samet
Frantisek Brabec
Gisli R. Hjaltason

Computer Science Department
Center for Automation Research
Institute for Advanced Computer Studies
University of Maryland
College Park, Maryland 20742

e-mail: hjs@cs.umd.edu
url: www.cs.umd.edu/~hjs

Copyright © 2001 Hanan Samet

These notes may not be reproduced by any means
(mechanical or electronic or any other) without the
express written permission of Hanan Samet

SAND

- GIS developed at the University of Maryland to deal with spatial and nonspatial data
- Adopts extended relational model
- SAND Browser provides a graphical user interface to facilities of SAND
- Supports ranking: enables finding objects in order of their proximity to other objects
- Tuples satisfying query are obtained in incremental order
- No need to wait long to get visual feedback

FEDSTATS

1. Statistical data of numerous federal agencies is on the web
2. Includes forecasts, projections, statistical tabulations, surveys, etc.
3. Data can be retrieved by topic, responsible agency, keywords, press release, etc.
4. Limited capability of retrieval of data by county or state
 - select a name from a menu, OR
 - click on desired area on map
5. Propose to add more power to users of FedStats by enabling users to:
 - define query area with more specificity by drawing the relevant objects or boundaries
 - retrieve data in order of distance from user-defined regions or user-defined data aggregates

REQUIREMENTS FOR INTERFACING FEDSTATS WITH THE SAND BROWSER

1. Make the SAND Spatial Browser available over the internet
2. Alternatively, provide a down loadable version of the SAND BROWSER
3. Enable users to import data into system
 - convert data to SAND format (e.g., from EXCEL)
 - specify attribute types
 - specify which attributes need an index

EXAMPLE

1. EXCEL files for EPA-regulated facilities that have Arsenic and Chlorine
2. EPA-ID, name, street, city state, zip code, latitude, longitude, and flags to indicate the participating programs
 - hazardous waste, wastewater discharge, air emissions, abandoned toxic waste dump, active toxic release
 - each program in a separate column in EXCEL file

SAND RELATION

- Spatial attribute 'location' indexed by point PMR quadtree
- Attribute tuple-id to make tuples unique
- How to implement 'program'
 1. names of programs
 - drawback: not first normal form
 2. five Boolean attributes 'is-program'
 - drawback: users must formulate queries in terms of the construct and its Boolean values
 - specification of Boolean value is not unique
 3. separate tuple for each program for the facility
 - OK if just a few programs per facility
 - drawback: facility is retrieved as many times as there are programs in which it participates (duplicate problem)
 - use 'group-by-attribute-name' (EPA-ID) to retrieve matching tuples simultaneously
 - not advisable to use the 'name' attribute for the grouping as there could be several facilities with the same name that differ by their location

SAMPLE QUERIES

1. Find all EPA-regulated facilities that have Arsenic that participate in the "Air Emissions" program in states from Georgia to Illinois, alphabetically
2. Find all EPA-regulated facilities that have Chlorine that participate in the "Air Emissions" program that lie within the state of Arkansas or 30 miles within its border
3. Find all EPA-regulated facilities that have Chlorine that participate in the "Air Emissions" program that lie within 30 miles of the border of Arkansas (i.e., both inside and outside Arkansas)
4. For each EPA-regulated facility that has Arsenic, find all EPA-regulated facilities that have Chlorine that are closer to it than to any other EPA-regulated facility that has Arsenic
 - use 'group by EPA-ID' mechanism to report each facility just once
 - like a discrete Voronoi diagram where the sites are the EPA-regulated facilities that have Chlorine