Geophysical Research Abstracts,
Vol. 11, EGU2009-6620, 2009
EGU General Assembly 2009

Recommendation Systems for Geoscience Data Portals Built by Analyzing Usage Patterns

C. Crosby, V. Nandigam, and C. Baru
University of California, San Diego, San Diego Supercomputer Center, La Jolla, United States (ccrosby@sdsc.edu)

Since its launch five years ago, the National Science Foundation-funded GEON Project (www.geongrid.org) has been providing access to a variety of geoscience data sets such as geologic maps and other geographic information system (GIS)-oriented data, paleontologic databases, gravity and magnetics data and LiDAR topography via its online portal interface. In addition to data, the GEON Portal also provides web-based tools and other resources that enable users to process and interact with data. Examples of these tools include functions to dynamically map and integrate GIS data, compute synthetic seismograms, and to produce custom digital elevation models (DEMs) with user defined parameters such as resolution.

The GEON portal built on the Gridsphere-portal framework allows us to capture user interaction with the system. In addition to the site access statistics captured by tools like Google Analystics which capture hits per unit time, search key words, operating systems, browsers, and referring sites, we also record additional statistics such as which data sets are being downloaded and in what formats, processing parameters, and navigation pathways through the portal. With over four years of data now available from the GEON Portal, this record of usage is a rich resource for exploring how earth scientists discover and utilize online data sets. Furthermore, we propose that this data could ultimately be harnessed to optimize the way users interact with the data portal, design intelligent processing and data management systems, and to make recommendations on algorithm settings and other available relevant data.

The paradigm of integrating popular and commonly used patterns to make recommendations to a user is well established in the world of e-commerce where users receive suggestions on books, music and other products that they may find interesting based on their website browsing and purchasing history, as well as the patterns of fellow users who have made similar selections. However, this paradigm has not yet been explored for geoscience data portals.

In this presentation we will present an initial analysis of user interaction and access statistics for the GEON OpenTopography LiDAR data distribution and processing system to illustrate what they reveal about user's spatial and temporal data access patterns, data processing parameter selections, and pathways through the data portal. We also demonstrate what these usage statistics can illustrate about aspects of the data sets that are of greatest interest. Finally, we explore how these usage statistics could be used to improve the user's experience in the data portal and to optimize how data access interfaces and tools are designed and implemented.