

CalSWIM Project: Progress Report

Water Resources Data Exploration Tool

As part of the CalSWIM project, we have developed an online tool that aggregates data from several data sources and allows users to browse the data visually on a map. This tool is publicly available on <http://nile.ics.uci.edu/calswim/>

Currently, we have included seven data sources. Data from other sources that publish their data publicly can be included in the future. The following is some of the data that is currently included:

- **Wikipedia water-related articles:** we have processed Wikipedia articles and have extracted those that are related to water. This includes articles about watersheds, rivers, aquifers, etc. Some of these articles also include the geographic coordinates of the concept they are describing. We have also extracted this information and this allows us to show Wikipedia's water related articles on a map.
- **JCOMMOPS data:** this dataset provides coordination at the international level for oceanographic and marine observations from drifting buoys, moored buoys in the high seas, ships of opportunity and sub-surface profiling floats.
- **Global Argo Data:** Argo is a worldwide network of free-drifting ocean profiling floats.
- **EPA STORET data:** STORET is a repository for water quality, biological, and physical data and is used by state environmental agencies, EPA and other federal agencies, universities, private citizens, and many others.

Measuring Reliability of Wikipedia Content

In this research, we have modeled user reputation in wiki systems. We have presented three reputation models and have tested the three models on English Wikipedia which is the largest online encyclopedia with an open editing model, allowing anyone to enter and edit content. Our experiments show that the three models can accurately assign reputation values to Wikipedia's users. Since wikis store history pages, the proposed models can be used in any intra-company or intra-organization wiki or in public wikis such as Wikipedia, Citizendium, and Scholarpedia. The proposed models can be used in wikis for rating users or as a decision support system for administrators. For example, they can be used for automatic vandalism detection, saving substantial amounts of time for wiki administrators. They can be integrated also in a quality assessment system that assesses the reliability of the content according to the reputation of its contributors. In the future work, we would like to analyze evolution of the content in wikis as another metric for quality assessment.

Highlighted research (among several):

"S. Javanmardi, C. Lopes and P. Baldi, *Modeling User Reputation in Wikipedia*, accepted for publication in *Journal of Statistical Analysis and Data Mining*, Winter, 2010".