CS410 - Project Proposal

Team Members: Zutao Yang - Coordinator (zutaoy2@illinois.edu), Nathan Nard (nnard2@illinois.edu), Shaun Phillips (shaunap2@illinois.edu)

Mapping Global Dams Through Text Searching/Mining

Background:

Despite the recognized importance of dams to local and regional social and natural systems, a global dataset describing their characteristics and geospatial distribution is incomplete. This is mostly because many countries still lack an inventory survey of their respective dams. Open sourced texts available on the public internet include a lot of geolocation information of dams, especially through tagged texts on geo-spatial platforms, such as Google Maps, GeoNames, and Open Street Maps. Here we propose to map the global distribution of dams through text searching/mining in the internet.

Main Goal

Generate a searchable database of global dams along with the important properties of the dams, and map them spatially.

Technique Outline:

1. Use web crawling to construct a database of global dams. We plan to include dam features in at least three major geospatial database/platforms that have dams tagged, i.e., OpenStreetMap, Google Maps, and GeoNames. This will allow us to have a database of dams with at least names and geolocation (i.e., latitude and longitude). This may include text search and filtering.

2. Applying text search and mining to the resulting database to find the most related document/webpages on the internet to each dam record, and find other important properties as much as possible. Other properties include construction year, main function, and water capacities (if it is used for building reservoir), and total electricity capacity (if it is used for producing hydropower).

Timelines:

Week 9: Group meeting to discuss tasks
Week 10 and 11: Database of dam names/location ready
Week 12 and 13: Coding for retrieving other properties of dam
Week 14: Code and presentation delivery