Midwest Big Data Hackathon

October 8-9, 2016
Iowa City IA, USA
bigdata.uiowa.edu

Datasets + Tools

- Twitter REST API's - Provide programmatic access to read and write Twitter data.
- Yelp - 2.7M reviews and 649K tips by Yelp's users and user businesses.
- BestBuy Open Product Data – Programmatic access to BestBuy's product and store data.
- The New York Times API's – Programmatic access to The New York Times' content. It includes articles, books, events etc.
- The Guardian API's - Programmatic access to The Guardian's content.
- GDELT Database - Most comprehensive database of human society.
- MIT Human Dynamics Lab Datasets – Several mobile data sets that contain dynamics of several communities of about 100 people each.
- Million Song Dataset - Collection of audio features and metadata for a million contemporary popular music tracks
- Wikipedia Dumps – Wikipedia database dumps in any language.

- [Google Books Ngram Dataset](#) – The dataset that used in Google Books Ngram Viewer.
- [International Cancer Genome Consortium](#) – Dataset that includes information about cancer patients and their genomes.
- [GeoNames Data](#) - Geographical database covers all countries and contains over eleven million place names.
- [NOAA Ocean CO_2 Dataset](#) – More than 30 years worth of surface ocean CO2 observations.
- [NOAA Climate Datasets](#) – Datasets that are provided by NOAA on climate.
- [The Movie DB API](#) – Allows programmatic access to TMDb that includes information about movies, tv shows, actors and actresses.
- [US Patent Citation Network](#) – Consists of all the utility patents granted between January 1, 1963 and December 30, 1999. totaling 3,923,922 patents
- [HAZUS](#) - Consists of hazard data, boundary map data, and a proxy for the general building stock (GBS) in the continental United States, Hawaii and the U.S. held Territories.
- [ACT](#) – Test events dataset

Project Ideas

PROJECT: LEEPFROG DATA VISUALIZATION CHALLENGE

Brief explanation:

Leepfrog is expanding its data visualization to have a formal data visualization position. Attached is a sample data file, "majorcourses.unl". The columns are (coursecode, majorcode, majorname); this is CourseLeaf's "course ecosystem" and is used to power the "Courses in a Concentration" option at: https://cab.brown.edu

It's fairly typical of the types of data we create visualization tools for: items and their relationships to each other. The project is to come up with a way to visualize the courses and the majors that reference that course. For example, MATH 0100 is used by 61 different majors at Brown and it would be helpful for the MATH chair to be able to see and understand those programs.

Expected results:

- Our primary visualization library is D3 (https://d3js.org/), but you can use a library of your choosing.
- The project should be lightweight.

Knowledge Prerequisite:

- No prerequisites

Attachments: Click here.

Technical Staff: Deb Walton / Leepfrog Technologies, Inc.

PROJECT: PROTOTYPE BIBFRAME 2.0 MODELING FOR LIBRARY INFO SPOTLIGHT "OPERA LAND"

Brief explanation:

Library linked data promises to meet libraries' need for agility in content delivery and user engagement. This project will demonstrate the prototype of BIBFRAME 2.0 modeling for work, instance, item, agent, topic, etc. from the local bibliographic records in Alma and external data sources, representing library info spotlight of operas in Opera Land, a collection of opera books,

videos, sound recordings, streaming media, etc. interwoven into user's online experience using Alma, LC MARC to BIBFRAME Transformation Service, RDF Translator, RDF Validator, RDFa Validator, Apache Jena Fuseki Server, etc.

The initial data set consists of six library bibliographic data in linked data model using BIBFRAME 2.0, representing an opera titled "The Marriage of Figaro." The sample files are in RDF/XML, JSONLD, RDFA, and N3 formats. More info about BIBFRAME 2.0 can be found in the overview of the LC BIBFRAME 2.0 Model, available from http://www.loc.gov/bibframe/docs/bibframe2-model.html

Expected results:

- Make library info spotlight of operas into linked data model using BIBFRAME 2.0 and extended vocabularies, visible to the web, and easily engaged with library users

Knowledge Prerequisite:

- Proficiency with programming languages such as PHP, Ruby, Python, jQuery, HTML, JavaScript and CSS
- Knowledge of triple stores, SPARQL, and RDF
- Experience performing data transfers utilizing software library or language APIs
- Working knowledge of Apache, Amazon Web Services, etc.

Attachments: Detailed Project Description

Mentor: Amanda Xu / University of Iowa Libraries

PROJECT: UNDERSTANDING STRUCTURAL AND GEO-SEMANTIC PATTERNS OF INTERPERSONAL COMMUNICATION ON TWITTER

Brief explanation:

Advances in human dynamics research and availability of geo-tagged communication flows provide an unprecedented opportunity for studying the interplay between online social communication, and geography. Among the most extensively studied topics in geographically embedded communication networks, are the effect of geographic proximity on interpersonal communication; the influence of information diffusion and social networks on real-world geographic events, such as demonstrations, and group activities; and revealing structural and geographic characteristics of the communication network. However, little is known on how

semantics and structural characteristics of communication vary across geographic space. For more information, please see the attached project description.

Expected results:

- Please see the attached project description.

Knowledge Prerequisite:

- Experience and prior knowledge in Cartography, Information Visualization, Network Analysis, and Geographic Information Systems
- Proficiency in programming languages such as Java, Python, HTML, JavaScript, and visualization libraries such as D3, Leaflet, and MapBox.

Attachments: Detailed Project Description , Datasets

Mentor: Caglar Koylu / Civil and Environmental Engineering, University of Iowa

PROJECT: AUGMENTED REALITY APPLICATION FOR STREAM GAUGE MEASUREMENT

Brief explanation:

The purpose of this project is to create a virtual sensor by using smartphone camera, geo-location of the user by GPS and a marker located on a bridge to provide a point of reference and virtual gauge to allow easy stream gauge measurements. Sample task list includes;

- Streaming webcam video with JavaScript and HTML5
- Tracking multiple markers with JavaScript
- Creating a virtual gauge with numbers near marker location using WebGL
- Creating additional 3D virtual objects in the scene using WebGL
- And more features with student proposals

Expected results:

- An augmented reality web/native application for smartphones/tablets to measure stream gauge.

Knowledge Prerequisite:

- JavaScript, HTML, Augmented Reality APIs, WebGL, Three.JS, Geo-location API, Google Maps API, or Android/iOS programming skills for native app development at medium skill level.

Attachments: None.

Mentor: Dr. Ibrahim Demir / IIHR – Hydroscience & Engineering, University of Iowa

PROJECT: IMMERSIVE REALITY APPLICATION FOR FLOOD SIMULATION

Brief explanation:

The purpose of this project is to create a virtual immersive 3D flood simulator using WebGL and latest web technologies. The project will support heads-up displays like Oculus Rift for immersive experience. The project will involve building terrain using high resolution lidar and elevation data, integration of river system/structure using hydrological models, and interacting with rainfall conditions and levee structures to simulate a flood event. Sample task list includes;

- Building a terrain with WebGL with textures, lidar data, and shaders
- Loading 3D models into scene (house, tree, car, river structure, levee, dam, terrain elements, and other objects in the scene)
- Integrating a JS physics engine for object collision
- Creating a water simulation using WebGL
- Adding interactivity for user controls - and more features with student proposals

Expected results:

- A web-based application to simulate flood event.

Knowledge Prerequisite:

- WebGL, Canvas, JavaScript, HTML, Google Maps API at medium skill level.

Attachments: None.

Mentor: Dr. Ibrahim Demir / IIHR – Hydroscience & Engineering

PROJECT: 3D VOLUMETRIC VISUALIZATION OF RADAR DATA

Brief explanation:

The purpose of this project is to create a volumetric interactive visualization of Hydro-Nexrad radar data. One volume of radar data can include around 19 different elevation scans of various parameters. It is challenging to display such large scale data and examine in real-time using a web browser. Programs. Sample task list includes;

- Creating a loader in JS for large scale volumetric radar data
- Creating interactive visualization using WebGL
- Adding interactivity for user interaction with the visualization and volume
- and more features with student proposals

Expected results:

- A web-based application for volumetric visualization of radar data.

Knowledge Prerequisite:

- WebGL, Canvas, JavaScript, HTML, Google Maps API at medium skill level.

Attachments: None.

Mentor: Dr. Ibrahim Demir / IIHR – Hydroscience & Engineering

PROJECT: DISTIBUTED COMPUTING ON THE WEB FOR FLOOD FORECASTING

Brief explanation:

The purpose of this project is to create a distributed computation framework using Javascript on the web for flood forecasting. The system will explore options for parallel computing using WebGL, WebCL, and pure JS. Sample task list includes;

- Streaming webcam video with JavaScript and HTML5
- Tracking multiple markers with JavaScript
- Creating a virtual gauge with numbers near marker location using WebGL
- Creating additional 3D virtual objects in the scene using WebGL
- and more features with student proposals

Expected results:

- A distributed computation framework for flood forecasting.

Knowledge Prerequisite:

- JavaScript, HTML 5, Web Workers, indexDB, WebGL, WebCL and programming skills for web development at medium skill level.

Attachments: None.

Mentor: Dr. Ibrahim Demir / IIHR – Hydroscience & Engineering

PROJECT: FLOOD INUNDATION SIMULATOR ON GOOGLE MAPS USING WEBGL

Brief explanation:

The purpose of this project is to create an interactive visualization of flood inundation data on Google Maps platform using WebGL. The platform will allow users to zoom in/out and navigate in Google Maps while realistic flood inundation is morphing from one level to another on maps. Sample task list includes;

- Creating a flood indation map polygon on Google Maps using WebGL
- Adding texture and live water effect to the flood map using shaders
- Adding morph effect for the transition of flood maps from one level to another
- Adding object to the map (levee, dam, etc.)
- Adding collision detection
- Adding interactivity to the map for user interaction with objects and scene
- And more features with student proposals

Expected results:

- A web-based application for visualizing flood inundation maps.

Knowledge Prerequisite:

- WebGL, GLSL, Shaders, Canvas, JavaScript, HTML, Google Maps API at medium skill level.

Attachments: None.

Mentor: Dr. Ibrahim Demir / IIHR – Hydroscience & Engineering