Team 3 - Text Geolocators

Samantha Juntiff, Adam McNabb, Weston Neal, Christopher Silvia, Jang Won, Jack Workman

Project Overview

- School of Sustainability
- Input text files, csv files, pdfs and generate a heat map
- What is needed:
  - Fast algorithm to parse through various types of files
  - Create a database holding coordinates of countries, cities, and landmarks
  - Create a JSON object displaying coordinates tied to name
  - From JSON object create a heat map

The Plan

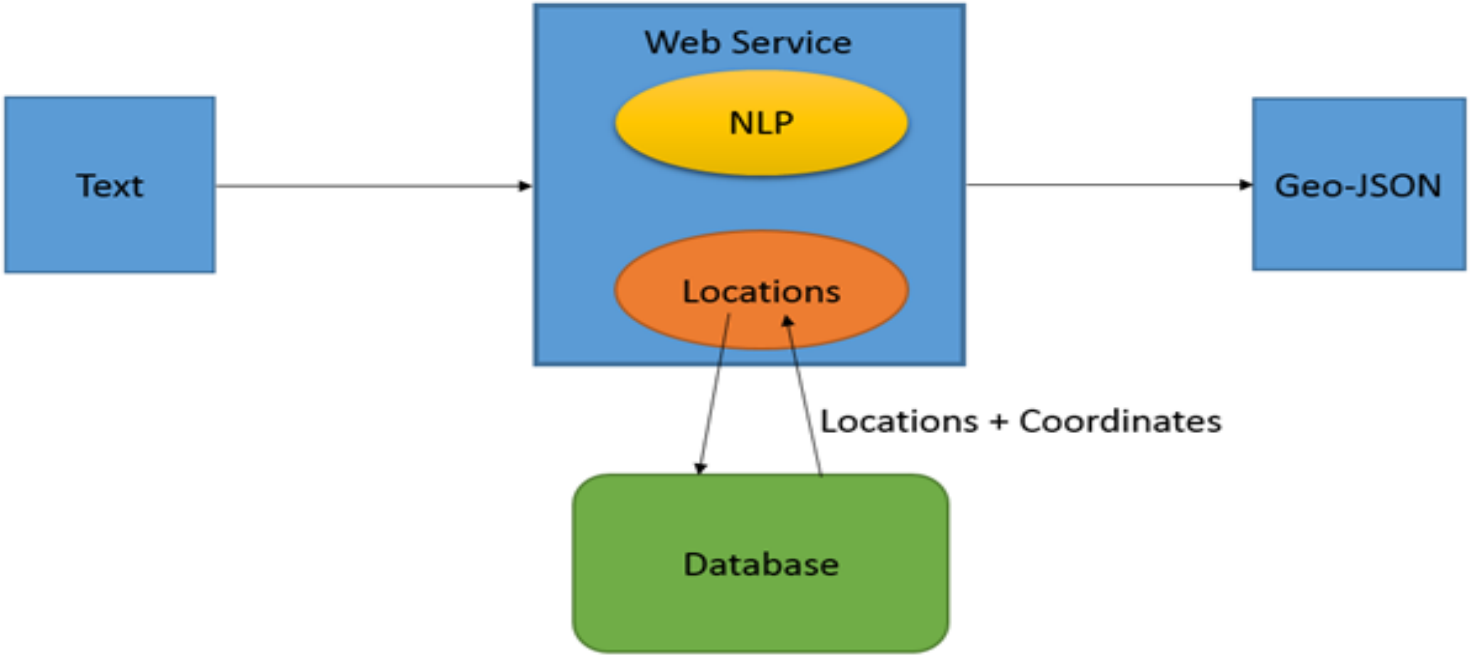
- Fall 2014
  - Create website interface for input files
  - Utilize and synthesize services for location data (JSON)
  - Database Implementation for matches

- Spring 2015
  - Use data to create a heat map of locations
  - Utilize DB data for Machine learning and accuracy ranking of results
  - Performance increases

Scope Changes

- Implementation vs production
  - Docker vs Server
    - Problem (native machine)
- Open source vs originality
- Team vs sponsors
  - How we wanted to design the project
  - everyone works together vs everyone works on separate problems

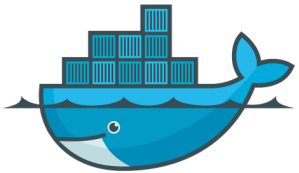
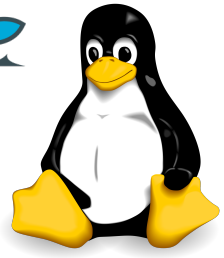
Design and Design Decisions

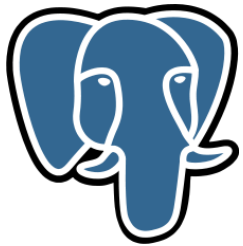


Technology Decisions and Rationale

Technology Stack

- Docker
- Linux
- Python
- PostgreSQL

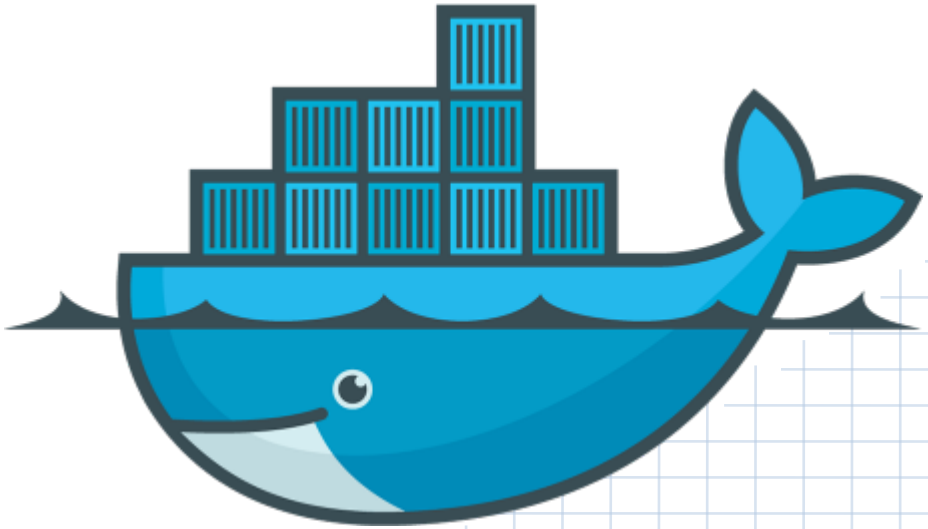


Tools and Libraries

- Flask
- NLTK - Natural Language ToolKit
  - Stanford NER (Named Entity Recognizer) tagger
- geojson

Why Docker?

- Open Platform
- Distributed
- Portable
- Reusable
- Scalable



Why NLTK?

What it is:

- Open Source
- Full Featured

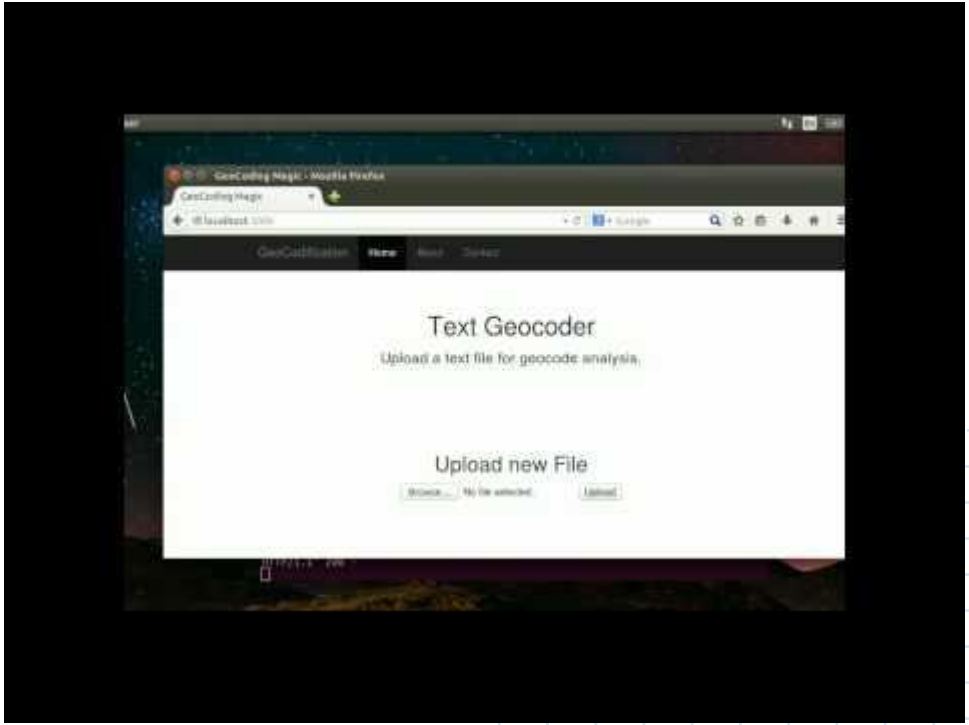
What the advantages are:

- Faster Turnaround Time
- Easier to learn
- Good Abstractions

Results

- Video
  - Logging in
  - Going to address of service
  - Browse for file
  - Upload
  - Return Geojson locations

- Future
  - Heat Map Implementation
  - Testing and Verification
  - Machine Learning
  - Refine our results



Lessons Learned

- New techniques and tools
  - NLP algorithms, Python NLTK, GeoJSON
- How to utilize everyone's skills
  - analyzing strengths/weaknesses, team work
- Communication skills
  - clear meeting notes, accountability
  - project sponsors
- Problem solving techniques
  - divide-and-conquer
  - agile-software model

Questions?