Cholera Database

Michael Roberto, Andrés García Solares, Gabby Alcantara, Hemakshi Sharma, Emily Croxall
CS 4624 Multimedia, Hypertext, & Information Access
Dr. Fox
Virginia Tech, Blacksburg, VA 24061
6 May 2020

Outline

- Background
- Deliverables
- Timeline
- Accomplishments
- Workflow Diagrams
- Testing & Assessment
- Lessons Learned
- Ideas for the Future
- Summary
- Acknowledgements

Background

- Build a database to track Cholera outbreaks worldwide over the last 10 years
- Usage of the database to analyze the data to identify trends and correlations in order to help predict when a cholera outbreak may occur
- To be used in future analysis by client

Deliverables

- Database of Cholera cases
  - Data includes number of cases, location, and dates
  - Data is organized by location or year
- Website to access data
- Ability to download data
- Pipeline of process to make project replicable

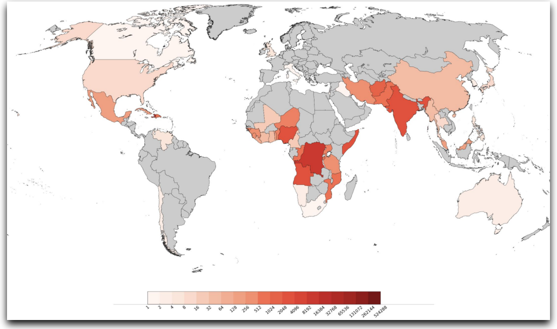

```
"0": {
  "promedRecord": {
    "location": "Somalia",
    "archiveNumber": "20191206.6824571",
    "text": "An outbreak of cholera has killed 7 people in Somalia's southwestern region of Bakol, an of
    "datePublished": "5 Dec 2019",
    "source": "https://www.garoweonline.com/en/news/somalia/cholera-outbreak-leaves-seven-dead-in-somali
  },
  "taggedText": "An outbreak of cholera has <CASE>killed 7 people<\\CASE> in Somalia's southwestern regi
  "totalCases": 7
  "numCases": 1
  "rankSeverity": 30
```

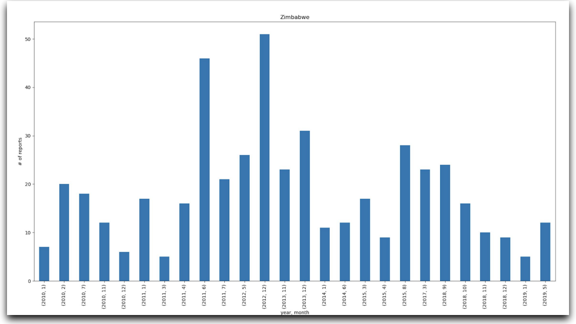
Timeline

- 1/28: Finalize project proposal
- 2/12: Data collection from one repository fully functional and a list of repositories to use
- 3/5: Prototype website with data from one online repository
- 4/1: Data collection from at least one more repository fully functional
- 4/14: Data collection from all repositories completed, website is being built and refined
- 4/23: Final project presentation -- work completed
- 4/30: User views finalized, data is downloadable
- 5/5: Reports and all other deliverables completed
- 5/6: Project due

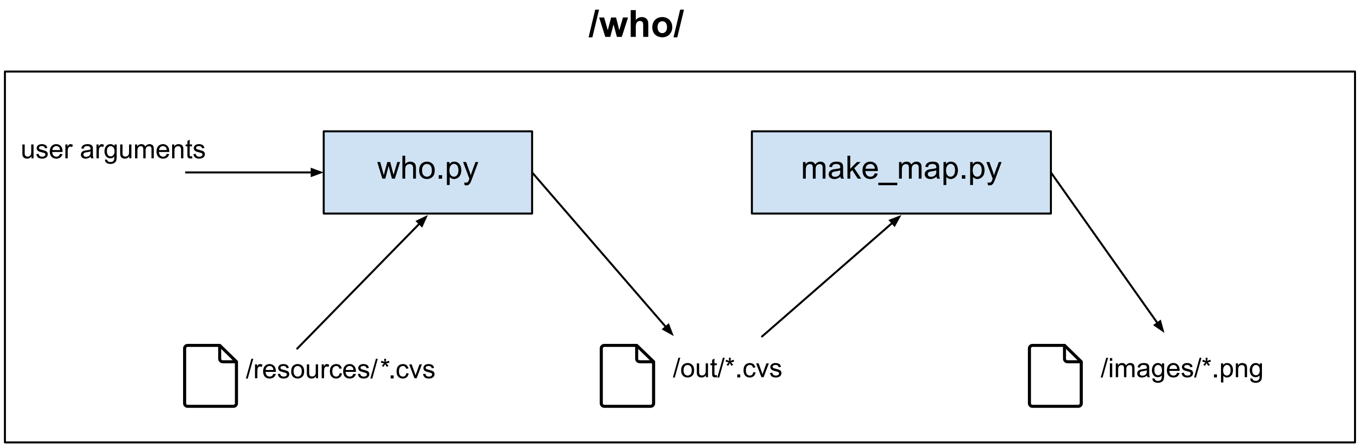
Accomplishments

- Database as CSV files per year
  - All data from WHO
  - Modified data from ProMED
- Website is live and active for 1 year
  - Maps from WHO
  - Graphs from ProMED
- Final report has user guide and developer guide for future users

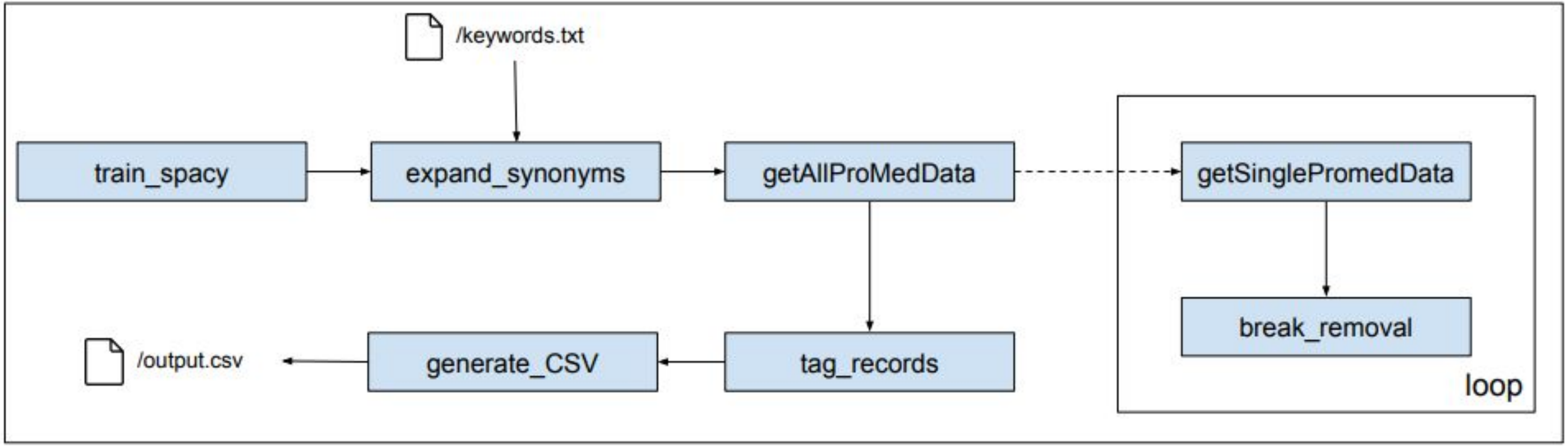




WHO - Workflow Diagram

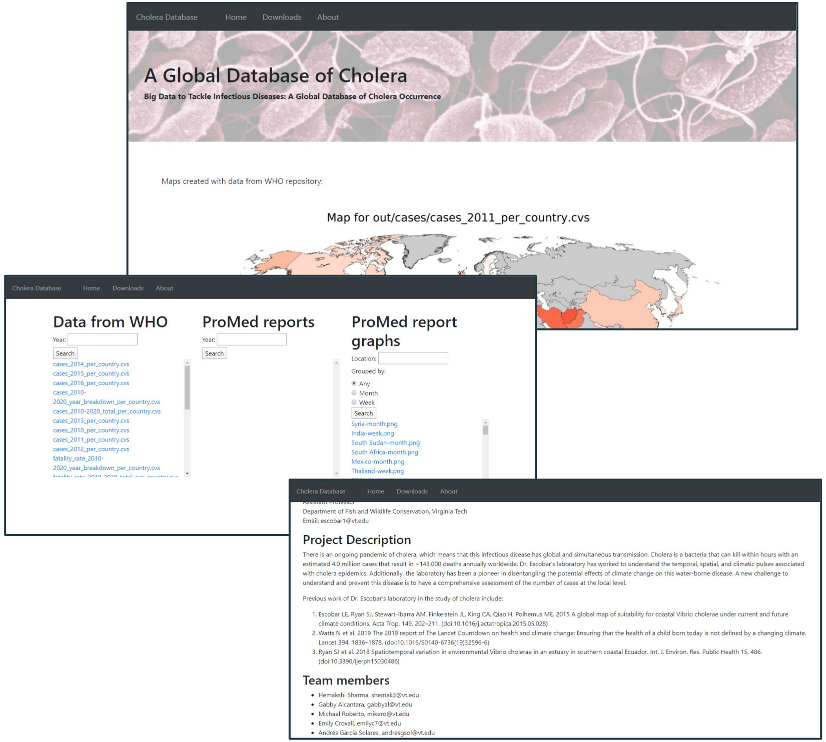


ProMED - Workflow Diagram



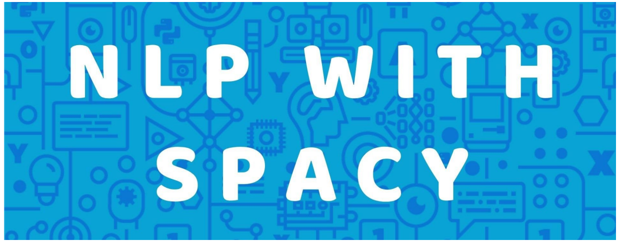
Testing and Assessment

- Static data allowed for us to test our code directly on all data collected
- Usability testing with test subject
    1. Locate the map for WHO data of cases in 2014 per country
    2. Download the fatality rate per country WHO data for 2011
    3. Download the monthly graph for Kuwait from the ProMed data
    4. Locate the team members' contact information
- Approval from client on website formatting



Lessons Learned

- Natural Language Processing
  - Parsing ambiguity
- Incremental development
  - Intermediate steps
  - Better planning
  - Communication over time zones



Ideas for the Future

- Gather data from HealthMap.org
- Develop better methods for extracting ProMED data
- Analysis for our client

Healthmap

Summary

- WHO data collected, sent to CSV, and data visualization maps created
- ProMED data collected, cleaned up, sent to CSV, and visualized with bar graphs
- All data is downloadable from our website
- Detailed document of process created so project is replicable
- Unable to completely extract data from ProMED; some manual analysis required
    - Improvements in our NLP to do this

Acknowledgements

Dr. Luis Escobar

- Department of Fish and Wildlife Conservation, Virginia Tech
- Laboratory
  - Analyze temporal, spatial, climactic pulses of cholera
  - Studies potential effects of climate change on water-borne disease

Dr. Edward Fox

- Our instructor

References

[1] *ProMED International Society for Infectious Disease*, 2020. Available: promedmail.org

[2] *World Health Organization*, 2020. Available: who.int

Questions?