Visualizing Collections of Internet Archives

John Berlin, Joel Rodriguez, Slobodan Milanko

Problem

- Quickly convey small to medium collection of archives

  - Inability to efficiently visualize collections
    - Lack of resource evolution
    - Lack of contextual comparison
    - Lack of statistics surrounding a collection

Goal

- Create a simple way to view collections of archives
  - Tags - words providing contextual meaning
  - Mementos - archive records
  - Time/date - timestamps of archive records
  - Domains - dominant resources of archiving
- User driven, clutter free, easy to use
  and efficient in processing moderate collections

Dataset Description

- We're abstracting the archive collection as a table
- Mementos map to categorical or ordered values
  - Tags - Categorical
  - Number of archives and domains - Quantitative
  - Domains of use - Categorical
  - Time/date of archive - Ordered
- The table approach
  - Facilitates ordering
  - Filtering of data

What-Why-How Framework

Idiom	**Web Archive Visualizer**
What: Data	Web archive collection
What: Derived	Table, attributes
Why: Tasks	Lookup, browse, explore and locate
How: Encode	Separate, Order, Align, Color Categories, Size, Area for Quantity
How: Manipulate	Navigate, Pan Zoom
How: Facet	Juxtapose view
How: Reduce	Zooming, filtering, aggregate
Scale	Attributes: half a dozen, Total items: several hundred

System Description

- Node JS and HTML 5 Application
- Google Sheets - Collaborative location for archives
- Extensive use of many javascript frameworks
  - D3 - Custom charts
  - D3 plus - Great bar chart support
  - C3 - Great bubble chart support
  - JQuery - Google Sheet parsing, slideshow view, navigation
  - Express, Jade

Problems

- Understanding what data is worth keeping
  - Archiver focus removed
- Deciding what framework to use for what chart
  - Some frameworks take away functionality
  - Some frameworks add too much complexity
- Predicting how the users will enter data for parsing
  - i.e. Comma delimited, one record per row vs many

Things Learned

- Use of too many idioms at once can create an overwhelming visualization
    - Split into multiple views; some hidden others visible

- Performance is an important factor to consider for any visualization

- Navigation should be powerful enough to allow transitions between views, but not too distracting

Demo

https://www.youtube.com/watch?v=yPe1t2ktT-Q