Selective Semantic Zoom of a Document Collection

Dustin Dunsmuir
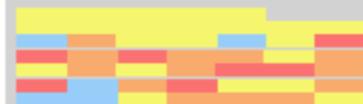
Dec. 14th, 2009

Context

- Intelligence analyst
- A document collection with extracted entities
  - VAST 2006 Challenge Data – Alderwood

- A visualization within the CzSaw system [1]
- Show "document contains entity" relationship through actual containment
- Use selective semantic zoom

Semantic Zoom

Old Zoom Algorithm – SHriMP [2]

1101163656339				
Jan. 12,	Kettle F			
Grandvi	June 25	John M	Grandvi	Melissa
two yea	Toppeni	Idaho	Melissa	1969
Grandvi	School	School	Arlingto	Board
Webma	Hawk P	year	School	Grandvi
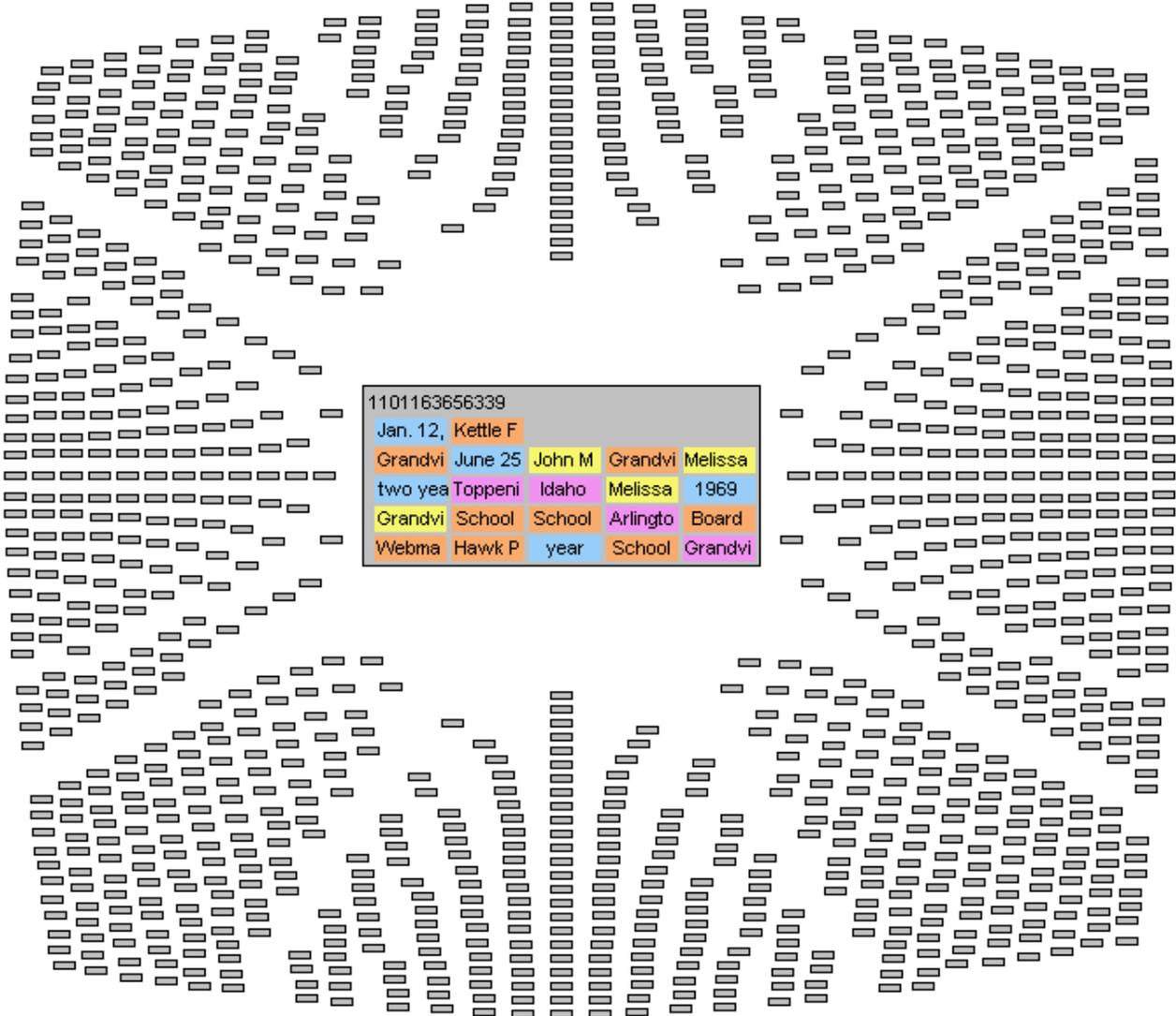
New Zoom Algorithm – Selective SHriMP

Selective Semantic Zoom

Search

- Query database for documents and entities
  - Contains "John" or exactly "John Panni"
  - Within full text or entity
  - Choose colour
    - New colours override old, but old are remembered

Results
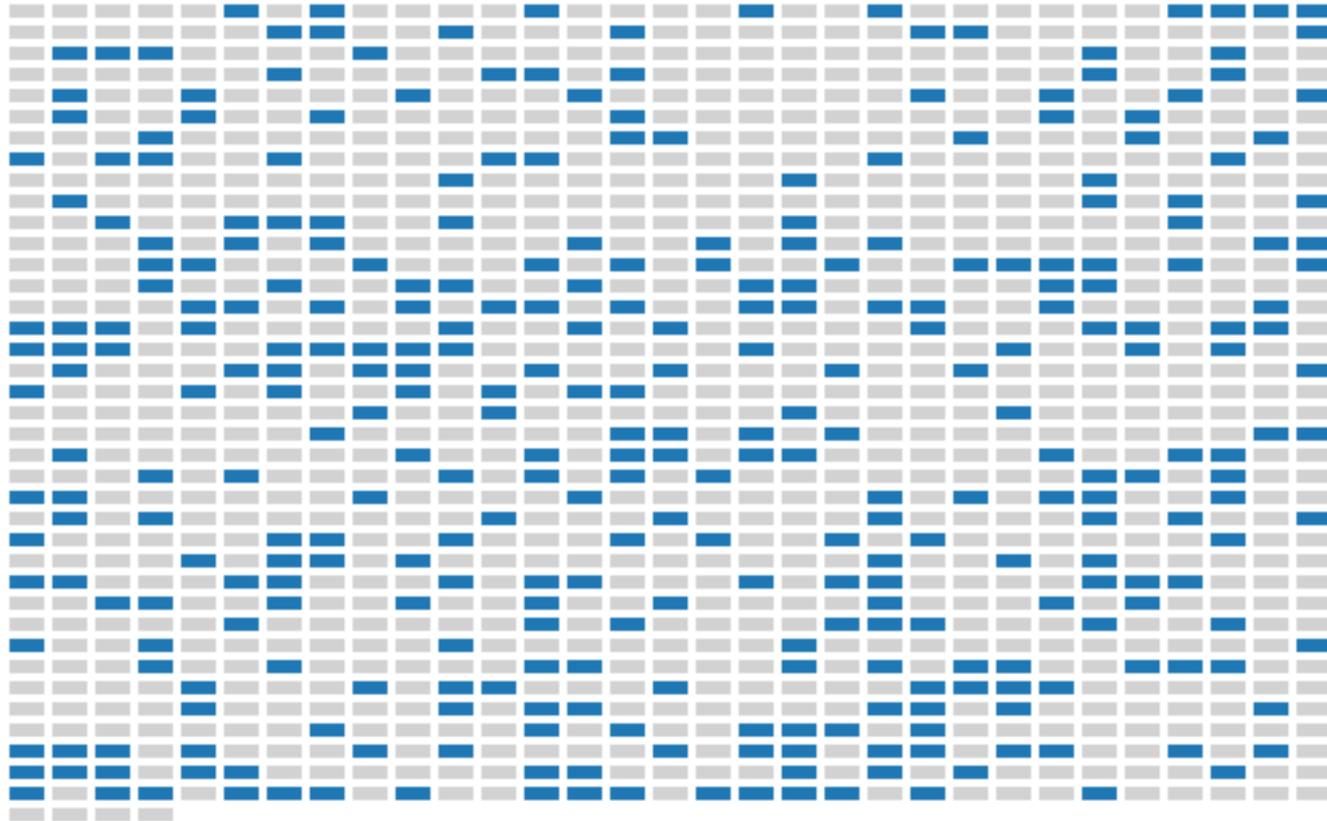
Multiple Search Results
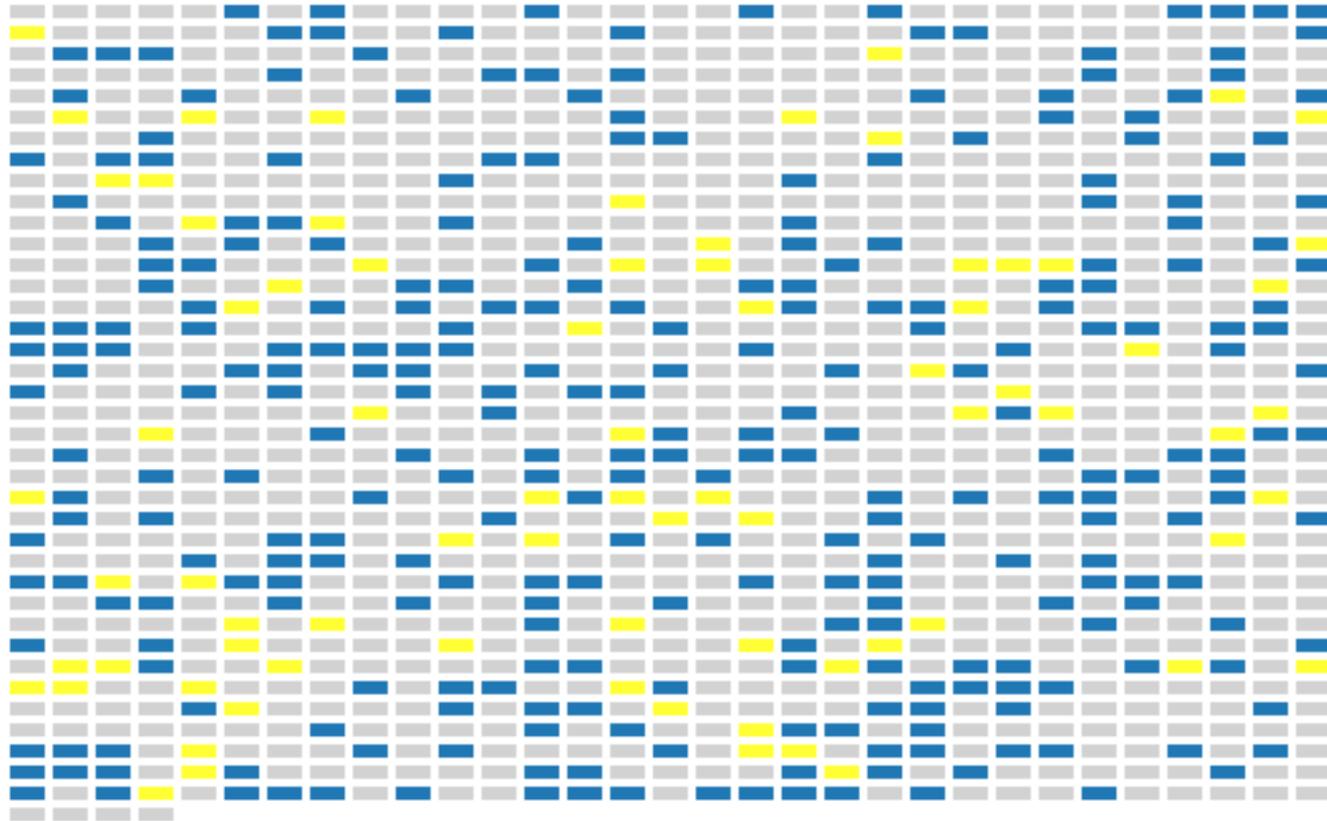
Entity Highlighting

- For search results
  - Highlight all matching documents
  - Highlight all matching entities if any

- By clicking an entity (brushing and linking)
  - Highlight all documents containing it
  - Highlight entity within each

Brushing and Linking

Layout by Date

- Date part of metadata of document, not entity
- Arrange in cells
  - Row is year
  - Column is month
  - Within cell layout like calendar
  - Diagonally stack same dates

- Date visualizations for finding temporal patterns found useful in Improvise [3]
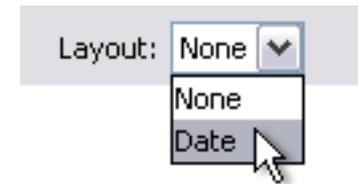
Date Layout
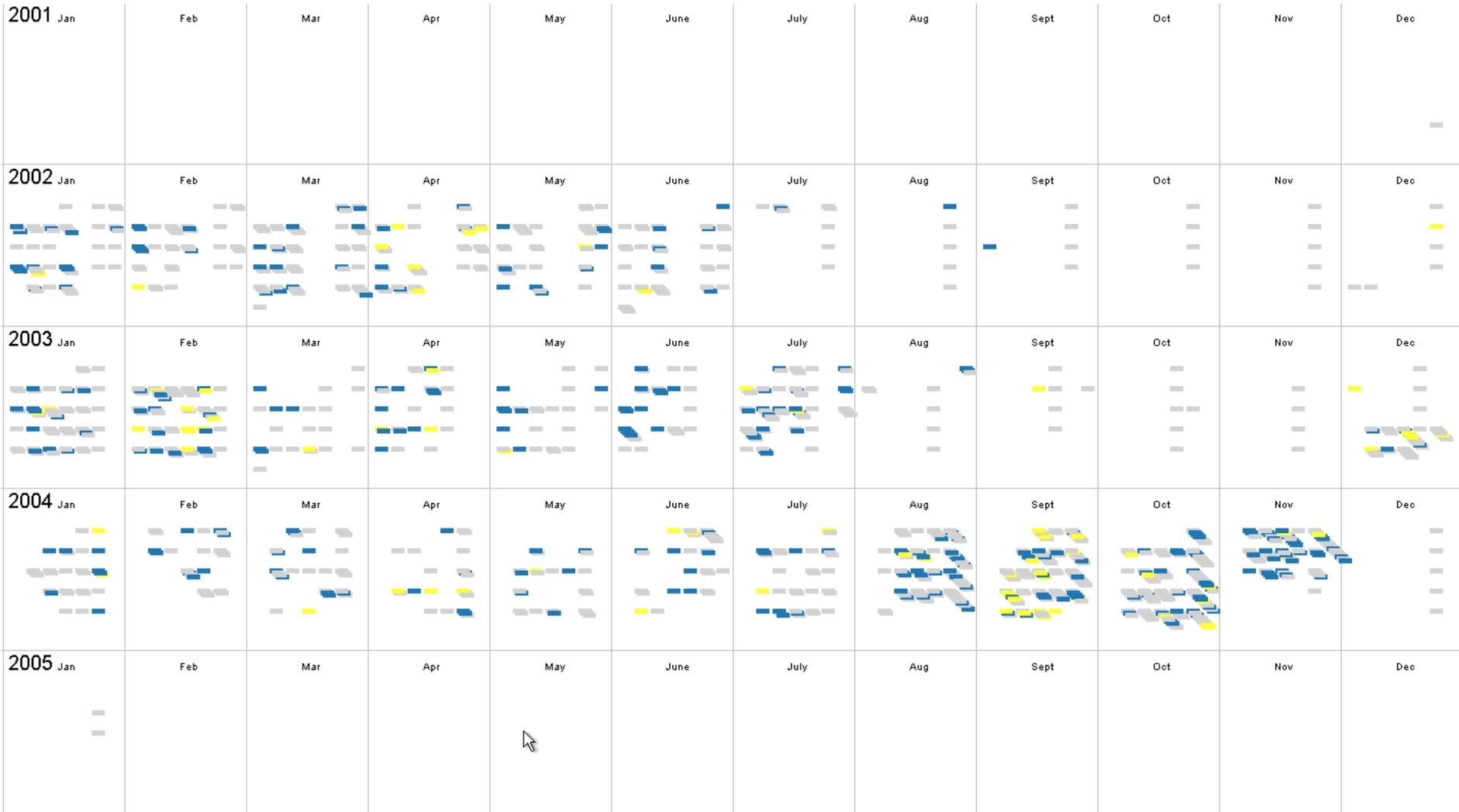
Actions on Active Documents

Actions upon active items: [Cluster] [Group] [Filter] [Unfilter All]

- Active documents are those highlighted in the active selection colour
- Filtering
  - Filter out all active documents
  - Unfilter all documents
  - Invert the selection and then filter

Cluster

- Applied to selected documents
- Choose a cluster location
  - Documents are clustered around the location
  - In the same relative positions
  - With the spaces between them removed
- Documents can be zoomed differently
- The movement is animated

- Useful for quickly gathering search results

Cluster Before
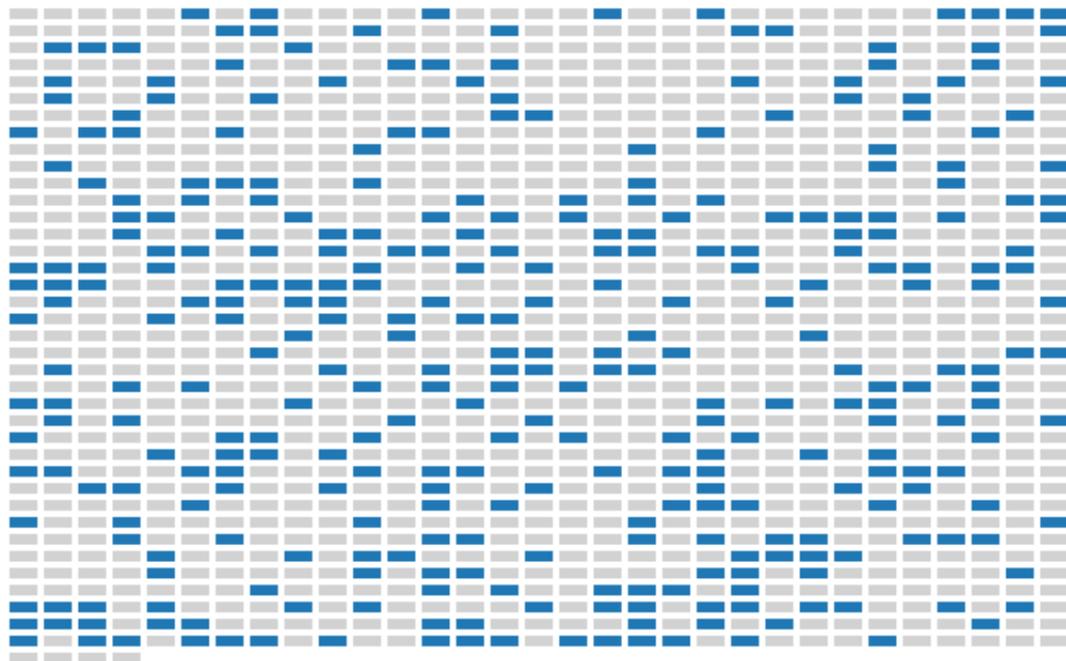
Cluster After

Cluster Before

Cluster After

Group

- Documents first clustered
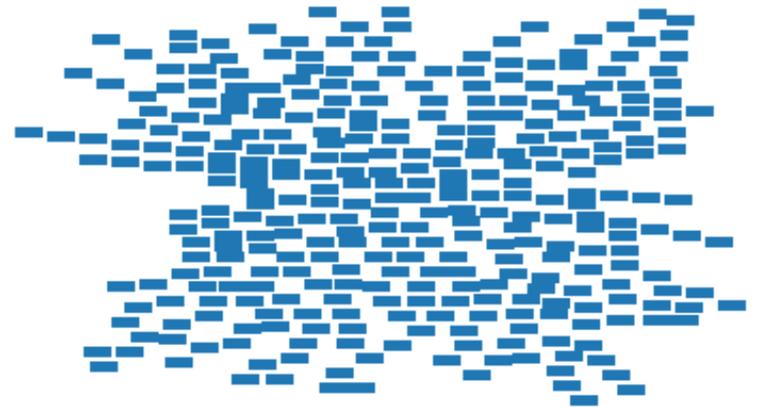
- Group has name and bounds

- Can be moved and resized

- Another level to the view
  - Dynamically resized based on contents
  - Make multi-level zoom algorithms

Group
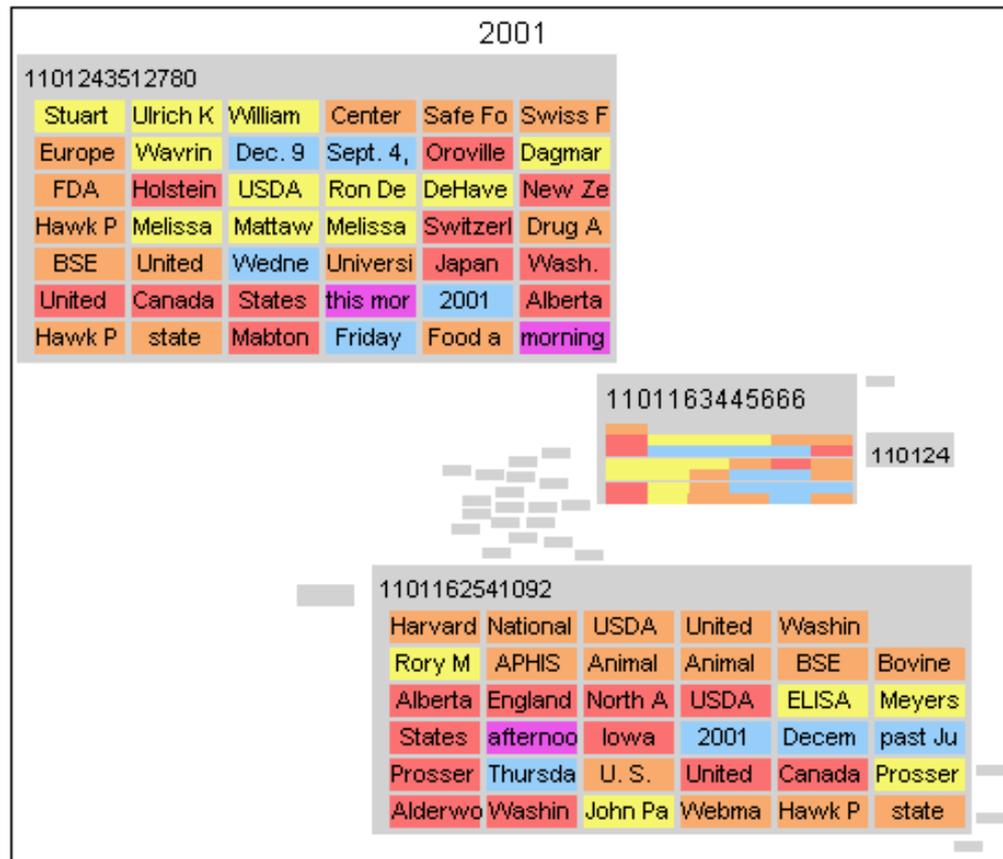
Future Work
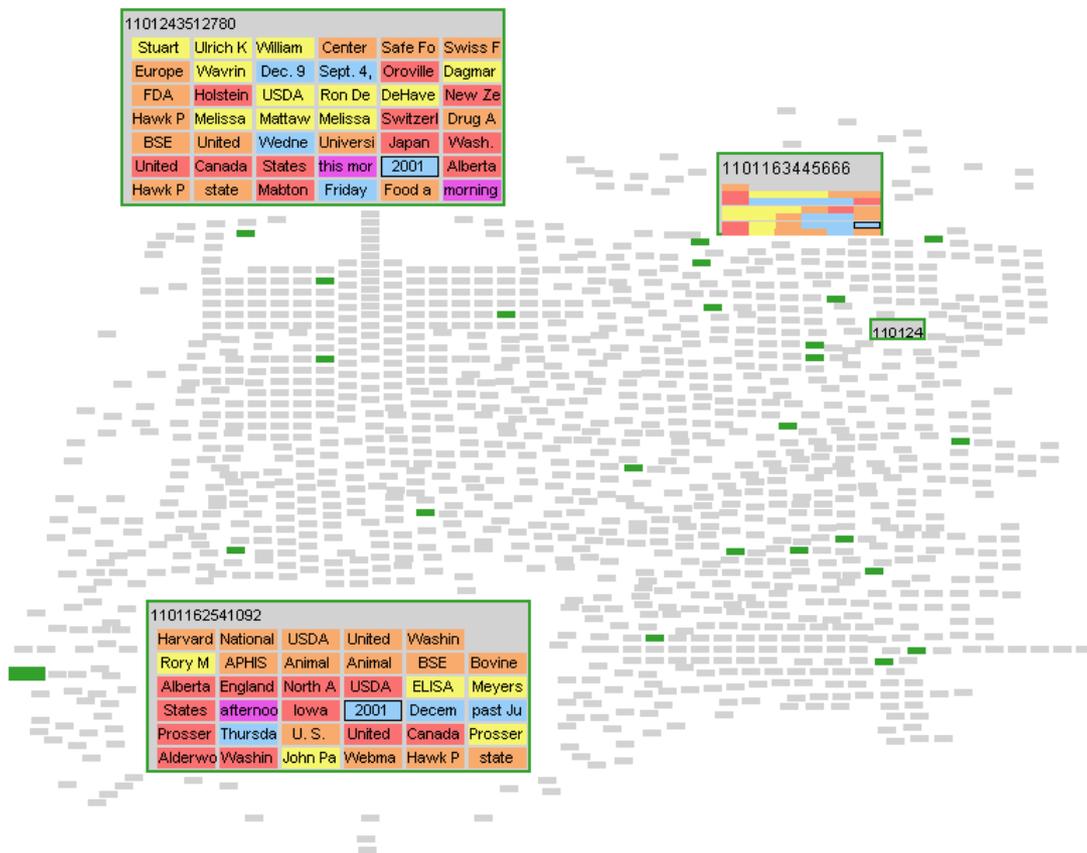
- This is my **Thesis Project**
  - More layouts, improve date layout flexibility
  - Zoom multiple documents at once
  - Combine documents and visualize entities
    - Treemap of entities within super document
    - Layout entities
    - Filter by entity type
  - Usability study!
    - What is a good initial layout?

References

- [1] Kadivar, N, Chen, V., Dunsmuir, D., Lee, E., Qian, C., Dill, J., Shaw, C., and Woodbury, R. (2009) Capturing and Supporting the Analysis Process. Proceedings of IEEE Visual Analytics Science & Technology 2009, Atlantic City, NJ, Oct 11-16, 2009, pp. 131-138.

- [2] Storey, M-A., Best, C., and Michaud, J. (2001) SHriMP Views: An Interactive Environment for Exploring Java Programs. International Conference on Program Comprehension.

- [3] Weaver, C., Fyfe, D., Robinson, A., Holdsworth, D., Peuquet, D., and MacEachren, A. (2007) Visual Exploration and Analysis of Historic Hotel Visits. Information Visualization, 6 pp. 89-103.