

Selective Semantic Zoom of a Document Collection

Dustin Dunsmuir

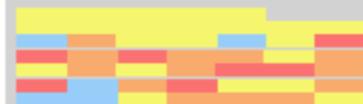
Context

- 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

110124

1101243402264



1101243402264

Roy Ma	Scott P	Teagan,	Treyton	Treyton,		
Parrish	Frank D	Mike Sc	Parrish,	Payton	Payton,	Roy Bro
1991	Big Ben	Zane W	Tammy	1986	Teagan	Chelan
Moses	Greyho	WEST	Grandvi	Hawk P	Benton	Prosser
Greyho	High	Yakima	Mid- Va	Selah	Hanford	Eastern
John	year	School	Grandvi	Grandvi	Workin	Grandvi
this yea	today	John Pa	High Sc	Hawk P	state	Panni

1101243402264

Story by: **John Panni**

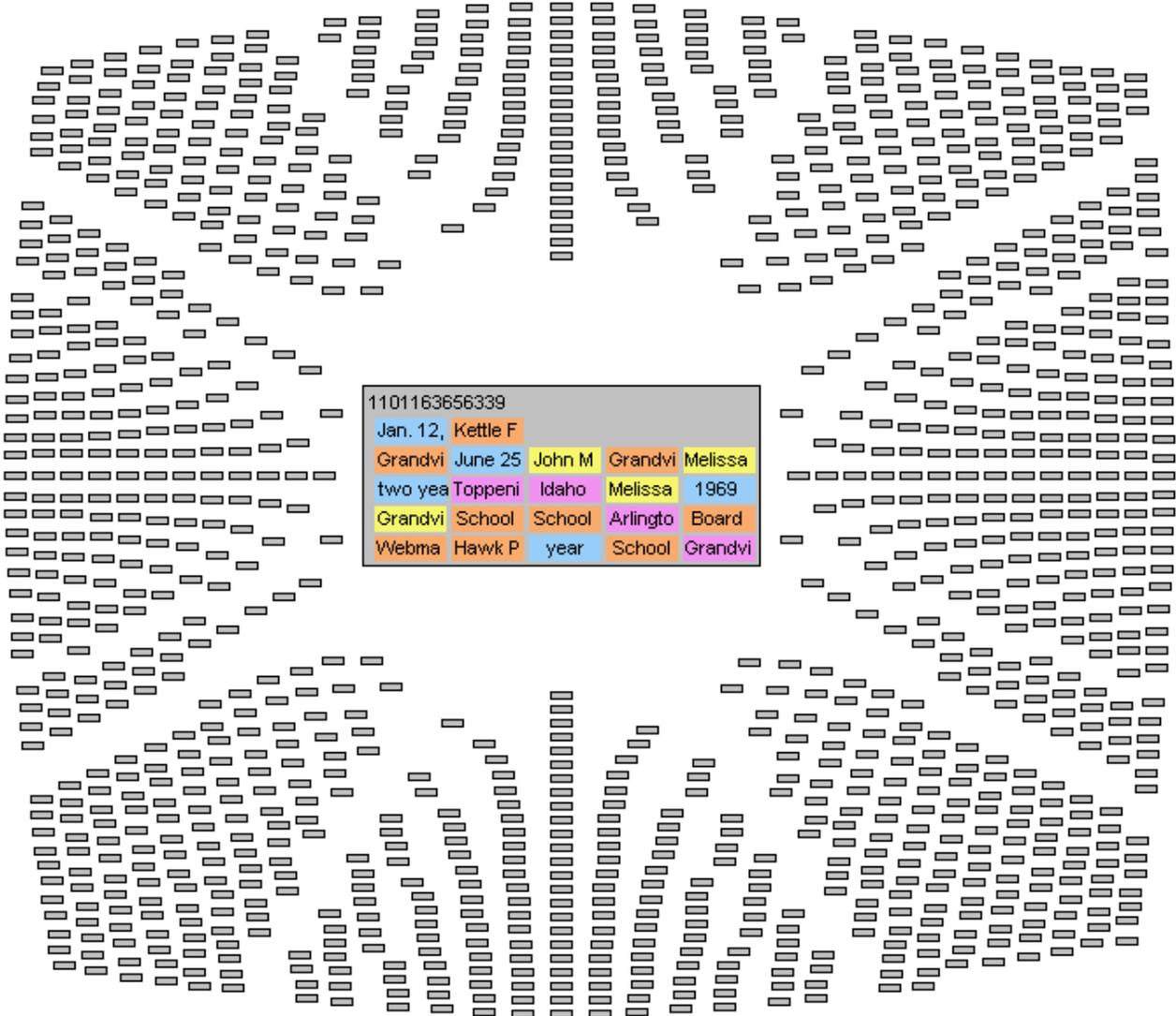
GRANDVIEW - Every family needs a father figure and the father figure for the **Grandview High School** boys basketball program is head coach **Scott Parrish**.

But **Parrish, 36**, will be the first to tell you none of the successes he has had as head coach of one of the most successful basketball programs in the **state** would be possible without those surrounding him.

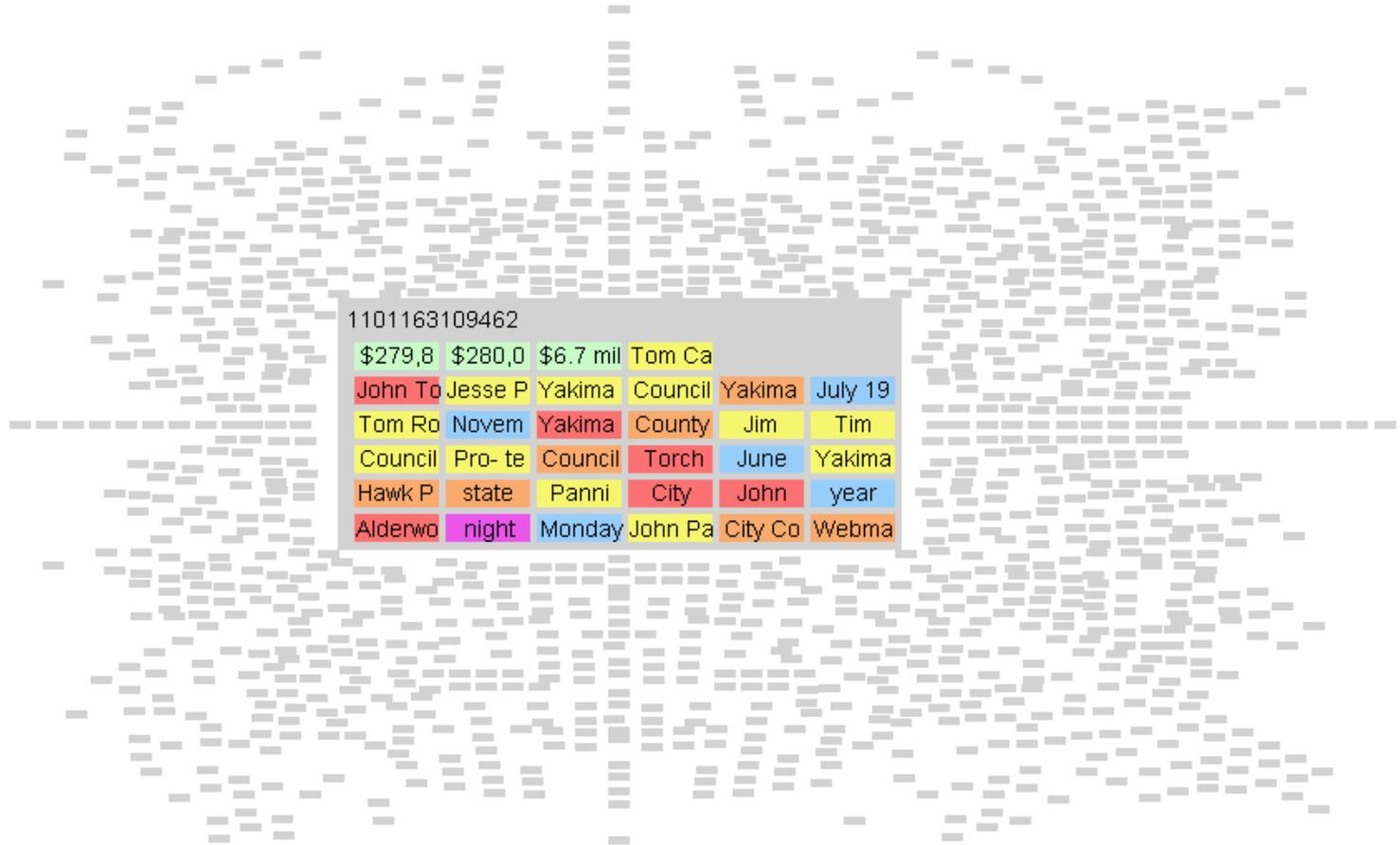
The one reason why **Parrish** might take such a pride with his family of players and the way they represent **Grandview** is he understands what it is to be a **Greyhound**.

A native of **Grandview**, **Parrish** played basketball for the **Greyhounds**, a three-**year** player on varsity. **Parrish** graduated from **Grandview High School** in **1986**, going on to **Big Bend Community College** in **Moses Lake** to play two **year**s of basketball. In **1991**, **Parrish** graduated from **Eastern Washington University** with a Bachelor's degree in

Old Zoom Algorithm – SHriMP [2]



New Zoom Algorithm – Selective SHriMP



Selective Semantic Zoom

1101243293654

last Apri	Feb. 11	past Se	Americ	Americ
Stockho	Ohio	Associa	Dallas	Cleavela
2003	Fair	last May	Kacee	Hawk P
Texas	Seattle	Washin	National	Last spr
Alderwo	High Sc	Hawk P	state	Jamie

11012433370

last Apri	Feb. 11	past Se	Americ	Americ
Stockho	Ohio	Associa	Dallas	Cleavela
2003	Fair	last May	Kacee	Hawk P
Texas	Seattle	Washin	National	Last spr
Alderwo	High Sc	Hawk P	state	Jamie

11011639

last Apri	Feb. 11	past Se	Americ	Americ
Stockho	Ohio	Associa	Dallas	Cleavela
2003	Fair	last May	Kacee	Hawk P
Texas	Seattle	Washin	National	Last spr
Alderwo	High Sc	Hawk P	state	Jamie

110116

1101162453186

up from the **2004** rate of **\$2.71** per **\$1,000** of assessed property value. The approved property tax increase will raise an additional **\$195,556** for city coffers. The city will raise an additional **\$625,970** in General Obligation Bonds. The payments on the bonds are being made from when the city refinanced together the **Law** and **Justice Center** and the pool measures voters approved. **Council** was split on the matter, with Mayor Pro-tem George Greenway and **Councilmen Tom Seeger** and **John Torch** wanting to scale back the property tax increase to **\$3** per **\$1,000** of assessed property value. Mayor **Rex Luthor**, **Councilwoman Jennie Angel** and **Councilmen Jim Sarducci** and **Raul Martinez** took the collective stance that the city needs to do something to raise money to prevent further reductions in city services. But not everyone agreed with the thoughts of **Council Monday night**. **Alderwood** resident **Bob Johnson** said he felt **Council** should stick to past practices of not raising taxes. **Johnson** said that with the cost of living and water and sewer rates

1101242635978

12 year	\$100 bil	\$6 billio	\$940	
Sept. 1	Washin	Hawk P	U.S	\$40,00
Senate	Murray	U.S.	afternoon	past Mo
Novem	United	Spokan	George	Patty M
Thursda	Lower	Mel Sac	U. S.	next yea
Alderwo	Washin	Hawk P	state	Tuesda

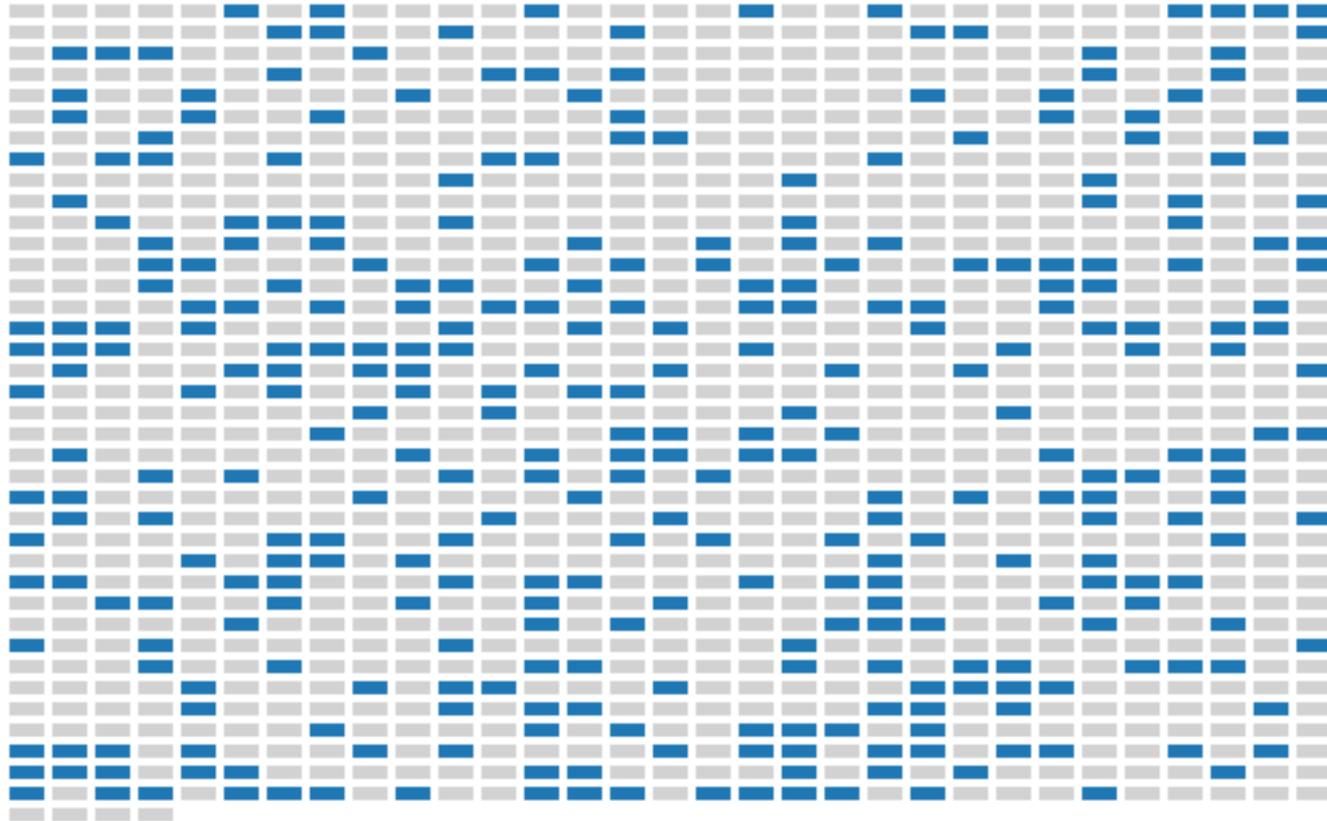
Search

The image shows a search interface with the following elements:

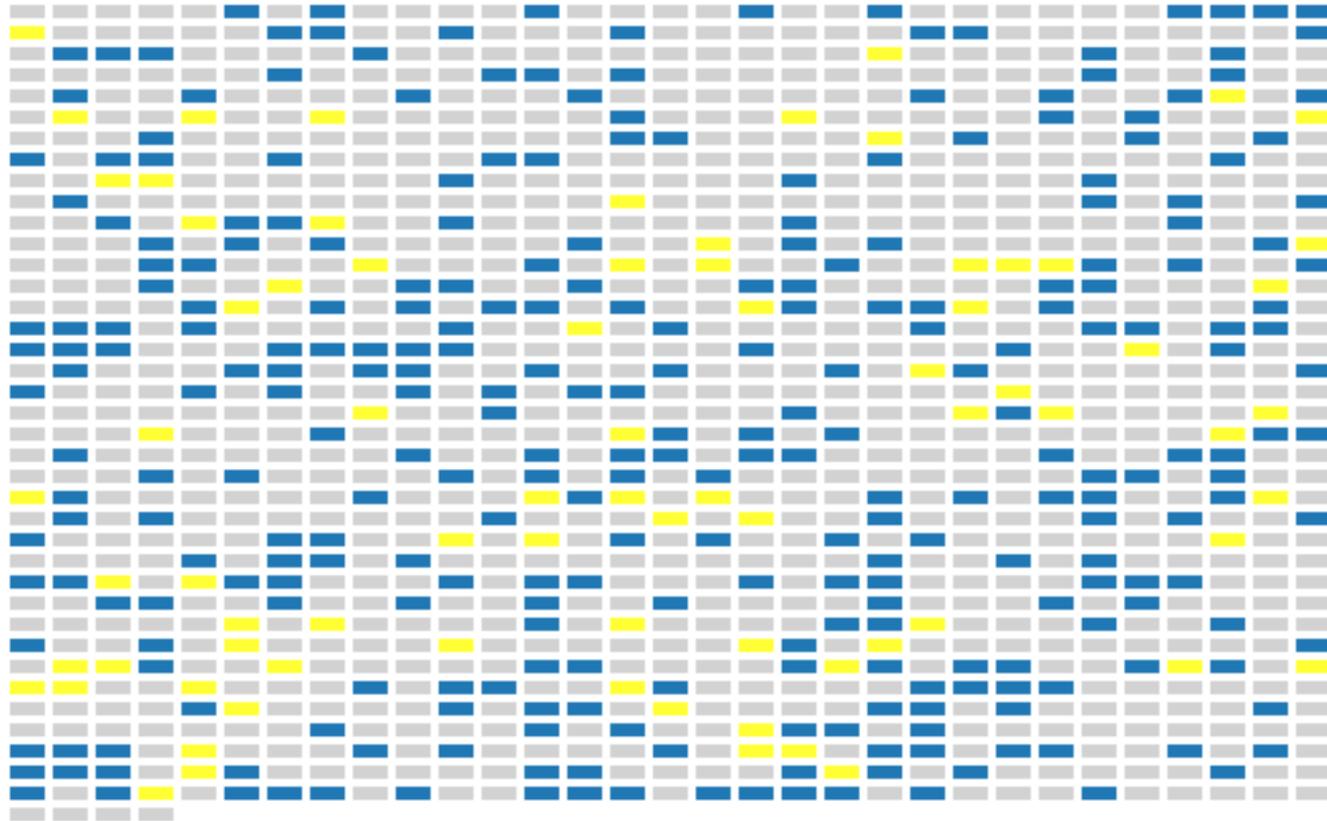
- Search: exactly in field
- Active colour:
- A dropdown menu is open under the 'report' field, listing: report, place, organization, person, date, money, time.
- A color palette is visible to the right of the 'Active colour' dropdown, showing a grid of colored squares: blue, grey, yellow, light green, light blue, and dark green.

- 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

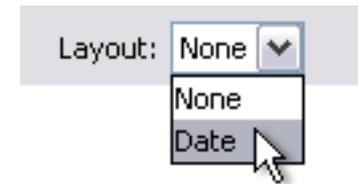


Entity Highlighting

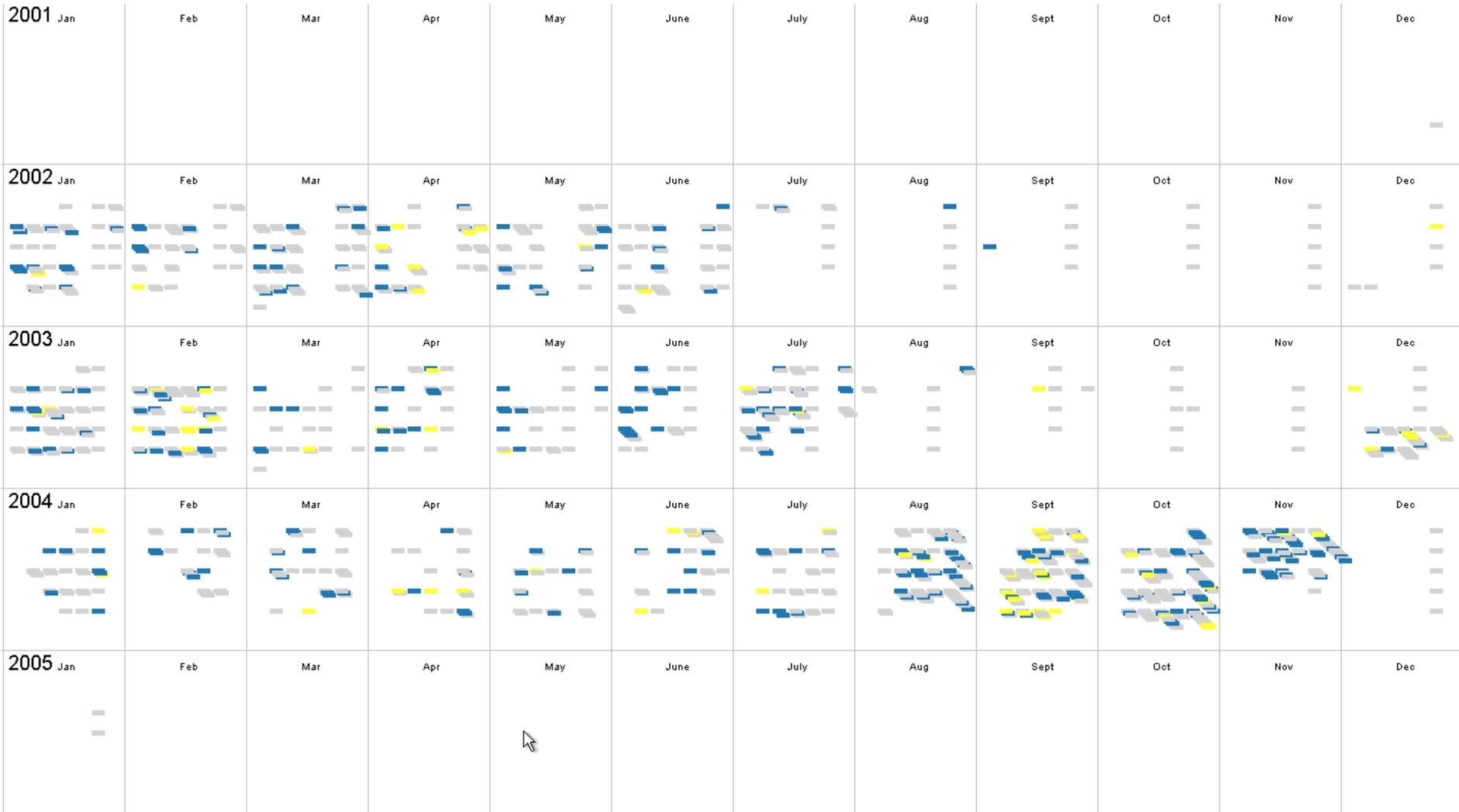
- For search results
 - Highlight all matching documents
 - Highlight all matching entities if any
- By clicking an entity (brushing and linking)
 - Highlight all documents containing it
 - Highlight entity within each

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

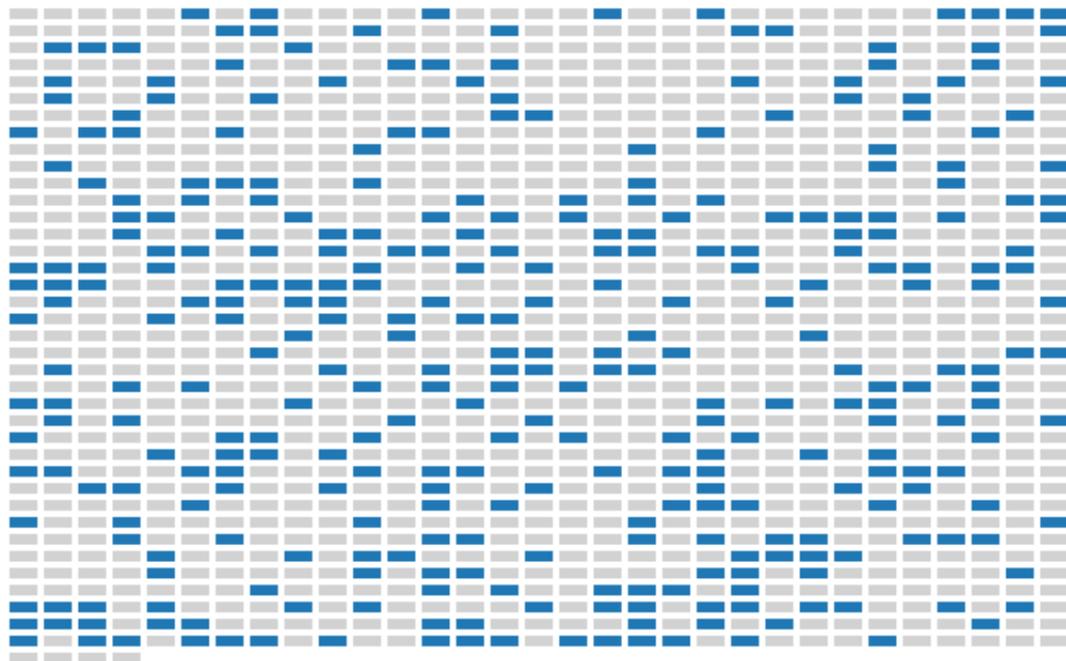


- Active documents are those highlighted in the active selection colour
- Filtering
 - Filter out all active documents
 - Unfilter all documents
 - Invert the selection and then filter

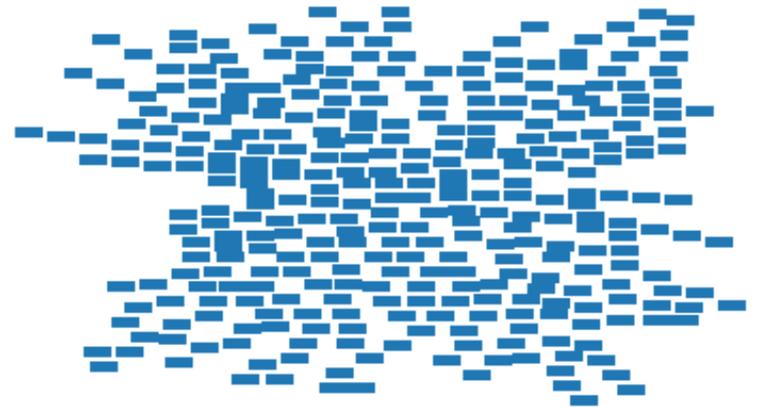
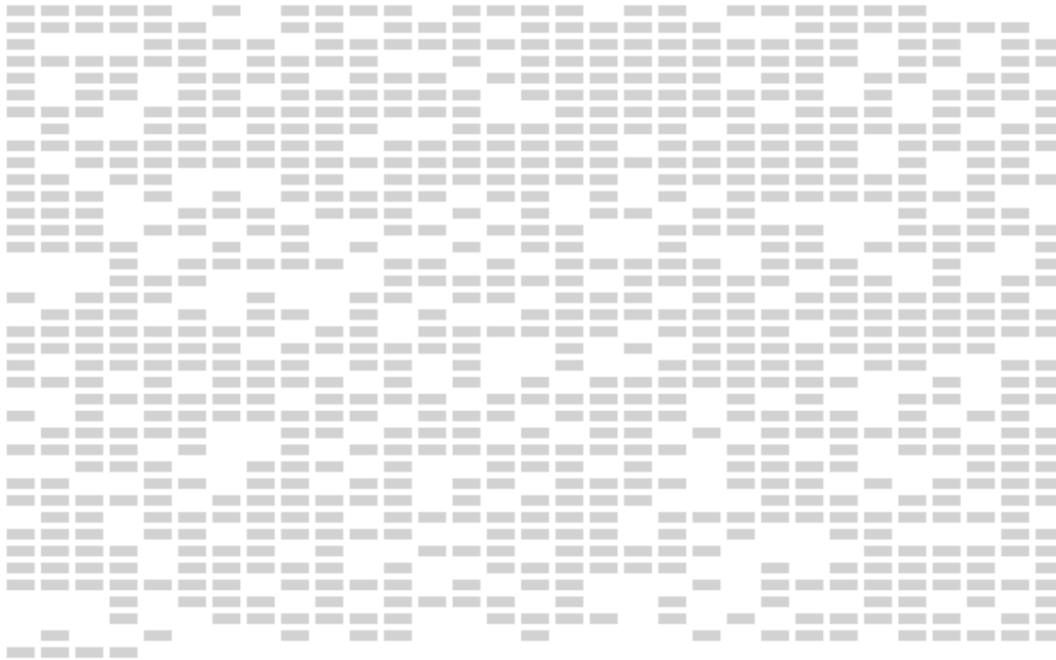
Cluster

- 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

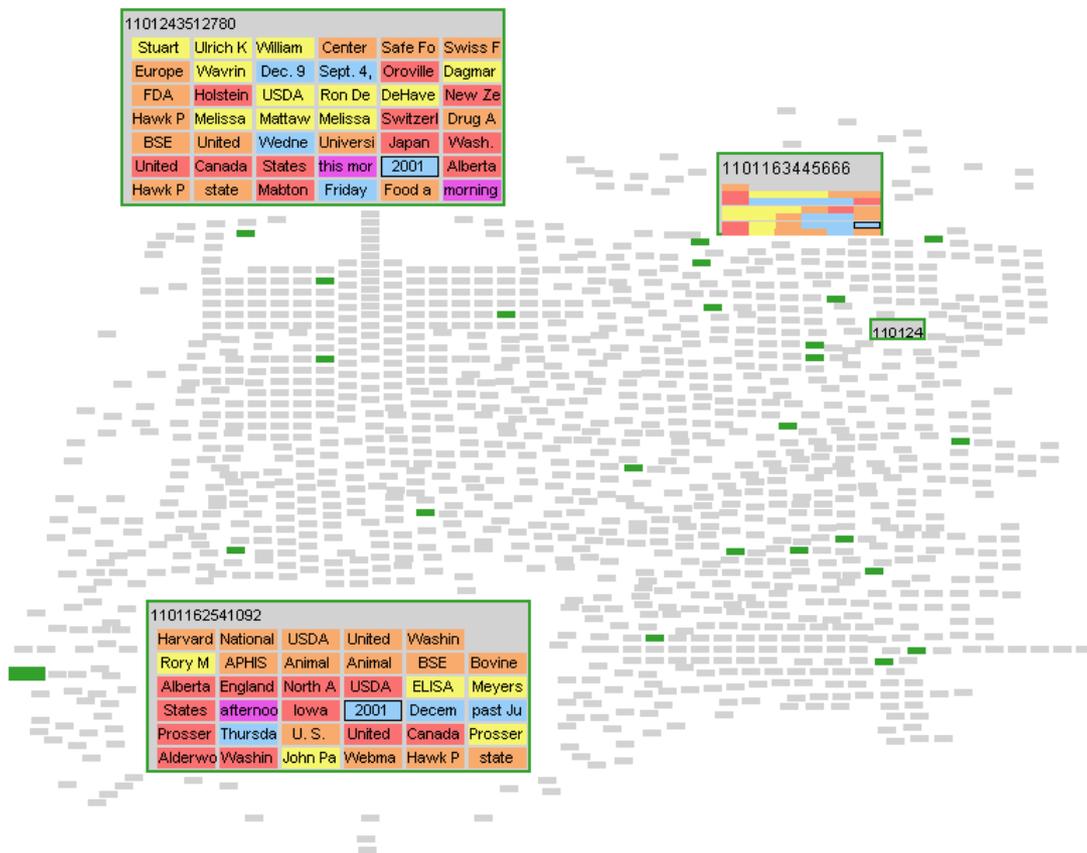
Cluster Before



Cluster After



Cluster Before



Cluster After



1101243512780

Stuart	Ulrich K	William	Center	Safe Fo	Swiss F
Europe	Wavrin	Dec. 9	Sept. 4,	Oroville	Dagmar
FDA	Holstein	USDA	Ron De	DeHave	New Ze
Hawk P	Melissa	Mattaw	Melissa	Switzerl	Drug A
BSE	United	Wedne	Universi	Japan	Wash.
United	Canada	States	this mor	2001	Alberta
Hawk P	state	Mabton	Friday	Food a	morning

1101163445666

110124

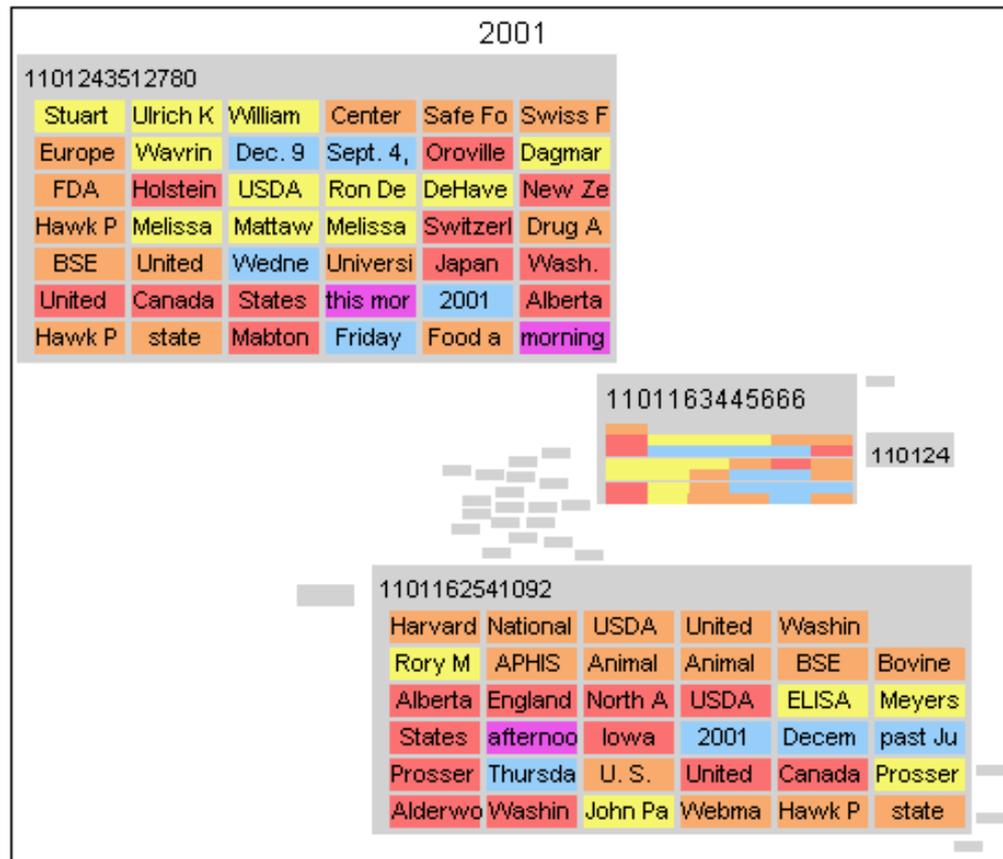
1101162541092

Harvard	National	USDA	United	Washin	
Rory M	APHIS	Animal	Animal	BSE	Bovine
Alberta	England	North A	USDA	ELISA	Meyers
States	afternoon	Iowa	2001	Decem	past Ju
Prosser	Thursda	U. S.	United	Canada	Prosser
Alderwo	Washin	John Pa	Webma	Hawk P	state

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



2001

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.