

# Text Vis and NLP

Anoop Sarkar

Natural Language Laboratory (<http://natlang.cs.sfu.ca>)

Simon Fraser University, Vancouver, Canada

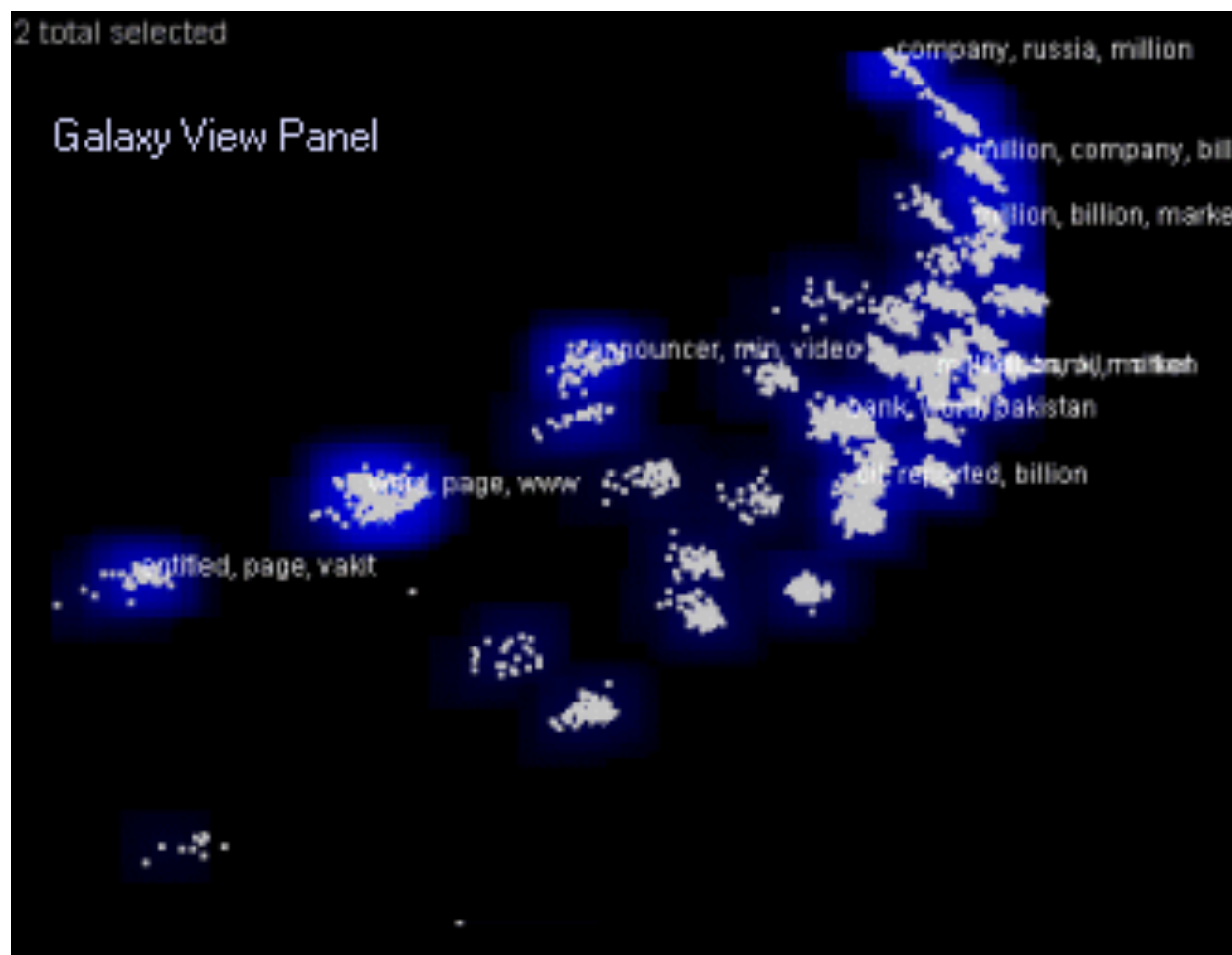
Explore new visualizations that exploit parsed language

# Lensing Language

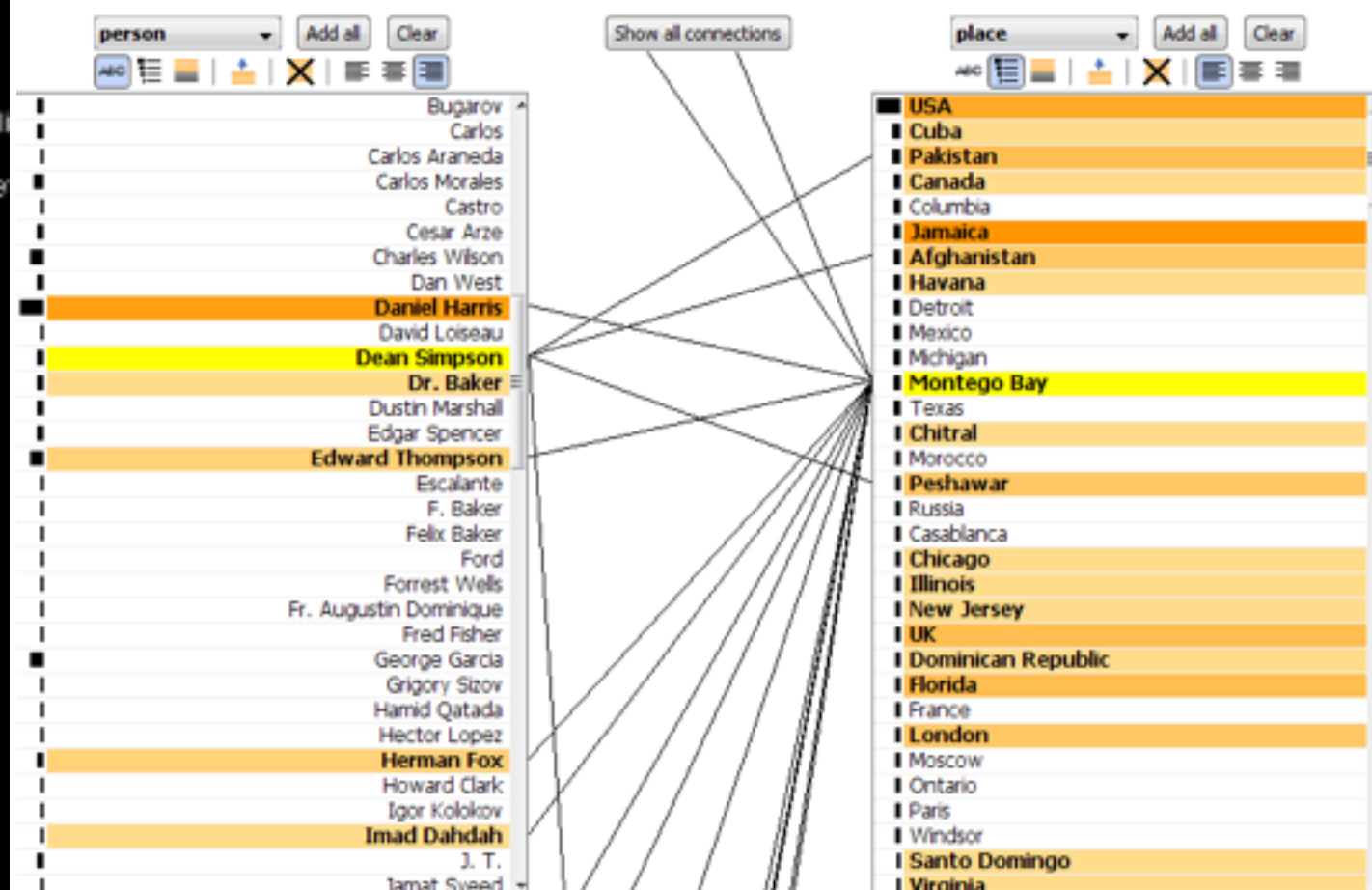
- Semantic parsing of natural language: going beyond topic models and clustering bags of words
- Exploit language understanding: *who* did *what* to *whom*, *where*, *when* and *how* ...
- "Embodied" visualization: place spatial, temporal and social entities into an intuitive low dimensional space

# Lensing Language

- Many visualization tasks involve language
- Current practice: ignore most of the structure.
  - Cluster a bag of words
  - Show entities and their connections



In-spire



Jigsaw

# Lensing Language

- Annotation of the language data is often left to the user
- Language is a lot more complex than most native speakers realize
- Native speakers of language do not have access to their knowledge of language
- Cannot use a “grammar book”!



# Lensing Language

- Goal: explore visualization of information in language
- Provide a proof of concept that more advanced NLP can help improve visualization of large document collections
- Engage with visualization and VA experts using a concrete example so that we have a concrete example
- Look for a task that can serve as proof of concept
- Language should be amenable to our goals of visualization and can be parsed accurately using modern statistical parsing methods
- Details of language analysis must be hidden from the user

# Lensing Wikipedia

- Provide a visualization of all of human history ... as represented in Wikipedia
- The information is all in natural language (English)
- The task: **visually browse through human history**
  - e.g. in order to write an essay about a fairly general topic: “Describe Roman interactions with Carthage between 200BC and 15BC”
  - e.g. to check Wikipedia coverage of various regions and time periods for completeness
  - e.g. is Wikipedia history biased towards violent events rather than peaceful ones?

# Lensing Wikipedia

104BC: Athenion starts a slave rebellion in Segesta

First Step: Parsing Wikipedia

Second Step: Project spatially

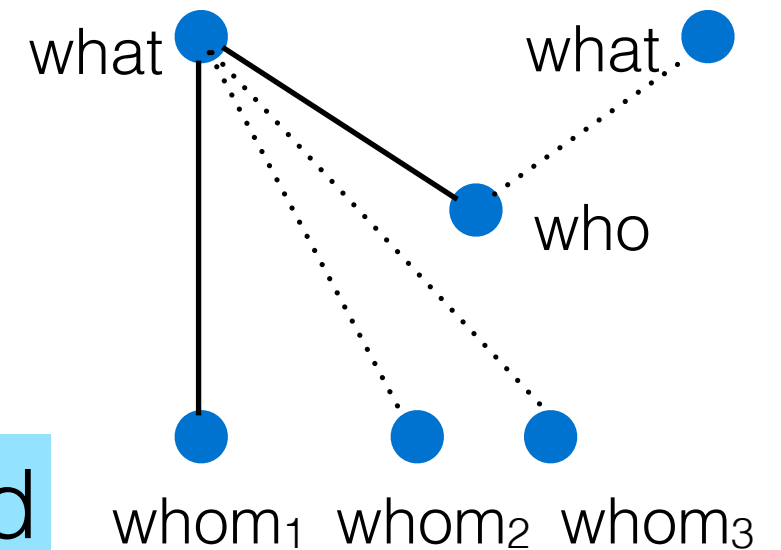
Predicate

Arg0: initiator

Arg1: thing\_started

Location

Time



(Sicily: 2D Plot, lat=37.93; long=12.83)

Plot 104BC on a 1D timeline



# Lensing Wikipedia



[Main page](#)  
[Contents](#)  
[Featured content](#)  
[Current events](#)  
[Random article](#)  
[Donate to Wikipedia](#)

▼ [Interaction](#)  
[Help](#)  
[About Wikipedia](#)  
[Community portal](#)  
[Recent changes](#)

[Article](#) [Talk](#)

## 1470s

From Wikipedia, the free encyclopedia

This is a list of events occurring in the 1470s, ordered by year.

### **Contents** [\[hide\]](#)

[1470](#) · [1471](#) · [1472](#) · [1473](#) · [1474](#) · [1475](#) · [1476](#) · [1477](#) · [1478](#) · [1479](#)

## 1470

### January–December

- **March 12** – *Wars of the Roses* – *Battle of Lose-coat Field*: The *House of York* defeats the *House of Lancaster*.
- **May 15** – *Charles VIII of Sweden*, who had served three terms as *King of Sweden*, dies. *Sten Sture the Elder* becomes Regent of Sweden.
- **October** – A rebellion orchestrated by King *Edward's* former ally, the *Earl of Warwick*, forces the King to flee England to seek support from his brother-in-law, the *Duke of Burgundy*.
- **October 30** – Warwick releases *Henry VI of England* from the Tower and restores him to the throne.

<b>Millennium:</b>	2nd millennium
<b>Centuries:</b>	14th century – <b>15th century</b> – 16th century
<b>Decades:</b>	1440s 1450s 1460s – <b>1470s</b> – 1480s 1490s 1500s
<b>Years:</b>	1470 1471 1472 1473 1474 1475 1476 1477 1478 1479
<b>Categories:</b>	<span><span>Births</span></span> – <span><span>Deaths</span></span> – <span><span>Architecture</span></span> <div><span><span>Establishments</span></span> – <span><span>Disestablishments</span></span></div>

[\[edit\]](#)

[\[edit\]](#)



# URLs from Wikipedia: ~2K

```
{  
  "arg0": "Mina de Ouro",  
  
  "arg1": "the chief center for the gold trade and a major source of revenue  
for the crown",  
  
  "description": "Portuguese sailors reach Mina de Ouro on the Gold Coast  
(present-day Ghana) and explore Cape St. Catherine, two degrees south of the  
equator. Mina de Ouro becomes the chief center for the gold trade and a major  
source of revenue for the crown.",  
  
  "event": "become",  
  
  "latitude": 5.5499977999999999,  
  "longitude": -0.249999,  
  
  "roleArg0": "Agent",  
  "roleArg1": "entity_changing",  
  
  "title": "Ghana",  
  
  "year": 1471  
}
```

Number of events: ~20K

746	found
593	defeat
570	become
494	kill
491	begin
360	take
325	lead
303	establish
294	make
282	capture
271	destroy
240	open
205	end
203	build
184	succeed
182	declare
171	invade
151	leave