



Interactive visualization of facts extracted from natural language

Anoop Sarkar
Simon Fraser University
Vancouver, Canada
anoop@sfu.ca



Jasneet Sabharwal



Rohit Dholakia



Andrei Vacariu

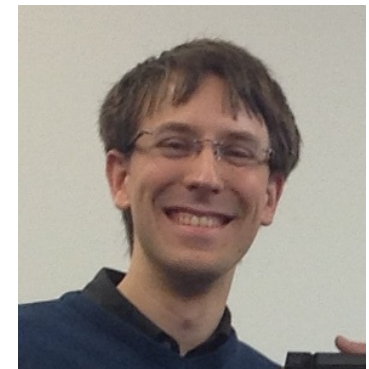
My collaborators on this project



Ramtin Seraj



Maryam Siahbani



Max Whitney

Summarization

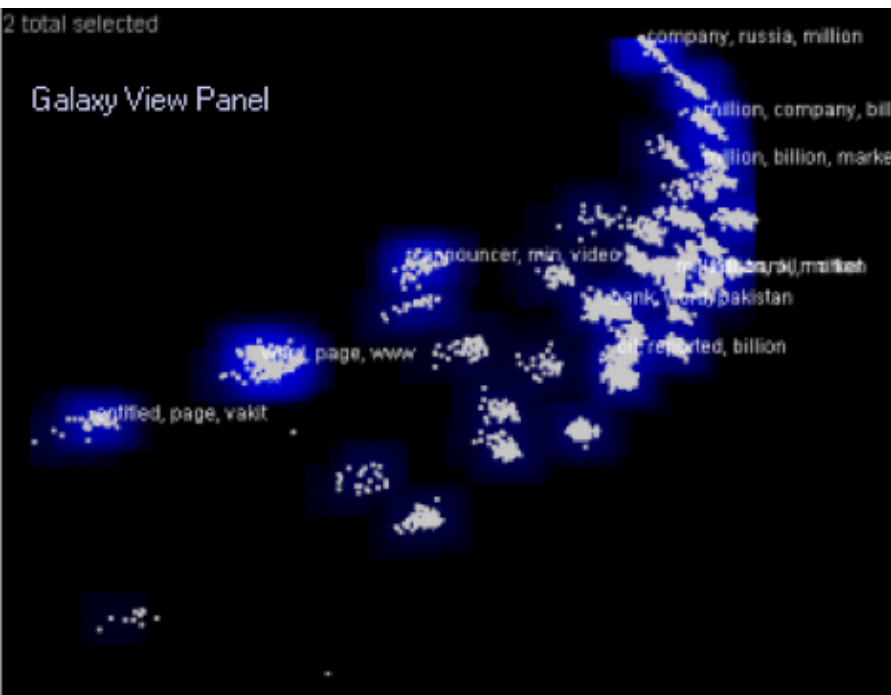
SQuASH: SFU QA Summarization System

Input: 25 news articles, Complex question **Output:** 250-word summary

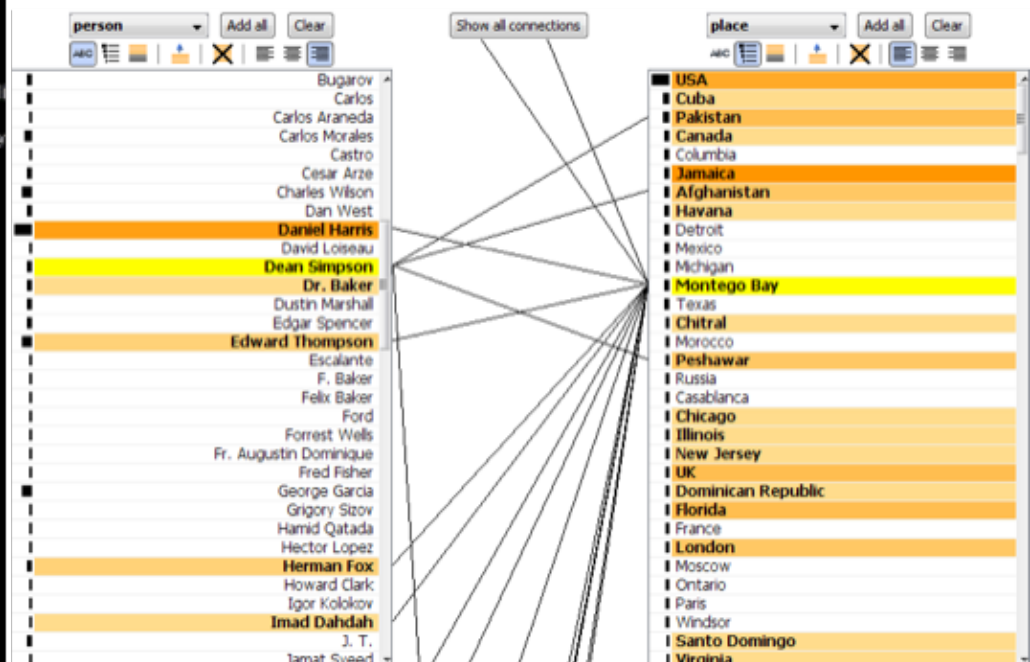
Q. Describe developments in the movement for the independence of Quebec from Canada.

A. Canadian Prime Minister Jean Chretien has ruled out the possibility that the French-speaking province of Quebec could declare independence even if a majority of Quebecers vote for secession in the next referendum. Premier Lucien Bouchard and the separatist Parti Quebecois have retained control of Canada's largest province, possibly setting the stage for a bitter confrontation with the rest of Canada, and possibly another referendum on independence. Canada's Supreme Court managed to rule on the question of Quebec secession without infuriating leaders on either side of a bitter ideological divide. ...

Summarization by exploiting visualization



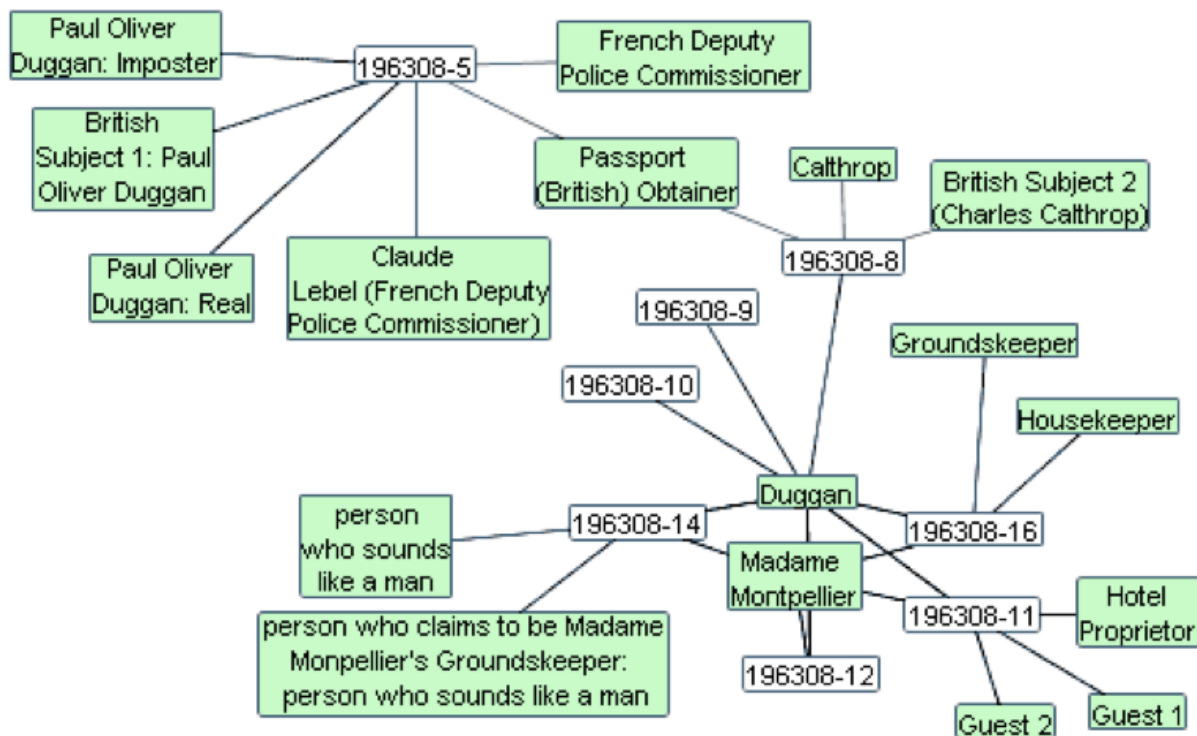
In-spire



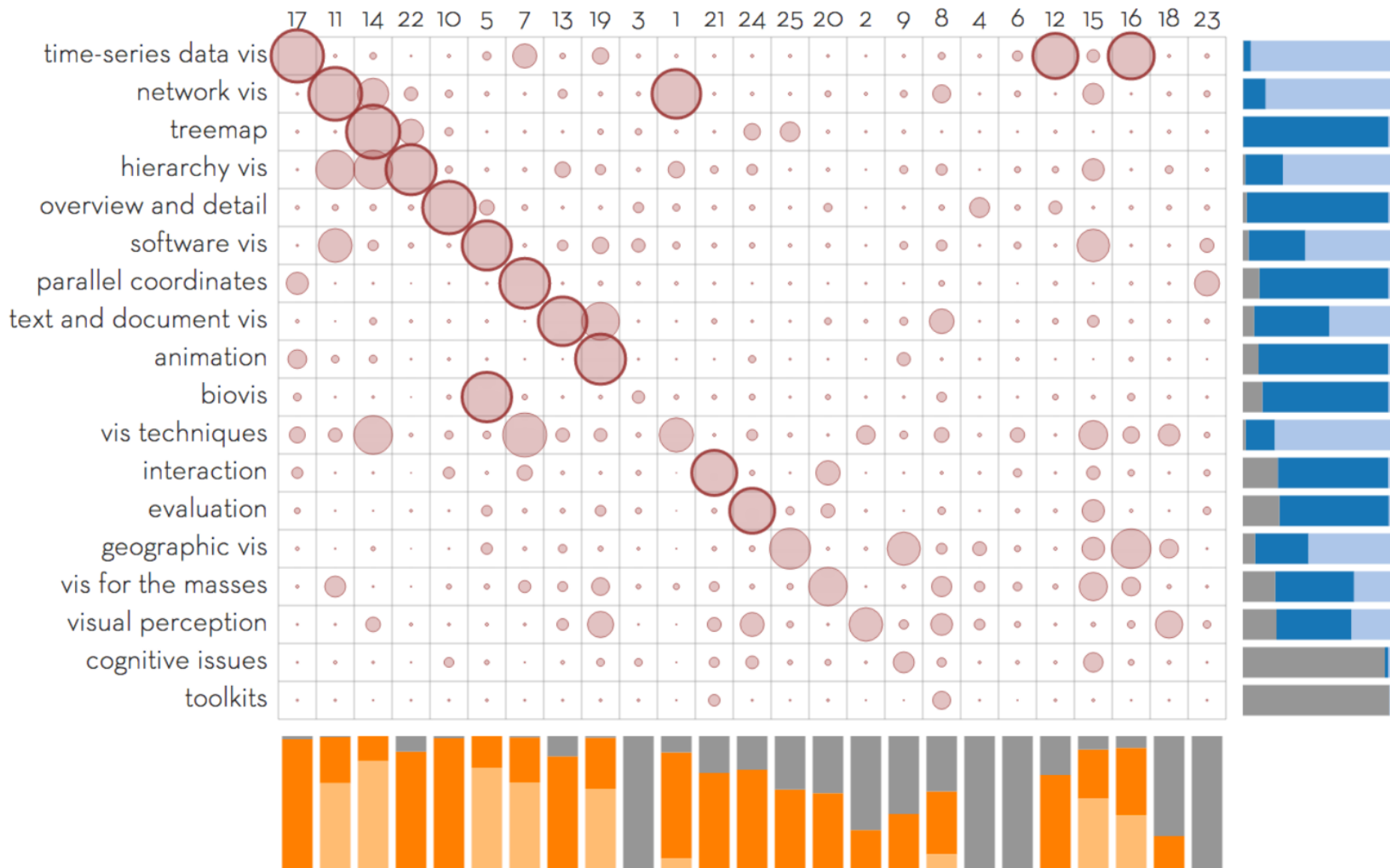
Jigsaw

| Documents | Entities |
|-----------|---------------------------|
| 196308-6 | Housekeeper |
| 196308-7 | Groundskeeper |
| 196308-8 | Madame Montpellier |
| 196308-9 | Duggan |
| 196308-10 | French Police |
| 196308-11 | Madame Montpellier's Home |
| 196308-12 | Aug. 17 |
| 196308-13 | Aug. 16 |
| 196308-14 | |
| 196308-16 | |
| 196308-15 | |

The **French Police** questioned **Madame Montpellier's Groundskeeper** and **Housekeeper** at **Madame Montpellier's Home**. Both said a person fitting the description of the bearer of the **Duggan** passport arrived after 6 pm **Aug. 16** to stay the night and that at 11 am **Aug. 17** they found the body of **Madame Montpellier** in her bed with her bedroom door locked from the inside and the bedroom window open. The groundskeeper said the person had arrived in a car with local license plates and that it was gone in the morning. **Madame Montpellier** had apparently been strangled to death.



CzSaw



Topic Models

Language is tough (to visualize)

SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC
SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC

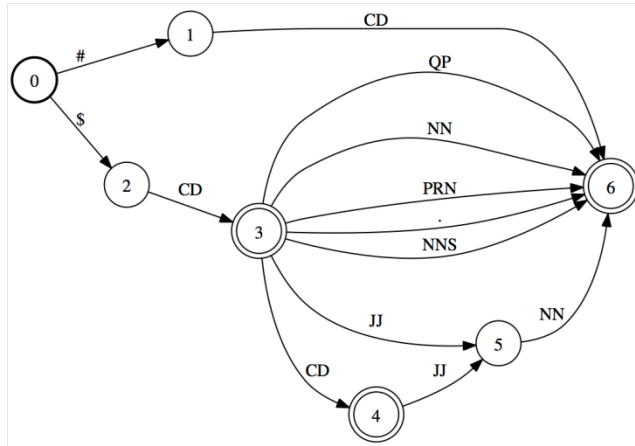
- the engine.
- the engine caught fire.
- Passengers **believed** the engine caught fire.
- Passengers **reported** they saw streaks of flames out of the engine and **believed** the engine had caught fire.



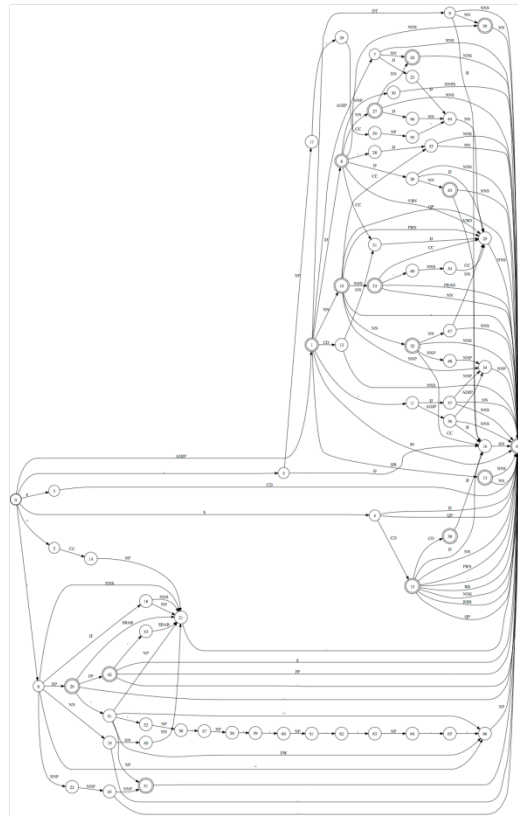
\$ 20 happy meal

\$ CD JJ NN

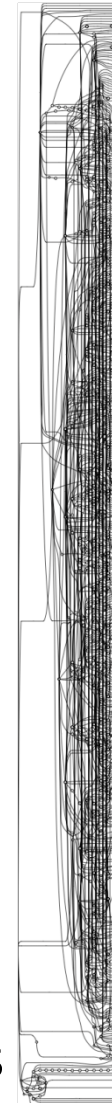
10 noun phrase types



100 noun phrase types



1000 noun phrase types



6K noun phrase types

Language is tough (to visualize)*

- Words to phrases: very high dimensionality
- Language is compositional and ambiguous
- Abstract concepts
 - So difficult to visualize
- **Not** pre-attentive
 - Must foveate to read
- Information can be divided into multiple facets

* i247: Information Visualization and Presentation by Marti Hearst

Our approach

Explore new
visualizations
that exploit
parsed
language

Lensing Language

- Semantic parsing of natural language
- Exploit language understanding: *who did what to whom, where, when and how ...*
- Visualize “Things, not Strings”¹: place spatial, temporal and social entities into an intuitive low dimensional space

1: <http://googleblog.blogspot.ca/2012/05/introducing-knowledge-graph-things-not.html>

Predicate-centric Ontology

- Semantic Role Labeling to extract predicate argument structures
- Verbose labeling
- Automatically populate a novel predicate-centric ontology
 - Killer
 - Corpse
- Using ontology as facet in visualization
 - Entity defeated
 - Entity victorious
 - Easily find uncommon facts
 - Micro-reading in contrast to macro reading

Semantic Role Labeling

- identifying semantic arguments for a verb of a sentence and defining their roles such as who did what to whom, when and where

The boy hit a ball

hit : Predicate

The boy: Agent

a ball: Patient

} Hitter
Semantic Role
Thing hit } Verbose labels

Predicate Sense Disambiguation

Mary gamely **kicked** in **\$5** to **John's** bail.

Kick.01

A0: kicker

A1: thing kicked

A2: instrument

Kick.03

A0: contributor

A1: contribution

A2: given to

Accuracy: 92%

Features:

Parse **Contextual**
tree (POS,...)

Proof of concept:
Visualize facts about human history

Crawl Wikipedia for facts



Article

Talk

Wikipedia Human History

Read

Edit

View

1471 **Time**

From Wikipedia, the free encyclopedia

Events [\[edit\]](#)

January–December [\[edit\]](#)

- [March 1](#) – Emperor [Le Thanh Tong](#) captures the Champa Capital, establishing new regions in middle Vietnam.
- [March](#) – The [Yorkist King Edward IV](#) returns to England to reclaim his throne.
- [April 14](#) – [Battle of Barnet](#): [Edward](#) defeats the [Lancastrian](#) army under [Warwick](#), who is killed.

Descriptions

**Entity Extraction
(NER & hyperlinks)**

Locations

Persons

Google reverse
geo-coding

Country

**Structured
Data**

| | |
|--------|------|
| kill | 2100 |
| found | 1801 |
| defeat | 1637 |

SRL

event1
event2
⋮

predicate
arg0
⋮

**Predicate Sense
Disambiguation**

roleArg0
roleArg1
⋮

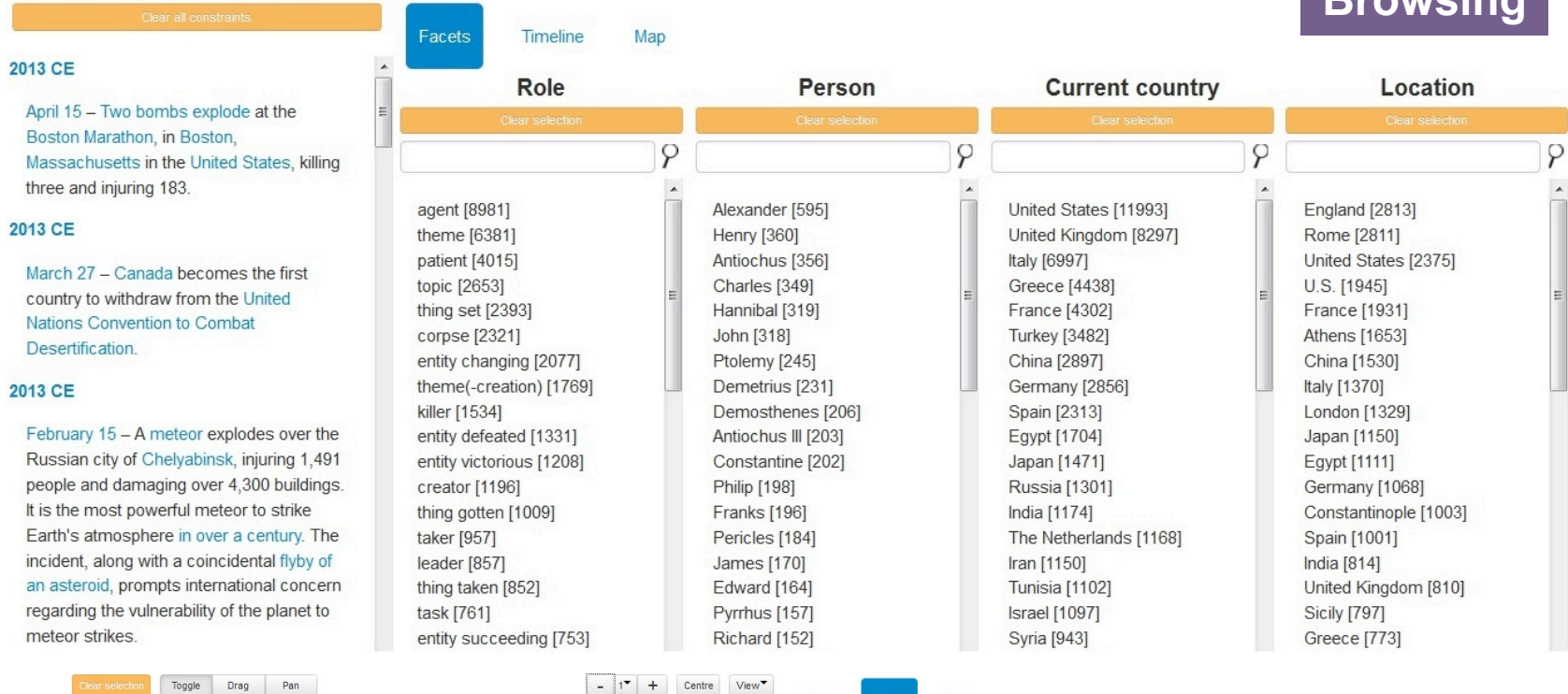
```
{  
  "arg0":      "Emperor Le Thanh Tong",  
  "arg1":      "the Champa Capital",  
  "event":     "capture",  
  "latitude":  21.03,  
  "longitude": 105.85,  
  "country":   "Vietnam",  
  "roleArg0":  "getter",  
  "roleArg1":  "thing gotten",  
  "year":      1471,  
  "person":    "Le Thanh Tong",  
  "location":  {"Champa Capital", "Vietnam"},  
  "description": "March 1 – Emperor Le Thanh Tong captures the Champa  
Capital, establishing new regions in middle Vietnam.",  
  "Wikipedia_categories": {  
    "Vietnamese poets", "Lê Dynasty emperors", "Southeast Asian  
countries", "15th-century monarchs in Asia",...}  
}
```

Domain adaptation

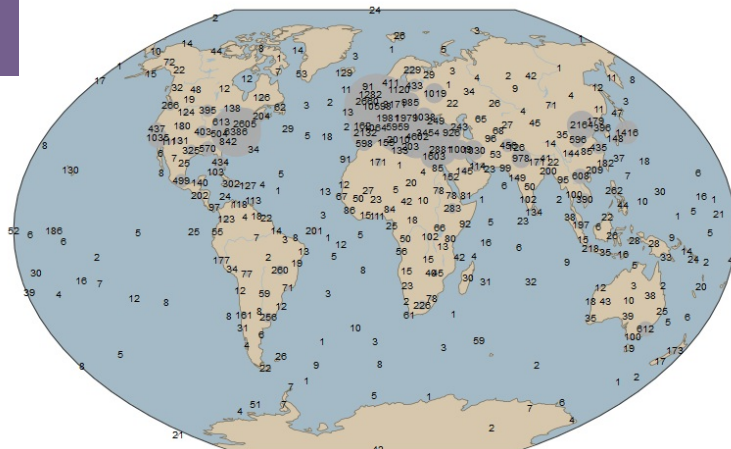
- Map to Wikipedia articles
 - Person
 - Wikipedia categories and infobox: “YEAR births”, “YEAR deaths”, “Kings of*”, “Born”, “Religion”,...
 - Location
 - Latitude & longitude
 - Organization
 - Wikipedia categories and infobox: “Established in*”, “Companies*”, “Founder”, “Headquarters”, “Employees”, ...
- Ongoing work on machine learning for domain adaptation using Wikipedia Categories

Visualization

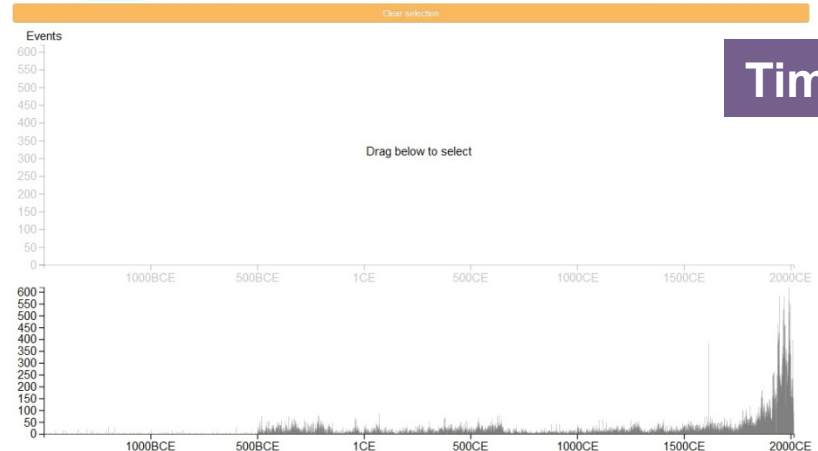
Faceted Browsing



Map

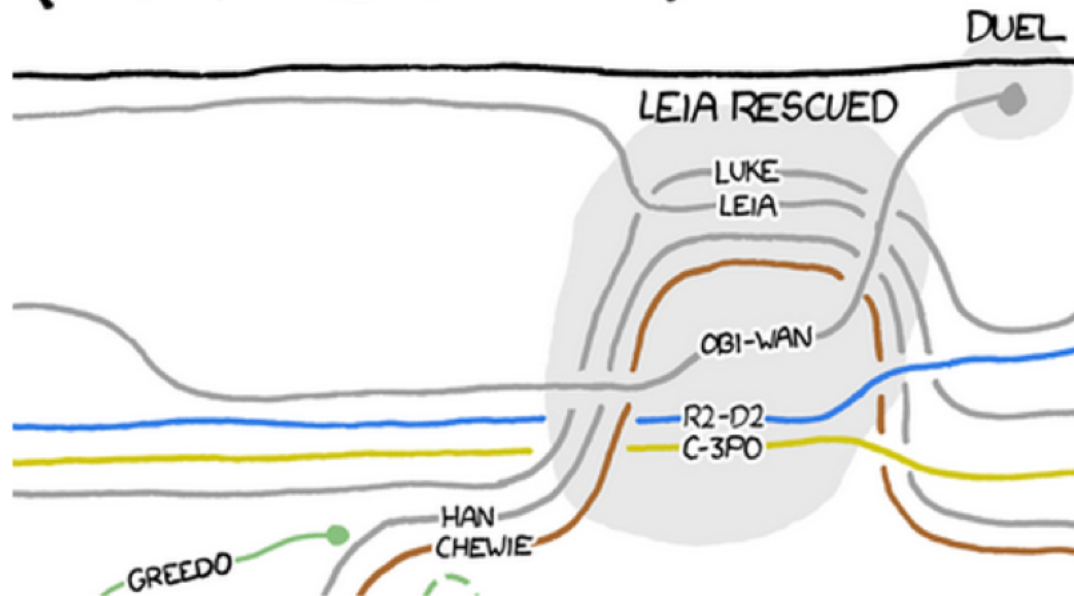


Timeline



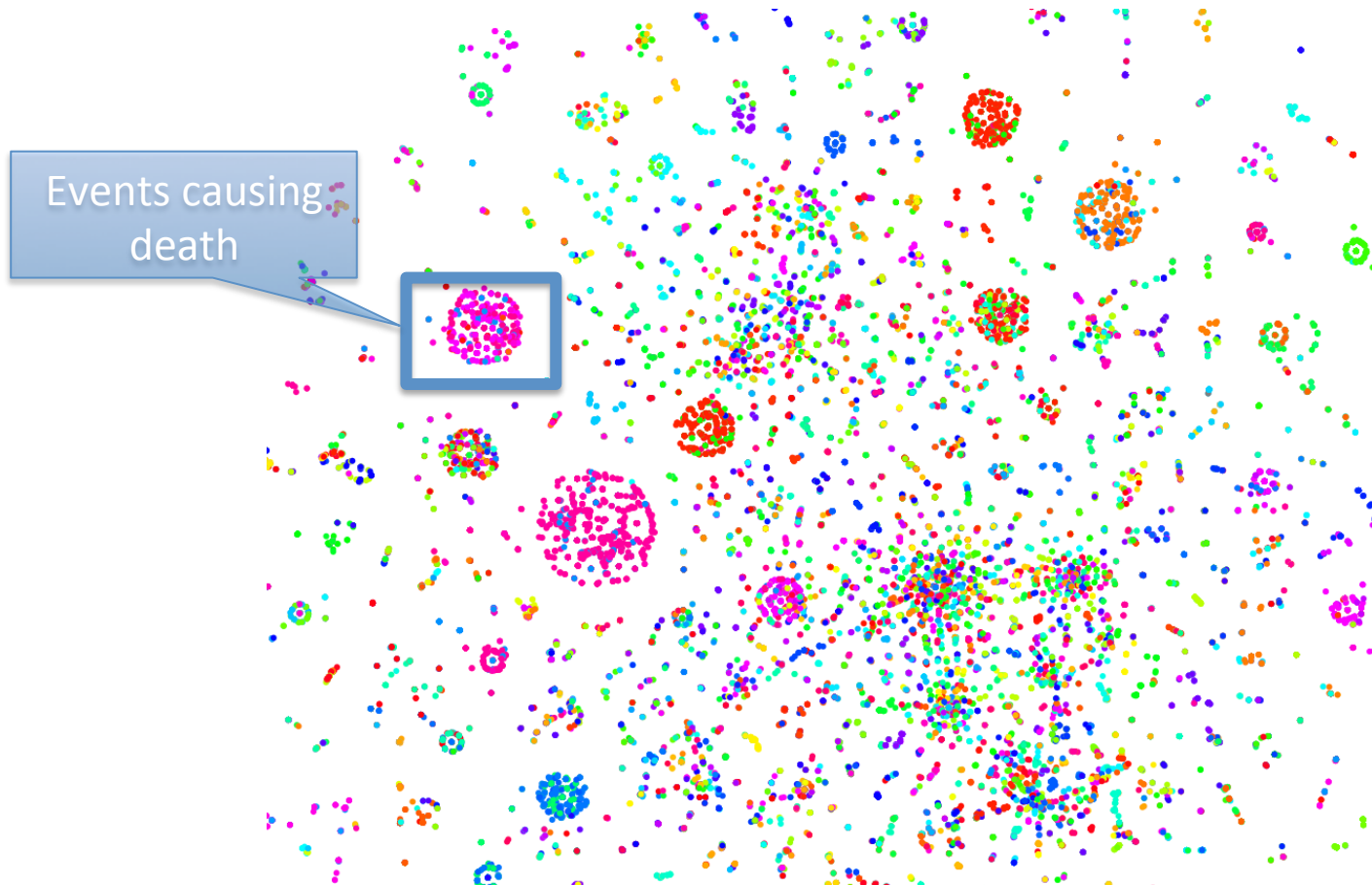
Storyline visualization

STAR WARS (ORIGINAL TRILOGY)

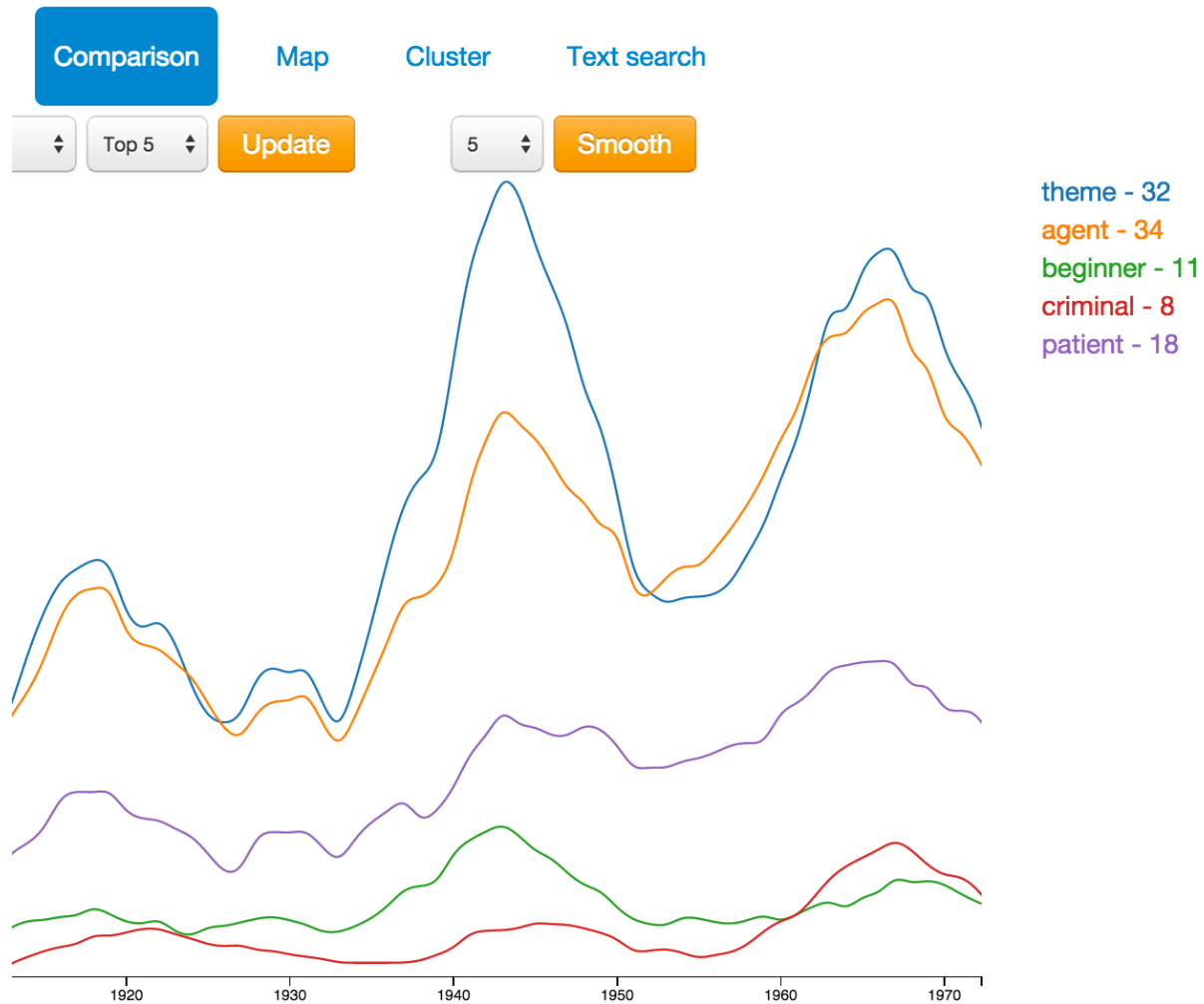


Narrative movie chart from <http://xkcd.com/657> showing entities arranged in a timeline in clusters based on their interactions

2D maps through embeddings



Comparisons



Anoop Sarkar, SFU Natural Language Lab, anoop@sfu.ca

- <http://lensingwikipedia.cs.sfu.ca>
- <http://lensingaviation.cs.sfu.ca>
- <https://github.com/sfu-natlang/lensingwikipedia>