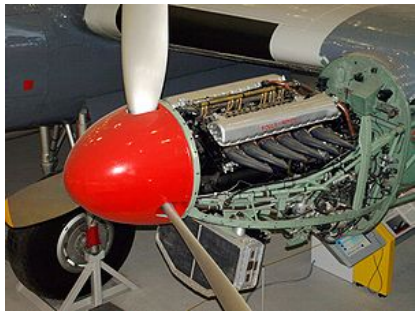
NLP and Text Visualization

SFU Natural Language Lab

http://natlang.cs.sfu.ca/

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.



?

Text is tough (to visualize)*

- Very high dimensionality
  - Tens to hundreds of thousands of features
- Language is compositional and ambiguous
  - Can be combined together in innumerable ways
- Abstract concepts
  - So difficult to visualize
- Not pre-attentive
  - Must foveate to read
- Unordered

* i247: Information Visualization and Presentation by Marti Hearst

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

Predicate-centric Ontology

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

Killer

Corpse

Entity defeated

Entity victorious

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

Semantic Roles

Hitter

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

Accuracy: 92%

Kick.03
A0: contributor
A1: contribution
A2: given to

Features:

Parse tree

Contextual (POS,…)

Visualization

Faceted Browsing

Clear all constraints

Facets | Timeline | Map

2013 CE

April 15 – Two bombs explode at the Boston Marathon, in Boston, Massachusetts in the United States, killing three and injuring 183.

2013 CE

March 27 – Canada becomes the first country to withdraw from the United Nations Convention to Combat Desertification.

2013 CE

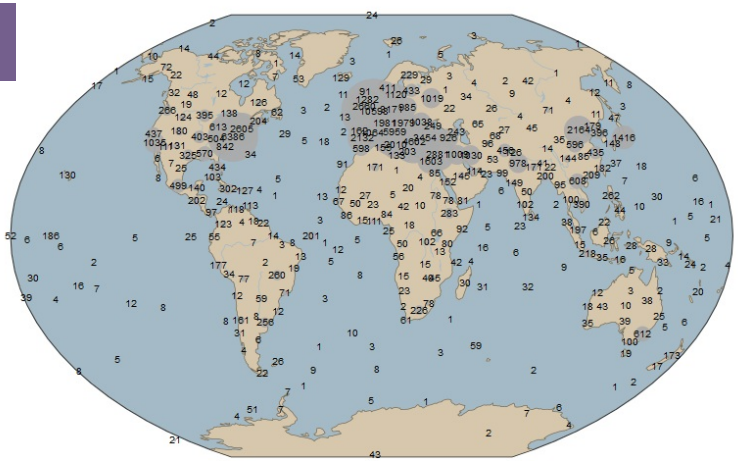
February 15 – A meteor explodes over the Russian city of Chelyabinsk, injuring 1,491 people and damaging over 4,300 buildings. It is the most powerful meteor to strike Earth's atmosphere in over a century. The incident, along with a coincidental flyby of an asteroid, prompts international concern regarding the vulnerability of the planet to meteor strikes.

Role	Person	Current country	Location
Clear selection	Clear selection	Clear selection	Clear selection
agent [8981]	Alexander [595]	United States [11993]	England [2813]
theme [6381]	Henry [360]	United Kingdom [8297]	Rome [2811]
patient [4015]	Antiochus [356]	Italy [6997]	United States [2375]
topic [2653]	Charles [349]	Greece [4438]	U.S. [1945]
thing set [2393]	Hannibal [319]	France [4302]	France [1931]
corpse [2321]	John [318]	Turkey [3482]	Athens [1653]
entity changing [2077]	Ptolemy [245]	China [2897]	China [1530]
theme(-creation) [1769]	Demetrius [231]	Germany [2856]	Italy [1370]
killer [1534]	Demosthenes [206]	Spain [2313]	London [1329]
entity defeated [1331]	Antiochus III [203]	Egypt [1704]	Japan [1150]
entity victorious [1208]	Constantine [202]	Japan [1471]	Egypt [1111]
creator [1196]	Philip [198]	Russia [1301]	Germany [1068]
thing gotten [1009]	Franks [196]	India [1174]	Constantinople [1003]
taker [957]	Pericles [184]	The Netherlands [1168]	Spain [1001]
leader [857]	James [170]	Iran [1150]	India [814]
thing taken [852]	Edward [164]	Tunisia [1102]	United Kingdom [810]
task [761]	Pyrrhus [157]	Israel [1097]	Sicily [797]
entity succeeding [753]	Richard [152]	Syria [943]	Greece [773]

Map

Clear selection | Toggle | Drag | Pan

\- | + | Centre | View



Timeline

Facets | Timeline | Map

Clear selection

Events

Drag below to select

1000BCE 500BCE 1CE 500CE 1000CE 1500CE 2000CE

Framework

Article Talk **Wikipedia Human History** Read Edit Vi

1471 **Time**

From Wikipedia, the free encyclopedia

Events [edit]

~41K descriptions

January–December [edit]

Descriptions

- 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.

Entity Extraction
(NER & hyperlinks)

Locations —(Google reverse geo-coding)→ **Country**

~12K locations
~12K persons

Persons

Knowledge Base

kill	2100
found	1801
defeat	1637

SRL

- event1
  - predicate
  - arg0
  - :
- event2
- :

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",...}
}

~83K events

Verifying Name Entities

- 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 and Wikipedia Categories

Maryam Siahbani

Max Whitney

Demo