



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

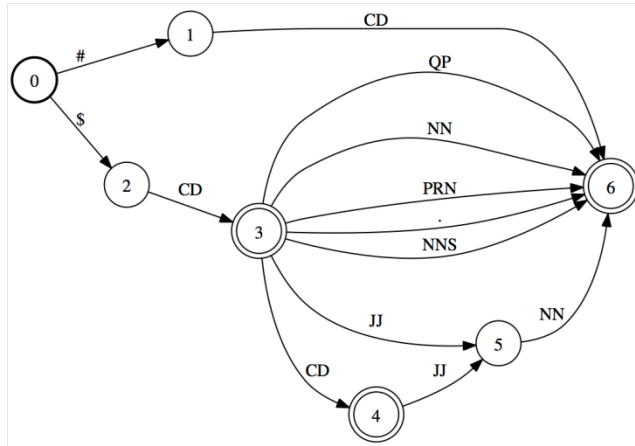
- the engine.
- the engine caught fire.
- Passengers **believed** the engine caught fire.
- Passengers **reported** they saw streaks of flames out of the engine and **believed** the engine had caught fire.



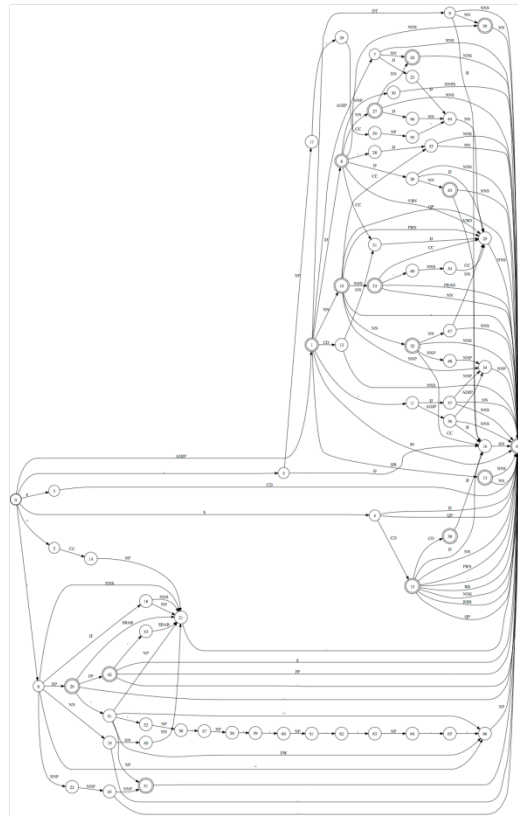
\$ 20 crappy meal

\$ CD JJ NN

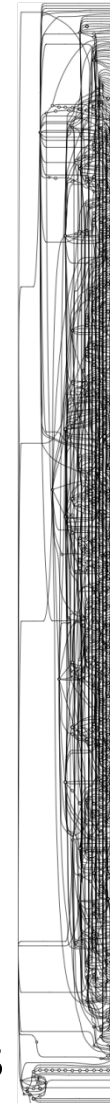
10 noun phrase types



100 noun phrase types



1000 noun phrase types



6K noun phrase types

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
 - Killer
 - Corpse
 - Entity defeated
 - Entity victorious
- Using ontology as facet in visualization
 - 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

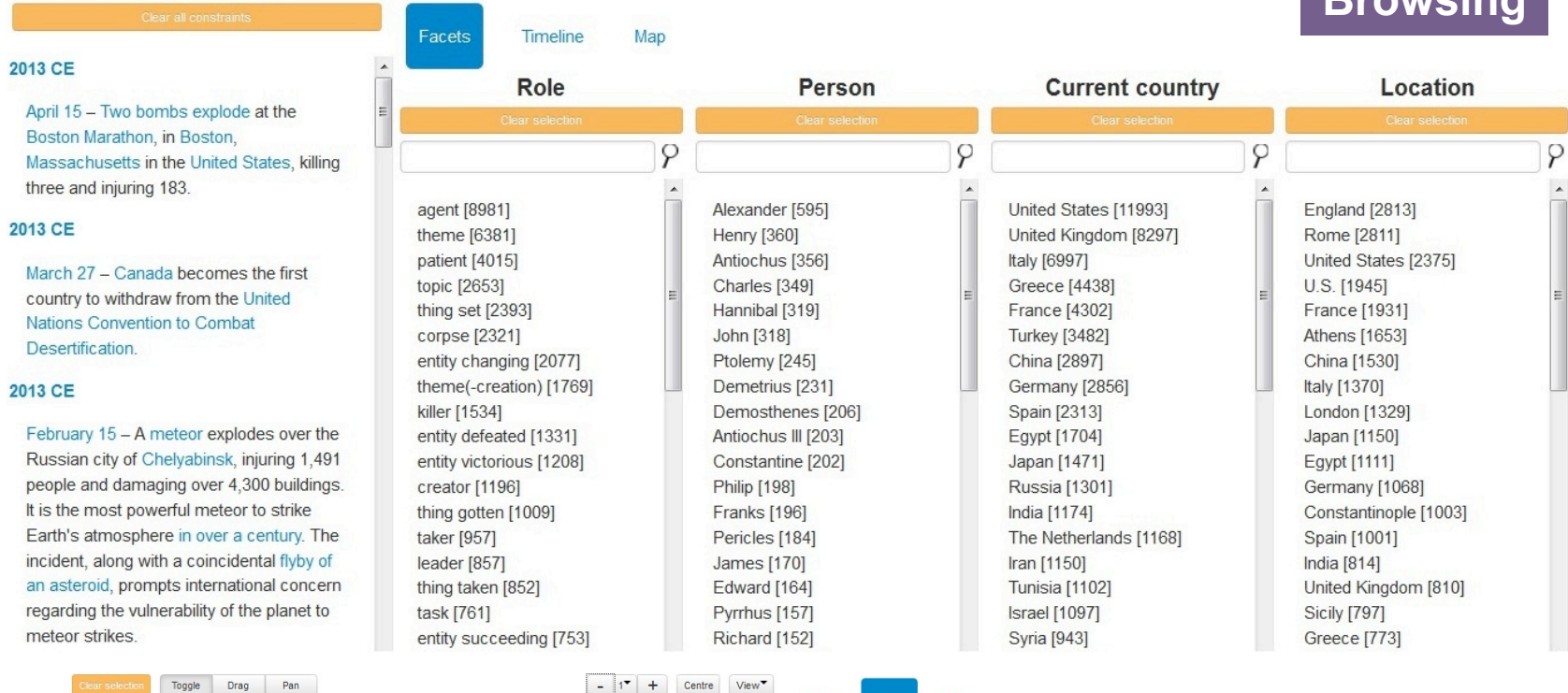
Accuracy: 92%

Features:

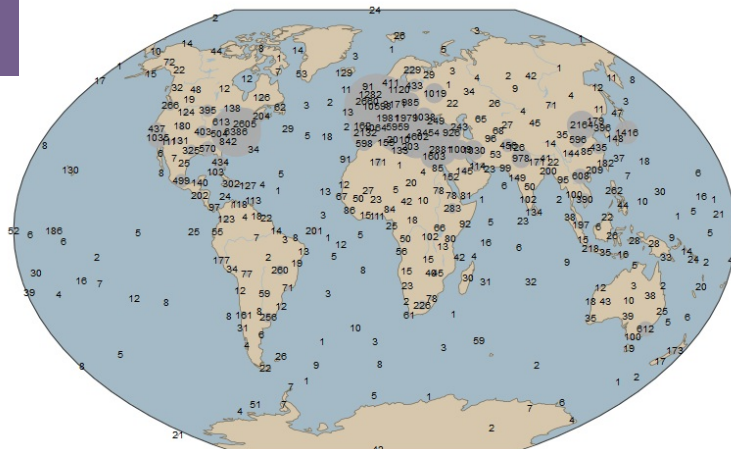
Parse **Contextual**
tree (POS,...)

Visualization

Faceted Browsing



Map



Timeline



Framework



Article

Talk

Wikipedia Human History

Read

Edit

View

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

~12K locations
~12K persons

Persons

Google reverse
geo-coding

Country

Knowledge
Base

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.”,
```

~83K events

```
  "Wikipedia_categories": {  
    "Vietnamese poets", "Lê Dynasty emperors", "Southeast Asian  
countries", "15th-century monarchs in Asia",...}
```

```
}
```

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