Machine Reading of Natural Language and Interactive Visualization

Anoop Sarkar
Simon Fraser University
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 Quebeckers 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.
What’s wrong with this picture?

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 Quebeckers 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. ...

• Difficult to evaluate.
• The summary is static. Not interactive.
• Completely new summary in response to new questions.
• Many different facets in the question and summary:
  • people, locations, predicates (cannot select facets)
• Can NLP help people create better summaries?
Interactive visualization for language summarization
One problem:
Language is tough to visualize
Language is not pre-attentive: must foveate to process the information
$ 20 happy meal
$ CD JJ NN

10 noun phrase types

100 noun phrase types

1000 noun phrase types

6K noun phrase types
• 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.
Our Approach
Summarization using Interactive Visualization

- Exploit (shallow) language understanding: who did what to whom, when, where and how.

- Visualize “things, not strings”*: exploit entity-linking of mentions in the text

- Map text spans to spatial, temporal, social views.

- From high dimensional words, phrases to low dimensional visualizations

*http://googleblog.blogspot.ca/2012/05/introducing-knowledge-graph-things-not.html
Lensing Wikipedia
## Semantic roles and Entities

<table>
<thead>
<tr>
<th>Role</th>
<th>Person</th>
<th>Organization</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clear selection</td>
<td>Clear selection</td>
<td>Clear selection</td>
<td>Clear selection</td>
</tr>
</tbody>
</table>
**Roles and Entity facets**

**1879 CE: attach**

June 1 – Napoléon, Prince Imperial (Napoléon IV), great-nephew of Napoléon Bonaparte, Bonapartist pretender to the French throne, is killed in Africa while attached to the British Army during the Anglo-Zulu War.

**1879 CE: kill**

June 1 – Napoléon, Prince Imperial (Napoléon IV), great-nephew of Napoléon Bonaparte, Bonapartist pretender to the French throne, is killed in Africa while attached to the British Army during the Anglo-Zulu War.

**1840 CE: place**

December 15 – The corpse of Napoleon is placed in the Hôtel des Invalides in Paris.

**1840 CE: bury**

He is buried in Les Invalides.
Location mentions in text

Text “Sparta” :-
Location (lat, long) :-
Map
Timeline of Events
Storylines

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

STAR WARS (ORIGINAL TRILOGY)
On May 12, 1797, Napoleon conquered Venice. The last ruler of Venice, referred to as the Doge, was Ludovico Manin. He was forced to step down as Doge by Napoleon. This event ended 1,100 years of independence for Venice.
From text data to visualization
March 1 – Emperor Le Thanh Tong captures the Champa Capital, establishing new regions in middle Vietnam
March 1 – Emperor Le Thanh Tong captures the Champa Capital, establishing new regions in middle Vietnam

<table>
<thead>
<tr>
<th>predicate</th>
<th>capture</th>
</tr>
</thead>
<tbody>
<tr>
<td>arg0</td>
<td>Emperor Le Thanh Tong</td>
</tr>
<tr>
<td>arg1</td>
<td>the Champa Capital</td>
</tr>
</tbody>
</table>
Semantic Role Labeling

- Semantic role labeling is a well studied NLP task
- Can be modelled as structured prediction
- First step: syntactic parsing of the text to obtain mention spans
- Second step: identify and classify semantic roles based on the syntactic structure
- Trained on the PropBank corpus (1M words of Wall Street Journal) — need for domain adaptation
March 1 – Emperor Le Thanh Tong captures the Champa Capital, establishing new regions in middle Vietnam.
Vietnam

From Wikipedia, the free encyclopedia

March 1 – Emperor Le Thanh Tong captures the Champa Capital, establishing new regions in middle Vietnam.
March 1 – Emperor Le Thanh Tong captures the Champa Capital, establishing new regions in middle Vietnam.
Mary gamely kicked in $5 towards John’s bail.
Predicate sense disambiguation

- Train a classifier to predict the **verbose labels** for each argument of a predicate, “arg0”, “arg1”, etc.

- Exploit metadata written by annotators on the PropBank corpus (1M words of Wall Street Journal).

- Our best classifier obtained 92% accuracy on the PropBank corpus. Performance is lower on Wikipedia.

- Creates a lightweight ontology that is predicate-centric

- Some of the human annotations are a bit strange, e.g. for predicate “resurrect”. 
Demo
User Study
## Interaction Types

<table>
<thead>
<tr>
<th>Interaction Type</th>
<th>Lensing Wikipedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>Faceted browsing of entities</td>
</tr>
<tr>
<td>Control</td>
<td>Remove constraints</td>
</tr>
<tr>
<td>Informational</td>
<td>Storyline, Timeline, Comparison, Map, Follow link to Wikipedia</td>
</tr>
<tr>
<td>Personalization</td>
<td>Return to document, Edit notes</td>
</tr>
</tbody>
</table>
The Task

- User study design based on Vakkari (2003)

- Write a 200 word outline for an essay on relationship between [primary person] and another person in history.

1. Find a small set of important people related to [primary person]. Pre-focus

2. Narrow focus down to one person and study the relationship to [primary person]. Focus formulation

3. Collect 5 Wikipedia articles and write the 200 word summary for the essay. Post-focus
Number of Interactions

- facets tab
- storyline tab
- timeline tab
- comparison tab
- map tab
- remove constraints
- follow link
- return to document
- edit notes

Number of interactions range from 0 to 12.
Useful Features

- facets tab
- storyline tab
- timeline tab
- comparison tab
- map tab
- remove constraints
- follow link
- return to document
- edit notes

Legend:
- Extremely Useful
- Very Useful
- Moderately Useful
- Slightly Useful
- Not Useful
- Not Used
Overall Utility

- **improved performance**
- **enhanced effectiveness**
- **useful for the task**
- **satisfied**

Legend:
- Green: Strongly Agree
- Light Green: Somewhat Agree
- Gray: Neutral
- Light Pink: Somewhat Disagree
- Red: Strongly Disagree
Knowledge gained by user

- John, King of England
- Franklin D. Roosevelt
- Chosen by Participant

None | A Little | A Moderate Amount | A Lot | A Great Deal
Conclusions from user study

• Participants self-reported positive knowledge gain over using text search only.

• Participants found the tool useful and were satisfied
  • Group features that are often used together
  • Re-evaluate the utility of the less-used features
  • Combine storyline, timeline and map for a unified view
  • Produce dynamically generated natural language summaries
Open-source software

github.com/sfu-natlang/lensingwikipedia

$ git clone <repo>
$ docker-compose build
$ docker-compose up
$ make staging
Launch web browser
Thank you!