Rich Interfaces for Reading News on the Web

Earl J. Wagner, Jiahui Liu, Larry Birnbaum and Kenneth D. Forbus

Intelligent Information Laboratory and Qualitative Reasoning Group
Northwestern University
IUI 2009
Results of 2007 AP Study of young-adult news readers

- **Looking for Back Story**: News is “here and now. If you want background, it's up to you.”

- **Want to See the Big Picture**: "Breaking news ...is kind of annoying sometimes. I don't like to get bits and pieces of information."
  
  Another reader “does not want to be fed bits. I want to know all the details at once”

- **Experience NEWS FATIGUE**: “debilitated by information overload”
  
  They “try to ascertain whole news stories but they regularly and repeatedly were left unsatisfied”

- **Want Story Resolution**: sports and entertainment news preferred because they offer “a beginning, middle and end, or clear next steps”
Meeting Information Needs through Structure

- Wikipedia Infoboxes: essential facts for people, organizations, products
- IMDb: artists’ film and television projects
- DBLP, ACM DL, CiteSeer: publications
- ZoomInfo.com: resumes
What Do Structured Views Offer?

- Genres of conventional information presentations
  - (+) Organize knowledge in familiar views
  - (+) Support easy navigation among views and to content elements
  - (-) Constructed manually - scalability?
  - (-) Reader has to go to them
- Easy to read about actor, then film, then director
- Similar flow through the news?
Overview

- **NEWS SITUATIONS**: limited and well-known sequences of causally related events commonly reported in the news

- **Demonstration** of Brussell functionality and user experience

- **How it works**
  - Uses models to automatically extract and aggregate information from articles on many news sites

- **Empirical questions**
Colombia rescues hostage Ingrid Betancourt

BOGOTA: French-Colombian politician Ingrid Betancourt, three Americans and 11 other hostages were rescued from leftist guerillas by Colombian troops on Wednesday after years in captivity, the government said.

All of the former captives were in reasonably good health after being held in secret jungle camps, Defence Minister Juan Manuel Santos said.
Colombia rescues hostage Ingrid Betancourt

Published: July 2, 2008

BOGOTA: French-Colombian politician Ingrid Betancourt, three Americans and 11 other hostages were rescued from leftist guerrillas by Colombian troops on Wednesday after years in captivity, the government said.

All of the former captives were in reasonably good health after being held in secret jungle camps, Defence Minister Juan Manuel Santos said.
Colombia rescues hostage Ingrid Betancourt

Published: July 2, 2008

BOGOTA: French-Colombian politician Ingrid Betancourt, three Americans and 11 other hostages were rescued from leftist guerrillas by Colombian troops on Wednesday after years in captivity, the government said.

All of the former captives were in reasonably good health after being held in secret jungle camps, Defence Minister Juan Manuel Santos said.

Kidnapping of Ingrid Betancourt person by FARC
Colombia rescues hostage Ingrid Betancourt

What happened in this kidnapping?
- About the victim, Ingrid Betancourt person
- About the kidnapper, FARC

Kidnapping of Ingrid Betancourt person by FARC
Colombia rescues hostage Ingrid Betancourt

Published: July 2, 2008

BOGOTA: French-Colombian politician Ingrid Betancourt, three Americans and 11 other hostages were rescued from leftist guerillas by Colombian troops on Wednesday after years in captivity, the government said.

All of the former captives were in reasonably good health after being held in secret jungle camps, Defence Minister Juan Manuel Santos said.
Colombia rescues hostage Ingrid Betancourt

Reuter

Published: July 2, 2008

BOGOTA: French-Colombian politician Ingrid Betancourt, three Americans and 11 other hostages were rescued from leftist guerrillas by Colombian troops on Wednesday after years in captivity, the government said.
Colombia rescues hostage Ingrid Betancourt

Reuter

Published: July 2, 2008

BOGOTA: French-Colombian politician Ingrid Betancourt, three Americans and 11 other hostages were rescued from leftist guerrillas by Colombian troops on Wednesday after years in captivity, the government said.
Sarkozy urges Betancourt release

French President Nicolas Sarkozy has made a direct appeal to Colombian rebels to release Ingrid Betancourt, a long-held French-Colombian hostage.

"I have a dream to see Ingrid among her family this Christmas," Mr Sarkozy said in the televised message.

Last week, Ms Betancourt was shown looking thin and haggard in a video seized by Colombian authorities.

After seeing it, Mr Sarkozy vowed to redouble his efforts to free her after a regional mediation effort failed.
Sarkozy urges Betancourt release

French President Nicolas Sarkozy has made a direct appeal to Colombian rebels to release Ingrid Betancourt, a long-held French-Colombian hostage.

"I have a dream to see Ingrid among her family this Christmas," Mr Sarkozy said in the televised message.

Last week, Ms Betancourt was shown looking thin and haggard in a video seized by Colombian authorities.

After seeing it, Mr Sarkozy vowed to redouble his efforts to free her after a regional mediation effort failed.
Organization: FARC

Collected by Brussell

- **Initials**: Farc (1573)
- **Name**: Revolutionary Armed Forces Of Colombia (2669)
- **Kind**: Terrorist Group (0)
- **Specific Kind**: Faction (1)
- **Military Status**: Militant (74)
- **Political Culture**: Leftist (29)
- **Nationality**: Colombian (18)
- **Home**: Colombia (18)
- **Leader**: Ivan Rios (1)

An aerial view of the former Revolutionary Forces of Colombia, or FARC camp in Angostura, on the Ecuadorean border with Colombia.
How it Works

- **Retrieves** news articles via RSS feeds
- **Selects** articles with situation-specific keywords
- **Extracts** events and facts from text: “X kidnapped Y on Z”
- **Records** textual source
- **Creates** situation instance for event
- **Merges** into set of all situations and participants
- **Links** with user’s news reading context
58% of 100 business articles had at least 1 common situation reference (entity and event verb)
Models Drive Analysis

- Instantiate situations and participants from models
- Constraints associated with models drive analysis at 3 levels
  - **IR**: document keywords: “kidnap*”
  - **IE**: text patterns: “X kidnapped Y on Z”
  - **Semantic**: RELEASED XOR KILLED
- Support for kidnappings, corporate acquisitions, legal trials
Merging Situations and Participants

- **Aggregates results** from multiple articles
- **Indexes situations** by profile: “kidnapping of X”, “A vs. B”
- Problem: Need to reconcile conflicting information
  - Conflicts when incorrect information reported and errors in extraction
  - Impediment to scaling model-based systems
  - Heuristic: correct information most often stated
- **Resolves conflicts by voting** on alternate values for slot
  - Model specifies what is voted on, and what counts as a vote
Scalability of Situation Extraction

- Wikipedia page: “Foreign Hostages in Iraq”
  August 2008
- Minimal descriptions - focus on recall
- 164 cases with named individuals
- 135 mentioned in corpus of 1M articles
- 101 found, recall of 74.8%
Evaluation Results

- Recall improves with more articles:
  5+ references significantly better than just 1
Future Work

- Support rapid authoring of situation models
- Machine learning of templates and text patterns
- Model author supplies semantic constraints
- ...or just confirms if suggested by pattern of occurrence
Background Work
News in Multiple Documents

- **News Summarization**: NewsBlaster, NewsInEssence
  - Differences in details: NewsJunkie
- **News Timelines**: Topic Detection and Tracking (TDT)
  - Google News Archive
- **Question Answering**: KnowItAll
Background Work Models and Interface

- Semantic Models of Situations: Scripts
- Event Extraction: Message Understanding Conference (MUC)
- Rich Interfaces for Viewing Context on Web
  - Get similar pages with Watson
  - View further links with Letizia
  - Browse Semantic Web views with Magpie
AP ethnography concludes: The news "audience is looking to be transported to relevant information no matter its source. Where online consumers once surfed and bookmarked news sites, users now wonder why a logical trail through the news can't simply unfold, link by link, across a multitude of sources."

Brussell presents big-picture views of situations and participants: **EXPLANATORY STRUCTURES**

- **Organize knowledge** of situation storylines and participant facts
- **Support navigation** among news articles about milestone events
- **Created automatically** from models by reading articles retrieved from multiple news sites
- **Links within browser** to user’s news reading context
<table>
<thead>
<tr>
<th>Participant Detail</th>
<th>Event Occurrence</th>
<th>Event Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quartile</td>
<td>Mean</td>
<td>Std Dev</td>
</tr>
<tr>
<td>1 (A)</td>
<td>0.21604</td>
<td>0.34483</td>
</tr>
<tr>
<td>2 (AB)</td>
<td>0.41532</td>
<td>0.39324</td>
</tr>
<tr>
<td>3 (BC)</td>
<td>0.62272</td>
<td>0.33386</td>
</tr>
<tr>
<td>4 (C)</td>
<td>0.73468</td>
<td>0.23512</td>
</tr>
</tbody>
</table>

Comparisons for all pairs using Tukey-Kramer HSD (q=2.61413, alpha=0.05)
Letters not connected by the same letter are significantly different.

http://www.nytimes.com/2008/07/02/world/americas/04Colombia.html
About 3750 article results for “Ingrid Betancourt kidnapping” in clusters
About 5050 dated article results for “Ingrid Betancourt kidnapping”

Top ten includes date of abduction and appeal for release among other results in non-chronological order
"Taking Topic Detection From Evaluation to Practice", Allan et al. HICSS'05
## Situation Model Comparison

<table>
<thead>
<tr>
<th>Situation Model</th>
<th>Kidnapping</th>
<th>Corporate Acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants</strong> (Focal)</td>
<td>Victim, Kidnapper, Rescuer, Finder</td>
<td>Acquiree, Acquirer, Regulator</td>
</tr>
<tr>
<td><strong>Events</strong></td>
<td>Abduction, Threat, Videotape, Appeal for Release, Release, Escape, Killing</td>
<td>Talks, Offer, Announce,Gov’t Approval, Shareholder, Approval, Close</td>
</tr>
<tr>
<td><strong>Semantic Constraints</strong></td>
<td>First: Abduction, Release XOR Killing</td>
<td>Gov’t Approval BEFORE Close</td>
</tr>
</tbody>
</table>