STORIES in time: a graph-based interface for news tracking and discovery

Bettina Berendt
and Ilija Subašić
K.U. Leuven, Belgium
www.berendt.de
Idea

• how does a "story" develop (in news, blogs,...)?
• graphical multi-document summarization (+ text)
• interactive, dynamic customisation of content and detail => individual sense-making
• couplings between story and underlying documents
• mash-up: Google News, Open Calais, SIMILE
Context

- Family of tools focusing on interactive sense-making helped by data mining
- PORPOISE: global and local analysis of news and blogs + their relations; some multilinguality
- CiteseerCluster: literature search + sense-making
- Damilicious: CiteseerCluster + re-use/transfer of semantics + diversity
Method

• Sources
• Who & What
• Basics
• The story unfolds
• English, please!
• Time-indexed web documents

**Sources**

- Time-indexed web documents
- Top TF.IDF terms/NEs, in a bursty co-occurrence
- Bursty := relative frequency of co-occurrence in t / relative frequency of co-occurrence overall > \( \theta \)
- Time slices \( t \) (may overlap)
- Top "facts" (semantic role labelling) indexed by edges
Sources * Who & What

- Time-indexed Web documents
- Top TF-IDF terms/NEs, in a bursty co-occurrence

Barack Obama health

birth congress
Sources * Who & What * Basics

- Time-indexed Web documents
- Top TF.IDF terms/NEs, in a bursty co-occurrence
- Bursty := relative frequency of co-occurrence in $t$ / relative frequency of co-occurrence overall $> \theta$

Barack Obama  health

birth  congress
Sources * Who & What * Basics * Story unfolds

- Time-indexed Web documents
- Top TF.IDF terms/NEs, in a bursty co-occurrence
- Bursty := relative frequency of co-occurrence in t / relative frequency of co-occurrence overall > θ

Barack Obama - health

birth - congress
Sources * Who & What * Basics * Story unfolds

- Time-indexed Web documents
- Top TF.IDF terms/NEs, in a bursty co-occurrence
- Bursty := relative frequency of co-occurrence in \( t \) / relative frequency of co-occurrence overall > \( \Theta \)
- Time slices \( t \) (may overlap)
11 Sep: Today President Obama delivered his 30th speech on health care reform following his speech to a joint session of Congress last night.

12 Sep: President Barack Obama says he'll be held responsible for any problems once a health care overhaul becomes law.
• Scan: Give me an overview of the Enron story!

(To understand the following screenshots based on the slides alone: check [http://en.wikipedia.org/wiki/Timeline_of_the_Enron_scandal](http://en.wikipedia.org/wiki/Timeline_of_the_Enron_scandal))
Reports from a bad time for Enron ...
Tool

- **Scan:** Give me an overview of the Enron story!
- **Zoom:** What happened that weekend?
• **Scan**: Give me an overview of the Enron story!
• **Zoom**: What happened that weekend?
• **Uncover**: More details please!
• **Scan**: Give me an overview of the Enron story!
• **Zoom**: What happened that weekend?
• **Uncover**: More details please!
• **Track**: Tell me more about the employees!
Tool

• Scan: Give me an overview of the Enron story!
• Zoom: What happened that weekend?
• Uncover: More details please!
• Track: Tell me more about the employees!
• Story search: Give me your sources!
The rapid decline of the Enron Corporation has devastated its employees' retirement plan, which was heavy with company stock, and has infuriated workers, who were prohibited from changing their investments as the stock plunged. Through the 401(k) retirement plan, employees chose to put much of their savings in Enron shares, and the company made contributions in company stock as well. But around the time Enron disclosed serious financial problems last month, the company froze the assets in the plan because of an administrative change. For several weeks, as the stock lost much of its value, workers stood by helplessly as their retirement savings evaporated. They were not allowed to switch investments at all -- even though the plan had far less risky choices. The unfortunate timing caps a year of pain for Enron's workers. At the end of last year, the 401(k) plan had $2.1 billion in assets, more than half invested in Enron, an energy conglomerate. Since then, the stock has lost 94 percent of its value. At Portland General Electric, the Oregon utility acquired by Enron four years ago, some workers nearing retirement have lost hundreds of thousands of dollars. The utility has lined up grief counselors to help them work through their problems. "We had some married couples who both worked who lost as much as $800,000 or $900,000," said Steve Lacey, an emergency-repair dispatcher for Portland General. "It pretty much wiped out every employee's savings plan." "Shortly after it was frozen, the articles started coming out about some of the questionable activities of Enron," Mr. Lacey added. "The stock took a tremendous drop, and we invested money and stock in the network. We contrasted the chairman's more than $12 million at the time of a stock buyout with the stock buyout of a second Enron company. We filed a legal protest, but Enron's legal team refused to listen. We were forced to sell the stock through our retirement plan. The lawsuit says that Enron schemed to pump up the price of the stock artificially and violated its fiduciary duty to its employees by failing to act in their best interests. "They were promoting Enron as a retirement investment vehicle and matching employees' contributions with Enron stock, when they knew the stock was overvalued, and that's a breach of their fiduciary duties," Mr. Berman said in an interview yesterday. What's more, he said, the assets were frozen on Oct. 17, with the stock at $32.20, even though Enron executives knew there would be imminent disclosures about the company's accounting practices. "They knew the worst news was about to come out, but they froze the stock," he said. Enron closed yesterday at $5.01, the company declined to comment on much of the allegations because of pending litigation. A spokeswoman, Karen Denne, said that the change in plan administrators had been in the works for a number of months and that she did not know the exact date the change was put into effect. A version of this article appeared in print on Thursday, November 22, 2001, on page A1 of the New York Times.
Tool

- Scan: Give me an overview of the Enron story!
- Zoom: What happened that weekend?
- Uncover: More details please!
- Track: Tell me more about the employees!
- Story search: Give me your sources!
- English, please: Give me a textual overview!
STORIES FACTS

- Traders snapped up retail stocks
- Baruch Lev asks accounting industry's watchdog group is examining this
- Daniel Ben-Ami
- Enron Creditors Forum a merger deal with Dynegy Inc. that is seen as Enron's last-ditch attempt to escape bankruptcy
Evaluation

- Major challenge: what is the ground truth?
  - Inter-rater agreement! - Wikipedia
  - Benchmark datasets (DUC, TREC, not completely applicable)
- Information retrieval: ~ 80% recall, ~ 33% precision
- Story-graph topological properties meaningful?
  - Global properties of the graphs (graph size, # connected components, existence of nodes with high degree centrality) useful for predicting “real” eventfulness
- Fact extraction: ~ current multidocument summarisation approaches
- Search:
  - search finds coherent subsets of documents
  - quality ≥ state-of-the-art clustering
  - Tool enables people to answer Qs on ground-truth events accurately and quickly
Outlook

• more advanced language processing
  – linguistic parsing, semantic role labelling for story graphs,...
• lexical resources and other background knowledge
• different sources of media bias/viewpoints.
• aggregation and analysis dimensions other than time, such as multilinguality
• evaluation: good results; further studies planned
Thanks

• ... also from the co-author!

Ilija Subašić