

Describing Story Evolution from Dynamic Information Streams

Stuart Rose, Scott Butner, Wendy
Cowley, Michelle Gregory, Julia Walker

Pacific Northwest National Laboratory

Kalpesh Padia

CS 796/896 - Spring 2011, Old Dominion
University

Introduction

- Information
 - Definition
 - Information Stream
- Dynamic information streams
- Document Corpus

Introduction

- Ever changing document corpus
- Few techniques exist to identify and understand trends

<http://newstimeline.googlelabs.com/>

- Burden left to user

The paper describes..

- A novel text analytic framework
- A data model for describing themes and tracking stories
- Visual representation of story flow

Related Work – Theme River

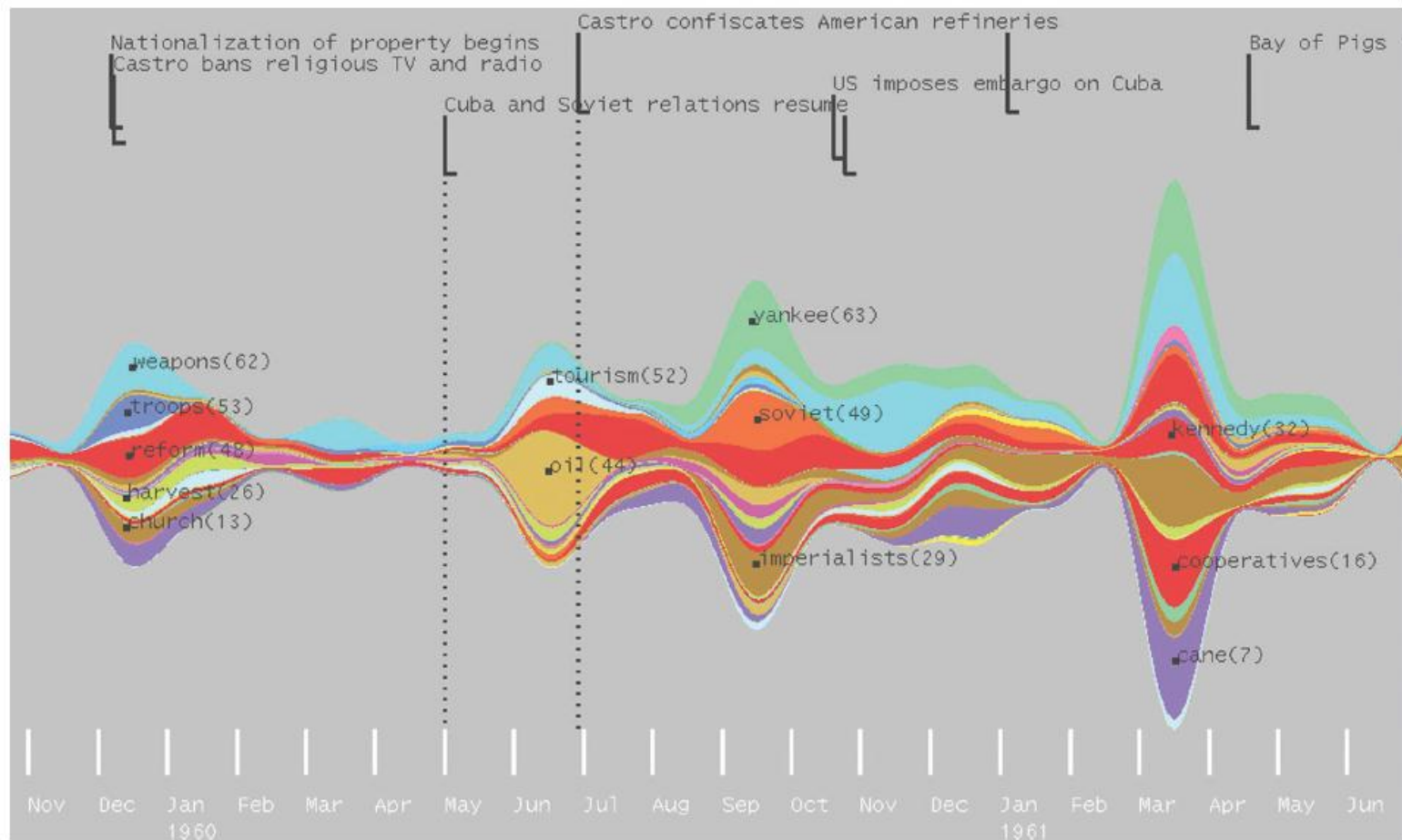


Fig. 2. ThemeRiver uses a river metaphor to represent themes in a collection of Fidel Castro's speeches, interviews, and articles from the end of 1959 to mid-1961.

Taken directly from Paper's reference [7]

Related Work – History Flow

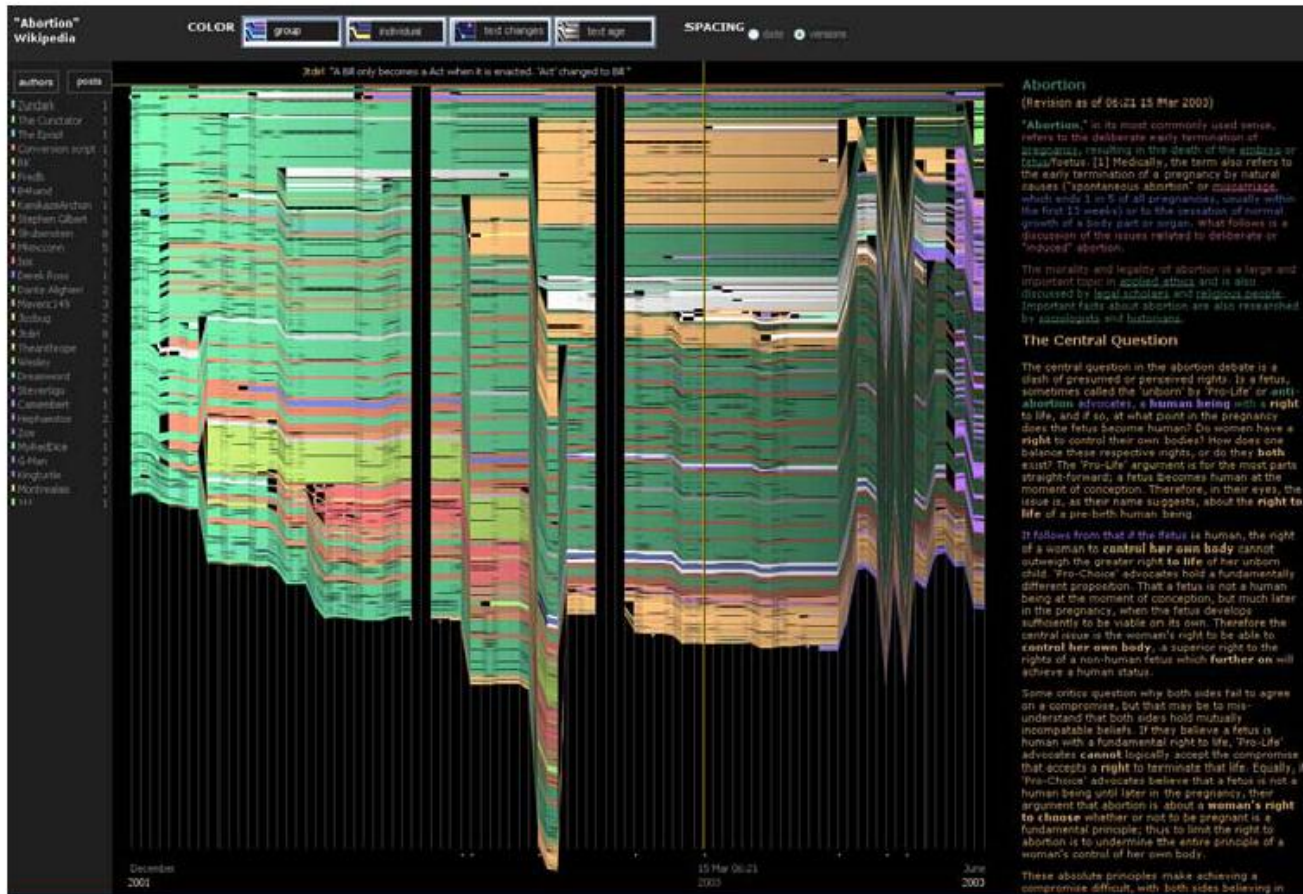


Figure 11-7. Editing history for "Abortion" showing versions equally spaced—black gutters represent "mass deletions," an act of vandalism whereby a user deletes all content in a given article

Paper's reference [17], Taken from Book "Beautiful Visualization" pp. 185

Related Work – Others

- Memetracker (<http://memetracker.org>) (paper's reference [11])
- Themail (paper's reference [16])
- Visualizing tags (paper's reference [3])

Analytic Methods

- Corpus Oriented Methods
 - Statistical discriminating features between documents
 - Likely to overlook features as they first appear
 - Not suited for dynamic corpus

Analytic Methods

- Document Oriented Methods
 - Keyword extraction on individual documents
 - Context-independent document features
 - Suited for dynamic corpora

Described Framework

Describing Stories

CAST

RAKE based feature selection

RAKE based feature selection

<http://www.voanews.com/english/archive/2008-02/2008-02-21-voa21.cfm>

Pakistani Opposition Parties Agree on Ruling Coalition

By Barry Newhouse

Islamabad

21 February 2008

The leaders of Pakistan's two main opposition parties have agreed to form a coalition in the national assembly that will control a majority of seats. VOA's Barry Newhouse reports the opposition leaders say they have come to an agreement on the crucial issue of reinstating the Supreme Court that Mr. Musharraf dismissed in November.

Nawaz Sharif, right, speaks to reporters as Asif Ali Zardari sit next to him at press conference after their meeting in Islamabad, 21 Feb 2008

At a joint news conference in Islamabad, Pakistan People's Party leader Asif Zardari and Nawaz Sharif of the Pakistan Muslim League-N told reporters the two parties would work together on what they called a government of national consensus. Zardari said they are focused on strengthening Pakistan's democracy.

"We intend to be together in the parliament," he said, "We have, insha'allah, a future of democracy in our grasp. We will strengthen the parliament, we will make a stronger Pakistan."

While the two parties had campaigned on a similar agenda that criticized President Pervez Musharraf and his ruling Pakistan Muslim League-Q party, analysts said there were indications they had different views on the issue of reinstating senior judges dismissed by Mr. Musharraf in November.

Nawaz Sharif read reporters a prepared statement on the position that both parties have agreed to.

"In principle there is no disagreement on the restoration of the judiciary. We will work out the modalities in the parliament," he said.

Lawyers who demonstrated in major cities in Pakistan had demanded the immediate reinstatement of the dismissed senior judges. Instead, it appears the parties will wait until they take control of parliament in the coming weeks before taking up the issue.

The judges' dismissals have been the central political issue for Pakistan Muslim League-N leader Nawaz Sharif. Earlier Thursday, he addressed hundreds of supporters outside the home of the fired Supreme Court Chief Justice, Iftikhar Mohammed Chaudry, saying President Musharraf's dismissal of the Supreme Court was illegal.

Chaudry has been under house arrest since November.

The PPP's Zardari also said the two parties are focused on building a broad coalition in parliament.

"We are trying to come up with a national consensus government with all political forces in and outside of the parliament," said Zardari.

The coalition brings the two parties closer to gaining a two-thirds majority in parliament. The super-majority is needed if the parties try to impeach President Musharraf.

Extracted Keywords

pakistan muslim league-n leader nawaz sharif,
pakistan people's party leader asif zardari,
pakistan muslim league-n told reporters, ruling pakistan muslim league-q party,
nawaz sharif read reporters, pakistani opposition parties agree,
fired supreme court chief justice, asif ali zardari sits,
reinstating senior judges dismissed, main opposition parties, nawaz sharif,
stronger pakistan, pakistan, dismissed senior judges,
criticized president pervez musharraf,
saying president musharraf's dismissal, supreme court

Figure 1: Sample document from Voice of America and its key-words extracted by RAKE.

Taken directly from paper

CAST

- Measure association of each keyword to documents (lucene index)
- Use Sorensen similarity coefficient to calculate similarity between each pair of keywords
- Ward's hierarchical clustering(paper's reference [18])
- Clusters with high coherence threshold (0.65 or greater) selected as candidate themes

CAST on MPQA Corpus

Table 1: MPQA Topics and CAST Themes. Each row contains the defined topic and description within the Defined Topics column and the labeled theme and top keywords within the CAST Themes column.

Defined Topics	CAST Themes
<i>mugabe</i> 2002 presidential election in Zimbabwe	<i>mugabe's re-election</i> (46) zimbabwe's election, mugabe, mugabe's, president mugabe, mugabe's government
<i>guantanamo</i> U.S. holding prisoners in Guantanamo Bay	<i>guantanamo prisoners</i> (43) detainees as prisoners, detainees prisoners, prisoners of war, war prisoners, prisoners
<i>kyoto</i> ratification of Kyoto Protocol	<i>kyoto protocol on climate</i> (40) kyoto protocol, ratification of the kyoto protocol, 1997 kyoto protocol
<i>venezuela</i> presidential coup in Venezuela	<i>venezuela's president hugo chavez</i> (37) venezuelan president hugo chavez, president hugo chavez, president chavez, hugo chavez
<i>settlements</i> Israeli settlements in Gaza and West Bank	<i>israeli</i> (34) palestinian, israel, occupied palestinian, israeli occupation, palestinian people's
<i>taiwan</i> relations between Taiwan and China	<i>china under which taiwan</i> (31) taiwan, united states and taiwan, taiwan issue, taiwan policy, taiwan affairs, taiwan strait
<i>axisofevil</i> reaction to President Bush's 2002 State of the Union Address	<i>axis of evil</i> (28) north korea as axis, iran and north korea, iran or north korea, iraq and north korea,
<i>humanrights</i> reaction to U.S. State Department report on human rights	<i>human rights</i> (28) human rights report, annual human rights report, human rights violations, rights
<i>spacestation</i> space missions of various countries	<i>space station</i> (28) international space station, space, international space, space shuttle, russian space mission
<i>argentina</i> economic collapse in Argentina	<i>argentina</i> (12) argentina's, help argentina, argentine government, argentine, worried argentina

Taken directly from paper

Describing Stories

- Select a set of keywords at end of current time interval
- Measure keyword document association
- Cluster keywords into themes and assign each document occurring over past n intervals to these themes
- Overlapping themes describe story evolution

Describing Stories - Example

Table 2: CAST Themes for 01/12/1998 and their document counts

Assigned Docs		Theme
< 01/12	01/12	
1	5	chuan government
0	3	serena williams who is playing
0	3	men's match
0	2	stabilizing japan's shaken financial system
0	1	women's race
1	2	news agency
3	2	five-kilometer race
3	1	northern ireland
5	1	japanese prime minister ryutaro hashimoto
6	1	world cup
7	1	president suharto
9	3	world swimming championships
15	4	hong kong

Taken directly from paper

Visual Representation

Table 3: Documents Assigned to CAST Theme *iraq* on 02/14/1998

- Tabular View
 - Present each theme as a chronological listing of its constituent documents
 - Lacks information on how the story/theme evolved

Day	Document Title
02-07	U.N. to excavate weapons dumps in Iraq
02-07	Kuwaiti government to distribute gas masks in two days
02-07	U.N. to excavate dumps where chemical weapons and warheads buried
02-08	Saudi Arabia will not back military strike against Iraq
02-08	Egyptian Foreign Minister upbeat on diplomatic end to US-Iraq
02-08	Russian FM: Dispute over Iraq has not damaged U.S.-Russian
02-08	U.S. won't ask Saudis to allow attacks from their territory
02-08	Iraq says talks on weapons destruction making progress
02-09	Citing regional tensions, Annan cancels visit to the Middle East
02-09	Canada: No decision yet on commitment against Iraq
02-09	byline
02-09	control on Yeltsin-Annan remarks, other new material
02-09	Iraq launches campaign to rally Arab support
02-09	Iraq crisis at 'critical stage,' U.N. chief pushes for peaceful
02-10	Report: Lebanon tells Iraq to cooperate with weapons inspectors
02-10	Iraqi foreign minister seeks Syrian support in standoff
02-10	Precede MAJDAL SHAMS
02-10	Iraqi ambassador says Baghdad cannot meet Annan's proposed sales
02-10	Israel short on gas masks for children
02-11	Iraqi daily says Washington always determined to use force
02-11	Britain rejects latest Iraqi offer on weapons
02-11	Iraq puts oil export potential at 1.6 million barrels daily
02-11	Gulf foreign ministers meet to discuss unified stand on Iraq
02-11	BC-Iraq-Opposition
02-11	Santer says EU position will change if Iraq stays defiant
02-12	Iraqi foreign minister calls U.S. rejection 'a bluff'
02-12	UN advises vacationing staff to stay away from Iraq
02-13	American, Russian defense chiefs hold cordial talks at military
02-13	Iraq accuses U.S. of psychological warfare
02-13	U.S. envoy meets with Japanese leaders on Iraq
02-13	Foreign Office urges 'higher degree of caution' for travelers
02-13	Iraq dispute underlines strain in U.S.-Russia ties
02-14	U.S. envoy seeks China's support on Iraq
02-14	Demonstrators at U.S. Embassy in Japan urge no military attack
02-14	Iraq urges diplomacy, releases Egyptian prisoners

Taken directly from paper

Visual Representation

- Use top keywords in a theme as theme's labels
- Use vertical columns to depict time intervals. Within each time interval, list identified themes in decreasing order of their size.
- Use the title of the document only once when it first appears in a theme. Thereafter, represent it using a line.

Visual Representation

- A line spanning across columns represents document flow and allows for visualization of story evolution
- Visualization presented in paper.

Conclusion

- Analytic framework for theme generation and visualizing story evolution.
- Novel story evolution visualization which conveys temporal context in dynamic document corpus

References

All figures are from the presented paper or references mentioned therein

- Presented Paper: Rose, S., Butner, S., Cowley, W., Gregory, M. and Walker, J., "[Describing story evolution from dynamic information streams](#)," In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*. October 2009, pp. 99 -106
- Addition figures: Julie Steele and Noah Iliinsky. Beautiful Visualization: Looking at Data through the Eyes of Experts. O'Reilly Media, 1 edition, June 2010
- <http://newstimeline.googlelabs.com/>

Comments

- Individual and small stories are difficult to identify
- Lack of filtering
- Consider reordering of themes for better visualization

Discussion

- Is it possible to predict the flow of the story and changes in theme?
- How to provide automatic summary of a story spanning over few time intervals?
- Proposed system works only for text documents. What is the value of adding support for multimedia? What might be the challenges?

Questions?