



Department of Computer Science
University of British Columbia

ConVis: A Visual Text Analytic System for Exploring Blog Conversations

Enamul Hoque, Giuseppe Carenini

{enamul, carenini}@cs.ubc.ca

NLP group @ UBC

Rise of Text Conversations

- ❑ People engage in asynchronuous conversations frequently
 - e.g., blogs, forums, twitter.



- ❑ Blogs:
 - More than 100 millions of blogs
 - The audience is rising exponentially

A Blog Conversation from Daily Kos

Politico strikes again, once again giving over op-ed pixels to the right-wing crazies to bleat about the evils of Obamacare. This time it's Matt Kibbe, president of FreedomWorks, who is **desperate to present** the burn-your-fake-Obamacare-card-as-a-meaningless-political-gesture campaign as a huge success.

Oh, where to begin?

I think we struck a nerve. Judging from the left's hysterical overreaction to FreedomWorks' "Burn Your Obamacare Card" campaign, this oppressive transfer of wealth from young Americans to the elderly appears to be the Achilles Heel of the new, insanely authoritarian progressive movement.

Here's Kibbe's first problem, he's confusing "hysterical" with "in hysterics." As in laughing hysterically at the idea of the FreedomWorks crowd emulating dirty fucking hippies. It is a priceless image. And one that Kibbe just can't resist.

When did the American left decide that it was cool to subsidize The Man on the backs of millennials struggling with student-loan debt and a job recession that never ends?

What would Jerry Rubin, the Yippie war protester, do? He would torch his Obamacare card without a moment's hesitation, chanting "Hell no, we won't go."

Hahahaha!!! Oh, lordy. But when did the American right decide that large corporations looking out for a profit were "The Man"? And didn't we just learn from a Kibbe cohort, on Politico's very op-ed pages, that affordable education was **a devious plan** by President Obama to enslave us all? See, that guy was arguing that being overburdened by student loan debt was Freedom! Now it's supposed to be a bad thing? Obviously, these guys have to work on coordinating their message.

Health and Human Services (HHS) chief Kathleen Sebelius calls FreedomWorks' ceremonial card-burning efforts "dismal." Is she projecting? HHS officials are so worried about their ability to conscript enough young able bodies into their master plan that they rolled out a \$700 million corporate PR campaign to convince twenty-somethings to buy an overpriced Edsel.

Obamacare

Student loan and job recession

Student loan

Buying over-priced Edsel

A Blog Conversation from Daily Kos (2)

103 COMMENTS

| PERMALINK

▼ Tip Jar (62+ / 0-)

"The NSA's capability at any time could be turned around on the American people, and no American would have any privacy left, such is the capability to monitor everything. [...] There would be no place to hide."--Frank Church

by Joan McCarter on Thu Sep 05, 2013 at 11:04:29 AM PDT

▼ But But But.....this is their job! (10+ / 0-)

If they didn't spend so much time muddying the water they'd have to get a job at WalMart.

by skillet on Thu Sep 05, 2013 at 11:10:19 AM PDT

▼ What they fail to come to terms with is that (7+ / 0-)

Obamacare is here to stay. Forever. There is absolutely nothing they can do to stop it. For every state that they may possibly be able to sabotage to the extent that Obamacare doesn't look like a clear success vs the previous system, there will be 2-3 states that are fully embracing it and making it a huge success for their constituents.

Which means, it's never going away. Period. Except for the constituents they're hurting in the process, I really don't give a fiddle fuck about their sabotage efforts in those handful of states. Or turning down federal funds for expanding Medicaid. Or their holding meaningless votes in the House repealing Obamacare. Or burning fake cards they made at home. Or writing OpEds at PolitHack. It does absolutely nothing to achieve their fantastical goal of getting rid of Obamacare. Nothing. So rage on Repugs. Waste your time.

by FraidKnot on Thu Sep 05, 2013 at 01:36:55 PM PDT

[Parent]

▼ Where can I get my Obama card? (2+ / 0-)

by immigradvocate on Thu Sep 05, 2013 at 03:15:31 PM PDT

[Parent]

▼ History Marches Back: the Generation Gap returns (15+ / 0-)

Long threads of discussion:

- Information overload (Jones et al. 2004)
 - Skip comments
 - Generate short response
 - Leave the discussion prematurely

Possible Solutions

□ InfoVis approaches

- Support the exploration of large amount of text
- Visual representation of
 - Metadata
 - Text analysis results

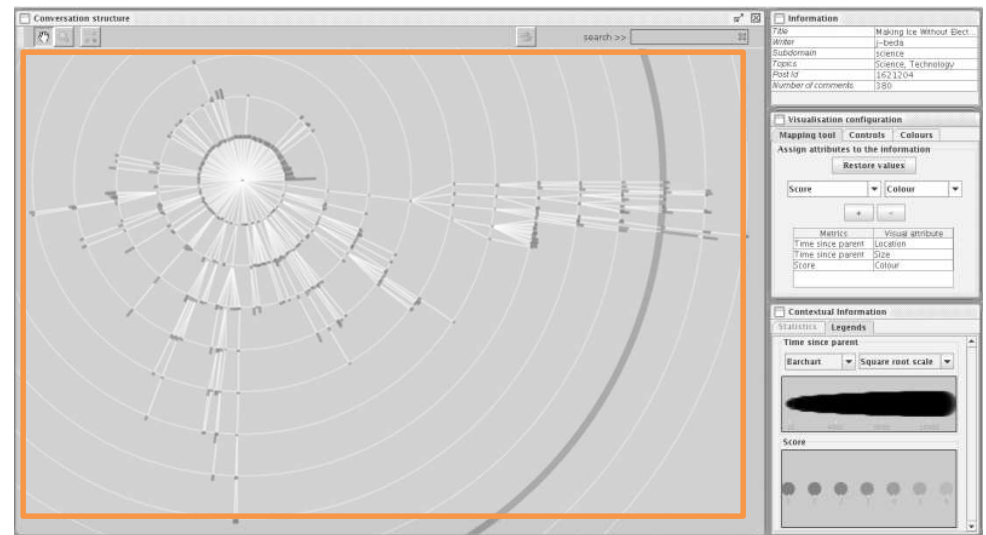
□ NLP approaches

- Extract content from conversations
 - Provide natural language summaries
- Very little efforts to integrate both **NLP** and **InfoVis** in a synergistic way

Visualization of Conversation Metadata

- thread structure,
- comment length,
- moderation score

No NLP



Thread Arc: Bernard Kerr (InfoVis 2003) Radial tree- based: Pascual-Cid et al. (InfoVis 2009)

- text analysis results (topics, opinions)

[illegible]

Topic Evolution Over Time

A Human-centered Design Approach

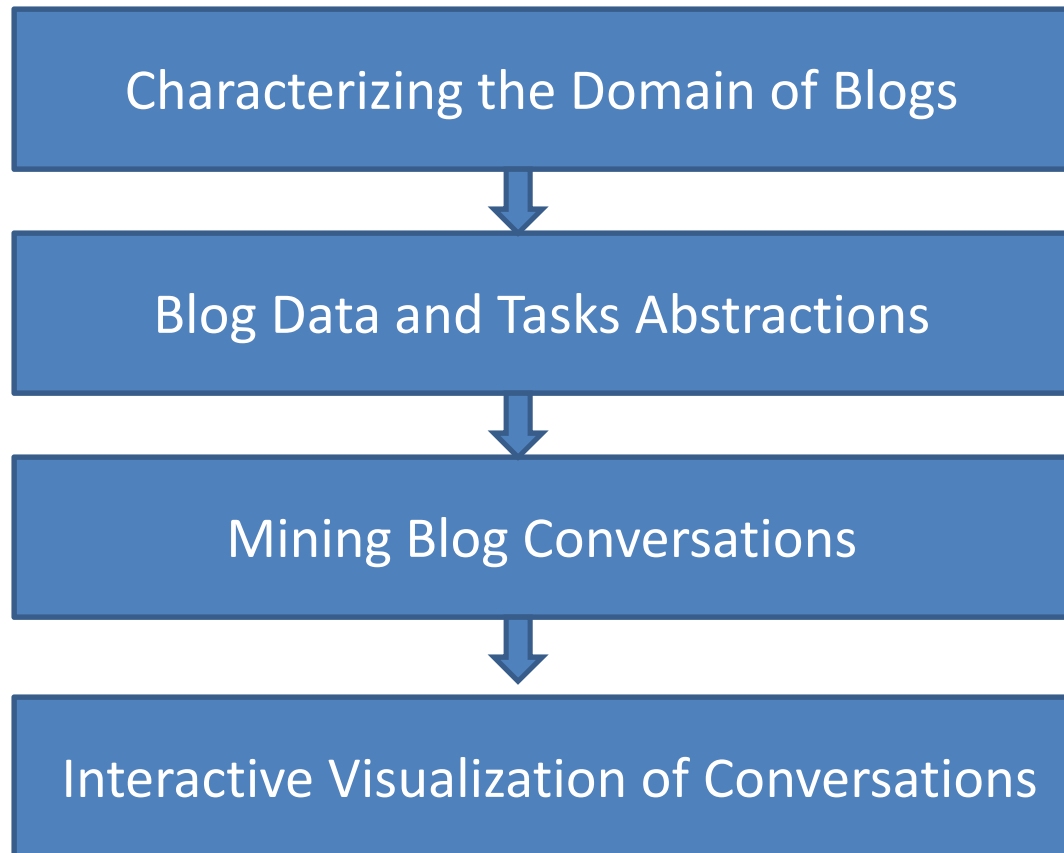
- How can we better support the user?
 - Need to integrate **NLP** and **InfoVis** techniques

- What NLP methods should be applied?
- What metadata are important?
- How the information should be visualized?

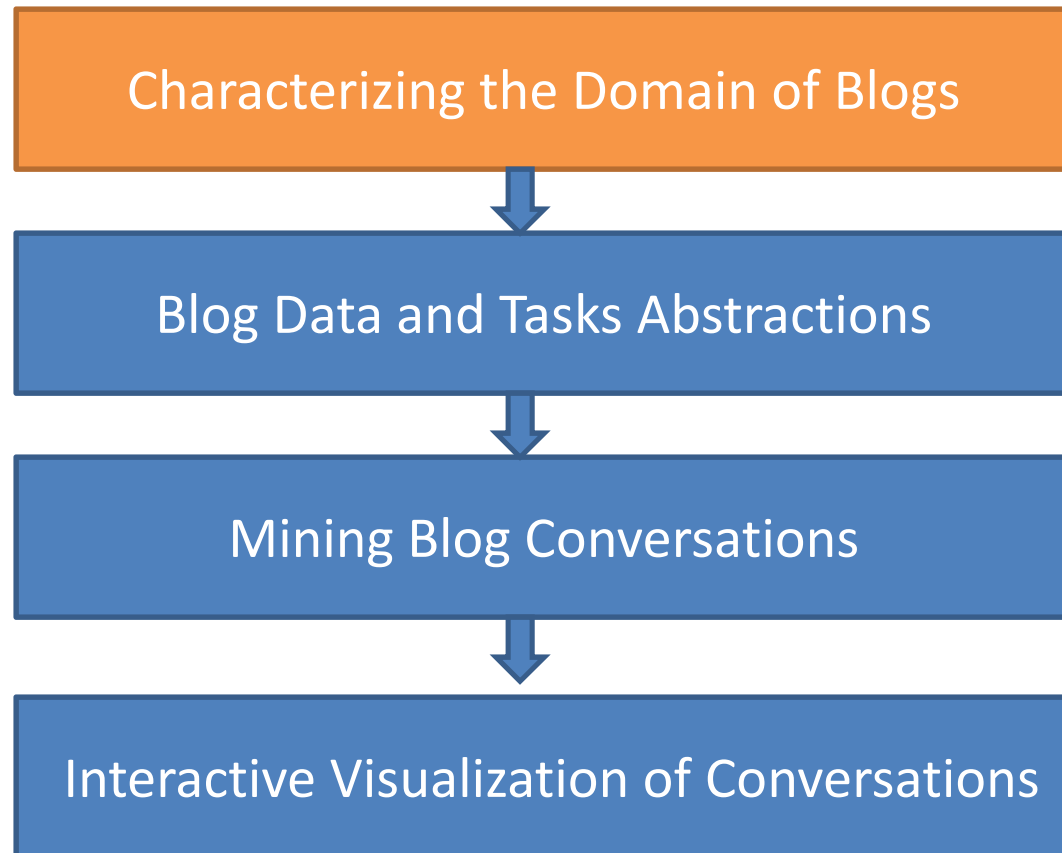


Human centered design approach
Nested Model [Munzner 2009]

Contributions

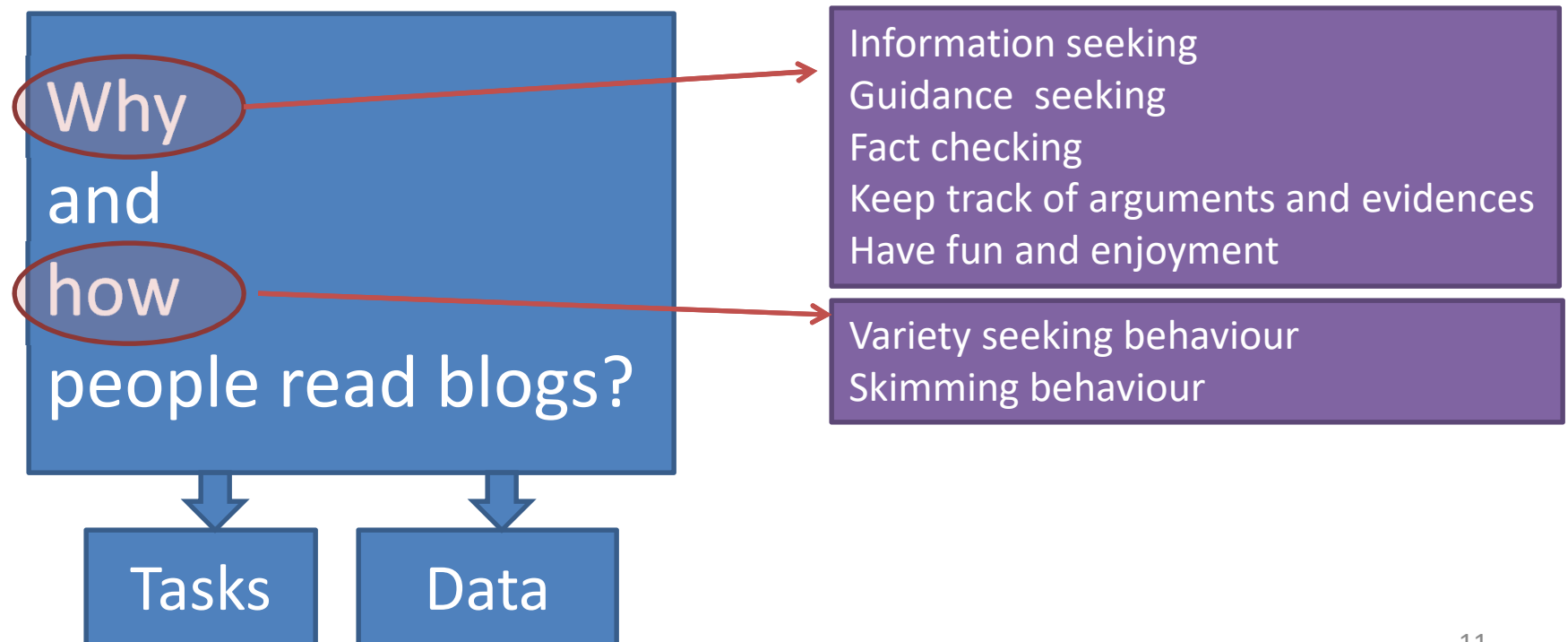


Contributions

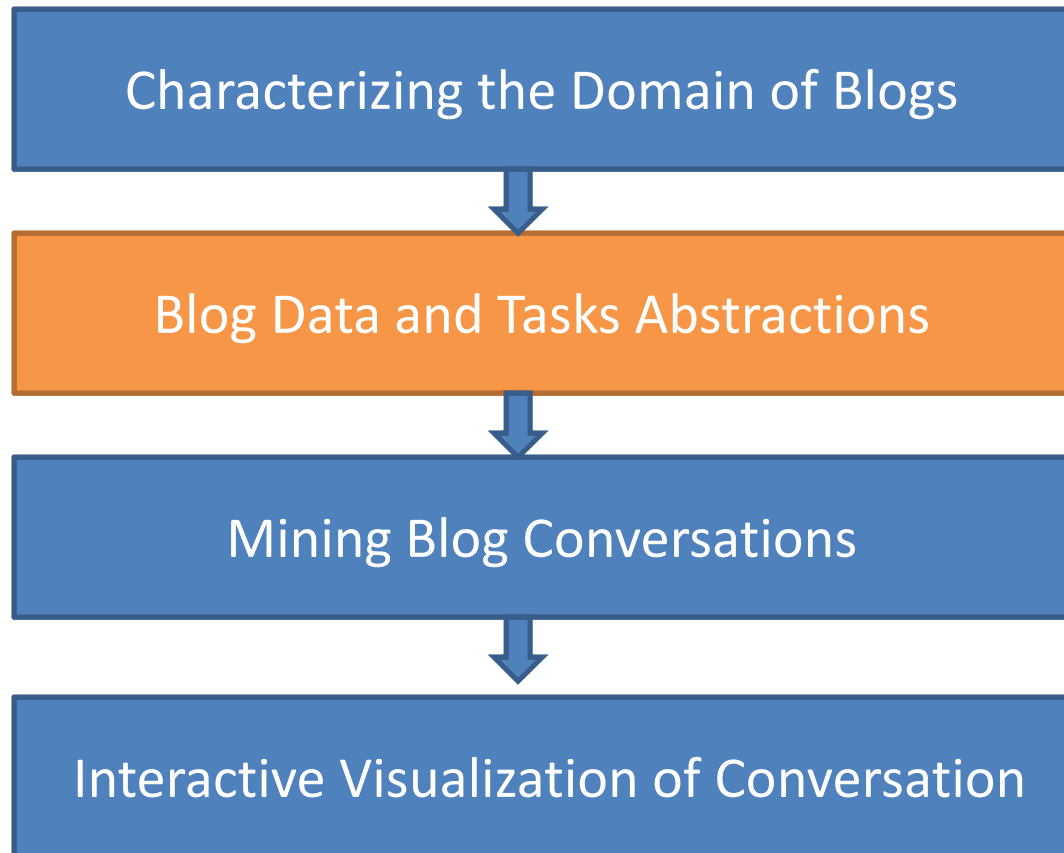


Characterizing the Domain of Blogs

- Computer mediated communications
- Social media
- Human computer interactions (HCI)
- Information retrieval



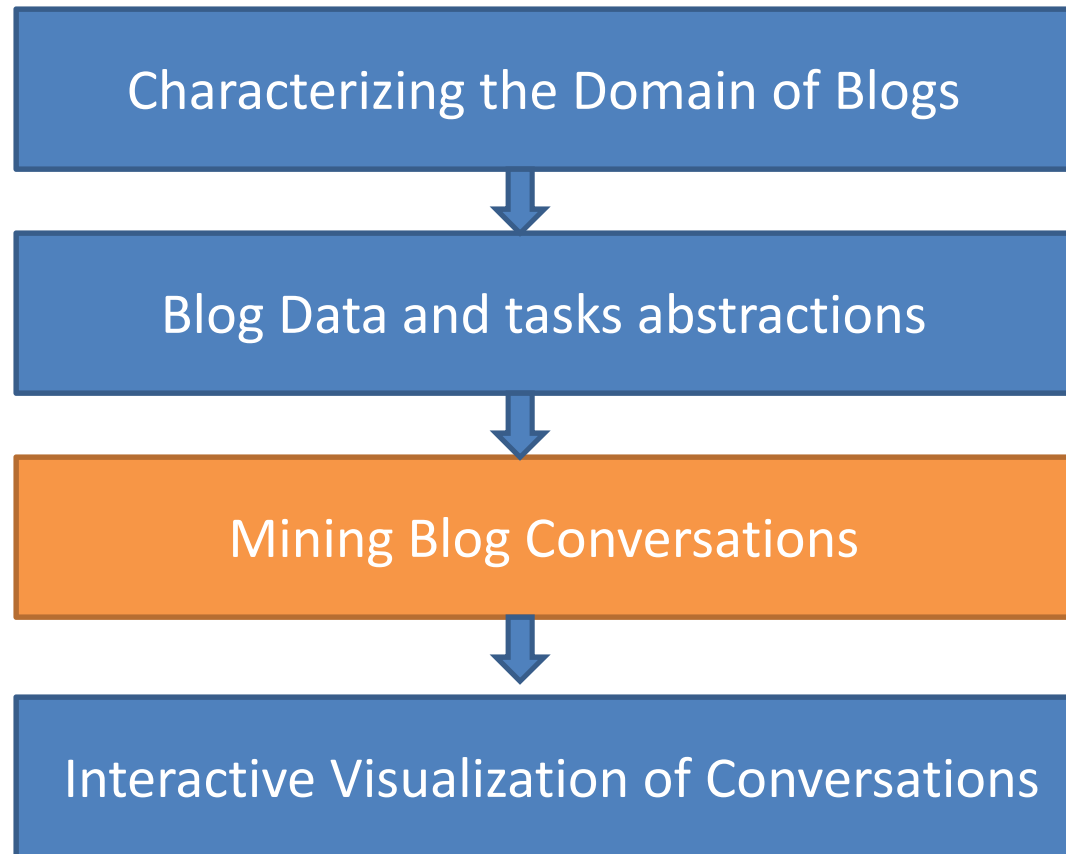
Contributions



Blog Data and Tasks Abstractions

TASKS	Data Variables				
	Topic	Author	Opinion	Thread	Comment
What this conversation is about?	x				X
Which topics are generating more discussions?	x				
What do people say about topic X?			X	x	X
How controversial was the conversation? Were there substantial differences in opinion ?	x	X	X	x	X
How other people's viewpoints differ from my current viewpoint on topic X?	x		X	X	X
Why are people supporting/ opposing an opinion ?			X	x	
Who was the most dominant participant in the conversation?	x	X	X	x	X
Who are the sources of most negative/positive comments on a topic?	x	X	X	x	X
Who has similar opinions to mine?	X	X	X		X
What are some interesting/funny comments to read?	X	X	x		X

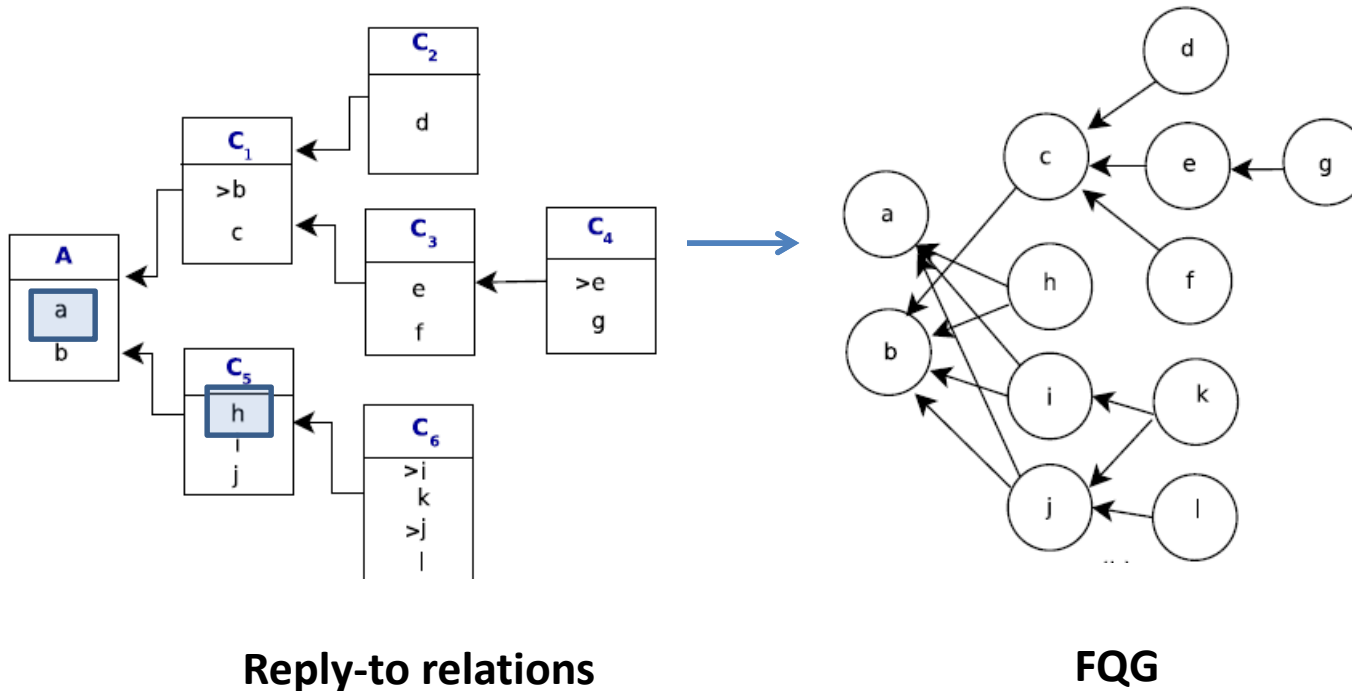
Contributions



Blog Mining: Topic Modeling

Taking advantages of conversational structure

- Fragment quotation graph (FQG) (*Carenini et al., WWW 2007*)



Blog Mining: Topic Modeling (2)

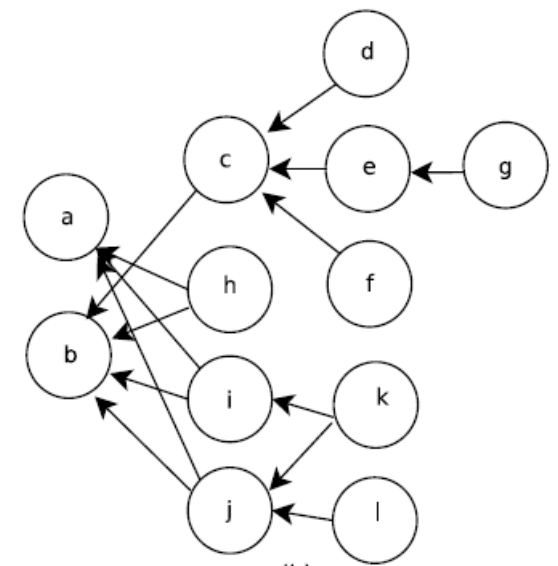
Segmentation: *(Joty et al., JAIR 2013)*

1. Apply Lexical cohesion-based segmentation on each path of the FQG
2. Graph-based technique:
Normalized cut criterion *(Shi & Malik, 2000)*

Labeling:

Generate k keyphrases for each segment

- Apply syntactic filter
- Co-ranking method
 - Based on FQG and information from leading sentences



Blog Mining: Sentiment Analysis

Semantic Orientation CALculator (SO-CAL):

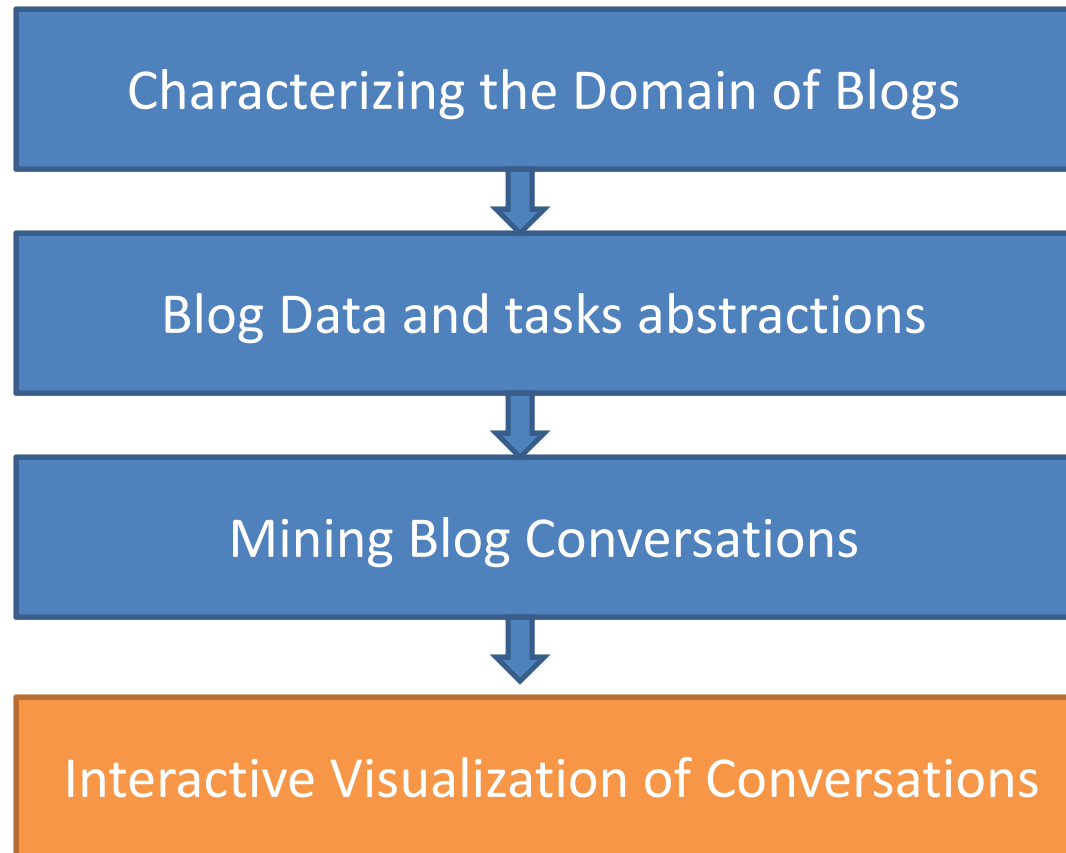
- Lexicon-based approach (*Taboada et al., JCL 2011*)

*Example: Usually Republicans are in lockstep on everything But they seem in **disarray** over this **issue**. (-2.5)*

Define 5 different polarity intervals [-2,-1,0,1,2]

- For each comment:
 - **Compute polarity distribution:** how many sentences fall in any of these polarity intervals

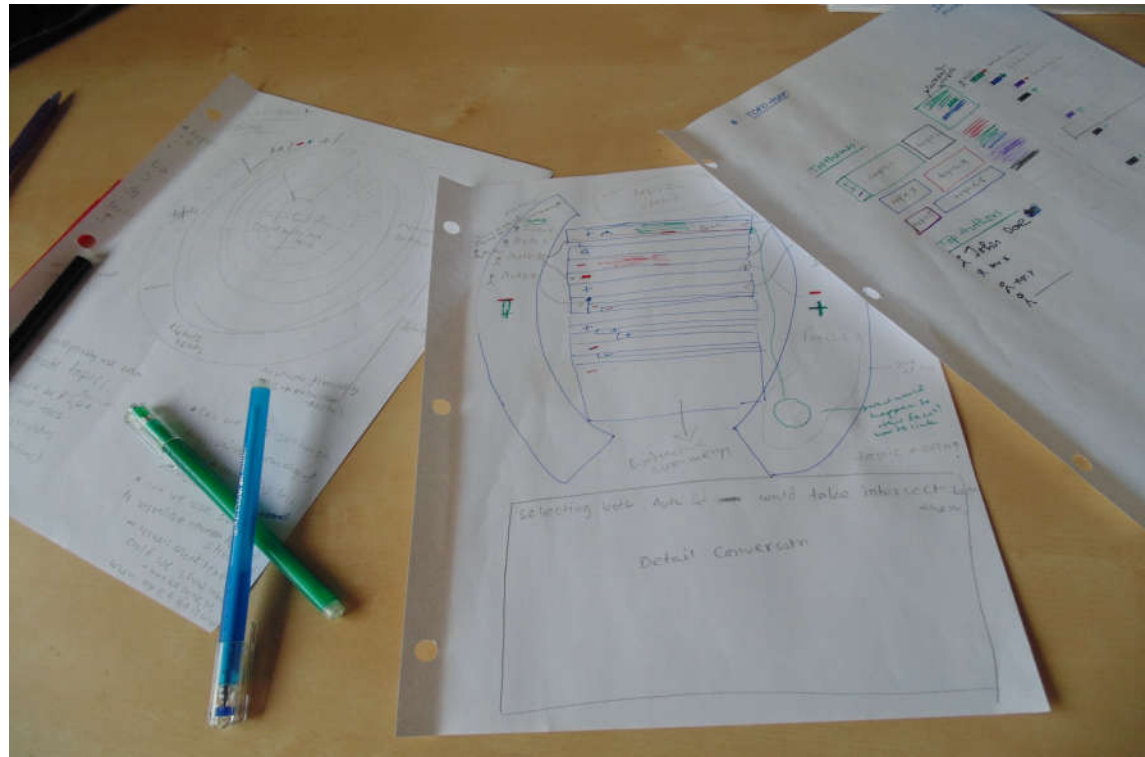
Contributions



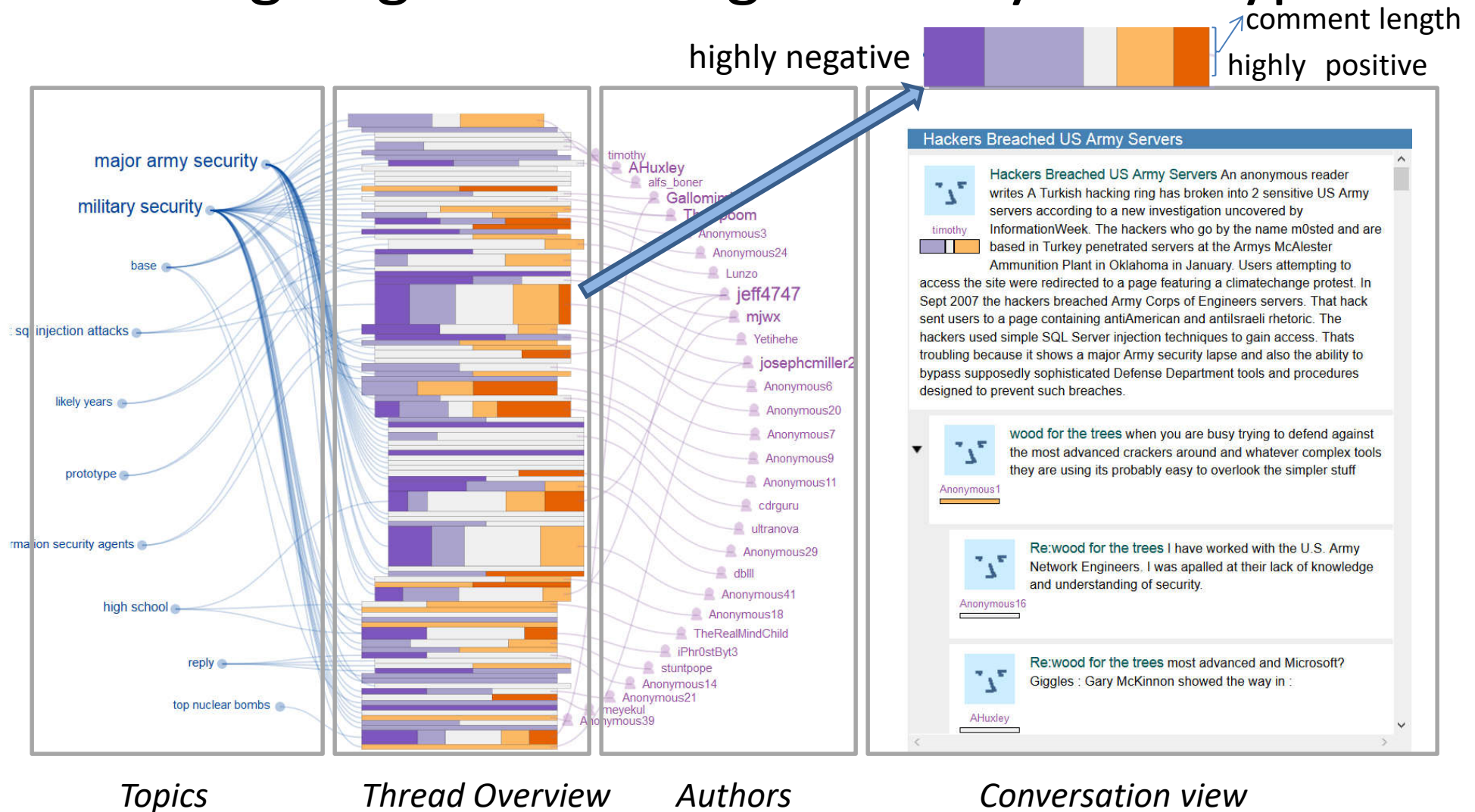
Designing ConVis: Low Fidelity Prototype

Integrate and extending Infovis to support:

- Show a comprehensive set of data
- Supporting multi-faceted exploration
- Interactive features



Designing ConVis: High-Fidelity Prototype



For particular tasks such as document comprehension, **overview + details** has been found more effective. (Cockburn et al. 2008)

Demo

<http://www.cs.ubc.ca/~enamul/convis/>

Informal Evaluation

Participants: 5 bloggers (age: 18-24, 2 female)

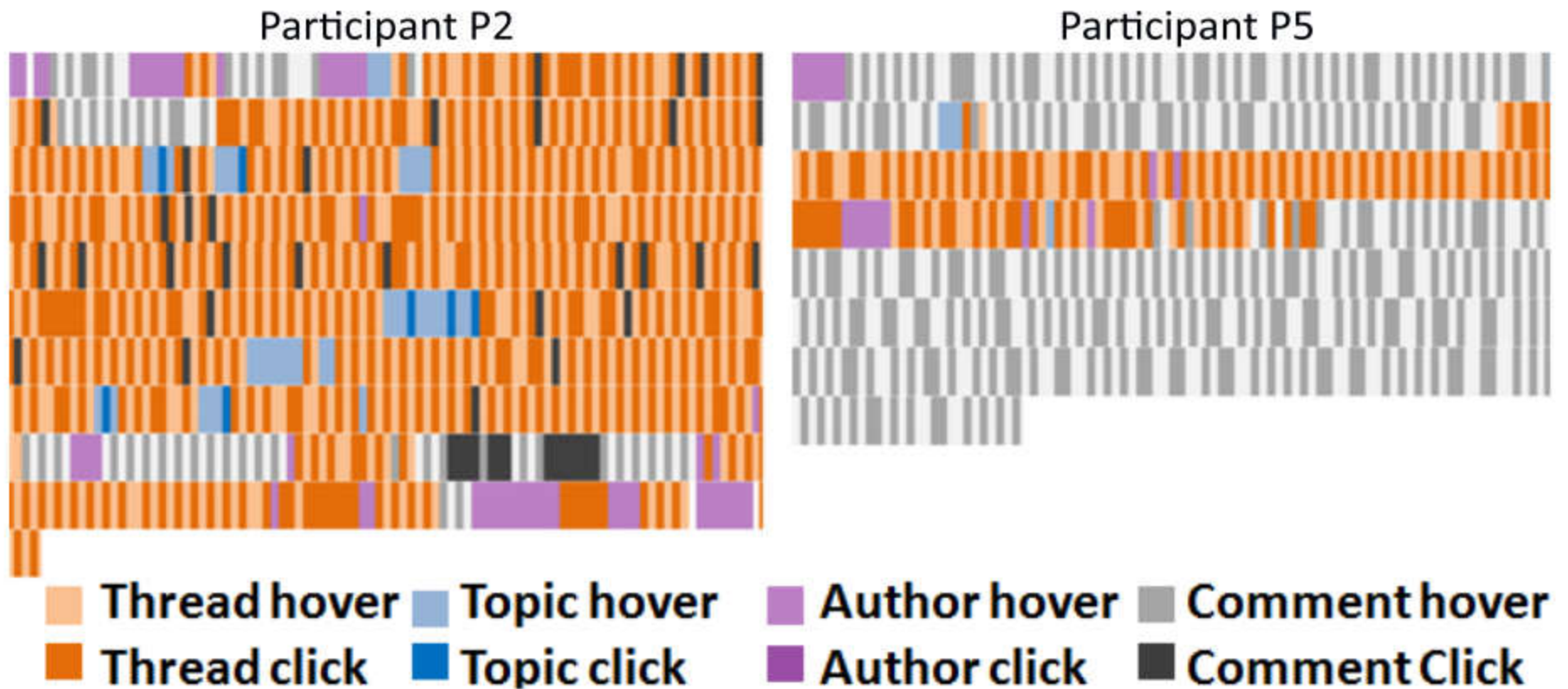
Exploratory tasks

Data Collection: Logs, observations and interviews

Results and Analysis

- ❑ How users perform their tasks?
 - 2 strategies: Explore by facets, skimming through comments
- ❑ What features worked/ didn't work?
 - Topic, sentiment, authors
- ❑ Ideas for improvements and enhancements

Usage Patterns



Explore by topic facets (Two Participants)

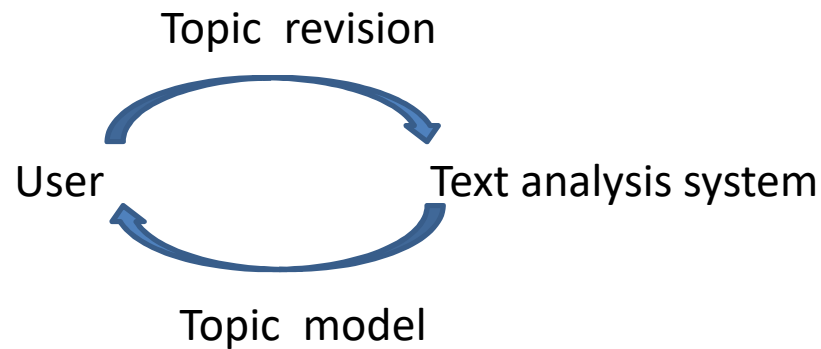
Scroll through the detail view (Three participants)

Users' Subjective Feedback

- ❑ **P1:** “Seeing the sort of pagination in current interfaces, you don’t get the overall. I have to read through all of them.” On the contrary, “Using ConVis **I would read more important parts of the conversation as opposed to just people talking**. I can navigate through the comments without actually reading them, which is really helpful.”
- ❑ **P2:** **It allows me to navigate through the most insightful stuffs out of five minutes which could take say 15 minutes otherwise.** Actually I found many comments to be interesting towards the end of conversations, which I probably wouldn’t notice if I would use my blog interface”.
- ❑ **P5:** I am so much used to scroll up and down in the list of comments, but using this additional visual overview, **I had a sense of where I am reading right now and what topic I am currently reading”**

Future Work

- ❑ Incorporate human feedback in computation



- ❑ Scalability
 - 1000 comments?
- ❑ Exploring Blogosphere

in address agenda barack byu certificate Chill define
education failures grocery
nspiring jumps kids obama obamas owns
potus ppl pres president prez
made responsibility rules shot socialism socialist
speech students teach

Acknowledgements



Tamara Munzner

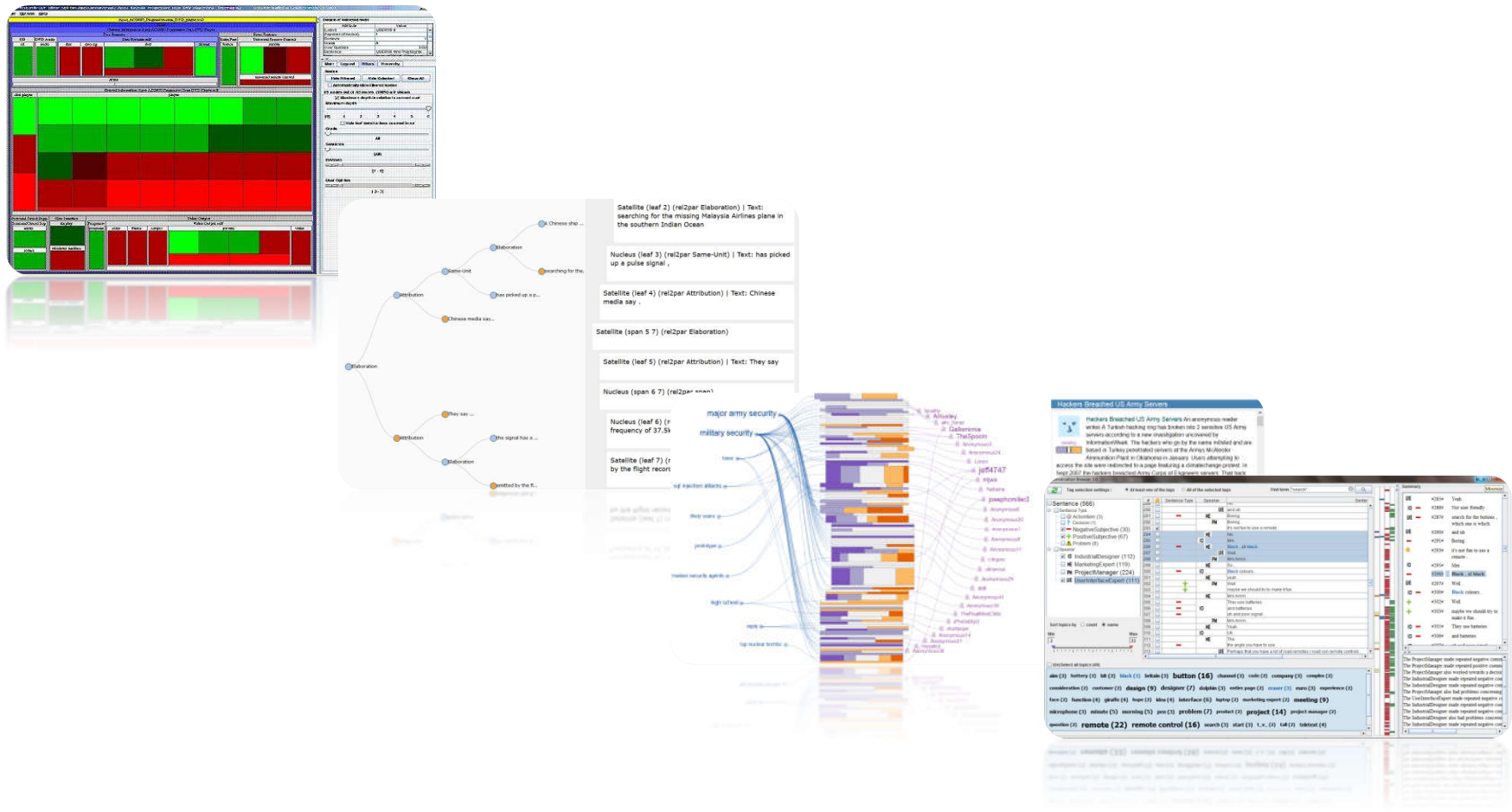


Raymond T. Ng



For More demos...

<https://www.cs.ubc.ca/cs-research/lci/research-groups/natural-language-processing/>



Selected References

- ❑ Baumer, E., Sueyoshi, M., and Tomlinson, B. Exploring the role of the reader in the activity of blogging. In Proceedings of the CHI '08 (2008), 1111–1120.
- ❑ Carenini, G., Murray, G., and Ng, R. Methods for Mining and Summarizing Text Conversations. Morgan Claypool, 2011.
- ❑ Hearst, M. A., Hurst, M., and Dumais, S. T. What should blog search look like? In Proceedings of the 2008 ACM workshop on Search in social media, ACM (2008), 95–98.
- ❑ Joty, S., Carenini, G., and Ng, R. T. Topic segmentation and labeling in asynchronous conversations. Journal of Artificial Intelligence Research 47 (2013), 521–573.
- ❑ Kaye, B. K. Web side story: An exploratory study of why weblog users say they use weblogs. AEJMC Annual Conference (2005).
- ❑ Kerr, B. Thread arcs: An email thread visualization. In IEEE Symposium on Information Visualization (2003), 211–218.
- ❑ Liu, S., Zhou, M. X., Pan, S., Song, Y., Qian, W., Cai, W., and Lian, X. TIARA: Interactive, Topic-Based Visual Text Summarization and Analysis. ACM Transaction on Intelligent System Technology 3, 2, 25:28.
- ❑ Munzner, T. A nested model for visualization design and validation. IEEE Transactions on Visualization and Computer Graphics 15, 6 (Nov. 2009), 921–928.
- ❑ Pascual-Cid, V., and Kaltenbrunner, A. Exploring asynchronous online discussions through hierarchical visualisation. In Information Visualisation, 2009 13th International Conference, IEEE (2009), 191–196.
- ❑ Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. Lexicon-based methods for sentiment analysis. Computational linguistics 37, 2 (2011), 267–307.