CS 6604 Middle Term Report
Computational Linguistics PJ
-Explore Correlation between Newswires and Twitter

Client: Mohamed Magdy Farag

***by** Tianyu Geng, Wei Huang, Ji Wang, and Xuan Zhang*

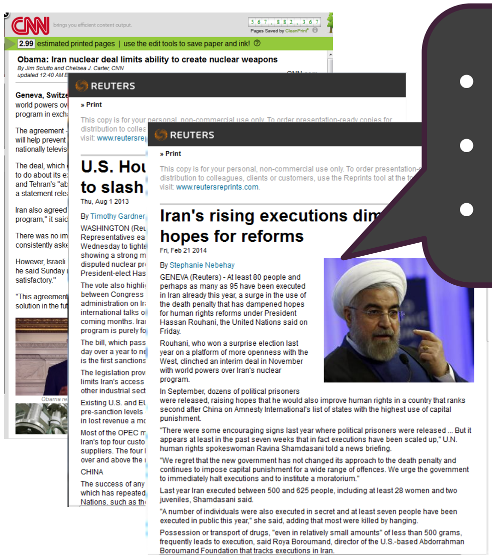
March, 6th, 2014
Blacksburg, VA

Table of Contents

- *Introduction*
- *Solution*
- *Progress*

Problem to Solve

Motivation: Much news, much tweets, little connection…



CNN brings you efficient content output.

5 6 7 , 8 8 2 , 3 6 7 Pages Saved by CleanPrint®

2.99 estimated printed pages | use the edit tools to save paper and ink!

Obama: Iran nuclear deal limits ability to create nuclear weapons

By Jim Sciutto and Chelsea J. Carter, CNN
updated 12:40 AM E

Geneva, Switze
world powers ov
program in excha

The agreement -
will help prevent
nationally televis

The deal, which
to do about its e:
and Tehran's "ab
a statement rele

Iran also agreed
program," it said

There was no im
consistently aske

However, Israeli
he said Sunday
satisfactory."

"This agreement
solution in the fu

Obama re

REUTERS

» Print

This copy is for your personal, non-commercial use only. To order presentation-ready copies for
distribution to collea
visit: www.reutersrep

U.S. Hou
to slash

Thu, Aug 1 2013

By Timothy Gardner

WASHINGTON (Reu
Representatives ea
Wednesday to tighte
showing a strong m
disputed nuclear pr
President-elect Has

The vote also highli
between Congress
administration on Ir
international talks o
coming months. Ira
program is purely fo

The bill, which pass
day over a year to ne
is the first sanctions

The legislation prov
limits Iran's access
other industrial sect

Existing U.S. and EU
pre-sanction levels
in lost revenue a mo

Most of the OPEC m
Iran's top four custo
suppliers. The four
over and above the

CHINA

The success of any
which has repeated
Nations, such as th

REUTERS

» Print

This copy is for your personal, non-commercial use only. To order presentation-
distribution to colleagues, clients or customers, use the Reprints tool at the to
visit: www.reutersreprints.com.

Iran's rising executions dim
hopes for reforms

Fri, Feb 21 2014

By Stephanie Nebehay

GENEVA (Reuters) - At least 80 people and perhaps as many as 95 have been executed in Iran already this year, a surge in the use of the death penalty that has dampened hopes for human rights reforms under President Hassan Rouhani, the United Nations said on Friday.

Rouhani, who won a surprise election last year on a platform of more openness with the West, clinched an interim deal in November with world powers over Iran's nuclear program.

In September, dozens of political prisoners were released, raising hopes that he would also improve human rights in a country that ranks second after China on Amnesty International's list of states with the highest use of capital punishment.

"There were some encouraging signs last year where political prisoners were released ... But it appears at least in the past seven weeks that in fact executions have been scaled up," U.N. human rights spokeswoman Ravina Shamdasani told a news briefing.

"We regret that the new government has not changed its approach to the death penalty and continues to impose capital punishment for a wide range of offences. We urge the government to immediately halt executions and to institute a moratorium."

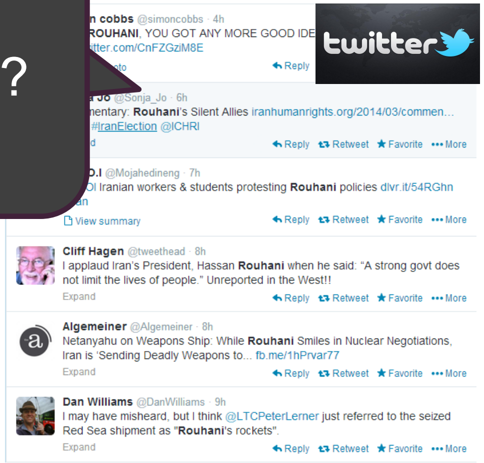
Last year Iran executed between 500 and 625 people, including at least 28 women and two juveniles, Shamdasani said.

"A number of individuals were also executed in secret and at least seven people have been executed in public this year," she said, adding that most were killed by hanging.

Possession or transport of drugs, "even in relatively small amounts" of less than 500 grams, frequently leads to execution, said Roya Boroumand, director of the U.S.-based Abdorrahman Boroumand Foundation that tracks executions in Iran.

Mainstream News

- Key points in news?
- Relation between news & tweets?
- Major attitudes of audience?



n cobbs @simoncobbs · 4h
ROUHANI, YOU GOT ANY MORE GOOD IDE
itter.com/CnFZGziM8E
oto Reply

a Jo @Sonja_Jo · 6h
nentary: **Rouhani**'s Silent Allies iranhumanrights.org/2014/03/commen…
#IranElection @ICHRI
d Reply Retweet Favorite More

D.I @Mojahedineng · 7h
Ol Iranian workers & students protesting **Rouhani** policies dlvr.it/54RGhn
an
View summary Reply Retweet Favorite More

Cliff Hagen @tweethead · 8h
I applaud Iran's President, Hassan **Rouhani** when he said: "A strong govt does not limit the lives of people." Unreported in the West!!
Expand Reply Retweet Favorite More

Algemeiner @Algemeiner · 8h
Netanyahu on Weapons Ship: While **Rouhani** Smiles in Nuclear Negotiations, Iran is 'Sending Deadly Weapons to... fb.me/1hPrvar77
Expand Reply Retweet Favorite More

Dan Williams @DanWilliams · 9h
I may have misheard, but I think @LTCPeterLerner just referred to the seized Red Sea shipment as "**Rouhani**'s rockets".
Expand Reply Retweet Favorite More

Tweets

Objects:

1. Summarize info in news and tweets
2. Explore correlation between news & tweets
3. Mine opinions in tweets

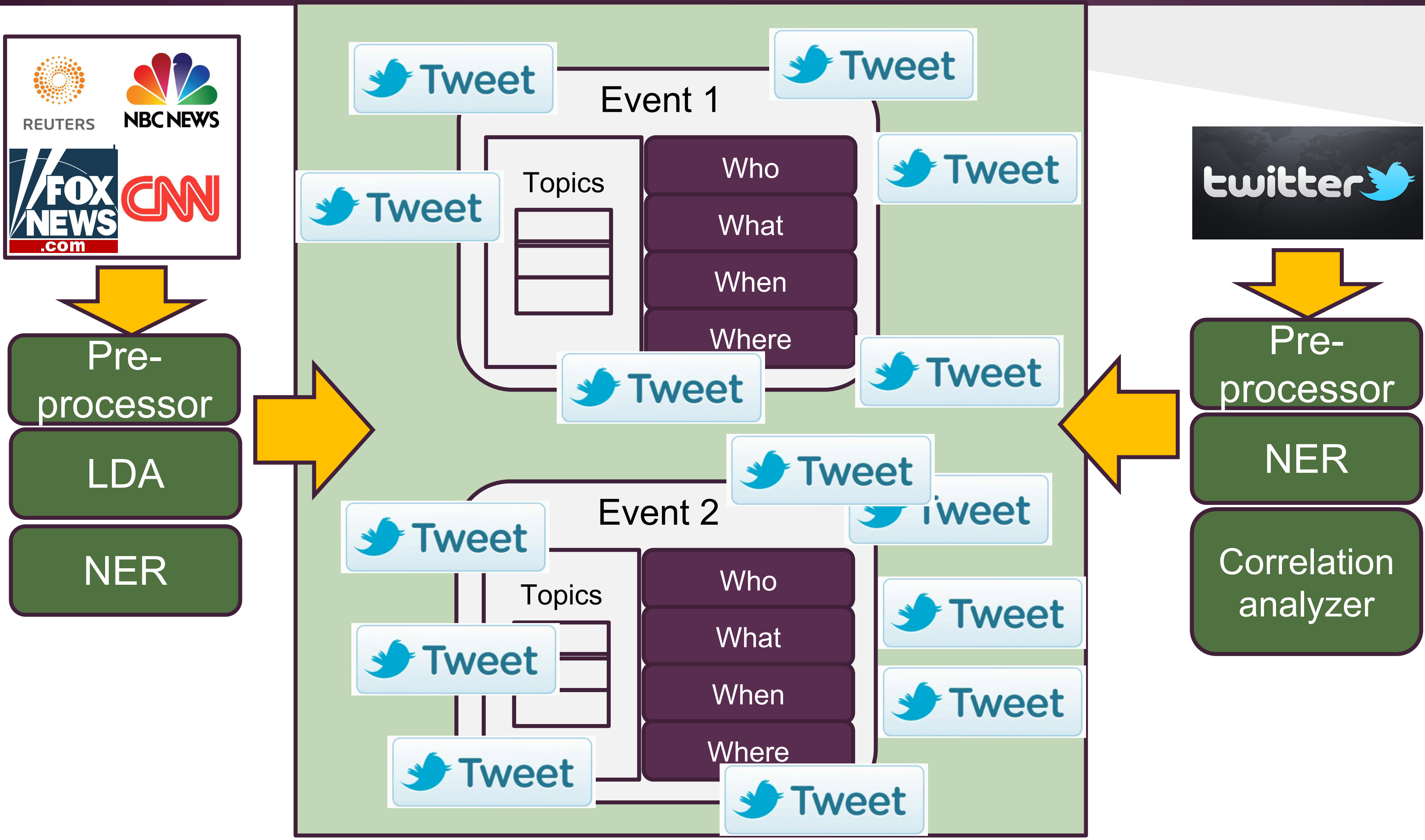
Solution Overview

1. Fetch text from news & tweets respectively
2. Preprocess texts: stemming, stop-word…
3. Extract events from news

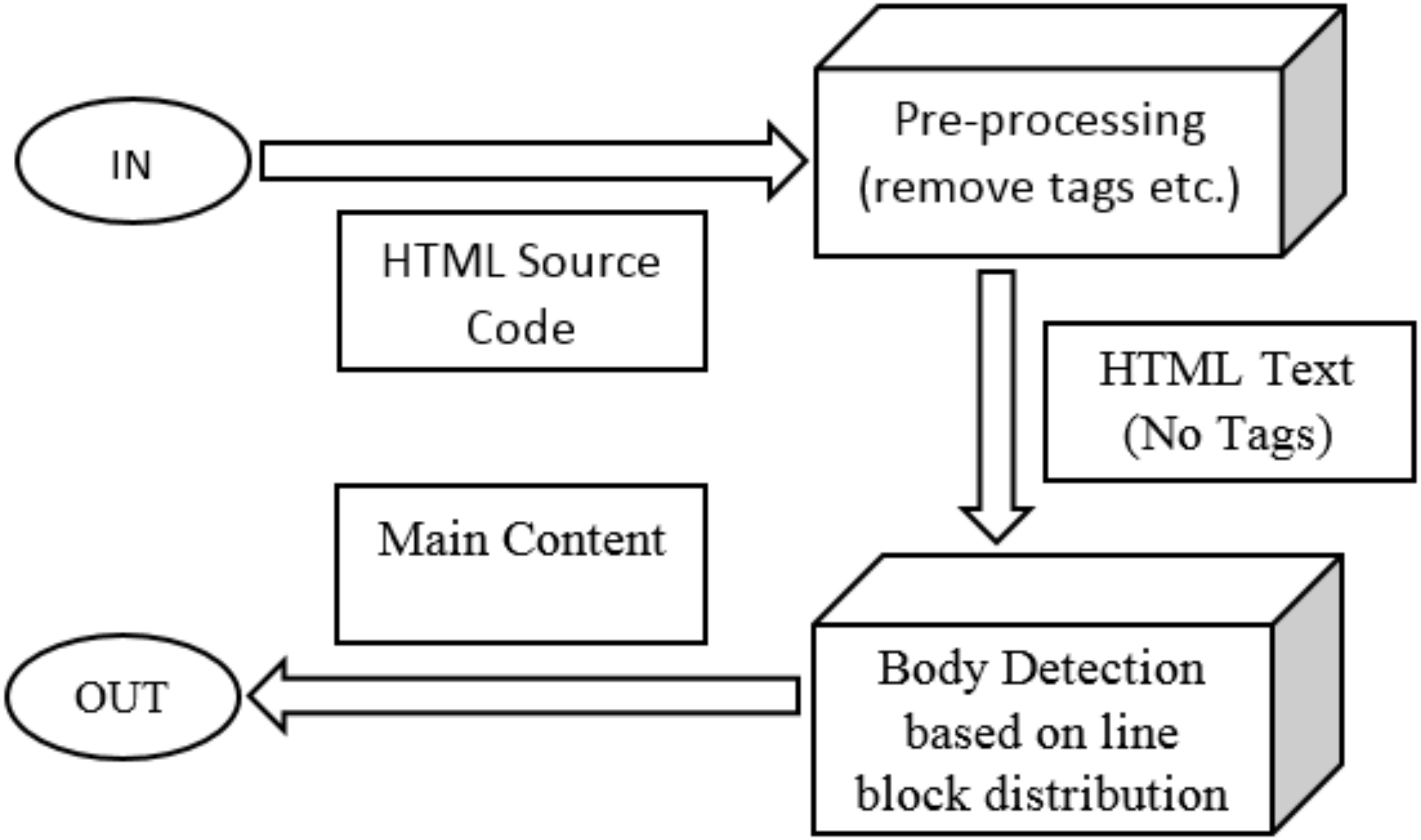
Event: [Topic, Named entities(who, what, where, when)]

4. Map tweets to events (correlation model)
5. Mine major opinions around events

Solution: Link Tweets to Events



Progress: Text Extraction



IN
HTML Source Code
Pre-processing (remove tags etc.)
HTML Text (No Tags)
Main Content
Body Detection based on line block distribution
OUT

Progress: News Analysis

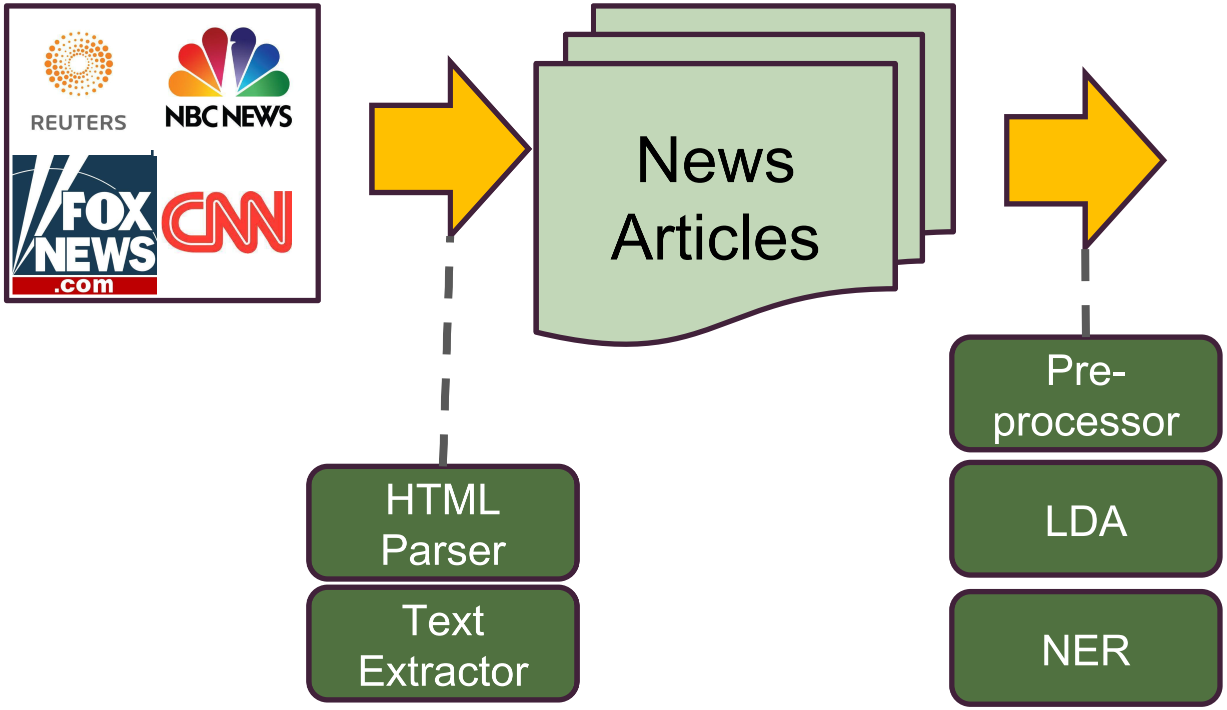
Tools: GibbsLDA, Stanford NLP

Dataset:

1) 2762 news about “Iran Election”.

--Only news titles used for topic modeling

2) News articles from CTRnet PJ



Event 1

Iran	Who: Obama
Sanction	What: N/A
Obama	When: N/A
warn	Where: Iran

Event 2

UN	Who: UN
urge	What: N/A
human	When: N/A
right	Where: Iran
Iran	

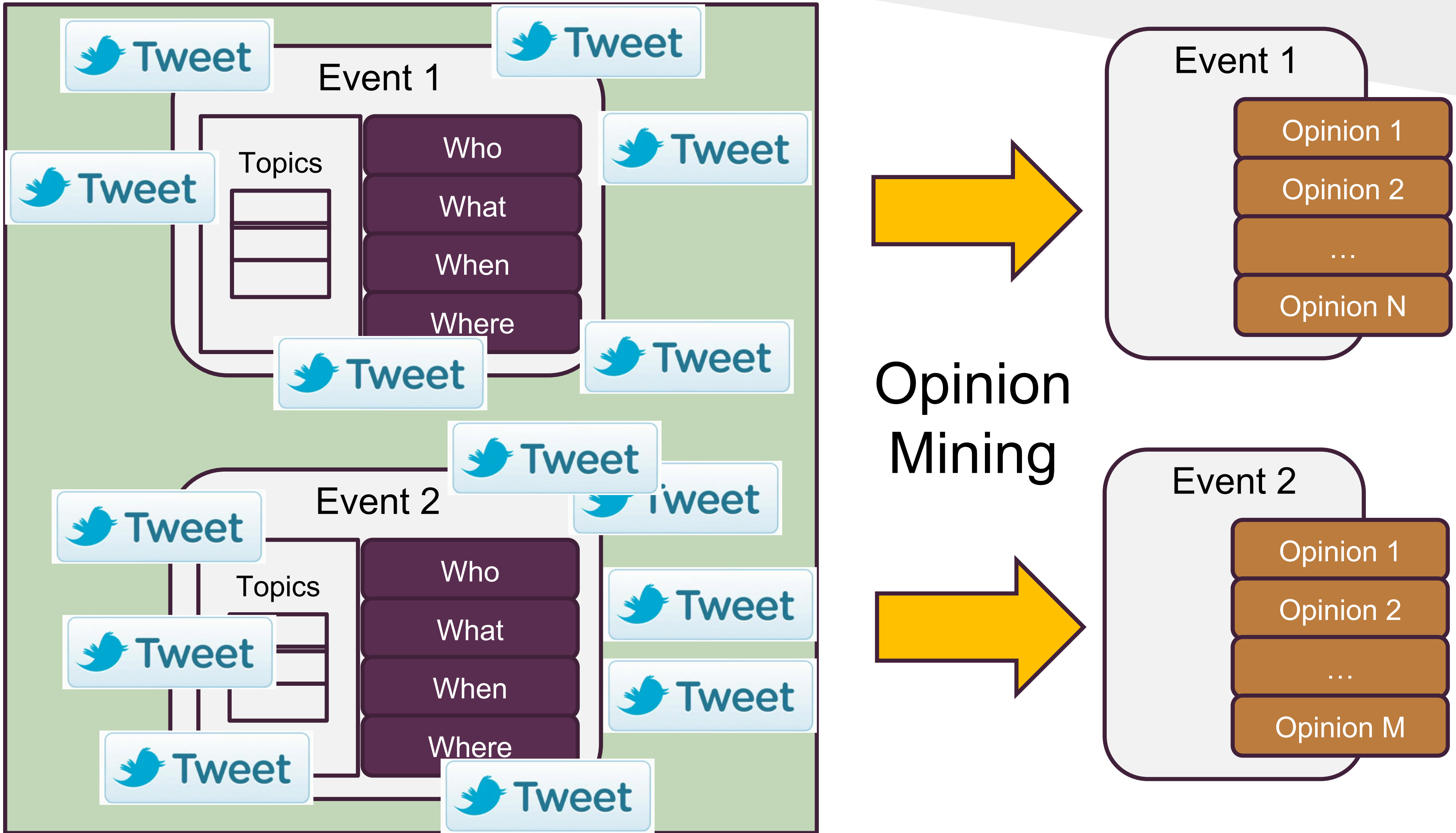
Progress: Tweets Analysis

We ran LDA on a sample of the #Iran collection from IDEAL

- 50,000 tweets
- Feb 13, 2013 23:58:30 ~ Feb 15, 2013 00:00:03
- 4 topics

Topic 0		**Topic 1**		**Topic 2**		**Topic 3**	
P		***P***		***P***		***P***	
	Keyword		***Keyword***		***Keyword***		***Keyword***
0.028	nuclear	0.027	iran		0.018	0.054	camp
0.018	weapons	0.019	time		iran	0.053	liberty
0.015	iranian	0.015	now	0.017	via	0.020	never
0.011		0.013	opposition	0.012	2help	0.015	martin
	stop	0.012	u	0.011		0.014	kobler
0.010	menlu	0.012	feb		killed	0.013	mr
0.008	iran	0.011		0.010	sanctions	0.012	adequate
0.008	executions		seeking	0.009	syria		

Future Work: Opinion Mining



Event 1

Topics

Who

What

When

Where

Event 2

Topics

Who

What

When

Where

Opinion Mining

Event 1

Opinion 1

Opinion 2

…

Opinion N

Event 2

Opinion 1

Opinion 2

…

Opinion M

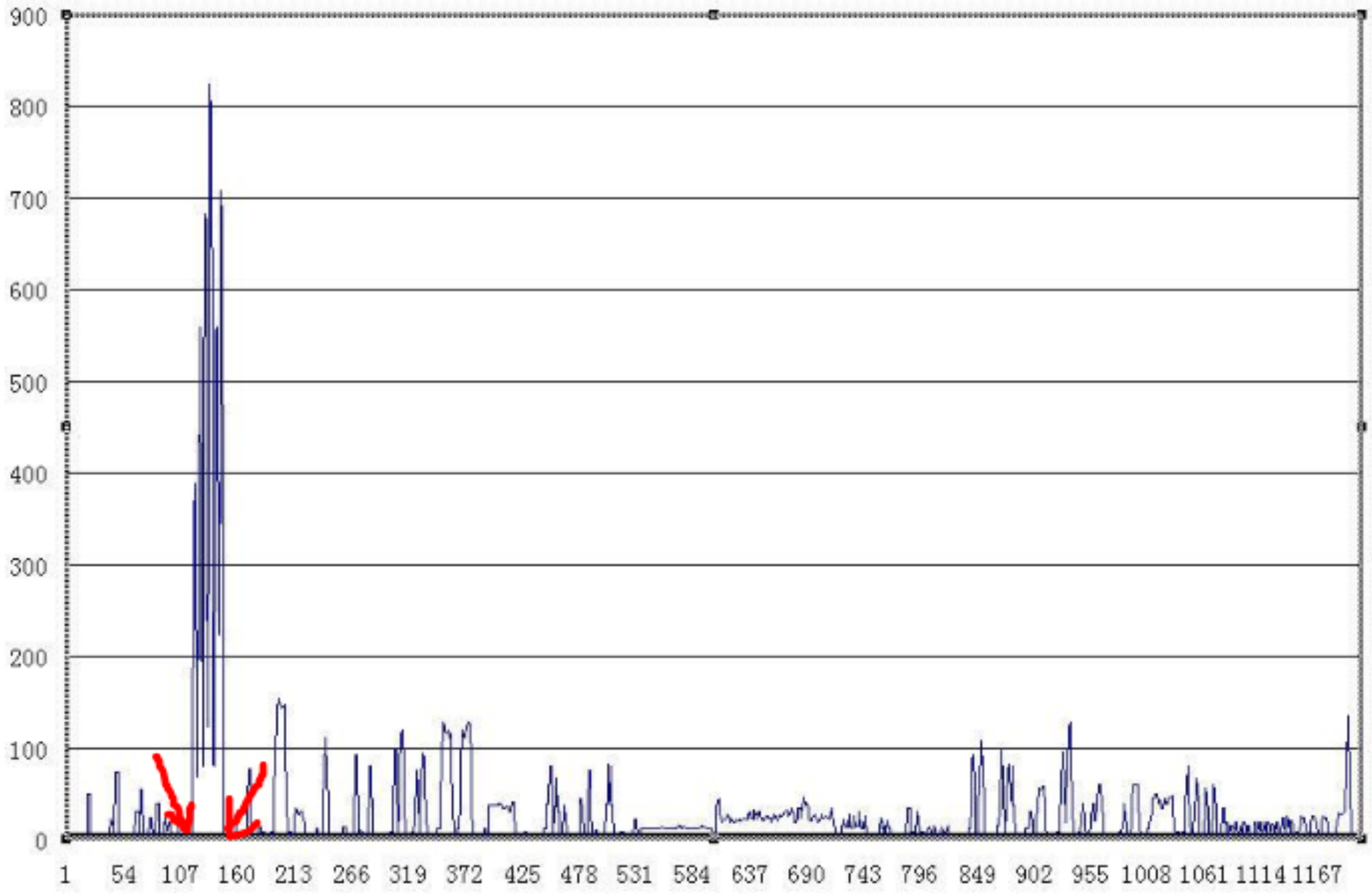
Appendix

Literature Review

1. Analysis between Tweets and News Articles
    - Fact: News providers report events earlier, but Twitter contains more details
    - Algorithm: LDA, cosine similarity, sentiment analysis

1. Summary based on Templates
    - Systems: SUMMARIST, Artequakt, etc.
    - Topic signature is used for selecting summarizing sentences
    - Using Apple Pie Parser, GATE and WordNet for knowledge extraction

The Characteristics of HTML



Reference

[1] Petrovic S, Osborne M, McCreadie R, et al. Can twitter replace newswire for breaking news[C]//Seventh International AAAI Conference on Weblogs and Social Media. 2013.

[2] Balahur A, Tanev H. Detecting Event-Related Links and Sentiments from Social Media Texts[J]. ACL 2013, 2013: 25.

[3] Lobzhanidze A, Zeng W, Gentry P, et al. Mainstream media vs. social media for trending topic prediction-an experimental study[C]//Consumer Communications and Networking Conference (CCNC), 2013 IEEE. IEEE, 2013: 729-732.

[4] Introne J E, Drescher M. Analyzing the flow of knowledge in computer mediated teams[C]//Proceedings of the 2013 conference on Computer supported cooperative work. ACM, 2013: 341-356.

[5] Hovy E, Lin C Y. Automated text summarization and the SUMMARIST system[C]//Proceedings of a workshop on held at Baltimore, Maryland: October 13-15, 1998. Association for Computational Linguistics, 1998: 197-214.

Reference

[6] Lin C Y, Hovy E. The automated acquisition of topic signatures for text summarization[C]//Proceedings of the 18th conference on Computational linguistics-Volume 1. Association for Computational Linguistics, 2000: 495-501.

[7] Luhn H P. The automatic creation of literature abstracts[J]. IBM Journal of research and development, 1958, 2(2): 159-165.

[8] Alani H, Kim S, Millard D E, et al. Automatic ontology-based knowledge extraction from web documents[J]. Intelligent Systems, IEEE, 2003, 18(1): 14-21.

[9] El Hamali S, Nouali O, Nouali-Taboudjemat N. Knowledge extraction by Internet monitoring to enhance crisis management[R]. CERIST, 2011.