

Using Natural Language Processing for Analyzing Disaster Management

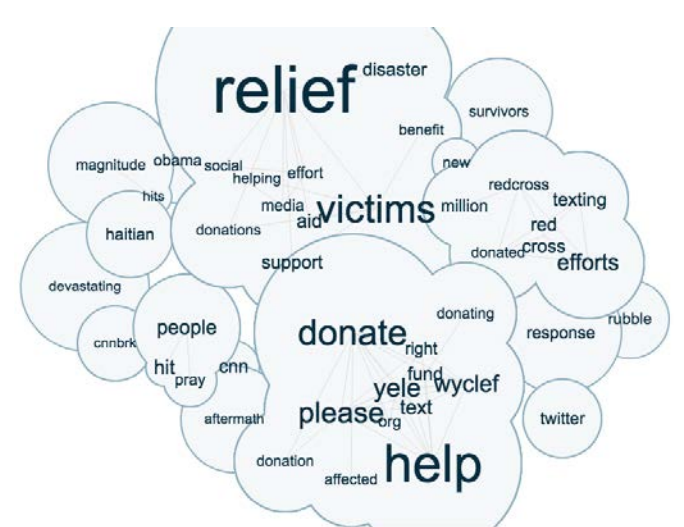
Samira Zad¹, Janina Sarol², Ly Dinh³, Jana Diesner³

¹Department of Computer Science, Florida International University

²Illinois Informatics Institute, University of Illinois at Urbana-Champaign

³School of Information Sciences, University of Illinois at Urbana-Champaign

Motivation



Humanitarian assistance and disaster relief (HADR) efforts can benefit from computational methods that aim to support the coordination of tasks between relief agents more rapidly and efficiently, and the collection and analysis of accurate and reliable information. Therefore, we aim to implement a Natural Language Processing (NLP) framework, using text data from 6 different information sources to develop conceptual and operational definitions of a ‘successful’ response. We then brought this work into an application context, namely the 2010 Haiti earthquake relief operations by the United States Coast Guard’s (USCG) and related organizations.

Data

Data sources and services used: Crimson Hexagon (Twitter & blogs), ProQuest (newspapers), DTIC.mil (government reports), Web of Knowledge (scientific journals), USCG points-of-contacts (interviews)

Data source	Sample	Timeframe
Social media (Twitter)	100,242 tweets collected	Response: January 12 – February 4, 2010
Expert blogs	54,662 blogs	Relief: February 5 – March 14
Newspapers	13,735 local & national newspapers	Restore: March 15 – April 14, 2010
Interviews	3 USCG Captains & 2 Vice Admirals	Recovery (short-term): April 15 – June 1, 2010
Government reports	397 SITREPS from DTIC.mil	
Scientific journals	489 peer-reviewed articles	January 12, 2010 – January 12, 2011

Figure 1: Data sources & samples used for this study

Results

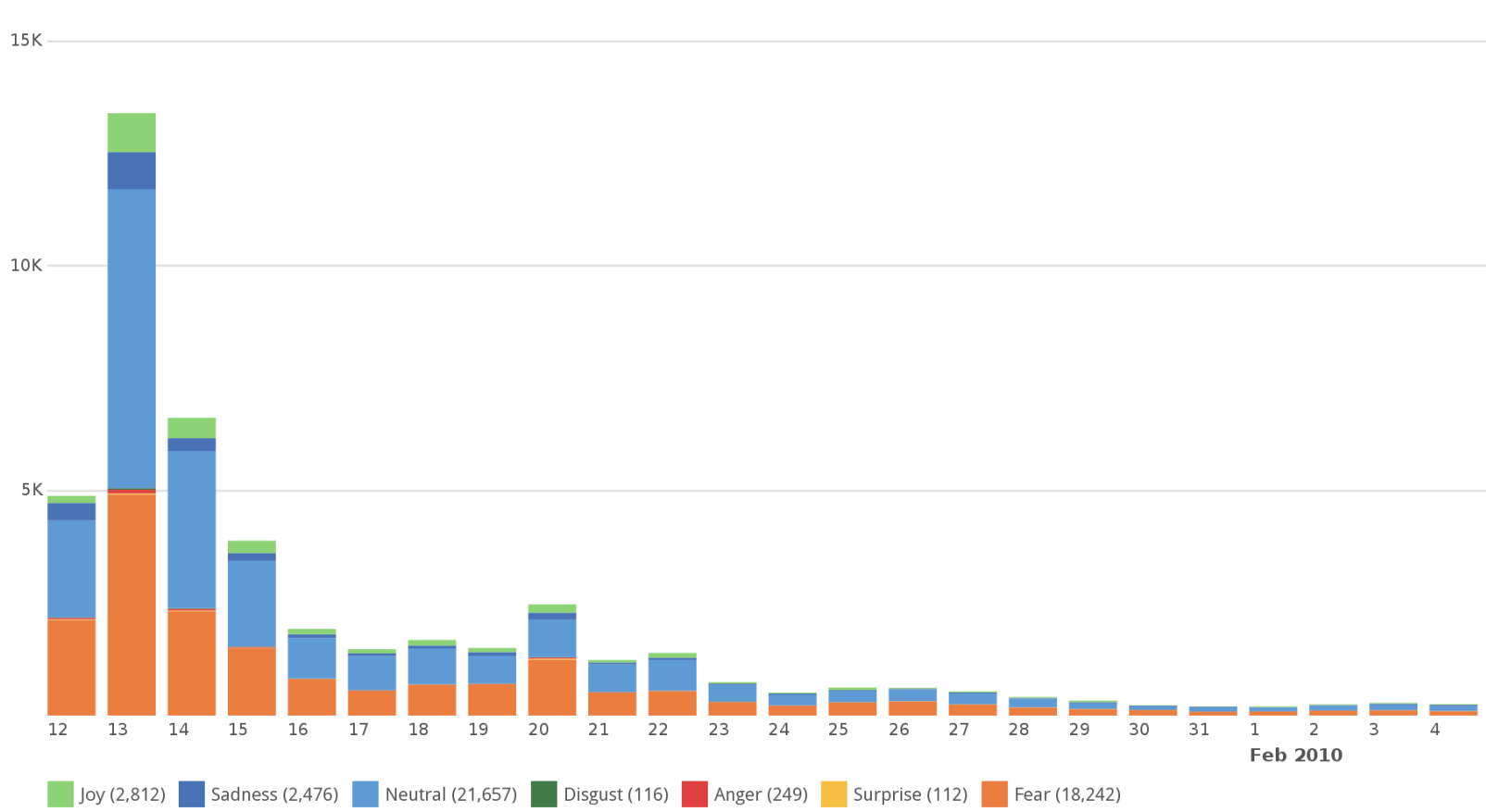


Figure 2: Emotion analysis for Response phase (Jan 12 – Feb 4, 2010)

Topic	Terms
Medical aid	hospital medical doctors haiti team health patients care earthquake field
Earthquake	earthquake quake magnitude haiti chile people buildings earthquakes damage tsunami
Haiti	people haiti earthquake port-au-prince haitians haitian food haiti’s aid country
Donations	haiti earthquake people money pounds raised appeal donations charity raise
School	school students community haiti university schools college education program work
International aid	haiti international government million president relief aid earthquake haitian clinton

Figure 3: Topic models (n=17) for newspaper data on all phases

Tools

Dictionaries & libraries that were used for preprocessing and analysis:

- MPQA lexicon (for subjectivity and sentiment analysis)
- MALLET (topic modeling)
- Scikit-learn (NLP)
- Python NLTK (for pre-processing, tokenizing, parsing, etc.)

Preliminary Conclusions

1) Preliminary insights from *quantitative* data analysis:

- Top TF-IDF terms vary by different phases of disaster management (i.e. *Response*: food, water; *Recovery*: plan, community)
- Prevalent topics extracted with topic modeling were: medical aid, donations, international aid, business/economics, and religion.

2) Preliminary insights from *qualitative coding* of each text source:

- Twitter: Peak activity on January 13. Many calls for donations and volunteers for help.
- Interviews: Situational awareness is difficult in the first 48 hours; inter-agency coordination difficult in first 72 hours.
- Newspapers: detailed anecdotes from heads of various agencies (USAID, USCG, many NGOs). Situational updates (death tolls, injured, missing persons).
- Blogs: Commentary on governmental/federal efforts for Haiti response. Link Haiti situation to past disasters.
- Government reports (SITREPs): First-hand situational updates from personnel on-the-ground.
- Scientific journals: Wide range of disciplines taking interest in Haiti response dynamics & coordination efforts during response, relief, restoration, recovery.

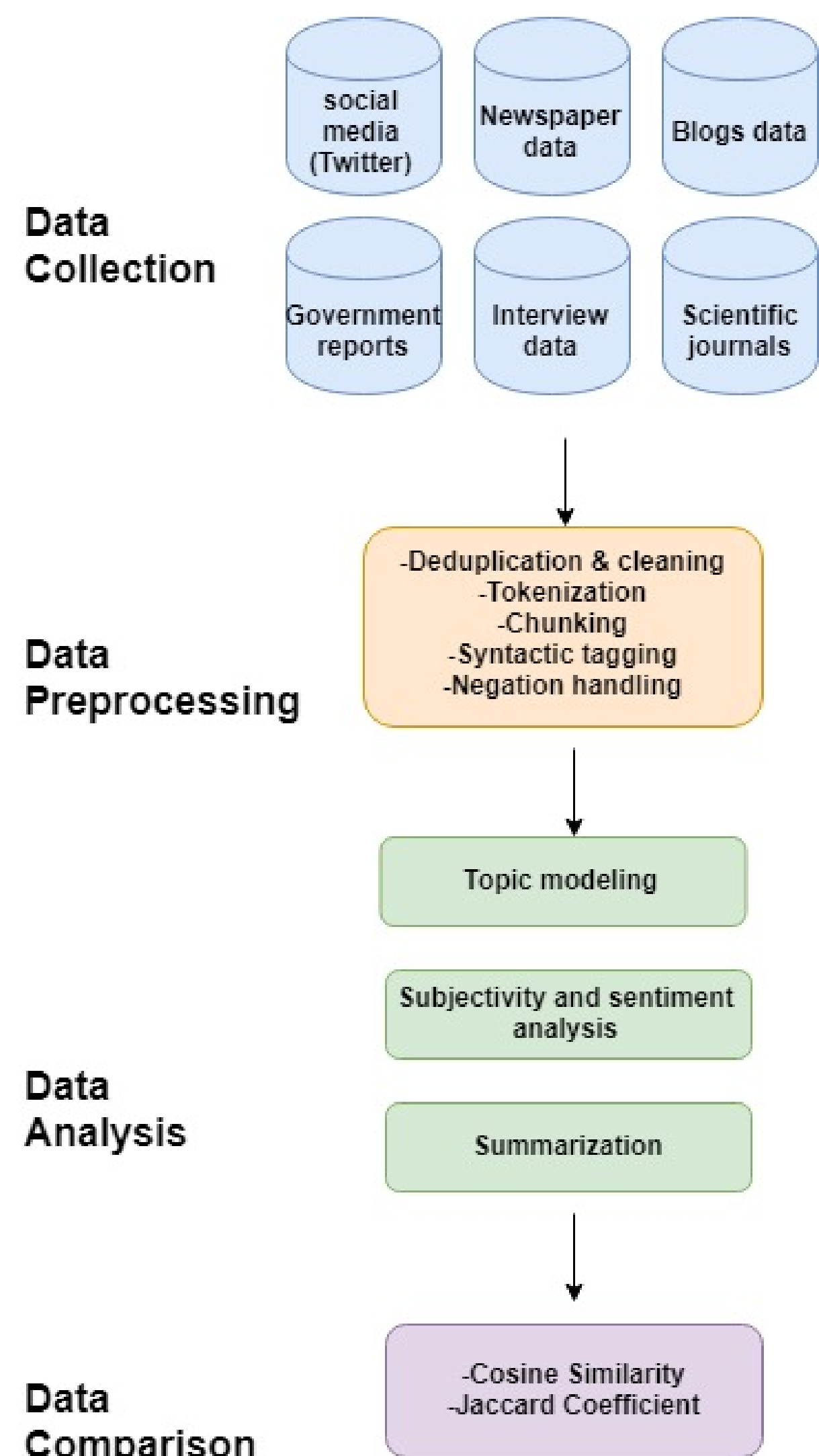


Figure 4: An NLP Pipeline used for this study

Future Work

- 1) Complete data analysis for remaining data sources (government data, scientific journals)
- 2) Summarization: compare results against each other against ground truth (the reconstructed timeline of events from interviews & secondary data)
- 3) Expand and apply pipeline to other disaster contexts; i.e. Chile earthquake & hurricane Irma.

Acknowledgments

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number, 2015-ST-061-CIRC01.