

# Variation in Situational Awareness Information due to Selection of Data Source, Summarization Method, and Method Implementation?

M. Janina Sarol\*, Ly Dinh\*, and Jana Diesner  
School of Information Sciences, University of Illinois at Urbana-Champaign

## RQ1: What, if any, are the differences in situational awareness information depending on source of information?

- **Most common topics (using LDA)**
  - **Blogs:** affect-related and donations
  - **News:** relief and international aid
  - **Twitter:** Haiti news and donations
- **Summary sentences**
  - **Blogs:** magnitude, needs, religion, relief efforts, school, and donations
  - **News:** magnitude, relief efforts, needs and casualties
  - **Twitter:** magnitude, requests, number of affected people, and an update link

Blogs	News	Twitter
assistance	resources	collecting
medical-assistance	medical-care	provide
food	assistance	accepting
needed	medicines	continue
need	clean-water	save
equipment	urgently-needed	support
distribution	equipment	auction
resources	medical-assistance	credible
medications	essential	pounds
medicines	stocks	pledge

Top 10 Resource Needs

## RQ2: What, if any, are the differences in situational awareness information depending on text summarization method?

- We compared two text summarization methods: COWTS and SumBasic
- Regardless of data source, the COWTS and SumBasic summaries contained more different information than similar information
- Twitter has the most information overlap, with many sentences in both summaries about requests for donations
- Blogs and news COWTS and SumBasic summaries are vastly different

<b>COWTS</b>	red-cross says 3 million people affected by the earthquake in haiti
<b>SumBasic</b>	over 3 million people have been affected by the earthquake in haiti.

Sample Summary Sentences from Twitter Data (Same Information)

<b>COWTS</b>	as a result a uk international search & rescue team has been deployed. the world's nations have pledged \$1 billion (pounds 616 million) in aid.
<b>SumBasic</b>	the relief system might not be working yet in haiti

Sample Summary Sentences from News Data (Conflicting Information)

## RQ3: What, if any, are the differences in situational awareness information depending on implementation of a text summarization method?

- We compared two implementations of COWTS, varying the selection of top words
  - **Implementation 1:** term frequency-inverse document frequency (TFIDF)
  - **Implementation 2:** keywords-in-context (KWIC)
- Regardless of data source, the TFIDF and KWIC implementations contained more different than similar information
- There is only a single case of the same sentence in both summaries (blogs)

<b>COWTS-TFIDF</b>	jose irazuzta, md support
<b>COWTS-KWIC</b>	project-hope's relief efforts in haiti donate today

Sample Summary Sentence from Blogs Data (Same Sentence in Two Summaries)

<b>COWTS-TFIDF</b>	right now they need money for food, water and medical supplies. despite great difficulties, god's people have found a way to help others.
<b>COWTS-KWIC</b>	help me raise money for mercy-corps' response to the haiti-earthquake. "human-beings are in some ways like bees," professor haidt said.

Sample Summary Sentences from Blogs Data (Different Information)

## RQ4: What are practical implications of choices for source, method, and implementation on resembling first-responders' accounts?

- We compared our results to first responder (FR) accounts, which were comprised of US and UN situational reports and interviews with first responders
- None of the organizations mentioned in the FR accounts were mentioned in the blogs, news, and Twitter summaries
- Blogs, news, and Twitter summaries contain different specific information from the FR accounts
- However, news data contains similar types of information to FR accounts (e.g. relief efforts of non-governmental agencies)
- Blogs and news need detection results correspond to the needs mentioned in the FR