

Towards Developing Methodology to Stem the Tide of Fake News

Kiran Kumar Bandeli, Muhammad Nihal Hussain, Samer Al-khateeb, and Nitin Agarwal

kxbandeli@ualr.edu

Department of Information Science, University of Arkansas at Little Rock, USA



Abstract

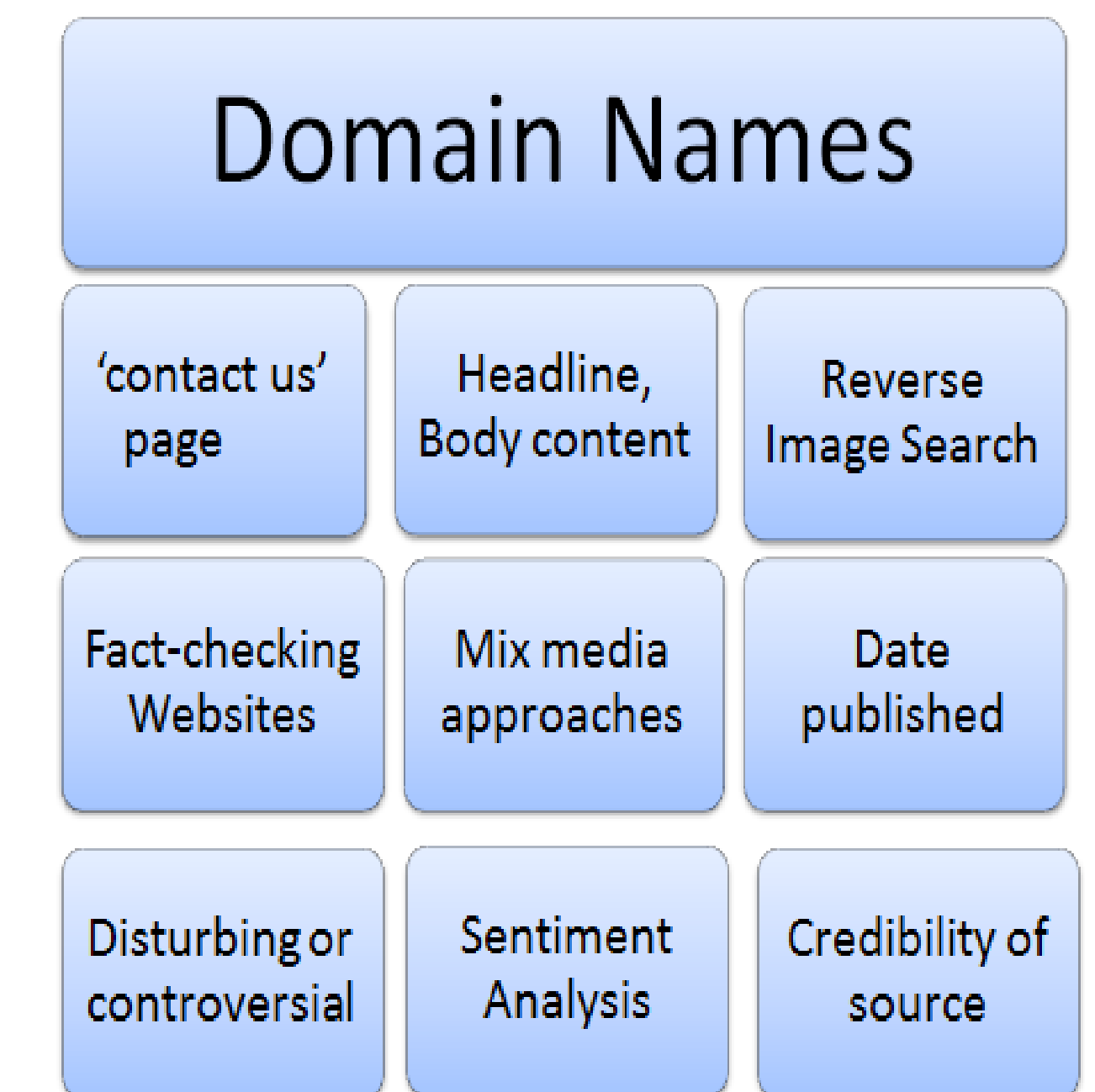
Numerous journalistic accounts have highlighted the prolific but disturbing use of social media to spread fake news. This phenomenon is fairly new with regards to the tactics, strategies, and procedures deployed to disseminate disinformation. Very little to no research has been conducted to systematically study the tactics, techniques, and procedures used to disseminate fake news. It is intriguing to see the way misinformation is disseminated along different social media channels.

For this data challenge, we explore the dataset provided to find answer to the following research question: Can we detect patterns or develop measures to help identify fake news? We started the experiment by loading the dataset into IBM Watson Analytics to have an insight about the data attributes, i.e., the data structure, in an effort to detect key patterns that will help us in finding an answer for the aforementioned research question. Then, we used off the shelf tools such as TouchGraph SEO Browser, MALLET (Machine Learning for Language Toolkit), LIWC (Linguistic Inquiry and Word Count), and Alchemy API to do a deep dive into the problem space. Later, we present our empirical observations based on the data exploration and provide heuristic measures for fake news source detection.

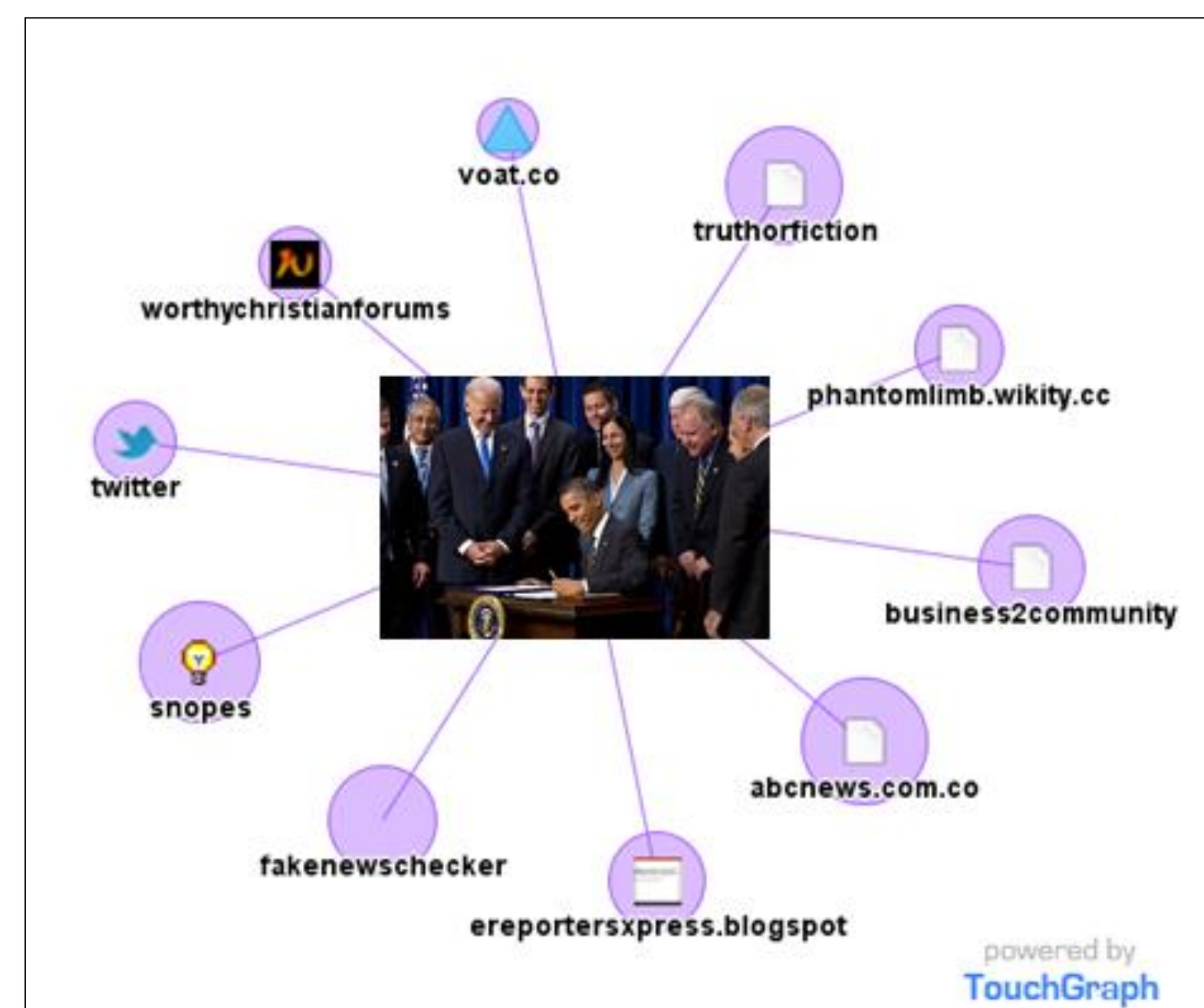
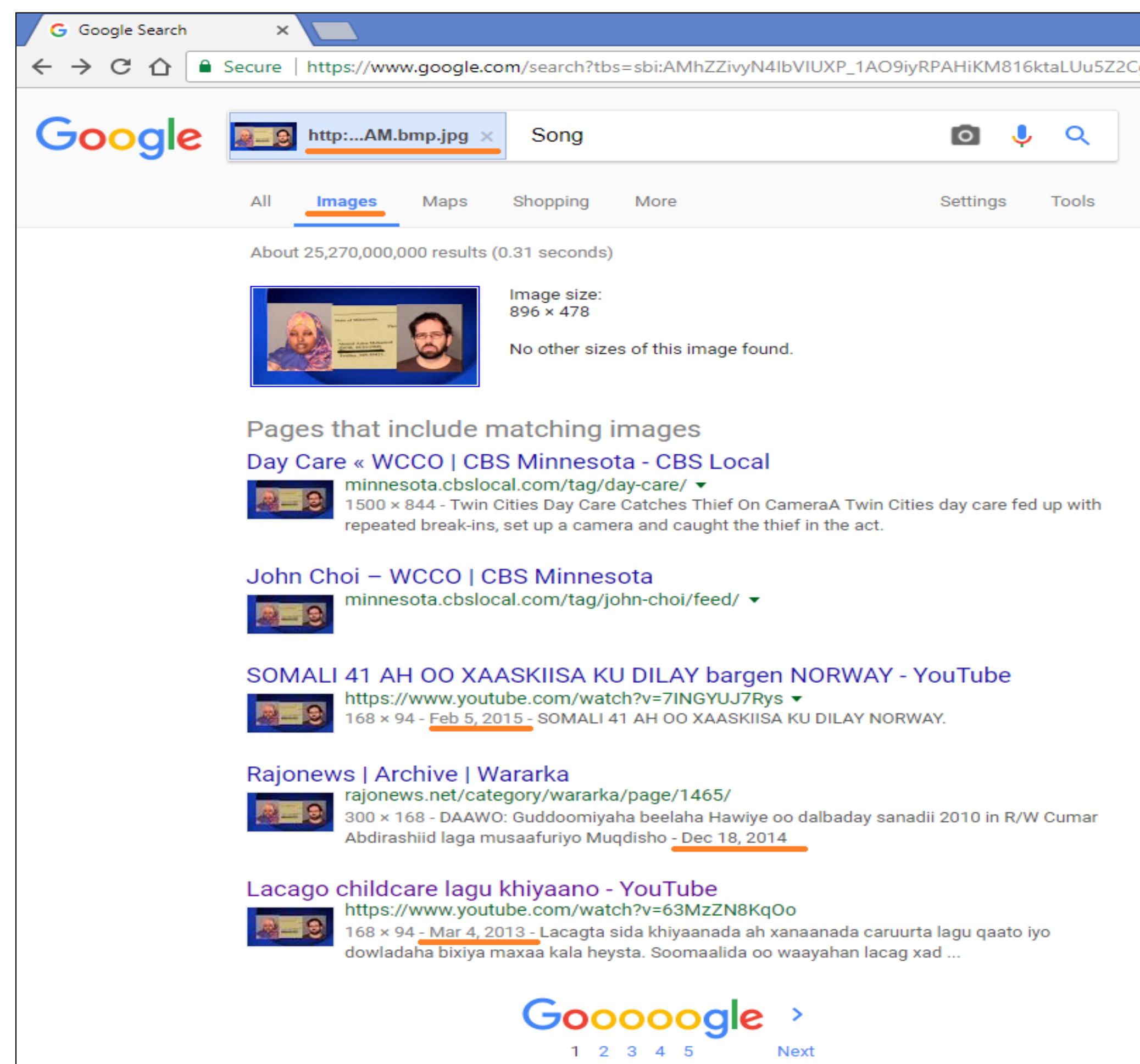
Challenges

- Knowing what is factually correct is very hard given the barrage of biased, satirical, or conspiracy theory-riddled stories on social media.
- Information deluge on social media challenges stemming the flow of mis/dis-information.
- Echo chambers quickly emerge on social media and envelope us making the problem even worse.

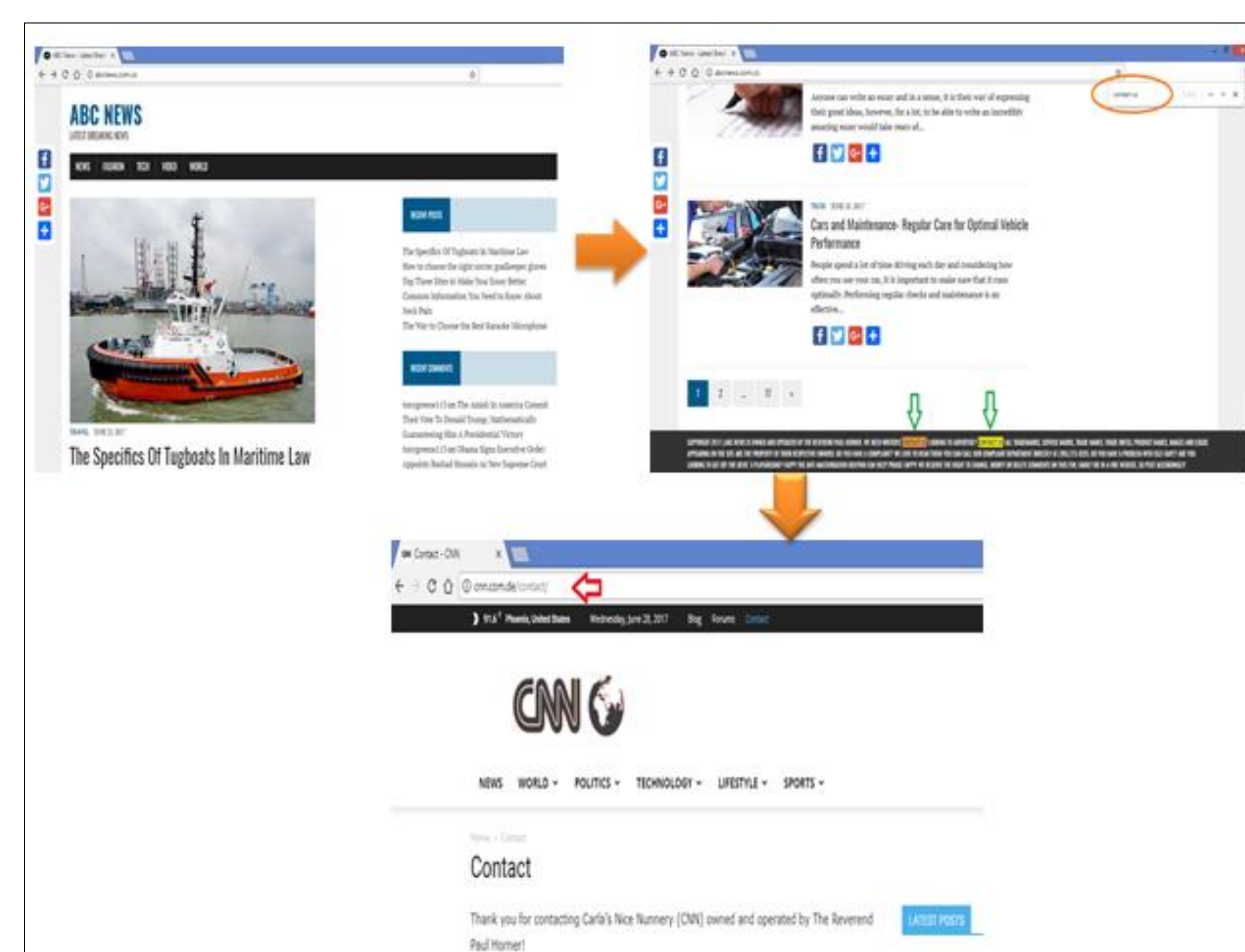
Proposed Methodology



Observations and Heuristics



The post is disseminated on different websites



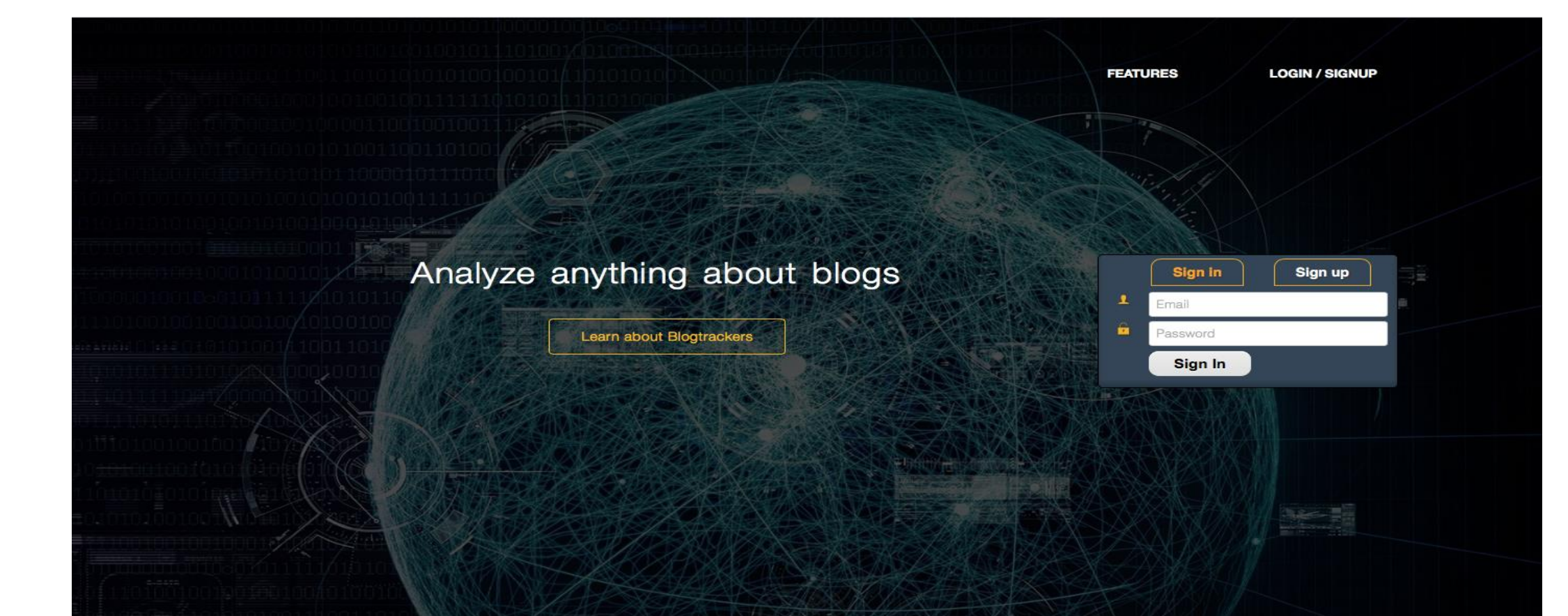
"Contact us" page redirects into another website

Reverse Image Search shows using same image with different narratives

Attribute	Attribute with Zero Values	Total Number of Posts	Percentage
comments	12,905	12,999	99%
likes	12,468	12,999	96%
replies_count	12,304	12,999	94%

Statistical data showing how posts are disseminated without any comments, likes, and replies.

Blogtrackers Features



- Posting Frequency
- Keyword Trends
- Sentiment Analysis
- Blogger Influence Score
- Blog Influence Score
- Data Export
- Various Metadata Reports
- Entity Network



Acknowledgments

This research is funded in part by:

- The U.S. National Science Foundation (IIS-1110868 and ACI-1429160)
- The U.S. Office of Naval Research (N000141010091, N000141410489, N0001415P1187, N000141612016, and N000141612412)
- The U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-16-1-0189)
- The U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059)
- The Jerry L. Maulden/Entergy Fund at the University of Arkansas at Little Rock.

Domain Names	Total Words Count	Affect	Positive Emotions	Negative Emotions	Anxiety	Anger	Sad
twitchy	1	100	0	100	100	0	0
info wars	2	50	0	50	0	50	0
anti war	2	50	0	50	0	50	0
abel danger	2	50	0	50	0	0	0
defend democracy	2	50	0	50	0	0	0
americas freedom fighters	3	66.67	33.33	33.33	0	33.33	0
the rebel media	3	33.33	0	33.33	0	33.33	0
mad world news	3	33.33	0	33.33	0	33.33	0
the last line of defense	5	20	0	20	0	0	0
anonhq	1	0	0	0	0	0	0
vdare	1	0	0	0	0	0	0

Domain names showing highly opinionated or clearly indicate a bias