Automating the Ranking of Article Visibility through Crowdsourced Trustworthiness

A Technical Report
presented to the faculty of the
School of Engineering and Applied Science
University of Virginia

by

Christine Baca

with

Sharon Bryant

date submitted in April 26, 2021

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Christine Baca

Technical advisor: Upsorn Praphamontripong and Daniel Graham, Department of Computer Science

Automating the Ranking of Article Visibility through Crowdsourced Trustworthiness

Capstone Project Proposal, 2021

Sharon Bryant[†]
Computer Science
The University of Virginia, School of Engineering and Applied Sciences
Charlottesville, Virginia USA
ssb7xx@virginia.edu

Christine Baca
Computer Science
The University of Virginia, School of Engineering and Applied Sciences
Charlottesville, Virginia, USA.
cab8xd@virginia.edu

ABSTRACT

Social media fosters confirmation bias that encourages the spread of misinformation. We propose a system that uses crowdsourced trustworthiness to give users a news feed of diverse articles/headlines to discourage the formation of echo chambers. We will address the question: in what ways can news distributors engineer machine learning applications and user interfaces to decrease the propagation of disinformation? Our findings will include evaluations of our proposed project's topic variance, media diversity, and the effectiveness of a news-shuffling button and compare user and fact-checker trustworthiness. The current name and branding of the application is "Rise and Grind".

1 INTRODUCTION

The spread of misinformation is rampant on the internet and social media platforms. Malicious agents utilize controversies and conspiracies to further polarize users on social media platforms within the United States. Misinformation spread contributes to growing anti-science sentiments, vaccine mistrust, and political polarization. Automated journalism may increase the amount of users that fall victim to confirmation bias. The Pew Research Center found that "62%" of U.S. adults "get news on social media, and 18% do so often"[1]. Users, particularly in the United States, consume a large amount of news media on social media and often on the same platform. Large tech companies such as Facebook therefore influence and contribute to issues of political polarization [2]. Selective exposure is the primary driver of "echo chamber" creation, and conspiracy newsreaders reacted very negatively to debunking posts, even interpreting them as acts of disinformation [3]. Debunking hurts more than helps the prevention of conspiracy theory and misinformation spread.

Social media frequently relies on labeling to prevent misinformation spread. Relying solely on labels is inefficient and difficult to execute as labels must prevent the continued influence effect, familiarity backfire effect, overkill backfire effect, and worldview backfire effect to be truly effective [4]. The continued influence effect is when "despite a retraction, people continue to rely on misinformation", and a label may not repetitively allow for an alternative account to fill in the gap left by the retraction of misinformation [4]. The worldview backfire effect, where "evidence that threatens worldview can strengthen initially held beliefs", parallels the effect of debunking posts [4]. Ineffective labeling can have similar effects to debunking posts that have been shown to fortify biases.

Currently, social media platforms and news distributors have developed multiple new ways to combat misinformation with varying levels of effectiveness. Facebook and Twitter both implement labeling to combat against misinformation. With the COVID-19 pandemic and the increase of vaccine misinformation, Facebook and Twitter have focused their efforts on improving their methods. Facebook utilizes "third-party fact-checking organizations" that are "non-partisan." Facebook's programs have several aspects such as "identifying false news," "reviewing content," "clearly labeling misinformation, and informing users about it," "ensuring that fewer people see misinformation," and "taking action against repeat offenders" [5]. Twitter implemented a "community-based approach to misinformation" called Birdwatch that "allows people to identify information in Tweets they believe is misleading and write notes that provide informative context" [6]. Twitter also implemented a strike system for repeat misinformation spreaders [7].

Due to unmoderated misinformation spread, dangerous beliefs about COVID-19 and other fringe theories such as QAnon have increased. The Pew Research Center found that though "four-in-ten Americans say COVID-19 came about naturally; about three-in-ten think it was created in a lab" [8]. The Pew Research Center also found that in a survey taken in a time frame between late February and early March of 2020, "about a quarter (23%) of U.S. adults said they had heard 'a lot' or 'a little' about QAnon [9]. By September, that number had increased to 47%", and on YouTube "5% of videos published by the 100 most viewed YouTube news channels at the time of the study included the word "QAnon."[9] Stocking et al. (2020) found that "about a quarter of all U.S. adults (26%) say they get news on YouTube" [10].

The interaction of users, social media platforms, and confirmation bias create the environment for misinformation spread and can negatively impact society. Negative impacts can range from election fraud beliefs destabilizing trust in democratic procedures to vaccine mistrust.

2 BACKGROUND

Online news personalization algorithms make consumers vulnerable to confirmation bias [11]. Ciampaglia and Menczer (2018) contend that confirmation bias, the propensity to believe information that supports one's beliefs and ignore those that discount them, make up a large part of misinformation spread. With machine learning (ML), engineers can apply artificial intelligence to communications media, introducing new advantages and new hazards. ML algorithms can automate validation of online news, and it can even automate journalism itself, through algorithms that "automatically generate news." Within the news industry, machine learning enables automated journalism: "the use of algorithms to automatically generate news" [12].

Filter bubbles and echo chambers exploit confirmation bias [13]. Created by Skider et al. (2020), a social network model demonstrated "that a lack of confirmation bias can ensure that small biased minorities much more easily hijack and dictate public discourse." A densely connected society promotes "a greater amount of biased reasoners" before polarization takes place [13]. Confirmation bias is unavoidable, but increasing awareness of confirmation bias in users and when designing systems of news distribution may help decrease the spread of misinformation. Pennycook & Rand (2019) found that, when discerning false news from mainstream news, user crowdsourced trustworthiness scores generally match professional fact-checker scores [14].

3 RELATED WORKS

Many systems rely on crowdsourcing to find solutions to different problems. Applications that aid in disaster reporting may use crowdsourced reports to provide real-time reporting from observers. Weaver et al. (2012) created a crowdsourcing disaster reporting web application and provided suggestions to improve the trustworthiness of the reports either through a selection or combination of group membership, crowdsourced rating, and machine learning [15]. An application that relies on crowdsourcing related to news distributions is Grasswire. Grasswire (2015) describes itself as "a community of over 1,200 people from all over the world who care about honest, accurate news." [16] Grasswire's method of news reporting is through collaborations specifically "online in an open Slack channel" where users can "pitch, source, verify, write and edit newsworthy, interesting and unbiased stories" [16].

Both systems apply to understanding the context of crowdsourcing applications and our application: Rise and Grind. The described disaster reporting algorithm relies on aggregating and changing the ranking of disaster reports through submitted observation reports, similar to how users can rate article trustworthiness in Rise and Grind to affect the visibility of the articles when being recommended. Grasswire's approach to crowdsourced news is a bit different. The information is crowdsourced rather than focusing on the recommendation of preexisting news articles. Though Grasswire's approach may be effective at producing reliable news, it is not as adoptable by established news distributors and doubtful to reliably compete in comparison to news organizations and independent news creators.

4 SYSTEM DESIGN

4.1 Review of System Architecture

The "Rise and Grind" system architecture would be modeled after a client-server web application and would provide the framework for a typical news recommendation system. The web application would consist of a client layer, application layer, web-server layer, and data layer. The client layer sends requests to and receives responses from other system components and provides users with an interface to use the application's services. The application layer connects the client and data layer and provides the

back-end framework for the application's services. The web-server layer passes the requests and responses between the clients and the applications and hosts the services. The current version of the application is hosted on localhost and the Google Cloud Platform. The data layer comprises the "Rise and Grind" databases accessed and modified by the methodologies of other layers. The recommender system layer uses machine learning models and the data layer to compute recommendations and outputs to the server. The recommendation system will consist of a simple scoring model that algorithmically disfavors news sources with low crowdsourced trustworthiness ratings to decrease the amount of misinformation and disinformation online. The algorithm, also, would weigh the ratings according to the publisher size and notability to protect smaller yet strong publications from excess distrust.

4.2 Web page Requirements

The application should prioritize the following design principles and usability criteria:

- To be a dynamic user interface.
- Easy learning curve and accessibility.
- Fast user interface performance.
- Consistent terminologies and usage.
- The minimization of user errors and maintainability in bug patching.
- Subjective satisfaction.

The content of the application should include:

- Project overall and specific objectives.
- Up-to-date information and credentials.
- Redirection links and credit to the sourced news articles.

The functionalities for the end-application should be provided by:

- News articles from a news API.
- The users themselves (by crowd-sourcing).
- The proposed recommendation algorithm.
- A sustainable server and host on the internet with data upload and download capabilities.

4.3 Key Components

The proposal has a prototype, a partially implemented website, and a relational database.

4.3.1 Project Website Development.

4.3.1.1 Prototype. A prototype of the site was created in Mockflow (https://www.mockflow.com/), a wireframing software on the cloud that facilitates collaboration between designers on user-interface prototypes for websites and software (Mockflow Prototype). There are three main scenarios to the "Rise and Grind" application: a home news feed, an article-viewing page with credibility rating functionality, and a profile page.

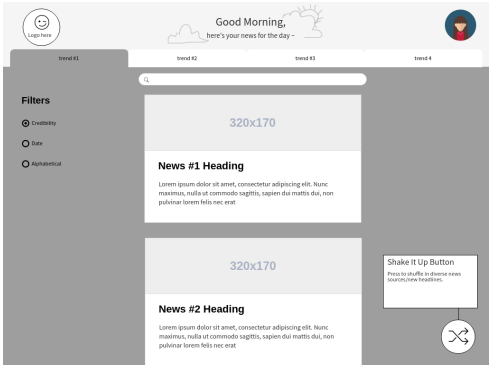
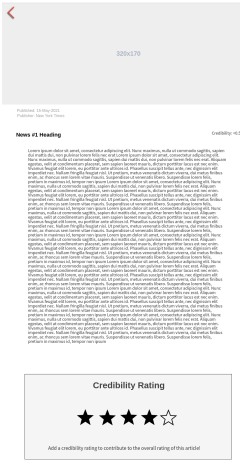


Figure 1: The home news feed, the first screen that users see when the application opens if they are already logged in or signed up.

The home screen will have an endless scrolling-style news feed with articles from multiple news publishers. Most users are comfortable with this style of media browsing, so we have also adopted it into our application to minimize the learning curve. The "shake it up" or shuffle button will be present on the home screen in the bottom right corner to encourage users to seek out more diverse media sources and articles. Users will also be able to filter articles on the



news feed that match specific queries and browse through different trending topics.

Figure 2: The article-viewing page with credibility rating functionality at the bottom of the page.

Once the user clicks on a news article, they are directed to an article viewing page. Users can focus solely on the article in a non-distracting setting. At the bottom of the page, when users finish reading, they can rate the credibility of the article. This credibility score gets calculated and incorporated into the crowdsourced credibility score for the article. The design of the rating system may change as the project develops.

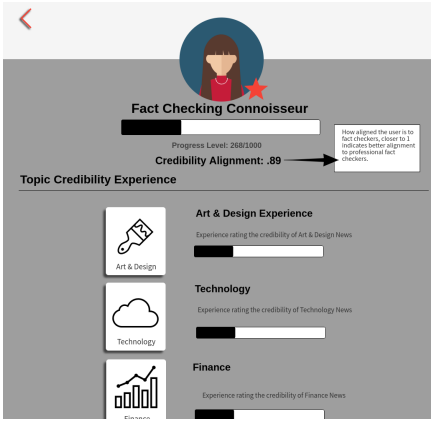


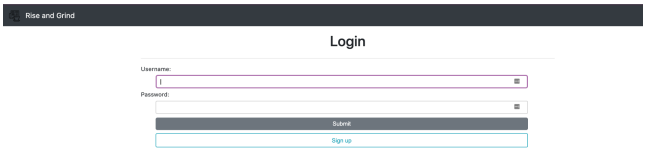
Figure 3: The profile page of the user that shows statistics on their credibility alignment to professional fact-checkers and by different article topics.

To create a sense of gamification, the profile page layout features different topics with corresponding reliability scores based on how users align with professional fact-checkers. We want to push users to continually participate in the crowdsourcing of credibility scores. The project utilizes achievements and a points system to maintain user participation. By giving users a credibility alignment score, they are rewarded for being as closely aligned to professional fact-checkers as much as possible.

4.3.1.2 Current Version. The current version of the proposed system is a web application created using HTML, PHP, JavaScript CSS, Bootstrap's free templates, XAMPP, SQL, and public assets and is currently hosted on the (Apache? XAMPP?) localhost and the Google Cloud Platform.

4.3.1.2.1 Website Content.

Figure 4: login-handler.php



The user can log into their account with an existing username and password. If the user does not have an account or wishes to make a new one, there is a button with a sign-up prompt that, when clicked, redirects the user to the sign-up page. For security reasons, if the user fails to enter a valid username and password three times, the submit button will be disabled.

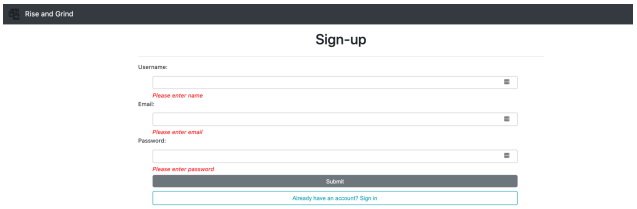


Figure 5: sign-up-handler.php

The user can sign up and create an account by submitting a unique username and password in the form. If the user already has an account, there is a button with a login prompt that, when clicked, redirects the user to the login page.

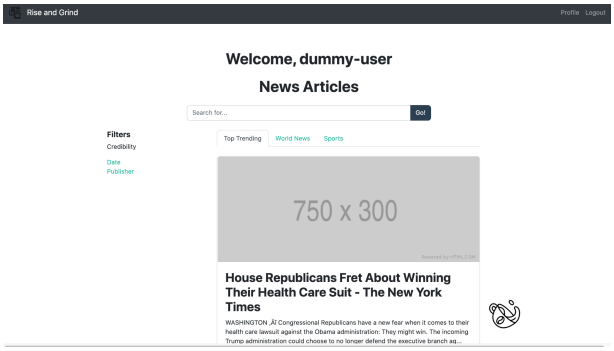


Figure 6: home.php

The home page presents the user with a welcome header. The user can scroll through the home page and preview a list of available articles. The articles can be shuffled by clicking on the coffee button in the bottom right corner of

the page, and filtered according to the tabs and filtering menu. The user can log out of their account and return to the login page by clicking the "Logout" link in the top right corner. By clicking the "Profile" link in the header, the user can be redirected to their profile page.

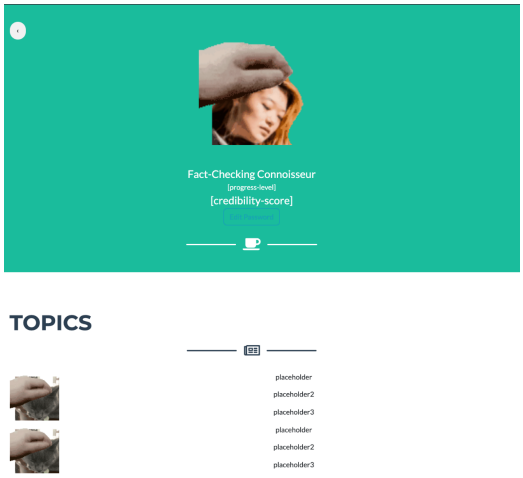


Figure 7: profile.html

The profile page provides the user with an illustration of their account components such as their profile picture and current points. The user can click the "Change password" button to be redirected to a password change page.

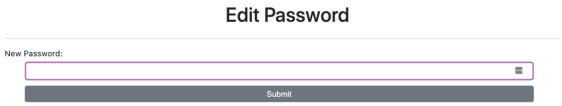


Figure 8: password-change.php

The user can update their password by filling out and submitting the prompt on the password page.

Figure 9: logout-handler.php

After clicking the logout prompt on another screen, the user's session will end and be directed to logout validation (shown above). After three seconds, the user will be redirected to the login page.

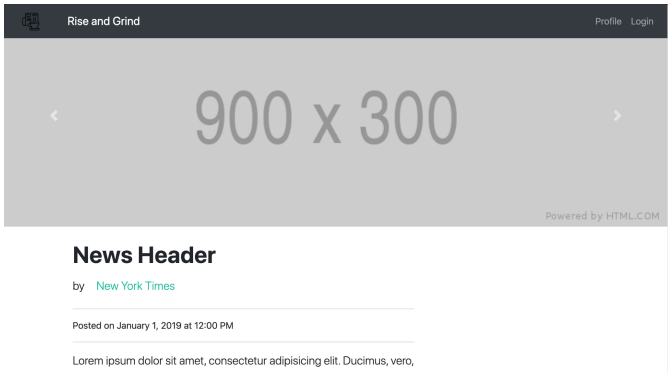


Figure 10: reading.html

4.3.1.2.2 Software components. The chart below depicts the software features each program file supplies.

Filename	Role	Features
login-handler.php	Creates a form that allows users to log into their "Rise and Grind" account.	- Provides a redirection button to the sign-up page. - If the user submits a correct form, redirects the user to the home page and starts the session
sign-up-handler.php	Creates a form that allows users to sign up for a "Rise and Grind" account.	- Provides a redirection button to the login page. - If the user submits a correct form, redirects the user to the home page and starts the session.
home.php	Creates visual and dynamic interface for user interaction on the home page	- Provides redirection links to other pages. - Displays an interactable list of available articles.
reading.html	Displays article and rating prompt.	- Provides redirection to other pages. - The user can rate the article.
profile.ph	Creates visual	- Provides a

p	and dynamic interface for user interaction on the profile pages.	redirection button to the change password page. - Displays user details (username, photo, statistics).
edit-password.php	Provides software for changing a specific user's password	- Changes user's password on the server and database if form submitted correctly
logout-handler.php	Provides software for logging out the current user and terminates the ongoing session.	- Validates user's logout request and redirects user to login page - Terminates user's session.
db-functions.php	External PHP file with user-defined functions that provides database helper functions.	- Interfaces with relational database - Protection against SQL injection. - Functions that, when executed, can connect to, select data from, and add or update the database.
start-session.php	External PHP file with user-defined functions.	- Uses server requests. - Instantiates a user's session.

4.3.1.2.3 Databases.

- DATABASE SERVER: MySQL 7.4.15

The application uses MySQL 7.4.15 database service. MySQL was chosen because it provides software developers with a simple, easy-to-learn, and flexible service for creating and utilizing databases. The database is named rg_db (Rise and Grind database) and contains two tables: articles and users. The users table contains information relating to each user. The articles table contains information about each accessible article.

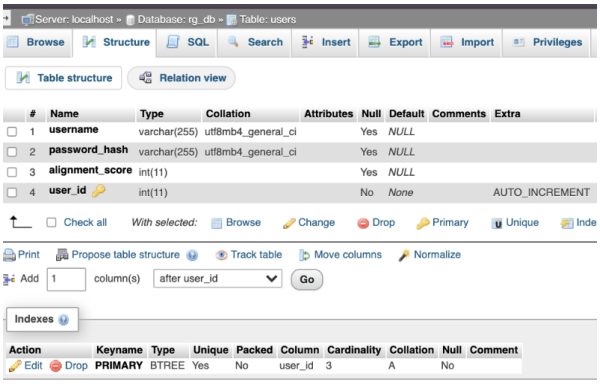


Figure 11: Structure of user table in mySQL database rg_gd

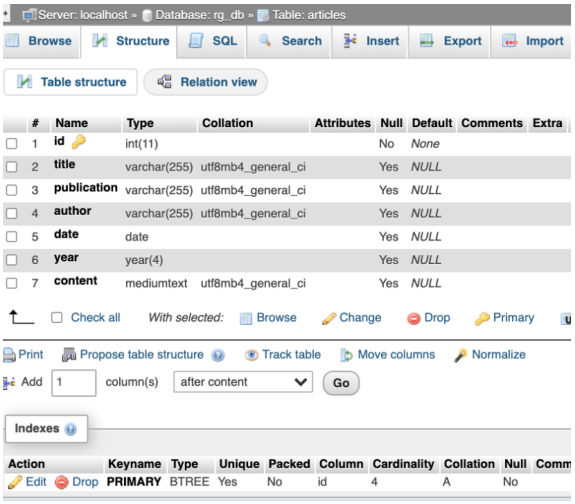


Figure 12: Structure of articles table in mySQL database rg_db

4.4 Challenges

Challenges in creating the project proposal included learning new concepts, and maintaining flexibility in a remote working environment. One group member had no prior experience in web development until starting CS4640 a few weeks before the initial submission of the project abstract. Both the class and the project partner, fortunately, were able to provide substantial guidance and assistance, helping the inexperienced member to overcome the learning curve and further develop the web components of the proposal. The ongoing COVID-19 pandemic may have caused a lack of time and resources that limited the scope and depth of the proposal; the team was unable to meet in

person with advisors and other experts in the relevant fields. The team, instead, had to become more flexible and familiar with online communication platforms like Zoom and Google Drive to overcome the challenges of remote work environments. Similarly, the pandemic has skewed the traditional Capstone schedule and project guidelines, and the team had to work through a different timeline to complete the proposal. To help resolve the schedule changes, the team created and shared a Google Doc to keep track of TODOs, action items, and general expectations to prevent any mismanagement of deadlines.

5 RESULTS

The project proposal and its latest version of the 'Rise and Grind' web application need additional work to undergo evaluation for topic variance, media diversity, and the effectiveness of a news-shuffling button and compare user and fact-checker functionality. With future work, the team hopes that the proposal can help create a system that uses crowdsourced trustworthiness in a recommendation algorithm in a web application that gives users a news feed of diverse articles/headlines to discourage the formation of echo chambers and extend the conversation of automated journalism.

6 CONCLUSIONS

We designed a system to mitigate the issues related to the spread of misinformation on digital media. The system should ideally decrease the amount of misinformation spread and increase media literacy to give users a tool to protect themselves from misinformation, echo chambers, and confirmation bias. The latest version of the project needs additional development, and we hope that our proposal extends the conversation on automated journalism and its potential to minimize the spread of misinformation and help create media literacy tools. Future work could be further extensions of this project such as a full and marketable implementation of the "Rise and Grind" web application and recommendation algorithm.

7 FUTURE WORK

Given additional time to work on this system, we would test, extend, and finalize the prototype to make a marketable implementation. Future add-ons could include an Angular implementation and extending the project to incorporate larger databases and News APIs as currently the project only accesses a small local database. We could also further implement the news-shuffling button and fact-checker features and make the website more user-friendly and educational in teaching users about news propagation on digital media.

REFERENCES

[1] Gottfried, J., & Shearer, E. (2016, May 26). News use across social media PLATFORM.

[2] Bode, L., Vraga, E.K., 2015. In Related News, That Was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media. Journal of Communication 65, 619–638.. doi:10.1111/jcom.12166

[3] Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H.E., Quattrociocchi, W., 2016. The spreading of misinformation online. Proceedings of the National Academy of Sciences 113, 554–559.. doi:10.1073/pnas.1517441113

[4] Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction. Psychological Science in the Public Interest, 13(3), 106-131. doi:10.1177/1529100612451018

[5] Facebook. (n.d.). Fact-Checking on Facebook.

[6] Coleman, K. (2021, January 25). Introducing birdwatch, a community-based approach to misinformation.

[7] Roth, Y., & Pickles, N. (2020, May 11). Updating our approach to misleading information.

[8] Mitchell, A., & Oliphant, B. (2020, March 18). Americans immersed in Coronavirus news; most think media are doing fairly Well covering it.

[9] Pew Research Center. (2020, November 16). 5 facts about the QAnon conspiracy theories.

[10] Stocking, G., Van Kessel, P., Barthel, M., Eva Matsa, K., & Khuzam, M. (2020, September 28). Many Americans get news on Youtube, where news organizations and independent Producers Thrive side by side.

[11] Ciampaglia, G. L., & Menczer, F. (2018). Biases Make People Vulnerable to Misinformation Spread by Social Media. The Conversation US.

[12] Graefe, A. (2016). Guide to Automated Journalism. Columbia Journalism Review. https://doi.org/10.7916/D80G3XDJ

[13] Skider, O., Smith, R. E., Vivo, P., & Livan, G. (2020). A minimalistic model of bias, polarization and misinformation in social networks. Scientific Reports, 10(1). doi:10.1038/s41598-020-62085-w

[14] Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. Proceedings of the National Academy of Sciences, 116(7), 2521-2526. doi:10.1073/pnas.1806781116

[15] Weaver, A. C., Boyle, J. P., & Besaleva, L. I. (2012). Applications and trust issues when crowdsourcing a crisis. *2012 21st International Conference on Computer Communications and Networks (ICCCN)*. doi:10.1109/icccn.2012.6289256

[16] Grasswire. (2015). About us.