Towards A Sustainable Model for Fact-checking Platforms: Examining the Roles of Automation, Crowds and Professionals

Naeemul Hassan[1], Mohammad Yousuf[2], Md Mahfuzul Haque[3]
Javier A. Suarez Rivas[1], Md Khadimul Islam[3]
[1]Department of Computer Science and Engineering, The University of Mississippi
[2]Gaylord College of Journalism and Mass Communication, The University of Oklahoma
[3]Meek School of Journalism and New Media, The University of Mississippi

ABSTRACT

A major challenge for fact-checking platforms to survive is a lack of resources. Platforms relying solely on professionals may not be able to become self-sufficient because of the amount of resources they require. We propose that a sustainable model requires incorporating the wisdom of the crowd and automated assisting tools into the process, which will increase efficiency and decrease costs. In this study, we examined a crowd curated political fact-checking platform, *reddit.com/politicalfactchecking*, and identified various roles crowds and professionals play in fact-checking. We've also developed an automated argument classification model and identified some steps in fact-checking, which could be automated.

1 MOTIVATION

Many professional fact-checkers are suspicious of the idea of crowdsourced fact-checking in which users verify factual claims [7]. Professionals often claim that users lack required skills, and are biased. We argue that contribution of crowds to fact-checking is essential in the networked media ecosystem where information is abundant and rumors spread like wildfire with resources for investigative journalism steadily plummeting [6, 14]. A sustainable model for fact-checking platforms would consist of crowds, professional fact-checkers, and automated assisting tools. Crowds can perform many mundane but important tasks under the guidance of professionals while programmers build tools to find credible sources and make sense of a large amount of user inputs. Crowds can also help better identify facts that people care about and identify sources that may lack credibility but still are popular. Professionals can play the 'moderator' and 'seminar leader' role in this process [16]. The purpose of this study is to explore such a crowdsourced fact-checking model where users check facts under the guidance of moderators. This study adds 'automation' to that model and identify roles of each component of this model. It examines posts and comments in a crowd curated political fact-checking platform of *reddit.com/politicalfactchecking* to find answers to the following questions: **i)** what claims do people want to see fact-checked? **ii)** what role does crowd play in fact-checking a claim? **iii)** what roles do moderators play in crowdsourced fact-checking? **iv)** how can computation play a role in scaling-up crowdsourced fact-checking?

To understand roles of the crowd in finding factual claims, we analyzed the posts in *reddit/politicalfactcheck* and identified popular topics that drew more user participation (e.g., number of comments), original sources of claims (e.g., mainstream media, alternative media, blogs), and type of claims (e.g., stat, figure). To understand the roles of crowd in fact-checking, we analyzed the comments and identified types of arguments users provide, originality of sources that users present to support or reject claims, and relation between crowd actions and professional fact-checking activities. To identify roles of moderators, we conducted a qualitative analysis on all posts and comments of moderators. We've also developed an automated argument classification model and identified some steps in fact-checking, which could be automated.

The findings suggest *Reddit's* model can lay a foundation for building a sustainable model for fact-checking organizations that lack resources. Roles of the crowd and automated assisting tools in fact-checking can increase efficiency and decrease expense and thus will enhance sustainability. Models solely reliant on professionals are not sustainable given the amount of resource they require to survive.

2 METHODOLOGY

2.1 Dataset

In this work, we study the crowd curated online political fact-checking platform *reddit/politicalfactchecking* [1]. By and large, the platform functions in the following way– **i)** a user creates a post to submit a claim, **ii)** the community presents evidence and arguments related to the claim using the platform's comment and reply features, **iii)** a group of moderators decides the appropriate flair for the claim based on the accumulated evidence and provides a justification. Note, both the post creator and the moderator group are members of the community and are eligible to participate in the second phase. A claim's flair can be modified later in case there is new evidence. Below, we provide a short description of each of the flairs. Detailed explanations are available in the platform's website. We say a claim is *fact-checked* if the corresponding post is assigned any of the flairs except *Please Verify* and *Mod Post/Meta*.

Confirmed: This post has enough evidence to support the claim.
Mostly True: The claim is accurate but needs clarification/context.
Half True: The claim is not entirely accurate, leaves out important information, or is out of context.
Mostly False: The claim contains some elements of truth but ignores critical facts that would change the reader's impression.
False: The claim is blatantly false.
Partisan Bias: This claim contains obvious political bias intended to edify one party or make another party look bad.
Unverifiable: This claim contains more opinion than fact. It can easily become a debate. This tag also covers doublespeak, instances of incorrect terminology, or other miscellaneous claims that cannot

[1]https://www.reddit.com/r/politicalfactchecking

Flair	# Posts	Avg. # Comments+Replies per Post
Confirmed	32	20
Mostly True	21	23
Half-True	18	27
Mostly False	13	18
False	70	23
Unverifiable	39	27
Partisan Bias	16	18
Please Verify	111	18
Mod Post/Meta	44	43
No Flair	179	13

Table 1: Distribution of the flairs

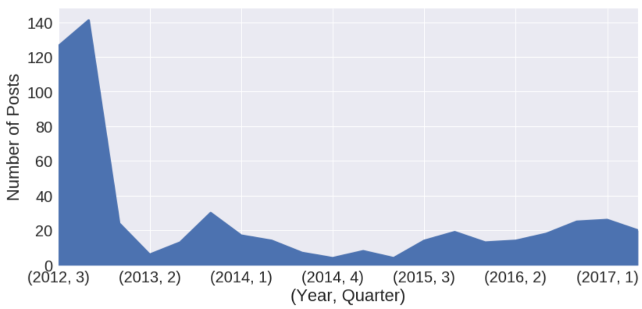

Figure 1: Number of posts per quarter

be proved true or false.
Please Verify: This post needs to be researched more in order to determine the validity of the claim.
Mod Post/Meta: This is a post by a moderator or a discussion about improving *reddit/politicalfactchecking*.

Using PRAW (Python Reddit API Wrapper) [2], a Python package which allows access to *reddit's* official API, we scraped all (everything before July, 2017) the posts, comments and their metadata (timestamp, author). In total, there are 543 posts, 2,835 comments, and 7,386 replies to the comments. Table 1 shows number of posts and the average number of comments and replies per post for each flair category. There are 209 posts which have been fact-checked, 111 posts in *Please Verify* category, 179 posts which didn't receive any flair, and 44 posts in *Mod Post/Meta* category.

Figure 1 shows number of posts created in each quarter since the platform's creation in September 2012 till July, 2017. It appears that the platform was more active at the initial stage. There were 10,083 members, 2,019 unique commentators, and 5 moderators in the community. About 80% of the comments and replies were created by 28% of the commentators. The average lengths of comments and replies are 477 and 331, respectively.

2.2 Coding and Analysis

We manually coded the posts and comments to answer research questions. The posts were coded for five variables: *post topic, post type, source of claim, nature of the claim*, and *fact-checked entity*. The comments were coded for three variables: *commenter action, argument type*, and *source of evidence*. We used a combination of analytical approaches to identify these variables and categories, as we did not find any study that analyzed *Reddit* posts for the purpose of understanding roles of crowds in fact-checking. We used a combination of deductive and inductive approaches [18]. We initially applied a deductive approach to identify the variables and the categories through a review of professional and scholarly literature on media content and typologies [13, 21]. Then, we used an inductive approach to refine those categories to fit the purpose of this study. The authors have had several training sessions to discuss and finalize the coding categories. One coder coded most of the posts while the comments were coded by four coders, three of whom were journalists with over 30 years of experience combined. Inter-coder reliability (Krippendorff's α)[4] ranged between 0.92 and 0.96 for comments, with an average of 0.94.

[2]https://praw.readthedocs.io/en/latest/

2.2.1 Posts. The first variable, *post topic*, included 12 categories: *Economy*, *Education*, *Election*, *Environment*, *Foreign Affairs*, *Health*, *Immigration*, *Equality*, *Media Bias*, *Religion*, *Security*, and *Other*. The second variable, *post type*, included three categories: *Fact-check Request*, *Seeking Information*, and *Judgement/Opinion*. We identify six categories for *source of claim*: *News/Information Media*, *Organization*, *Alternative Media*–websites where majority of contents is created by contributors, *User Generated Content (UGC) Platforms* (e.g., social media), *Personal Contact*, and *Others*. The fourth variable, *nature of the claim*, was coded for four categories: *Non-statistical* (e.g., quotes), *Statistical*, *Media* (e.g., photo or video), and *Other*. The fifth variable, *fact-checked entity*, included *Person*, *Organization*, *Event*, *Policy/Issue*, and *Other*. Note that a post may be coded with multiple categories for a variable.

2.2.2 Comments. The same categories used for the variable *post type* and *source of claim* in posts were used for the variable *argument type* and *source of evidence*, respectively. The *commenter action* variable was coded for six categories: *Providing Argument*, *Seek Clarification*, *Check Verifiability*, *Contact Claim Source*, *Assign Flair*, and *Post Irrelevant*.

2.2.3 Qualitative Analysis. To understand the roles of the moderators, we qualitatively analyzed the posts and comments published by the moderators. We applied an approach developed by Altheide et al.[1], which is widely used in various fields.

3 RESULTS AND DISCUSSION

Following the above explained methods, we analyze the *political-factchecking* dataset, present the results and discuss the findings in this section.

3.1 Role of Crowd in Finding Claims

There are 499 posts in the dataset published by the crowd. Table 2 shows topic distribution of the claims. 70% of all the claims belong to the top 4 topics– *Economy*, *Election*, *Security*, *Health*. For each topic, we identified the posts which have been fact-checked (*Confirmed*, *Mostly True*, *Half True*, *Mostly False*, *False*, *Unverifiable*, *Partisan Bias*) and which have not been fact-checked (*Please Verify*, *No Flair*). Table 2 shows number of fact-checked posts, number of not fact-checked posts, and the percentage of fact-checked posts for each topic. Among the topics having at least 10 posts, *Health*, *Foreign Affairs*, and *Economy* related posts have significantly higher probability of getting fact-checked than the *Immigration* and *Equality* related posts.

Topic	# Posts	# Fact-checked	# Not Fact-checked	% Fact-checked
Economy	130	62	68	47.69
Election	83	34	49	40.96
Security	68	24	44	35.29
Health	63	30	33	47.62
Other	39	12	27	30.77
Foreign Affairs	34	16	18	47.06
Media Bias	27	10	17	37.04
Equality	20	6	14	30.00
Immigration	14	3	11	21.43
Religion	6	4	2	66.67
Education	5	4	1	80.00
Environment	4	2	2	50.00

Table 2: Distribution of topics

Among all the posts, 352 of them are request for fact-checking claims, 108 are about seeking opinion/information (e.g., *where can I find ...*), and 49 posts are giving judgment regarding an issue without requesting any fact-check. About half of the fact-checking requests are statistical in nature (e.g., check a number, check degree of a phenomenon) and other half are non-statistical (e.g., check if someone said something). There are 13 requests for checking authenticity of photos and videos.

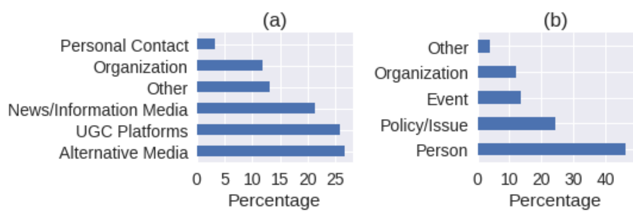

Figure 2: Distribution of source categories (a) and checked entities (b)

We investigated the sources of the claims which crowd wants to see fact-checked. Figure 2a shows the distribution of source types. We observe that *Alternative Media* and *UGC Platform* are the major categories of sources the crowd wants to fact-check. These two source categories together cover 53% of all the claims. The most common UGC platform is *imgur.com* [3]. We also study what entities the crowd wants to fact-check. Most (45%) of the posted claims are about *Person* (e.g., Barack Obama, Donald Trump) and *Policy/Issue* (e.g., Medicare, Tax rate).

3.2 Role of Crowd in Checking Claims

Fact-checking is a complex task which consists of multiple subprocesses. It's interesting to see how the crowd self-organize itself to accomplish this task by contributing to the subprocesses. We analyzed the comments and replies to study the different types of actions performed by the crowd. Table 3 shows the distribution of actions in comments. Note that a comment can be long and may contain multiple actions. In such cases, we labeled it with multiple actions. That is why the percentages do not sum up to hundred. About (70%) of the comments are used to provide argument. The crowd also identifies if a posted claim is verifiable or not and seek for information if further clarification is required. Some users also post irrelevant (ads, jokes) comments. We observe that contacting source of a claim, a primary task in professional fact-checking, is not practiced by the crowd. Only one comment was found where the user mentioned contacting the source.

Action	# Comments	Percentage
Provide Argument	1942	69.56
Post Irrelevant	402	14.40
Seek Clarification	355	12.71
Assign Flair	141	5.05
Check Verifiability	48	1.72
Contact Claim Source	1	0.04

Table 3: Distribution of actions performed by the crowd

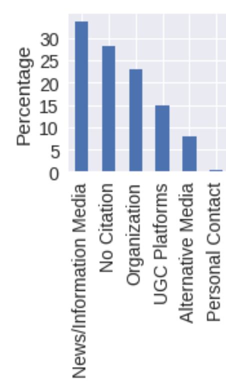

Figure 3: Source category distribution

Domain	# Citations
wikipedia.org	298
politifact.com	150
reddit.com	111
washingtonpost.com	91
nytimes.com	78
youtube.com	68
whitehouse.gov	51
factcheck.org	40
huffingtonpost.com	34
imgur.com	32
forbes.com	29
cnn.com	27

Table 4: Top-12 cited domains

We further studied the nature of the provided arguments. There are 1,080 arguments which contain factual evidence. 701 of these contain additional information leading to a fact-supported inference, either supporting the claim or refuting it. We observe that some commenters give judgment or opinion without factual justification. There are 860 such cases. We also analyzed the source of the factual evidence provided by the crowd. Figure 3 shows distributions of the categories. The top category is *News/Information Media*. There are 305 cases (28%) where the commenter presented factual evidence but didn't provide citation. Table 4 shows the top-12 cited domains. In many cases, the commenters cite evidence from popular fact-checking platforms such as *politifact.com* [4] and *factcheck.org* [5]. Table 5 shows the average number of comments containing factual evidence per post for each flair category. It is observed that, in general, *Unverifiable*, *Please Verify* flair categories receive less factual evidence and more opinionated comments compared to the conclusive (e.g., *Confirmed*, *Mostly True*) categories.

We measured the number of days required for the crowd to fact-check a claim. The *Reddit* API doesn't provide the exact flair assignment timestamp. A simple heuristic gives us an approximation of the flair assignment timestamp. We observe that before assigning a flair to a post, the moderators publish a comment mentioning and justifying the flair to be assigned. We use the timestamp of the latest comment made by a moderator containing the assigned

[3] http://imgur.com/

[4] http://www.politifact.com/

[5] http://www.factcheck.org/

Flair	Avg. # Factual Evidence	Avg. # Opinion/Judgment
Mostly True	3.22	1.83
Half True	3.17	1.67
Confirmed	3.10	1.56
False	2.89	2.59
Mostly False	2.75	1.86
Partisan Bias	2.73	1.90
Please Verify	2.59	2.40
Unverifiable	2.38	3.33
None	1.74	3.10

Table 5: Average number of factual evidence and opinionated judgments per flair

flair's words in the text as an approximation of the flair assignment timestamp. Subtracting the post creation timestamp from this gives us an approximation of the required fact-checking time. The mean, median and standard deviation of the required fact-checking time is 7, 1, and 20 days, respectively.

3.3 Role of Moderators

A qualitative analysis [1] of the moderator posts revealed several roles that moderators play to keep discussions on topic and fact-based. At the submission level, a moderator's roles can be compared to those of a gatekeeper who defines rules of a group and decides which posts qualify for fact-checking. At the comment level, moderators acted as "seminar leaders" in which they took part in verification, analysis, and evaluation of evidence posted by users.

The most prominent role of moderators at the submission level, defined by the frequency of posts with similar themes, emerged as reminding users of the rules and regulations, and issuing warnings. For instance, one post starts as: "I've seen posts in this thread increasingly devolving into political bickering, off-topic arguing, and..". The post ends with several warnings and a possible consequence of violation of rules: "...Get your act together, keep things on-topic, and keep things fact-based. Otherwise, your posts or comments may be removed". The second most prominent role of moderator is to encourage users by thanking them for participation and showcasing impact of their works. For instance, part of a post reads: "so many of you make this a fantastic little sub with a lot to offer. Please keep it up." Another post adds, "...Presidential debate drew over 16,000 page views here and more than 4,330 unique viewers." The third prominent role is to announce new events and topics to be covered. Other roles include seeking suggestions on various topics (e.g., how to cover a live debate). Some of the moderator posts were short while others provided more details. The number of words in a post ranges between one and 673.

The moderators posted 381 (13.44%) out of 2, 835 comments. In addition to assigning flairs based on available evidence, moderators took on several roles at the comment level. We find 165 instances where the moderators themselves presented source with factual evidence. Such sources include links to books, columns, datasets, news articles, press releases, research papers, and transcripts. One noticeable role of moderators appeared to be changing flair as more evidence is presented. For instance, one moderator had labeled a fact as mostly true: "Marking this as Mostly True. Yes, 4.5M jobs were created, since the lowest point of the recession. a net gain of 300K since the start of Obama's administration." As more evidence came in, another moderator changed the flair to half true: "Marking this one as 'Half True' based on everyone's research here. The 4.5 million jobs created seems factually verified. Whether it was better than the Bush recovery from the first recession seems unclear. But Cutter seems to have suggested she was in error on the Reagan recovery.". Another role was to ask for clarification about vague posts. We find 69 such occurrences.

3.4 Role of Computation

After observing the way *politicalfactchecking* operates, we identified several places where computation techniques can work with the crowd hand in hand and produce a faster, scalable and sustainable fact-checking model.

Argument Classification: One critical role of the moderators is to go through the arguments provided by the crowd and produce a conclusive decision based on the evidence. We argue, machine learning techniques can facilitate this task by automatically separating factual evidence from opinionated or evidence lacking judgment. Even though the dataset at hand is small in size, nonetheless, we built a binary classifier using Gradient Boosting algorithm and trained it over the manually coded comments (coding details in section 2.2.2 and 3.2). Word tokens were used as features after removing the stop words. The model was evaluated using 5-fold cross-validation. Its average precision of identifying comments containing *factual evidence* and *evidence lacking judgment* is 85% and 82%, respectively (i.e., the model is accurate 85% of the times when it says that a comment contains factual evidence). In a crowdsourced environment, where the number of moderators is significantly smaller than the crowd size, such an automated argument classifier can greatly reduce the workload of the moderators by straining the evidence lacking judgments from large number of comments.

Stance Detection: Another important task in fact-checking is to juxtapose supporting and opposing evidence. The moderators scrutinize the comments and identify the supporting/opposing evidences. We argue, like the argument classification task, identifying stance of an evidence can also be automated using computation techniques; particularly, natural language processing, computational reasoning, and machine learning. Given a claim and an evidence, the goal should be automatically detecting the stance of the evidence towards the claim. Our preliminary investigation over the *reddit/politicalfactchecking* dataset suggests that stance detection is a harder problem than argument classification. [17, 19] have studied the stance detection problem where the target is an issue rather than a claim. Recently, *Fake News Challenge*[6] have released a large dataset of news headline-body pairs with manually labeled stance information. We plan to continue investigating the stance detection problem.

In addition to the above mentioned tasks, there are other rooms for automation as well. For instance, given a claim, identify existing fact-checks or related evidence from the web, predicting the veracity of a claim based on provided evidence, soliciting arguments from users based on expertise, and so on.

[6] http://www.fakenewschallenge.org/

4 RELATED WORK

Professionals and scholars described fact-checking as a complicated process consisting of various steps that can be grouped into three major categories–(i) selection of facts to check, (ii) collection of evidence, and (iii) decision [3, 5, 7]. Each of these steps comprises of multiple sub-steps. For instance, fact selection includes choosing claims "from countless public utterances", separating facts from opinion, separating newsworthy/check-worthy facts from all facts, and filtering verifiable facts [7]. In the literature on news media, fact-checking is narrowly defined and refers to as verification[5]. It is believed to be a job of trained professionals. Graves noted that professional fact-checkers practice a type of "intertextual and annotative journalism" that lies "within the framework of a larger political critique" [7]. Fact-checking sites run by professionals "operate very self-consciously as hybrids of old and new media practice, organizing veteran print and broadcast reporters around a modernizing genre meant to update political journalism" [7].

With the advent of digital technologies, research on fact-checking started to broaden its horizon. Mark Little [15] asked journalists to "get comfortable with risk, transparency and collaboration" and find "the wisdom in the crowd". Cohen et al. [2] envisioned a system called a *cloud for the crowd*, which combines computational resources as well as human expertise to support more efficient and effective investigative journalism. In a recent white paper [7], Mevan et al. presented the state of automated fact-checking and described how fact-checking can be scaled up dramatically using existing technologies. Hassan et al. suggested a *Holy Grail* towards computational fact-checking by depicting the roles of professional fact checkers and automation [8]. They also studied how automated claim detection techniques perform against professional fact-checkers [11]. A number of tools have recently been proposed to fully automate fact-checking using natural language processing, machine learning, knowledge graph query, and question-answering techniques [9, 10, 12, 20]. However, there is a lack of research on studying the active role of crowd in fact-checking. According to the best of our knowledge, this is the first work which systematically studies fact-checking in a crowdsourced environment.

5 CONCLUSION AND FUTURE WORK

Despite of many limitations of crowdsourced fact-checking on *reddit/politicalfactchecking*, this model outlived many fact-checking sites [8]. It uses less resources compared to most professional fact-checking sites and utilizes the wisdom of the crowd to check hundreds of facts. Yet, it received negligible attention from scholars and professionals. Our findings show strong potential of this crowdsourced fact-checking model. The results contradict traditional beliefs that crowd is unable to fact-check and they provide opinions, not facts [9]. We found that more than 50% of the user comments providing an argument in support or against a claim had contained factual evidence. This contradiction may result from the contribution of moderators who often intervene to keep discussion on topic and fact-based. This study has revealed that crowds do more than fact-checking. They, among other roles, help find facts, play a role in deciding which facts are worth checking as well as which facts are verifiable. What professional fact-checking sites can learn from this study is that a small number of trained fact-checkers can help check a large number of facts if they can properly lead crowds in the right direction. This study also adds to previous research on automated fact-checking tools which could strengthen the crowdsourced model and improved its efficiency in multitudes. In conclusion, a future fact-checking model must incorporate all three components–automated assisting tools, crowds and professionals–to survive and thrive in the current media ecosystem. Each of these components complements one another. In future, we plan to investigate the comment replies and understand how discussion forms in a crowdsourced fact-checking model. We also intend to build robust stance detection techniques to facilitate the fact-checking process.

[7] https://fullfact.org/blog/2016/aug/automated-factchecking/

[8] http://www.poynter.org/2016/why-do-fact-checking-sites-close-and-how-can-new-ones-avoid-that-fate/390651/

[9] https://www.journalism.co.uk/news/could-crowdsourcing-help-you-fact-check-your-data-/s2/a557045

REFERENCES

[1] David L. Altheide and Christopher J. Schneider. 2013. *Qualitative media analysis.* Sage.

[2] Sarah Cohen, Chengkai Li, Jun Yang, and Cong Yu. 2011. Computational Journalism: A Call to Arms to Database Researchers.. In *CIDR*, Vol. 2011. 148–151.

[3] David A. Craig and Mohammad Yousuf. 2013. Excellence in journalistic use of social media through the eyes of social media editors. (2013).

[4] Knut De Swert. 2012. Calculating inter-coder reliability in media content analysis using Krippendorff's alpha. *Center for Politics and Communication* (2012).

[5] Michael Dobbs. 2012. *The rise of political fact-checking.* New America Foundation.

[6] Benjamin Doerr, Mahmoud Fouz, and Tobias Friedrich. 2012. Why rumors spread so quickly in social networks. *Commun. ACM* (2012).

[7] Lucas Graves. 2013. *Deciding what's true: Fact-checking journalism and the new ecology of news.* Columbia University.

[8] Naeemul Hassan, Bill Adair, James T Hamilton, Chengkai Li, Mark Tremayne, Jun Yang, and Cong Yu. 2015. The quest to automate fact-checking. *Proceedings of the 2015 Computation+Journalism Symposium* (2015).

[9] Naeemul Hassan, Fatma Arslan, Chengkai Li, and Mark Tremayne. 2017. Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster. *23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (2017).

[10] Naeemul Hassan, Chengkai Li, and Mark Tremayne. 2015. Detecting check-worthy factual claims in presidential debates. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management.* ACM, 1835–1838.

[11] Naeemul Hassan, Mark Tremayne, Fatma Arslan, and Chengkai Li. 2015. Comparing automated factual claim detection against judgments of journalism organizations. *Proceedings of the 2016 Computation+Journalism Symposium* (2015).

[12] Naeemul Hassan, Gensheng Zhang, Fatma Arslan, Josue Caraballo, Damian Jimenez, Siddhant Gawsane, Shohedul Hasan, Minumol Joseph, Aaditya Kulkarni, Anil Kumar Nayak, and others. 2017. ClaimBuster: The First-ever End-to-end Fact-checking System. *Proceedings of the VLDB Endowment* 10, 7 (2017).

[13] S.I. Hayakawa and Alan R. Hayakwa. 1990. *Language in thought and action.* Houghton Mifflin Harcourt.

[14] Brant Houston. 2010. The future of investigative journalism. *Daedalus* (2010).

[15] Mark Little. 2012. Finding the wisdom in the crowd. *Nieman Reports* (2012).

[16] Wilson Lowrey and Peter J Gade. 2011. *Changing the news: The forces shaping journalism in uncertain times.* Routledge.

[17] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiao-Dan Zhu, and Colin Cherry. A Dataset for Detecting Stance in Tweets.

[18] Kimberly A. Neuendorf. 2002. *The content analysis guidebook.* Sage.

[19] Parinaz Sobhani. 2017. *Stance Detection and Analysis in Social Media.* Ph.D. Dissertation. Université d'Ottawa/University of Ottawa.

[20] Brett Walenz, Y Wu, S Song, Emre Sonmez, Eric Wu, Kevin Wu, Pankaj K Agarwal, Jun Yang, Naeemul Hassan, Afroza Sultana, and others. 2014. Finding, monitoring, and checking claims computationally based on structured data. In *Computation+Journalism Symposium.*

[21] Mohammad Yousuf. 2016. *Niches in the networked media ecosystem: Functionalities, gratifications, and a typology of online news and informational media.* Ph.D. dissertation. The University of Oklahoma.