언론정보학과 석사학위 논문

The Validity and Credibility of Crowdsourced Fact-checking

크라우드소싱을 이용한 팩트체킹의 유효성과 신뢰도

2017 년 2 월

서울대학교 대학원

언론정보학과

홍 나 희

Abstract

Demand of fact-checking has increased together with the amount of information. However, fact-checking is often neglected due to lack of time and labor source. Also, development of digital technology has contributed to the acceleration of information spread. As online information is overabundant, it is difficult for professional journalists to fact-check all information before publishing it. Many researchers are developing algorithms for fact-checking, but they are not practical yet. In order to assist those deficiencies of fact-checking algorithms, collective intelligence was considered as an alternative method. With the help from the public, journalists collect, classify, and analyze data, and also widen their perspectives.

This research aims to examine the validity of crowdsourced fact-checking and its credibility level compared to professional journalism fact-checking results. Also, the interface elements to enhance credibility of crowdsourced fact-checking results were observed. The results show that crowdsourced fact-checking process is promising except for ambiguous and partially true

claims. The Mechanical Turk workers provided their deliberate opinions and critical evidence. Their rationale implies the possibility of public discussion and their capability of narrowing down broad statements to verifiable sentences. The credibility level of traditional journalism fact-checking was generally higher than crowdsourced results but for certain categories crowdsourced condition had a higher credibility level due to the features of social media. Moreover, the reputation of the user was more influential to the credibility than social information disclosure level. The findings from current research implies the future design of fact-checking platform and how could the current fact-checking algorithms could benefit from the collective intelligence.

Key words: fact-checking, crowdsourcing, journalism, credibility, online information, digital media, collective intelligence

Student Number: 2015-20247

Table of Contents

Table Index

Figure Index

1. Introduction

The development of media technology has changed the process of producing and distributing information. Digital information can be copied and pasted easily with simple clicks which allowed the information to reach people around the globe in a short period of time. Information can empower readers with knowledge but excessive amount of information raises several issues. Quality and accuracy of information should be questioned consistently and people should precisely know what they want to find when searching for necessary knowledge in the overabundant information era (Nyhan & Reifler, 2014; Magdy & Wanas, 2010; Chen, Conroy, & Rubin, 2015).

Evaluating the importance and credibility of information is one of the journalists' important roles and they have investigated the veracity of information before publishing it on the newspaper or broadcasting it on television. Just 20 years ago, when newspapers and televisions were practically the only source that people could acquire information, 'gatekeeping' was a sort of privilege that journalists held. The introduction of online platforms in the digital

environment caused the spread of information to speed up and expert journalists cannot keep up with this speed when writing news articles. Thus, journalists are always under pressure to meet the deadline (Johnson & Kaye, 2010; Flanagin & Metzger, 2007) and their lack of time and labor resources causes gatekeeping to be performed carelessly (Pavlik, 2000; Backett & Mansell, 2008; Coddington, Molyneux, & Lawrence, 2014).

With digital technology, Me-media, blogs, and social network services were introduced and their usage spread across the world. These media platforms allowed individuals to produce and publish information, making it easier to communicate their opinion with others (Woodly, 2008; Backett & Mansell, 2008). Using these online media, people reproduce news articles published by traditional media, exchange opinions on these articles, and sometimes even write one themselves. These changes in media transformed the journalistic process since anyone can generate and deliver information. This diminished the power of journalism, which was once called the fourth estate of democracy. Compared with the past where journalism was solely responsible for agenda setting and gatekeeping, public can now carry out those roles

with the development of communication media and help from media technology.

The size of the Internet media and its share in the entire media industry has exceptionally increased but most people still learn news from traditional media such as televisions and newspapers (Horrigan, Garrett, & Resnick, 2004; Woodly, 2008). However, as mobile population is increasing, more information is expected to be acquired from the Internet and mobile media (Backett & Mansell, 2008; Horrigan et al., 2004). Due to these changes in the media environment and people's life style, the speed of news production and distribution process has been accelerated and fact-checking is often excluded from the journalistic practice. Consequently, unverified information can reach the public easily and online information has credibility issues despite its enormous amount of knowledge. In 2015, the fear of epidemic Ebola in West Africa was also raged in the United States when some cases were reported domestically. An online article that reported the Ebola virus is transmitted through air escalated the fear but this article was later revealed as false (Chen et al., 2015; Mikkelson, 2014). Because of these unreliable

online rumors, the responsibility of readers grew bigger. When reading an online information, it is important for readers to judge what is useful information and it is their responsibility to filter out inaccurate and low-quality information. As it is not possible to fact-check all of online information, it is easy to find false information (Chen et al., 2015). The development of digital technology and the Internet media demand users to fact-check information individually before accepting it.

'Fact-checking' is the process of checking the veracity of a piece of information, which means checking if the information is stating only truthful claims. This process is considered as one of the important roles in journalism because 'accuracy' is an important virtue in journalism profession (Hanitzsch, 2007). In the past, when most of information distribution was only responsible by the established media, fact-checking was a must procedure before publishing the information. However, it becomes challenging as the volume of information explodes. There are a few reasons that caused fact-checking difficult in digital information era. First, comprehensive research is mandatory in fact-checking and it requires a certain amount of time and

professional labor. Therefore, this process is often omitted in order to catch up the distribution speed of information. Also, the number of journalists checking the veracity of information is insufficient compared to the amount of online information. Likewise, in a fast-changing digital environment, scandalous and episodic news articles are preferred by both journalists and readers instead of analytic articles with profound interpretations (Woodly, 2008). Thus, fact-checking process is neglected in order to generate as many articles as possible with prompt and interesting topics. Lastly, online information covers topics of diverse field but journalists or experts who investigate on the truthfulness of the information cannot have the competency in all those fields (Dunwoody, 1982).

The emergence of fact-checking website is the evidence of increased fact-check demands. These web pages, including 'FactCheck', 'Politifact', 'Snopes.com'[1)], professionally fact-check online rumors and articles especially political ones. These sites are operated by media institutes and press such as The

1) The URLs of the web pages are (1) factcheck.org, (2) politifact.com and (3) snopes.com. Professional journalists of the website fact-check online information, rumors, and news articles.

Washington Post, Annenberg Public Policy Center of Pennsylvania, Tampa Bay Times, and Poynter Institutes. Many of the fact-checking sites usually check claims of politicians particularly during the election period. According to Poynter Institute, a journalism organization, 37 countries are planning 96 fact-checking initiatives as of 2016[2]. Moreover, a census from the Duke University Reporter's Lab found that 64 fact-checking sites are active as of January 2015 which is up from active sites in May 2014[3].

Recently, during the 45th presidential election of the United States of America, fact-checking was a major subject for many newspaper media and some newspapers introduced real-time fact-checking platforms during the presidential debates[4].

---

2) Alexios Mantzarlis. (Feb. 16, 2017) "There are 96 fact-checking projects in 37 countries, new census finds."
http://www.poynter.org/2016/there-are-96-fact-checking-projects-in-36-countries-new-census-finds/396256/.

3) Bill Adair & Ishan Thakore.(Jan. 19, 2015). Fact-Checking Census finds continued growth around the world
http://reporterslab.org/fact-checking-census-finds-growth-around-world/.

4) Major news media including The New York Times, The Washington Post, and NBC provided real-time fact-checking for the president. The examples can be found at:

- http://www.nytimes.com/2016/09/27/us/politics/fact-check-debate.html?_r=0

However, fact-checking is a difficult and laborious job and its demand is increasing with the amount of information online. Therefore, some researchers and corporations are developing fact-checking algorithms to automate the process in order to meet the demands of readers (Wu et al., 2014; Ciampaglia et al., 2015; Magdy & Wanas, 2010). These algorithms extract verifiable claims from natural language sentences and search for related information (Ennals, Trushkowsky, & Agosta, 2010; Conroy et al., 2015; Hassan, Li, & Tremayne, 2015). Other systems analyze networks to determine the veracity of information (Ciampaglia et al., 2015). Despite their effort to automate fact-checking process, these algorithms are far from being practical. The fact-checking algorithms can answer only simple questions using information from websites such as Wikipedia or Twitter. The currently applicable systems can verify claims of limited knowledge field and select check-worthy or disputed claims (Ennals et al., 2010; Hassan et al., 2015).

- https://www.washingtonpost.com/politics/2016/live-updates/general-election/real-time-fact-checking-and-analysis-of-the-first-presidential-debate/trump-opens-debate-with-inaccurate-statement-about-ford/?tid=sm_fb&utm_term=.28b0545cea16

Therefore, these systems seek help from people. During the 2016 presidential election of the United States, a lot of false information and fake news articles spread through social network services and other online communities. Google and Facebook are making effort to filter out these false news articles using artificial intelligence collaborating with professional workforce. However, professional and expert labor is still expensive and this process is a top-down, elite centered procedure. Thus, some organizations sought help from the public in addition to experts and algorithms. As experts can be found everywhere, a little effort from the public can make a big outcome, which is also known as 'collective intelligence.' The attempts of using collective intelligence to solve complex problems are now prevalent in online platforms such as Amazon's Mechanical Turk or StackOverflow[5]. When the crowd participate in the fact-checking process, even if it is not fully voluntarily done, they are carrying out their independent right on acquiring proper information to realize one of important features of democracy. Many people are actually fact-checking online information themselves on their personal blogs or social network

5) The URLs of these websites are (1) mturk.com, (2) stackoverflow.com.

services and online communities. However, those self fact-checked data are scattered all over online websites, blogs and social media and in order to make fact-checking with public knowledge possible, all these data should be gathered in one platform. Moreover, engaging many people in the fact-checking, the process can speed up and empower the public with the role of experts. It is also meaningful that it improves efficiency in cost and time for fact-checking.

The general fact-checking process includes extracting words from natural language sentences, classifying myriads of these words into factual claims, opinions, and beliefs (Hassan et al., 2015), and lastly finding implications from those results. This process needs certain amount of repetitive and tedious labor and this could be done with collective intelligence model or perhaps even work better than one journalist inspecting all information by oneself. For instance, The Guardian, with the help from the public, investigated 700,000 receipts from individual MP (Member of Parliament) published by The Commons which contained 5,500 PDF files of 646 members of parliament in order to analyze four-years' worth of expenses and claim (The Guardian, 2009).

The public classified suspicious receipts for journalists to investigate them further. This case is a successful collaboration model of the public and journalists which engaged readers in journalistic process and gained media attention (Anderson, 2009).

The digital media environment is suitable for collaborative work since it has no limits on time and space. In order to meet the increased demands of fact-checking, there was an attempt to use crowdsourcing in the process using online website. The result of this effort was encouraging but have no empirical evidence (Florin, 2010). Regardless of the accuracy, crowdsourced fact-checking results have credibility issues. It is important to know how people perceive the trustworthiness of the results. The credibility level of crowdsourced results are lower compared to the results of journalists' fact-checking because many people participate in the process and it decreases responsibility of an individual. Therefore, it is critical to understand the factors affecting the credibility of crowdsourced fact-checking.

Thus, this paper aims to identify the possibility of crowdsourced fact-checking and its validity. In addition, by

comparing the credibility level of crowdsourced fact-checking results and traditional journalism, the interface elements of online communities are explored to enhance credibility. By applying different weight to those elements, an online platform could be designed to improve the reliability of crowdsourced fact-checking results.

2. Related Work

2.1 Journalism in Digital Environment

Journalism is highly associated with media technology since it uses media to deliver information and its articles to the public. From Gutenberg's printing press to the newest digital technology, the development of communication technology has brought huge transformation in the media industry. Journalistic process has changed with technology advancements and new media journalism is emerging as an alternative to traditional journalism (Pavlik, 2000). Journalists follow specific procedures when writing articles which reminds the conveyor belt of a factory. Therefore, newspaper press was also called as 'News Factory' (Woodly, 2008; Bantz, McCorkle, & Baade, 1980). According to a survey in 1999 by Dan Middleberg and Steve Ross, 93% of journalists collect data from the Internet (Pavlik, 2000) and the way of presenting these data to readers also has changed from text-based articles to multimedia format (Woodly, 2008). Furthermore, as information distribution becomes faster, time for fact-checking is reducing (Pavlik, 2000). In order to upload their

news articles before deadline, journalists hastily finish writing articles (Bantz et al., 1980) and spend less time on data analysis and storytelling to concentrate on delivering only simple information (Johnson & Kaye, 2004).

The significance of new media journalism has increased as the credibility of traditional news press such as newspapers and television has decreased (Johnson & Kaye, 2004). Traditional journalism handles scandalous and episodic events that could absorb the public attention and treats news articles as profitable goods. In this custom, news articles are written in top-down procedure, thus biased towards the opinions of limited sources and elites (Backett & Mansell, 2008). With these rising doubts of traditional journalism practice, people started questioning the 'objectivity', one of major features of journalism (Cunningham, 2003; Woodly, 2008). The concept of objectivity in journalism was introduced in the 19th century America. This characteristic of journalism separates values from facts and it identifies American journalism from European journalism (Schudson, 2001). There are many hypotheses for the reasons that introduced objectivity in journalism and two of acknowledged reasons are development of

communication technology and business profit of the media companies. With the advancement of communication technology, the pressure for rapid and accurate report increased and in order to keep business neutrality to receive advertisements from as many corporations as possible (Schudson, 2001), the newspapers kept objectivity in the tone in their articles.

Objectivity is surely one of the important factors of journalism for balanced reporting. However, this objectivity encourages 'lazy reporting' of journalists instead of deep analysis and explanation (Cunningham, 2003). Journalists simply deliver facts and hide behind the objectivity to avoid responsibility of their own reporting. On the contrary of balanced reporting of traditional media, new media disagrees with traditional news reporting as it does not follow the editing procedure of conventional journalism. New media seek direct communication with the public without governance of editors, which contributed to increased preference of new media (Johnson & Kaye, 2004).

One of the important roles of journalists is to guide readers in the flood of information. In order to perform this role properly,

analyzing a specific phenomenon and applying it to an appropriate context to explain it to the readers are a must ability of a journalist (Hanitzsch, 2007; Beckett & Mansell, 2008; Cunningham, 2003). This is why it is important for journalists to develop expertise in order to find the right information in large data and make judgement through analysis. According to survey results of Pew Research Center in 1999, more than half of newspaper journalists answered that interpretation of information is the fundamental element of journalism (Cunningham, 2003). Another research results showed that clearly biased news articles aroused aversion from readers, but people preferred articles that agree or even disagree with their opinion to the articles that have no perspectives at all (Horrigan et al., 2004). This indicates the needs for in-depth reporting instead of snippets of news.

The changes of the media environment with digital media technology brought changes in tasks of journalists as well. The readers also had to change their way of reading information since the role of journalists in digital media era has diminished and anyone can easily produce and distribute information. The possibility of public agenda setting has risen with digital

technology and blogs and social network services inversely influence the traditional media (Beckett & Mansell, 2008; Woodly, 2008). Newspapers, the classical traditional media, also provide online platform together with offline paper news. As online traffic on these platforms are associated with profit, it is more critical to write news articles of topics that attract more people (Woodly, 2008).

Network journalism which uses new media technologies has been introduced and by using network journalism, public agenda setting and discussions on these agenda are possible without difficulty (Woodly, 2008; Kriplean et al., 2014; Beckett & Mansell, 2008). These transitions in media caused active communication between journalists and readers (Pavlik, 2000). Still, traditional media reaches out to larger audience but there are some study results that online discussions can change attitudes of political elites (Woodly, 2008; Beckett & Mansell, 2008). Since the public can participate in agenda setting as well as in the process of opinion formation for politicians and journalists, political elites are now more cautious about their statements and claims than before. More participation from the public makes journalism to shift

towards reader-centered through increased interaction in journalistic process (Beckett & Mansell, 2008).

These changes that were brought by technology advance expanded participation for those who are highly interested in politics. The users of new media usually have high level of political knowledge and they are likely to participate in politics and vote. These people also learn diverse information from traditional media as well (Hill & Hughes, 1999). The changes in the media environment did not dramatically change political behavior but altered interaction between political elites and their related people (Hill & Hughes, 1999; Nyhan & Reifler, 2015). According to Pew Research Center for the People & The Press, over 50% of voters acquired information about election from the Internet for the 2000 presidential election. Pew Internet & American Life Project, together with The University of Michigan School of Information, carried out research which concluded that the online media users are more likely to be exposed to political controversies and perspectives and they are more aware of the opposite opinions despite the common notion that the Internet users would only search for information that supports their stance

(Horrigan et al., 2004). These results confirm the roles of the Internet media as a discussion forum (Woodly, 2008; Beckett & Mansell, 2008; Kriplean et al., 2014). National Conference of Editorial Writes (1996) explained that Internet media provides readers with an opportunity to participate in discussions of public issues and encourages interaction between themselves.

Table 1. Changes in digital media and its influence in journalism

Subject	Features of Digital Media	Changes in Journalism
Producer (Journalists)	Amount of Information	Data collection process
	Multimedia	The way of presenting information
	Speed of Spreading Information	Increased reports on episodic events
		Delivering merely simple information and less analysis and interpretation on articles
Consumer (Readers)	Accessibility	Public agenda setting
	Interactivity	

In sum, fast-paced production and distribution of information are pushing journalists to write episodic news snippets. However, readers are demanding more analytic articles that could explain the backgrounds of those episodes. This is the reason why more readers are turning towards new media journalism. The

interactivity of digital media allows participation of readers in journalistic process. Especially in political journalism, people are establishing a new forum and actively search for necessary information.

2.2 Fact-Checking

2.2.1 Journalism and Fact-checking

'Accuracy' is one of the important features in journalism (Hanitzsch, 2007). Nyhan and Reifler (2015) explained that 'accuracy' in journalism does not mean the correctness of information but how well the journalist has delivered the words from the source. This definition of 'accuracy' is related to another feature of journalism which was mentioned above, 'objectivity.' Journalists make effort to avoid questioning truthfulness of controversial claims even if they are verifiable (Nyhan & Reifler, 2015). In order to maintain balanced and impartial reporting, journalists should embrace both sides of opinions and write articles with sentences such as "he said", "she said" (Dobbs, 2012; Cunningham, 2003; Amazeen, 2013). However, some people question the "fairness" of delivering information that includes

unverified claims or that does not reflect the reality properly. For instance, when 70% of letters from readers are opposed to Afghanistan war, searching for letters that agree with the war to deliver the opinion in the same ratio would not be called "fair" even if it is for balanced reporting (Cunningham, 2003). Furthermore, Michael Dobbs (2012) said that since journalists are not stenographers, they should be taking more responsibility than just transcribing words from politicians and celebrities. He asserted that journalists should not only deliver the right information but they should make effort to seek the truth behind it.

Fact-checking has a significant role in American politics. When Ronald Reagan was the president of the United States, many newspapers began fact-checking because he said lots of incorrect statements (Dobbs, 2012). Afterwards, fact-checking in journalism usually verifies the truthfulness of politicians' statements and claims (Graves & Glaisyer, 2012; Graves, Nyhan & Reifler, 2016). This is one of the reasons why fact-checking is active during election periods. Fact-checking became popular during the 2004 presidential election period and many organizations opened

fact-checking websites around that election period. These websites including 'FactCheck', 'Politifact' and 'The Fact Checker' were at their climax in the 2010 election campaign period (Spivak, 2011). These fact-checking websites follow different ways of verifying information just like journalists persist their own method of fact-checking. Thus, these fact-checking websites do not have a consistent way of investigating truthfulness of statements or claims. For example, 'Politifact' has a few steps for editorial process. If an article is written by a reporter, one editor primarily checks the appropriateness of the topic then 3 additional editors revise the article. On the other hand, 'FactCheck' has 6 editors and 90% of their articles are reviewed by at least 4 of them before being posted on their web page (Amazeen, 2013). After the concept of 'Big Data' has been introduced, this fact-checking process has been depreciated as analysis results with big data are perceived as absolute truth. Some people have raised doubt on empirical method for verifying the truth. Yet, many study results indicate that fact-checking and its process enable democratic discussions and promotes political credibility (Dobbs, 2012; Nyhan & Reifler, 2014).

The effect of fact-checking is usually assessed with interviews or anecdotes, still there have been attempts to quantify the fact-checking effect (Graves & Glaisyer, 2012; Florin, 2010). The measurements include the number of citations by other media press, how much the public has changed their attitude, how much the fact-checking results influenced the journalistic process and political conversations (Graves & Glaisyer, 2012; Amazeen, 2013). New American Foundation explored about general fact-checking outcomes in 2012, and journalists were most influenced by the fact-checking results (Amazeen, 2013). Despite the contradictory results of fact-checking effect, fact-checking generally affects the behavior of political elites such as politicians and journalists (Woodly, 2008; Beckett & Mansell, 2008; Hill & Hughes, 1999; Nyhan & Reifler, 2015). Politicians in particular are sensitive to their reputation and if they are aware of being fact-checked they are more likely to be cautious on their statements about unverified facts (Nyhan & Reifler, 2015). However, fact-checking cannot prevent politicians from asserting false claims and they sometimes do not withdraw their statements even if they were confirmed as false (Amazeen, 2013). The objective of

fact-checkers is not to change politicians' attitude or political practice but to deliver the right information to the public. Hence, fact-checking is still important even if politicians persist with their unverified statements (Nyhan & Reifler, 2015; Amazeen, 2013).

One of the important components of fact-checking sites is the scale that indicates the truthfulness of the statements. Currently, there are two major methods to illustrate the veracity of politicians' assertions. 'FactCheck', presents the fact-checked results in article format. Thus, it is difficult to grasp the result at a glance but it can present diverse perspectives in the article. On the other hand, 'PolitiFact' and 'The Fact Checker' of the Washington Post use their original scale to display the truthfulness visibly. PolitiFact created 'Truth-O-Meter' that evaluates statements in 6 levels: 'True', 'Mostly True', 'Half True', 'Mostly False', 'False', and 'Pants on Fire.' 'The Fact Checker' assesses the 'fact' with illustrations of Pinocchio. The number of Pinocchio represents the untruthfulness of the claim and 4 Pinocchios is the maximum. These scales allow readers to easily determine the reliability of the statements but opens a

room for controversies by simplifying the complex reality. These websites and newspapers are asserting that they are making effort to present the truth of statements without simplifying (Dobbs, 2012) but still discussions are needed since these scales can convey subjective point of view.

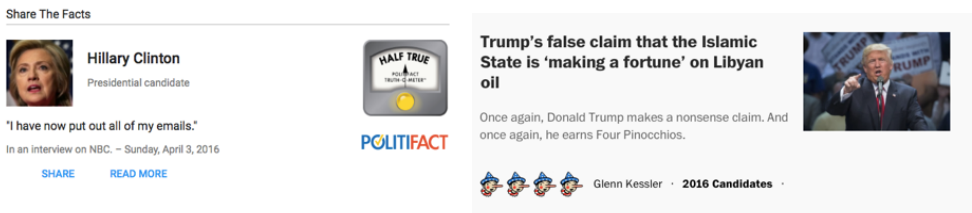

Figure 1. The scales used in fact-checking websites (Right) Politifact (Left) The Fact Checker

2.2.2 Automated Fact-checking Algorithms

Recently, IT corporations including IBM and Google are highly interested in fact-checking. They are developing automated fact-checking algorithms to prevent spreading of false information. 'ClaimBuster' is one of these algorithms that extract verifiable statements from speeches and judge their value of fact-checking. There are algorithms that collect information from Wikipedia to fact-check. These algorithms can answer simple questions such

as ‘Is Michelle Obama the first lady?’ Michigan University developed a system called ‘Rumor Lens’ which analyzes Twitter contents to figure out how fast a rumor spreads and how it is corrected. ‘Fact Minder’, a web extension, displays the background information of a personal figure that users are reading on the website. IBM introduced a beta version of ‘Watson Angels’ which analyzes 5.5 million news articles to check their facts[6]. Google’s latest feature highlights websites in Google News section that are suitable for fact-checking by adding fact-check tag for publishers. This tag will help readers to filter out fake news[7].

There are two major methods for fake news detecting: linguistic approach and network approach (Conroy, Rubin, & Chen, 2015). Analyzing linguistic features of information such as syntax,

6) Alan Greenblatt. (Apr. 4, 2016). “What does the future of automated fact-checking look like?”
http://www.poynter.org/2016/whats-does-the-future-of-automated-fact-checking-look-like/404937/
Alan Greenblatt. (Mar. 31, 2016).“Fact-checking 2.0: Teaching computers how to spot lies.”
https://www.poynter.org/2016/fact-checking-2-0-teaching-computers-how-to-spot-lies/404501/

7) Richard Gingras (Oct. 13, 2016). “Labeling fact-check articles in Google News.”
https://blog.google/topics/journalism-news/labeling-fact-check-articles-google-news/

semantic features, or rhetoric structure is linguistic approach. This includes using "bag of words" and discourse analysis (Ennals et al., 2010; Conroy et al., 2015; Hassan et al., 2015). On the other hand, network approach analyzes linked network or social networks to detect false information (Ciampaglia et al., 2015; Conroy et al., 2015). As both approaches are highly accurate only in limited domain, Conroy, Rubin, & Chen (2015) suggest a hybrid system. However, detecting verifiable claims is possible with current systems. Although the algorithms are not able to check veracity from the statements it is possible to highlight disputed claims using web extension and extract check-worthy claims from natural sentences from presidential debates (Ennals et al., 2010; Hassan et al., 2015).

The contribution of these automatic fact-checking algorithms is significant because fact-checking process is a highly laborious task. Many researchers and companies are working on algorithms in order to make automatic fact-checking possible (Wu et al., 2014; Ciampaglia et al., 2015; Magdy & Wanas, 2010). Since fake news are usually entertaining and attractive, they are more likely to spread rapidly through digital media without limitations on time

and space. These algorithms also help journalists and experts to save time for them to search for additional data. Above all, the ultimate goal of automatic fact-checking is making fact-checking possible for anyone before misinformation spreads.

However, these fact-checking algorithms are not practical yet. There have been trials to seek alternative method of fact-checking to save time and effort. Previously, news database was not accessible to the public but today, news articles and archives are easily searchable with the Internet. The accessibility of raw data and news archives brought public participation in fact-checking process (Dobbs, 2012). In fact, fact-checking websites and even traditional newspaper companies are seeking to incorporate public knowledge in their process and pull out readers' cooperation (Florin, 2010; Van der Haak, Parks, & Castells, 2012). One example of collaborated fact-checking process is 'The Times' where it selects the fact-check item with the participation from their readers.

2.2.3 Crowdsourced Fact-Checking

The word 'crowdsourcing' was first mentioned by Jeff Howe in

a magazine called 'Wired' in 2006[8]. The idea of this concept comes from 'outsourcing' and it uses cognitive ability from many anonymous people, in other words, collective intelligence to solve complex problems. Crowdsourcing is widely used in online communities by employing many people temporarily and rewarding them with their work. This allows complex work to be finished in a short period of time with a small amount money[9]. Similarly, 'social computing' and 'human computation' are used in solving problems through collective intelligence (Quinn & Bederson, 2011). Journalism also employs people's collaboration and effort for its works, especially in investigative journalism. As investigative journalism needs a large number of data collection and also thorough research on them, it is efficient to engage as

8) Jeff Howe. (Jun. 6, 2006). "The Rise of Crowdsourcing." https://www.wired.com/2006/06/crowds/.

9) One example is a paper published in 'Nature', "Space-time wiring specificity supports direction selectivity in the retina (2014)" by Kim, J.S. This paper used 'collective intelligence' in analyzing thousands of pictures of retina neurons. People participated in analyzing by playing a game called EyeWire. The 'citizen' scientists who participated in the game have their names as authors in the paper. (Cameron Scott. (May 14, 2014). "Crowd-Sourced Science Project Discovers How The Eye Perceives Motion." http://singularityhub.com/2014/05/14/crowd-sourced-science-project-solves-mysterious-function-of-vision.

many people as possible in the process. The receipt investigation of the Guardian that was mentioned in the Introduction section is a good example. Also, there was a case in Finland where short selling and internal trading of board members and executives in bank were investigated with the help from the crowd. The Finnish government disclosed the list of stock transactions and with that data, the public investigated the crimes of bank executives (Vehkoo, 2013).

Journalism process can be described in 4 steps: 1) collecting data, 2) analyzing and interpreting data, 3) storytelling, and 4) distributing the story (Beckett & Mansell, 2008). These works can be efficiently accomplished with the public's support by connecting the nodes of individuals. This collaboration of individuals is also known as 'network journalism' which changed linear journalistic process to network form (Beckett & Mansell, 2008; Van der Haak et al., 2012). News archiving and searching became simple in the digital information era and with this simple news searching task, people participate in fact-checking easily. Journalists and journalism media are making effort to find roles of individuals by using network and crowdsourcing journalism. For

investigative news reports, some journalists use social network services to collect data or encourage crowd to collect data themselves (Van der Haak et al., 2012; Vehkoo, 2013). Digital media enabled the participation of the public in journalistic process. The public now participate in journalism by asking questions and suggesting topics and also in technical support, editing, analyzing, and news storytelling process (Beckett & Mansell, 2008; Van der Haak et al., 2012).

News Trust, a nonprofit organization, launched a platform called 'Truthsquad' which checked the possibility of crowdsourced fact-checking. 'Truthsquad' added gamification factors to encourage people's participation but their participation and interest in this crowdsourced fact-checking platform did not last long because of motivational issues. However, with this platform, News Trust found that there are demands for crowdsourced fact-checking as well as fact-checking in general. The results of the pilot test conducted through this platform were also encouraging that crowdsourced fact-checking results generally matched with the results from the experts. However, identifying a reasonable reward other than educational benefit still remains as

an issue to be solved (Florin, 2010).

In order to encourage collaboration between the public and experts, Kriplean and his colleagues (2014) designed an interactive fact-checking framework. The purpose of designing this framework was to encourage public discussions and promote mutual relationship. In this research, a platform using this interactive framework was used to provide linkage between librarians and the general public. Librarians, as experts, give people advice on what information to look for and classify the fact-checking subjects into three groups: 'accurate', 'unverifiable', and 'questionable.' This collaborative platform helped promoting people's fact-checking attitude.

Marco Rubio, a politician in the United States of America, during a Republican debate for the presidential candidate said that his rival, Donald Trump would use illegal labor to build the wall along the U.S.-Mexican border that is meant to prevent illegal immigrants. Rubio said that people would easily find related information on Donald Trump and illegal workers in his company and told the public to 'Google it'[10]. Rubio's statement

demonstrates that anyone can perform fact-checking. The foundation for democratic citizen actions already exists and this event could be the implication of crowdsourced fact-checking. The importance of information quality increases with the amount of information together with demands for fact-checking. Collective intelligence could be one solution to fulfill the supply deficiency of fact-checking.

2.3 Credibility of Online Information

2.3.1 Media Credibility

Many studies on media credibility have examined the difference between traditional media and new media (Johnson & Kaye, 2010; Kiousis, 2001). Fogg & Tseng (1999) defined credibility as a 'perceived' and subjective concept which is determined by the receivers, not the information itself. They classified credibility into four categories: 'presumed', 'reputed', 'surface', and 'experienced.' Previous studies have defined credibility as a

10) Michael Lynch. (Mar 9, 2016). "Googling is Believing: Trumping the Informed Citizen."
http://opinionator.blogs.nytimes.com/2016/03/09/googling-is-believing-trumping-the-informed-citizen/?_3=0.

variable perceived by media users and credibility of information is not a characteristic of information that could be measured objectively (Sundar, 1998; Fogg et al., 2001; Freeman & Spyridakis, 2004). This means that there could be different credibility levels on the same piece of information. Information itself may not possess credibility, but many components of information can affect the credibility of information (Flanagin & Metzger, 2007).

The results of previous studies on media credibility have discrepancies as the approach method differs (Johnson et al., 2007; Kiousis, 2001). Regular surveys on media credibility use holistic approach while individual researchers usually use multi-dimensional approach. Furthermore, the researchers had difficulties in agreeing about key components of media credibility (Johnson & Kaye, 2004). Discussions on the credibility of online media have been continued since the introduction of the Internet. Information on the Internet is extensive and its topics are diverse, thus consensus on what aspect should be measured for its credibility is questionable. Moreover, media credibility is sometimes misunderstood as source or communicator reliability

and people are confused about credibility of media with credibility of television newscaster or newspaper company (Kiousis, 2001).

The advancement of media technology that brought changes in media platform did not diminish the impact of mass media during election period. Most people still read and hear about political information from traditional mass media rather than new media (Woodly, 2008; Johnson & Kaye, 2000). This is because voters who trust the established political system tend to believe traditional mass media. However, the credibility of existing news media is decreasing and influence of the Internet media has grown as people consider online media is independent from financial issues and objectivity arguments (Johnson & Kaye, 2004). According to research by Johnson and Kaye (2004), users with high internet usage consider online news more reliable than other users and other study results also indicate that usage of media including traditional and online media, leads to high credibility of online information (Johnson & Kaye, 2000; Flanagin & Metzger, 2000). A more recent study has identified the floor effect of the Internet as the number of users exploded and most people are now comfortable with online media usage (Flanagin &

Metzger, 2000).

According to the survey by Korea Press Foundation, the ratio of the Internet is increasing on 'The Most Reliable Media on Reporting on The Same Issue' (Chang, Ha, & Kim, 2014). This means that more people are depending on the Internet for acquiring public knowledge than before. According to the media credibility survey results in 1996, 2000, and 2004 on the presidential elections in the United States of America, the Internet received the highest credibility compared to the rest of traditional media in 2000, but has decreased in 2004. This is because people are able to recognize the fairness, accuracy, and believability of online information as they get familiar to this new media (Johnson & Kaye, 2004). Online information has its advantages on delivering news rapidly using network all over the world and this is useful when reporting disaster news in various parts of the world. Unlike traditional media, online media have higher interaction with the readers that allows 'self-purifying' effect (Chang et al., 2014). This means that misinformation can be corrected by readers with the interactivity feature of online media and the Internet news is considered as an alternative of

traditional media news which has lost trust from the public. Despite these positive aspects, the Internet media does not have systematic verification procedure such as gatekeeping thus vulnerable to deceptive information (Flanagin & Metzger, 2007).

2.3.2 Source Credibility

The absence of professional gatekeepers and vague boundary of information genre in online platforms bring down the credibility of online information (Flanagin & Metzger, 2007; Chen et al., 2015). Users have difficulty distinguishing a news article from a blog post. It is the user's responsibility to judge the credibility of the information. The contents from online platforms including Wikipedia and YouTube are created with the participation of many anonymous users. This makes it hard to infer the reliability of information just with the media credibility. Therefore, users make use of alternative information to figure out the credibility of online information, such as opinions from other users, social information and reputation of the source (Giudice, 2010). This background information and experience becomes the base for rating quality of information.

In accordance with Asch's conformity experiment, if people do not have background knowledge or previous experience, then they tend to follow other people's opinion (Petty & Brinol, 2010; Kim & Srivastava, 2007). In Asch's experiment, people changed their answers according to other people's answers even if they knew that answers were wrong (Asch, 1951). The results of this experiment are in line with the study results that concluded by deriving social agreement, it is possible to reduce the uncertainty of message and increases its reliability. The research by Sundar and Nass (2001) showed that compared to the news articles that were perceived to be selected by editors and computer, the articles that were perceived to be selected by other users were considered more representative and of a higher quality. In another experiment on website credibility, the size of users who wrote feedback did not have impact on the credibility but the types of feedback did (Giuidice, 2010). The credibility study on online shopping and its review system had the same results on the relationship between the reviews and the credibility of the shopping website (Flanagin et al., 2011). Regardless of truthfulness of information, the act of sharing the information can

generate social consensus (Fogg & Tseng, 1999; Wang et al., 2008). However, users evaluate the credibility of the information by using complex calculation considering involvement of other users, the rating of the users themselves, and also the feedback style, such as star rating, text, or survey (Giudice, 2010; Flanagin & Metzger, 2007).

The results of collective intelligence often have a low credibility level because of its origin of information is unclear. Online information users usually evaluate the reliability level of the content by combining the expertise and objectivity of the source but since online media do not provide enough evidence on these features, online information is often considered low in quality and credibility level. The source credibility in collective intelligence works becomes significant because perceived quality and usefulness of the outcome depends on source credibility. Hence, it is important to recognize the source identity and by identifying the social information of the source (Donath 1999; McKenna & Bargh, 1999; Ma & Agarwal, 2007), people feel it is easier to form social relationship with other users (Ren, Kraut, & Kiesler, 2007). In online communities, people tend to organize

groups with other people by sharing similar preferences and interest (McKenna & Bargh, 1999) and the more they have common concerns, the further their relationship develops (Jensen, Davis, & Farnham, 2002; Wiesenfeld, Raghuram, & Garud, 1999; Ren et al., 2007). In order to actively participate in online groups, users need to disclose their social identity. Social information of other users helps people to establish the first impression of other users by using cognitive shortcut (Forman, Ghose, & Wiesenfeld, 2008; Sussman & Siegal, 2003). Using social information, users can easily infer about other users and focus on information itself rather than its source. The relationship between social information disclosure and usefulness of information was often examined in the studies on online shopping. The usefulness of product reviews changed with the writer's behavior and social information disclosure level. People usually rated higher on negative reviews than positive reviews. However, when the reviewer's identity was provided, both types of reviews were rated equally high (Kusumasondjaja, Shanka, & Marchegiani, 2012). Social information such as name and pictures can lead to positive evaluation on the reviews (Fogg et al., 2001) and the location of the writer led to

the actual purchase to bring higher sales rate (Forman et al., 2008).

The reputation of the users also influences the information they provide. The reputation is made through the user's past behavior and this affects the perceived quality of his/her work (Dellarocas, 2001). Reputation is formed with multi-dimensional factors (Cho, Kwon, & Park, 2009) and Chen and his colleagues (2007) used social network analysis method to calculate reputation by taking social relationship into consideration. In order to figure out a user's reputation, the feedback history written by the user and other user's feedback on the user, the number of positive feedback, and the user's experience with the platform should all be taken into account (Chong & Abawajy, 2007; Wu, Li, & Kuo, 2011). A previous study also used expertise, credibility, and similarity with other users to calculate a user's reputation (Cho et al., 2009).

One of the main reasons that degrade the online information credibility is the low reliability of the source. The source identity becomes more ambiguous compared to other information since

some online information is written by many people together. The reviews on online shopping are directly associated with sales, therefore many researchers tried to identify factors that create user's credibility to pull up the credibility of the reviews (Fogg et al., 2001; Forman et al., 2008; Kusmasondjaja et al., 2012; Dellarocas, 2001). The credibility of the user or the source determines the perceived quality of information and the user credibility is formed with the disclosure level of social information and the user's past behavior.

Digital environment is suitable for crowdsourced fact-checking since it has no limit for time and space, thus a large number of people can participate in the process. Therefore, if an online platform is to designed for crowdsourced fact-checking, the credibility level of the results is critical. However, these anonymous users do not have ground for their expertise since their social information is limited. Thus, in order to increase the credibility level of collective intelligence works, using reputation and regulating social information disclosure level would be useful to design an evaluation system. This evaluation system can be used to enhance source credibility of online crowdsourced

fact-checking platform by adjusting each component's contribution level.

3. Research Methods

3.1 Validity of Crowdsourced Fact-checking

Advancements of digital technology have changed journalistic process and the importance of fact-checking has risen with the increased amount of information. With the absence of professional gatekeepers, it has become the user's responsibility to evaluate the quality and truthfulness of online information. The acceleration of fact-checking process is required as the distribution of information speeds up, but fact-checking experts do not have enough time to look through all information that are shared on the Internet. Many researchers are developing algorithms to help this job but they lack practicality. Network journalism has become an alternative way of cutting down cost and time for fact-checking as number of people can participate in the process and journalists can acquire data they have overlooked. This would even help journalists to have a wider perspective of the world. Unlike traditional journalism, digital technology has helped the public to find fact-checking topics by themselves, collect data and do research on their own. Considering the interactive features of

online media, crowdsourced fact-checking is not a new process.

However, the quality issue still remains. The quality of crowdsourced outcomes is not guaranteed and many previous studies on crowdsourcing investigated the methods to improve the quality (Quinn & Bederson, 2011; Hansen et al., 2013; Lease, 2011). Fact-checking with the public's participation can minimize the responsibility of each individual. Some people may intentionally mislead others with incorrect information and some may disclose their political position and criticize others with inflammatory words. Despite these concerns, people who participate in crowdsourced fact-checking are the ones with high political interest and knowledge since it is not a simple task. A pilot test by NewsTrust on 'Truthsquad' showed a promising result for crowdsourced fact-checking: the fact-checking results of the crowd generally matched the results of the professional journalists (Florin, 2010). Moreover, some participants provided links for critical evidence. Yet, no empirical data was given for this result so this paper aims to observe empirical data for crowdsourced fact-checking and its validity.

Research Question 1: Is crowdsourced fact-checking possible?

Amazon's Mechanical Turk was used to collect data in this study. Mechanical Turk is an online marketplace that provides a venue for requesters to describe their request and recruit workers. On the other hand, the workers upload their work on this platform and receive reward[11]. Requesters publish batch of HITs (Human Intelligence Task)[12] with short descriptions and the amount of reward and workers voluntarily participate in the work to collect reward. Mechanical Turk is used in various fields, such as photo/video processing, data cleaning/verification/processing, information collection, and even in artistic projects[13].

In order to find the validity of crowdsourced fact-checking,

---

11) Amazon Mechanical Turk FAQ.
https://www.mturk.com/mturk/help?helpPage=overview#what_is

12) A Human Intelligence Task, or HIT is a term used in Amazon's Mechanical turk that represents a single, self-contained task and a question that needs an answer. A Worker of Mechanical Turk can work on a HIT to collect a reward.
https://www.mturk.com/mturk/help?helpPage=overview#what_is_hit

13) Wikipedia. Amazon Mechanical Turk.(Nov. 5, 2016)
https://en.wikipedia.org/wiki/Amazon_Mechanical_Turk#cite_note-24
AaronKoblin &Takashi Kawashima. Ten Thousand Cents(2008).
http://www.tenthousandcents.com/top.html

10 sentences were chosen for verification. The sentences cover various topics including politics, world, science and general news. Table 2 describes the sentences that were used for fact-checking. The participants earned $0.10 for completing the

Table 2. Sentences used for verification of crowdsourced fact-checking

Category	Sentences	Answer
World	A human trafficking survivor who escaped from Japan to Canada, completed world's longest triathlon in 2014 in order to cope with trauma.	TRUE
Politics	Donald Trump called pregnant employees 'an inconvenience'.	TRUE
Technology	Twitter will increase its per-tweet character count from 140 to 10,000.	FALSE
General	Bananas will disappear in 5~10 years because of epidemic in the Philippines.	TRUE (Partial)
General	Smoking in a car with children will become illegal.	FALSE (Partial)
Technology	Google has admitted the self-driving car accident on 2016 March was on their fault. For 6 years, Google's self-driving car had total 17 accidents but never admitted their fault before.	TRUE
Politics	Vladimir Putin sent a message of congratulations to Donald Trump on his victory in the US presidential election. In addition to his congratulation message, he said that he is looking forward to resolving issues on international agenda including the nuclear issue of North Korea at the East Asian six-party talks next year.	FALSE (Partial)
Politics	Hillary Clinton became the first "first lady" to win an elected office after winning the U.S. Senate seat for New York in 2000.	TRUE
World	European Union flag will be losing a star after the Brexit vote.	FALSE
Science	Eating chocolate while studying helps the brain retain new information easily.	TRUE

HIT by answering True or False for the 10 sentences which were randomly presented. The participants had to choose between 'True' or 'False' to proceed to the next sentence and they had to finish responding the 10 sentences for their reward. Potential bonus rewards up to $0.50 were given for additional links or reason for their answer.

3.2 Perceived Credibility of Crowdsourced Fact-checking

The validity of the crowdsourced fact-checking does not lead to the usefulness of the result. Public should embrace this fact-checking result in order for it to be practical and useful. Therefore, how much the public is willing to accept the result is crucial in crowdsourced fact-checking and the usefulness of information depends on the credibility level of the information. As previous studies have determined, credibility is not one of the properties of information but a 'perceived' concept by the readers. With the second research question, this paper addresses the comparison between the credibility of fact-checking results of the professional journalists and the collective intelligence. In order

to observe a user's perception on news articles, Sundar and Nass (2001) gave participants identical news articles to read but divided participants into different conditions. In their study, the participants in different conditions believed that their articles were recommended by different agents: a system, editors, or other users. Then the users evaluated credibility, quality, and representativeness of the articles. This crowdsourced fact-checking study also provides identical stories to experiment participants but in different format for them to assess the difference in perceived credibility of fact-checking results by different agents.

Research Question 2: What is the credibility level of crowdsourced fact-checking result compared to traditional journalism fact-checking result?

A 2 x 3 x 4 experiment was designed to investigate the credibility difference of traditional journalism and crowdsourced fact-checking (See Table 3). Mock up pages of 'The New York Times' and 'Facebook' were made for each condition of the experiment: professional journalist and crowdsourced

fact-checking results. For traditional media styled fact-checking results, mock up page of 'The New York Times' was used in the experiment whereas for social media styled fact-checking results, mock up page of 'Facebook' was used.

Three categories of information were selected for credibility evaluation: 1) politics, 2) non-political hard news (economics and world), and 3) non-political soft news (life style). Two stories were chosen for each category for true and false condition. Moreover, each story was written differently for two conditions of verification results. In other words, 4 stories were generated for each of three categories, thus 12 stories were created in total. All stories were written in the form of news article and social media posts.

Table 3. Conditions used to measure perceived credibility of fact-checking results

	The New York Times				Facebook (Crowdsourced)			
Actual Fact-check Result	TRUE		FALSE		TRUE		FALSE	
Presented Fact-check Result	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
Name of Condition	True Positive	False Positive	False Negative	True Negative	True Positive	False Positive	False Negative	True Negative

Previous studies show that the credibility level of the Internet media is increasing as people consider it as new and alternative news source to the traditional news media that has lost the trust from the public. Still, most people learn news from established news media, and the uncertainty of information genre and source of online media decreases the credibility of their content (Johnson & Kaye, 2004). A hypothesis comparing the credibility level between the professional journalists and crowdsourced results can be developed based on this knowledge.

H 2-1. The general credibility level of traditional media styled fact-checking results will be higher than social media styled fact-checking results.

However, there are some news categories that are often exposed in social media. For these subjects, crowdsourced fact-checking results would receive higher credibility level than professional fact-checking results. Many news articles links that deliver technological or general life information are shared through 'Facebook' and other social network services. When evaluating web-based information, types of information are

important factors to be considered (Metzger et al., 2003). Moreover, the more familiar the readers are with the online information and the web site, the credibility increases (Lowry, Roberts & Higbee, 2007). Also, high familiarity with a certain situation leads to high efficacy expectations and high accuracy in credibility judgement (Reinhard, Scharmach & Sparer, 2012). Thus, soft news articles which are generally read through social network services will receive higher credibility level in 'Facebook' condition compared to 'The New York Times' condition.

H 2-2. The credibility level of social media styled fact-checking results will be higher than traditional media style for soft news articles.

On the other hand, traditional news articles such as political and hard news are often read in traditional news media. Therefore, the fact-checking results of professional journalists would receive higher credibility level compared to crowdsourced fact-checking results for hard news and political articles.

H 2-3. The credibility level of traditional media styled fact-checking results will be higher than social media style

for hard news and political articles.

As mentioned above, new media often disagrees with traditional news reporting and usually denies the original reporting which engaged more people to read its news articles (Johnson & Kaye, 2004). Therefore, news articles read through social network services and crowdsourced platforms, are mostly the articles that opposes the earlier news articles from traditional media. Thus, False-Positive articles which incorrectly rejects certain statements would receive higher credibility for crowdsourced condition than professional journalists condition.

H 2-4. The credibility level of social media styled fact-checking results will be higher than traditional media style for False-Positive condition.

Likewise, traditional media news delivers the statements or events straightforwardly in its articles. Therefore, the articles that tells the truth in its article and does not disagrees with anything would receive higher credibility level for 'The New York Times' condition compared to the 'Facebook' condition.

H 2-5. The credibility level of traditional media styled fact-checking results will be higher than social media style for True-Positive conditions.

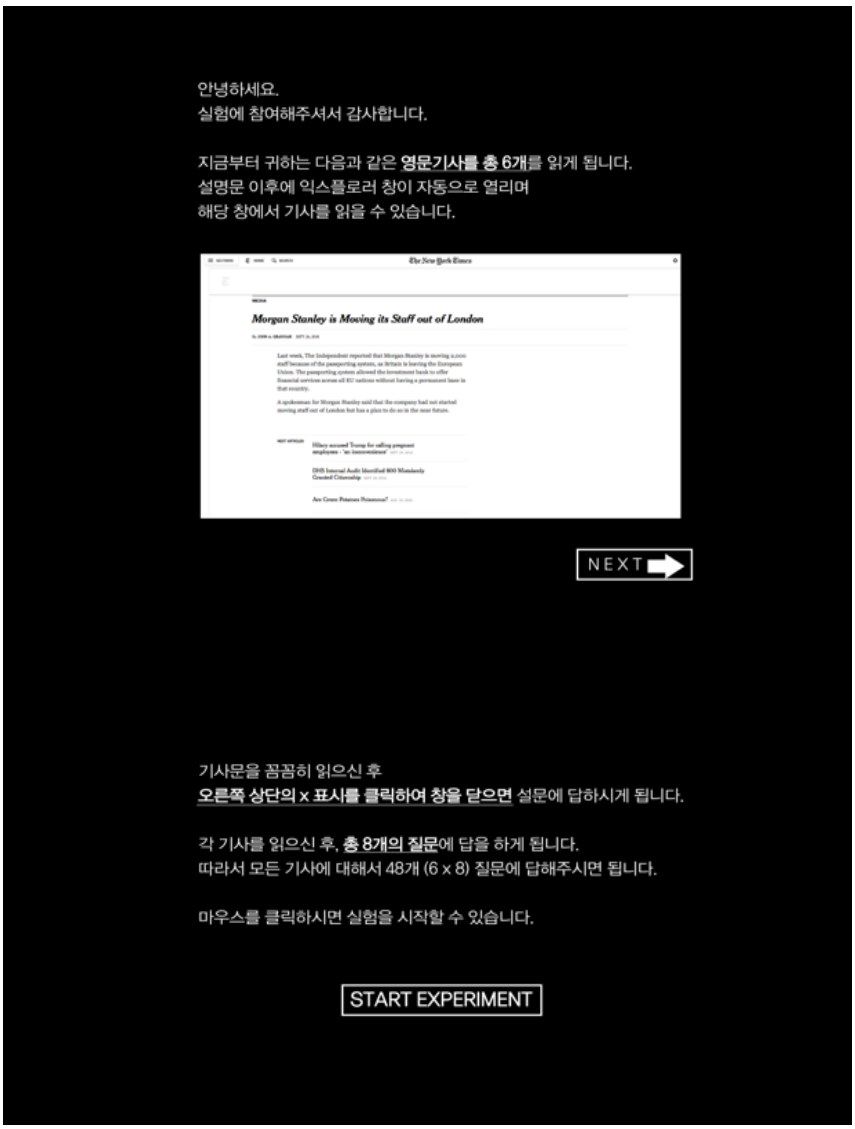

Figure 2 Instruction pages used for 'The New York Times' condition

Table 4. Topics used for each condition

	True -Negative	False -Negative	False -Positive	True -Positive
Political	WikiLeak's DNC Email Calling Voters "White Trash" Found False	WikiLeak's DNC Email Calling Voters "White Trash"	Hillary Clinton accused Trump for calling pregnant employees - 'an inconvenience' which Trump never did	Hillary Clinton accused Trump for calling pregnant employees - 'an inconvenience'
Non -political (Soft)	Are Green Potatoes Poisonous? - It is not	Are Green Potatoes Poisonous? - Yes, it is	Google's Self-Driving Car Crashed and It is not Google's fault	Google's Self-Driving Car Crashed and it is Google's fault
Non -political (Hard)	Morgan Stanley is Not Moving its Staff out of London	Morgan Stanley is Moving its Staff out of London	DHS Internal Audit Report Revealed as a Fake One	DHS Internal Audit Identified 800 Mistakenly Granted Citizen ship

In order to test the hypothesis, a lab study was conducted. Participants were recruited from a website and social network site (N = 51, 31 males and 20 females). They were randomly divided into two conditions ('The New York Times' and

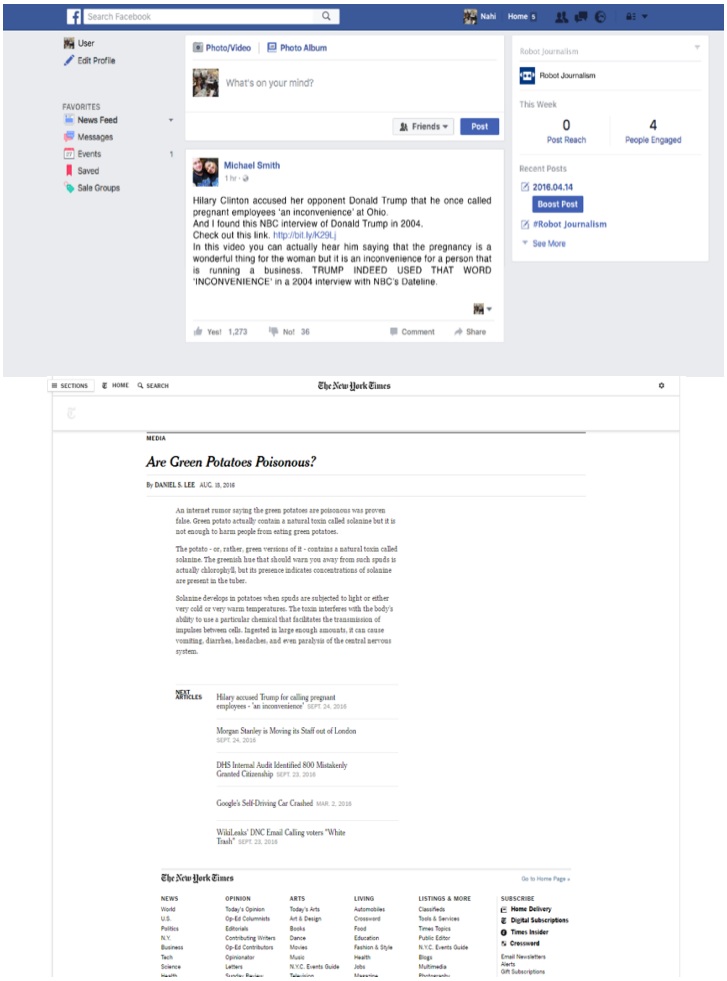

Figure 3 Examples of mock up sites used in Experiment 2 for
(1) 'Facebook' condition (top)
(2) 'The New York Times' condition (bottom)

'Facebook') and were assigned to one of the two sets which included 6 stories each. The topics used in the articles are listed in Table 4. The participants received 10,000 KRW for reward and the experiment took approximately 15 to 25 minutes including the time for instruction.

To address the credibility level of each article, media credibility measures from the previous studies were selected. Sundar and Nass (2001) measured 6 adjectives of credibility variable using Likert type scale. They used 'accurate', 'believable', 'biased', 'fair', 'objective', and 'sensationalistic' to measure credibility. Clerwall (2014) divided perceived quality in credibility and readability and credibility is composed of 'informative', 'trustworthy', 'objective', and 'descriptive.' Kaa and Krahmer (2014) defined perceived credibility as trustworthiness and journalistic expertise and trustworthiness consists of 4 components which are 'reliability', 'honesty', 'accuracy', and 'fact-based.' Based on previous studies, Graefe (2016) used 5-point Likert scale to measure credibility of news articles, and he used 4 adjectives: 'accurate', 'trustworthy', 'fair', and 'reliable.' Other studies considered source, message, and media dimensions

to measure credibility of news, using adjectives such as ‘believable’, ‘trustworthy’, ‘comprehensiveness’, ‘relevance’, ‘unbiased’, ‘accuracy’, and ‘completeness’ (Chung, Nam, & Stefanone, 2012; Flanagin & Metzger, 2000). In this study, 8 adjectives were selected to measure credibility of fact-checking results using 7-point Likert scale. The journalistic expertise aspect and other source related adjectives are excluded because the professionality of the source is not considered in this study. This study focuses on how participants perceive the credibility of

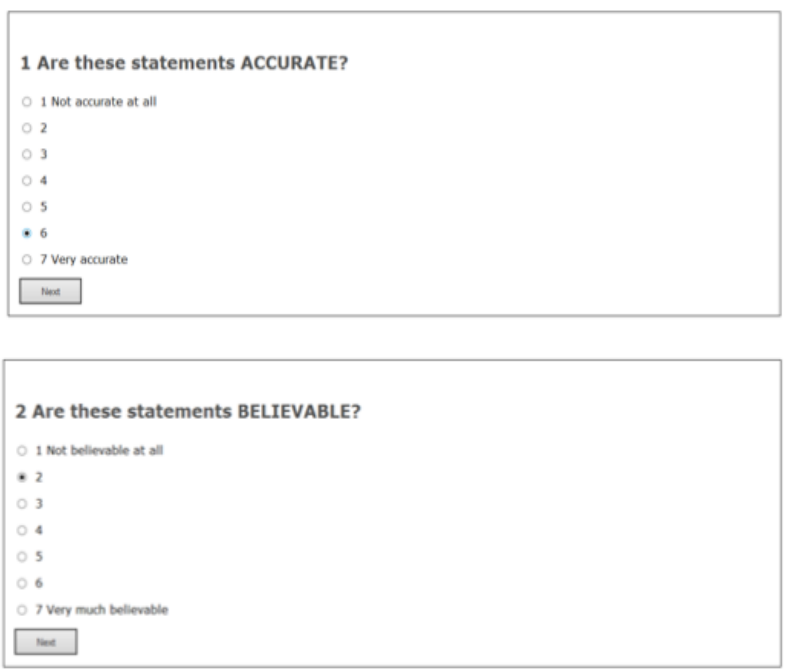

Figure 4 Examples of question forms used in credibility survey

content that were evaluated or written by people who do not have expertise compared to professionals. The adjectives used in this study are 1) accurate, 2) believable, 3) biased, 4) reliable, 5) written with completeness, 6) objective, 7) fact-based, and 8) trustworthy.

3.3 Designing Online Platform Elements

There are difficulties of traditional fact-checking method, such as lack of labor source and time when there is enormous amount of information. Automated fact-checking algorithms are being developed, however, as fact-checking involves natural languages, they are not yet applicable in various fields. Fact-checking of online information cannot meet the demand of fact-checking that has increased dramatically. Therefore, many people are checking the veracity of online information by themselves. To encourage the collaboration of these pople and make use of these fact-checked results of each individual, an Internet forum is suitable platform. Therefore, designing an Internet platform for crowdsourced fact-checking that is asseible to many people is an important procedure.

The usefulness and credibility of information depends on the source credibility. When many people participate in fact-checking process, their source credibility influences the perceived credibility of crowdsourced fact-checking results. Thus, in order to maintain a high credibility level of crowdsourced work in online forum, management of source credibility is necessary. If an online platform for crowdsourced fact-checking is to be developed, the interface elements and user information exposure level should be carefully designed to enhance the source credibility which leads to information usefulness. Many people participate in crowdsourced work, but as credibility level are different for each individual, their contribution level should be different on the crowdsourced output. In other words, when an individual with a high credibility level asserts that a certain claim is 'true' while another individual with a low credibility level insists that it is 'false', then it is likely that this claim is a 'fact.'

Wikipedia is a good example of results using collective intelligence. Wikipedia assigns different editing authority depending on the participation and social information disclosure level of the users. For instance, an unregistered user needs to input

CAPTCHA (Completely Automated Public Turing Test to tell Computers and Humans Apart) and he/she cannot upload multimedia file. On the other hand, a user who participated in editing for more than 4 days and 10 times, can edit semi-protected data without inserting CAPTCHA[14]. Wikipedia provides more editing authority to users according to their participation and social information disclosure level. In line with Wikipedia's regulations, several studies on online shopping malls have developed user reputation system that grants different weight on a user's component to distinguish reviews with high credibility (Dellarocas, 2001; Chong & Abawajy, 2007; Wu et al., 2011). Therefore, identifying the factors that constitute the credibility of users is important to determine the credibility of fact-checking results.

Research Question 3: What are the factors that affect the credibility of crowdsourced fact-checking?

In order to raise credibility of crowdsourced fact-checking

14) Wikipedia. Wikipedia: User access levels. (Dec. 4, 2016). https://en.wikipedia.org/wiki/Wikipedia:User_access_levels#Autoconfirmed_and_confirmed_users.

results, it is important to manage source credibility components and determine their influence on the results. By applying different weight on user components, this research aims to level up the credibility of the crowdsourced results. Social information disclosure level and reputation of users are two main categories that influence the credibility of the source and online information.

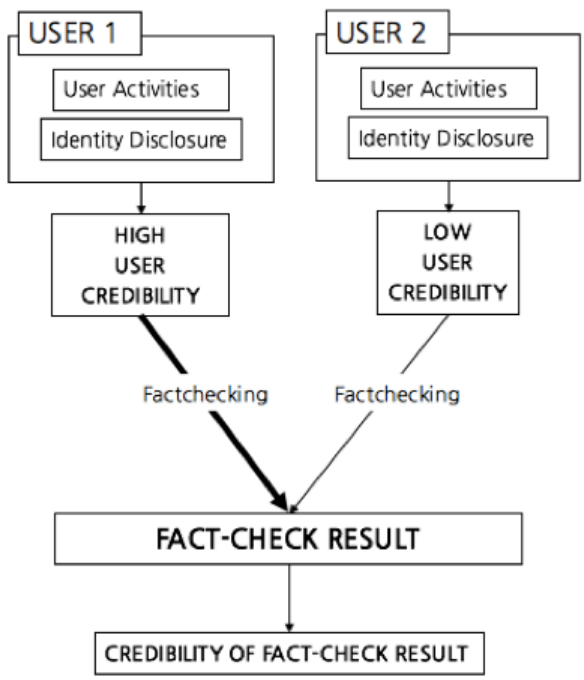

Figure 5 Different weight on fact-check results for user components

Based on the results of the previous studies, 6 variables that affect online information credibility were selected to be measured.

For social information disclosure category, name, picture, and location were chosen as variables and the user's participation level, consensus on the user's claim, and number of comments were included in reputation category. Table 5 displays the category and variables of online source credibility and also conditions for each variable.

Table 5. Variables that influence source credibility of online information

Category	Variables	Conditions
Social Information Disclosure	Name	Real name / Nickname
	Picture	Picture with face / No picture
	Location	Open / Close
Reputation	Participation Level	Low / High
	Consensus Rate	High 'Yes' Rate / High 'No' Rate
	Number of Comments	High / Low

Considering the variables mentioned above, total 64 scenarios are generated to observe credibility level of crowdsourced fact-check results. For each scenario, a mock up page was given to participants and they evaluated the credibility of the information. A 7-point Likert scale was used to evaluate the credibility of information and the same 8 adjectives that were

used in the Experiment 2 were measured. Again, Mechanical Turk was used to gather data. The participants received $0.10 for reading the mock up post and answering 8 questions. After the instruction and agreement page, users examined the online post, then answered 8 questions of credibility measures. When answering the survey question, users were allowed to read the post again if they desire.

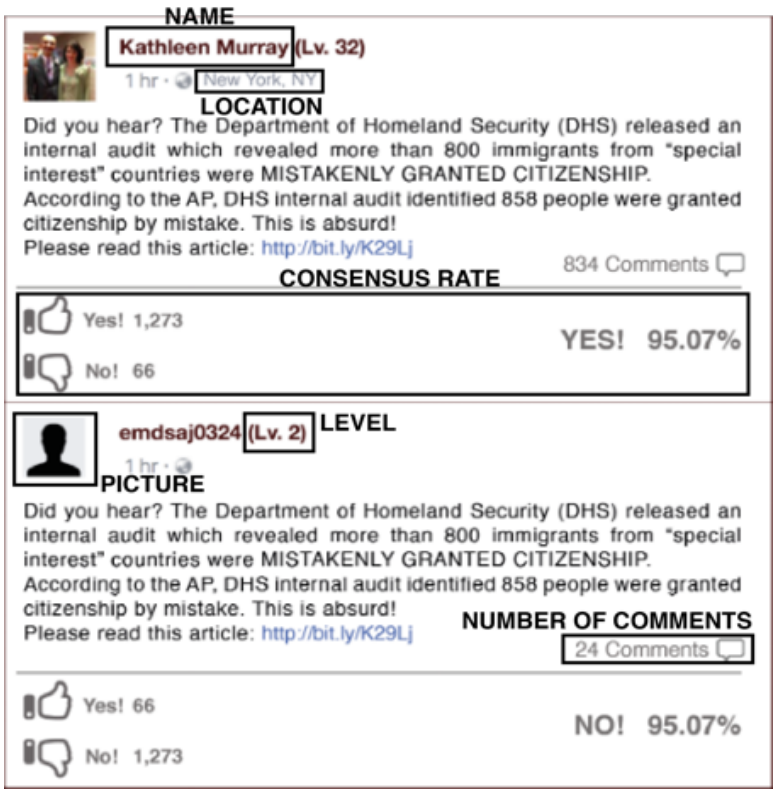

Figure 6 An online post with all positive variables (Top)
An online post with all negative variables (Bottom)

4. Results

4.1 Validity of Crowdsourced Fact-checking

Total 209 Mechanical Turk workers participated in the fact-checking HIT. The average time for the HIT was 23 minutes 4 seconds. Most of the workers inserted at least one additional piece of information or a reason for their answer except for 27 people which means that 182 people (87.08%) provided additional links or their thoughts on the statements. The average percentage of the correct answer rate is 69.20%. “Smoking in a car with children will become illegal” had the lowest correct answer rate with 43.48% while “Hillary Clinton became the first “first lady” to win an elected office after winning the U.S. Senate seat for New York in 2000” had the highest answer rate with 86.96%. The complete correct answer rate of 10 statements are listed in Table 6. ‘True’ statements had a little higher correct answer rate than ‘False’ statements. When partially true statements are separated, the correct answer rate for ‘True’ statements increases, while the correct answer rates for ‘False’ statements decreases.

The 3 statements that earned the lowest correct answer rates are partially true which means that the statements include some facts. Twitter has no plan to increase its per-tweet characters to 10,000 and the character limit remains 140, but the way Twitter counts its characters will change by excluding links and user names in the character count. Also, a pre-existing false news report of the Wall Street Journal[15] caused the low correct answer rate of this sentence. Many of the participants provided the link of this outdated article, which was later found false, as evidence for their answer. Likewise, smoking in a car with children is illegal in some countries such as the United Kingdom but no such plans in the United States yet. The participants were confused about the subject of the sentence and some users explicitly stated the countries that have plans to make it illegal and the links to the news articles. That Putin has congratulated Trump on his winning presidential election is true, but he never mentioned about North Korea and six-party talks. Moreover, since this event is very recent which has been only a month, people had not

15) Yoree Koh. (Jan. 5, 2016). "Twitter to Expand Tweet's 140-Character Limit to 10,000"
http://blogs.wsj.com/digits/2016/01/05/twitter-to-expand-tweets-140-character-limit-to-10000/

Table 6. Correct answer rate for each sentence

Category	Sentences	Answer	Correct Answer Rate
World	A human trafficking survivor who escaped from Japan to Canada, completed world's longest triathlon in 2014 in order to cope with trauma.	TRUE	84.78%
Politics	Donald Trump called pregnant employees 'an inconvenience'.	TRUE	86.96%
Technology	Twitter will increase its per-tweet character count from 140 to 10,000.	FALSE	44.93%
General	Bananas will disappear in 5~10 years because of epidemic in the Philippines.	TRUE (Partial)	76.09%
General	Smoking in a car with children will become illegal.	FALSE (Partial)	43.48%
Technology	Google has admitted the self-driving car accident on 2016 March was on their fault. For 6 years, Google's self-driving car had total 17 accidents but never admitted their fault before.	TRUE	61.59%
Politics	Vladimir Putin sent a message of congratulations to Donald Trump on his victory in the US presidential election. In addition to his congratulation message, he said that he is looking forward to resolving issues on international agenda including the nuclear issue of North Korea at the East Asian six-party talks next year.	FALSE (Partial)	47.10%
Politics	Hillary Clinton became the first "first lady" to win an elected office after winning the U.S. Senate seat for New York in 2000.	TRUE	86.96%
World	European Union flag will be losing a star after the Brexit vote.	FALSE	75.36%
Science	Eating chocolate while studying helps the brain retain new information easily.	TRUE	84.78%

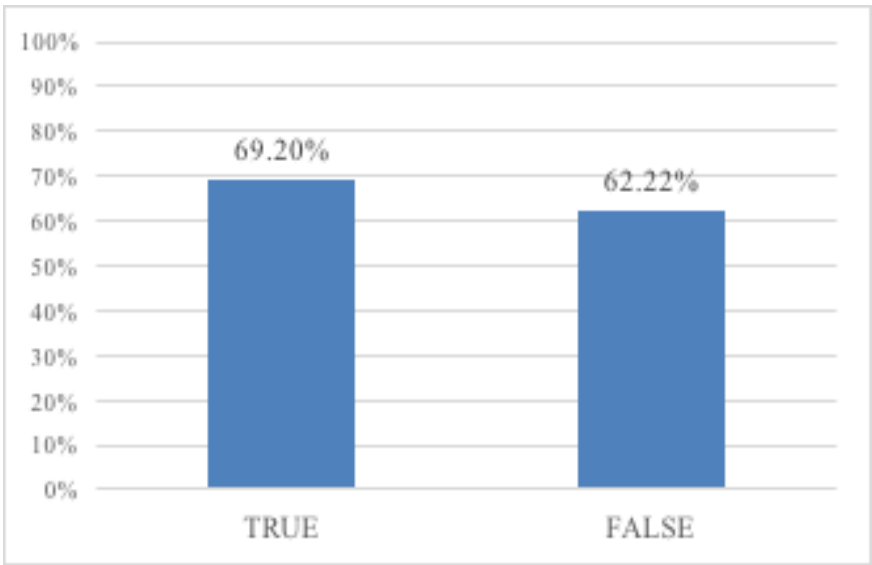

Figure 7 The correct answer rate for True/False statements

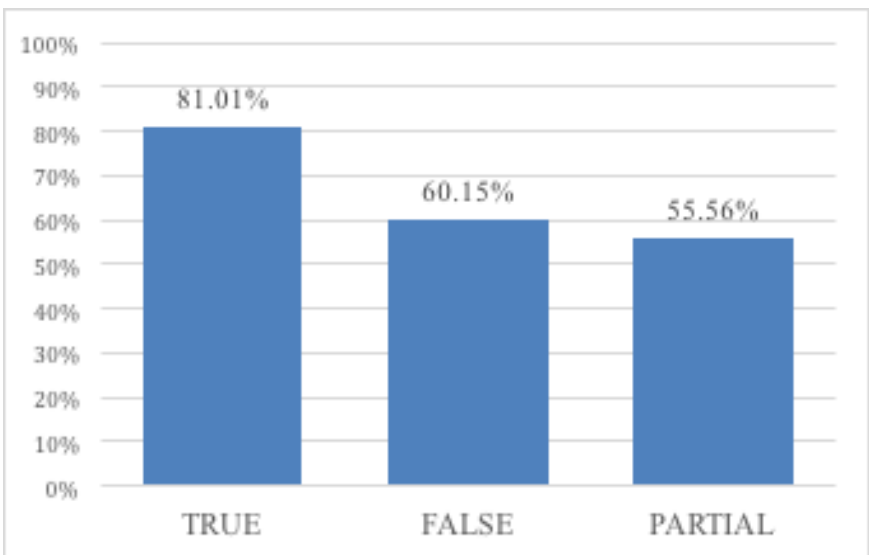

Figure 8 The correct answer rates for True/False/Partially True statements

enough time to read related news articles. Mostly the statements that were published in major news articles or rumors that has been around for months have higher correct answer rates compared to the other sentences.

The maximum amount of bonus reward was $0.50. Some participants received $0.30 and $0.10 for bonus according to the quality and the number of reasoning they provided. Total 180 participants received bonus reward and 134 (74.44%) of them received the maximum amount, which equals to $0.50. 27 participants (15.00%) received $0.30 and 19 (10.56%) received $0.10. 2 participants did not receive any reward because of insincerity of their answer. Few people wrote their random thoughts or just criticized politicians. However, most of the links that were provided were earnest and some participants even wrote their opinions and rationale together with the links.

4.2 Perceived Credibility of Crowdsourced Fact-checking

4.2.1. Overall Perceived Credibility

First, a one-way ANOVA was conducted to test the mean differences of credibility assessments between 'The New York Times' and 'Facebook' conditions. 'The New York Times' had higher credibility level overall (M = 4.72, SD = 1.47) while the credibility level of 'Facebook' condition was slightly lower (M= 4.54, SD = 1.47) than traditional journalism as expected in H 2-1. The results showed a significant difference between credibility of two conditions of traditional journalism and crowdsourced fact-checking results (F(1,2445) = 8.83, p = .002).

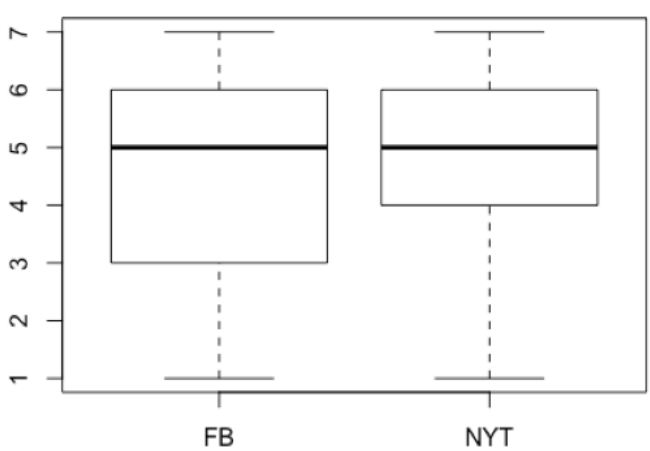

Figure 9 Credibility difference between 'Facebook' and 'The New York Times'

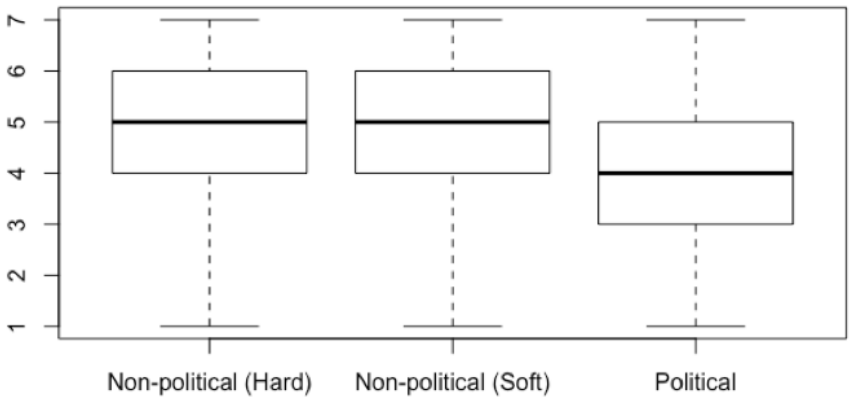

Figure 10 Credibility difference per category

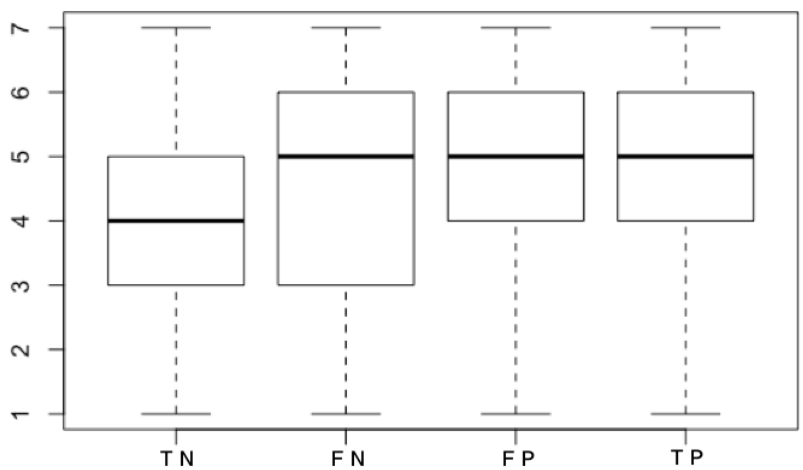

Figure 11 Credibility difference among story types

The credibility level difference among categories showed significance as well ($F(2,2445) = 30.87$, $p < .001$). Political news had the lowest credibility level ($M = 4.319$) compared to other

news categories. Non-political soft news had a little higher credibility level (M = 4.870) than non-political hard news (M = 4.703). The results for story types also indicated that there were significant differences ($F(3,2444 = 17.5, p < .001$). Among four stories of all conditions, True-Negative had the lowest credibility (M = 4.317) even though it delivered the truth in the stories. When original stories presented were true, they were considered more credible than other stories. False-Positive and True-Positive conditions had M = 4.831, M = 4.822 respectively while False-Negative had M = 4.554.

The post-hoc test results using Tukey HSD show that the political category news had significant difference with other news categories. For different story types, False-Positive and True-Positive stories had the biggest mean difference with True-Negative stories with significance. Except for True-Positive and False-Positive stories, all story types had some significant mean differences. This means that the political news articles received low credibility level that generated significant difference. The reason for this difference could be that recently many fake political news articles were spread throughout online during U.S.

presidential election and many people lost trust on those news articles. For story types, True-Negative stories brought the most difference with other story types. From this result, it could be inferred that even if the story is delivering the truth, people tend not to believe the articles that states that certain event is false.

Table 7. Post-hoc test for credibility level differences among categories

Pair-wise Comparison	Difference	95% Confidence Interval		p-value
		Lower Bound	Upper Bound	
Non-political (Soft) -Non-political (Hard)	0.1667	-0.0022	0.3355	0.0539
Political- Non-political (Hard)	-0.3848	-0.5536	-0.2160	< 0.0001
Political -Non-political (Soft)	-0.5515	-0.7203	-0.3826	0

Table 8. Post-hoc test for credibility level differences among story types

Pair-wise Comparison	Difference	95% Confidence Interval		p-value
		Lower Bound	Upper Bound	
FN-TN	0.2377	0.0236	0.4518575	0.0226
FP-TN	0.5146	0.3012	0.7280491	0
TP-TN	0.5058	0.2917	0.7199496	0
FP-FN	0.2769	0.0628	0.4910322	0.0050
TP-FN	0.2681	0.0533	0.4829304	0.0074
TP-FP	-0.0088	-0.2230	0.2053392	0.9996

4.2.2. Credibility Difference Between Traditional Media Style and Social Media Style

The overall credibility difference between 'Facebook' and 'The New York Times' showed slight difference and the professional journalists fact-checked results had a higher credibility level. However, when credibility comparison is separated into different categories, 'Facebook' condition had a higher credibility level for non-political soft news. Those news articles include 'Are Green Potatoes Poisonous?' and 'Google's Self-Driving Car Crashed'.

The mean value of credibility for each category of 'The New York Times' did not differ much ($F(2,1197) = 8.582$, $p = .0001$). Political news had the lowest credibility level ($M = 4.504$) and non-political soft news articles had a higher credibility level compared to that ($M = 4.728$). Non-political hard news had the highest credibility level ($M = 4.930$). On the other hand, compared to 'The New York Times' condition the mean value of credibility level for 'Facebook' condition fluctuate with categories ($F(2,1245) = 38.782$, $p < .001$). In 'Facebook' condition, the political category received the lowest credibility level ($M = 4.142$) which is the same as 'The New York Times' condition. However, non-political

soft news had the highest credibility level (M = 5.007) and the credibility level of non-political hard news was in the middle (M = 4.486).

The mean difference of 'The New York Times' condition for story types showed significant difference with F(3,1196) = 11.649, p <.001 and 'Facebook' condition displayed F(3,1244) = 29.554, p <.001. 'The New York Times' had a higher credibility level compared to 'Facebook' for three out of four story types. For False-Positive stories 'Facebook' had a higher credibility level (M = 5.144) than 'New York Times (M = 4.510). It was surprising that among all stories and conditions, 'Facebook' False-Positive had the highest credibility level. The lowest credibility level was also 'Facebook' and the story type was True-Negative (M = 4.170) and for the same story 'The New York Times' had a comparably low credibility level (M= 4.467) as well. The second lowest was 'Facebook' with False-Negative story (M = 4.276) while 'The New York Times' with the same story had relatively high credibility with M = 4.848. True-Positive story had the highest credibility for 'The New York Times' condition (M =5.068) and the second highest for the 'Facebook' condition (M = 4.590).

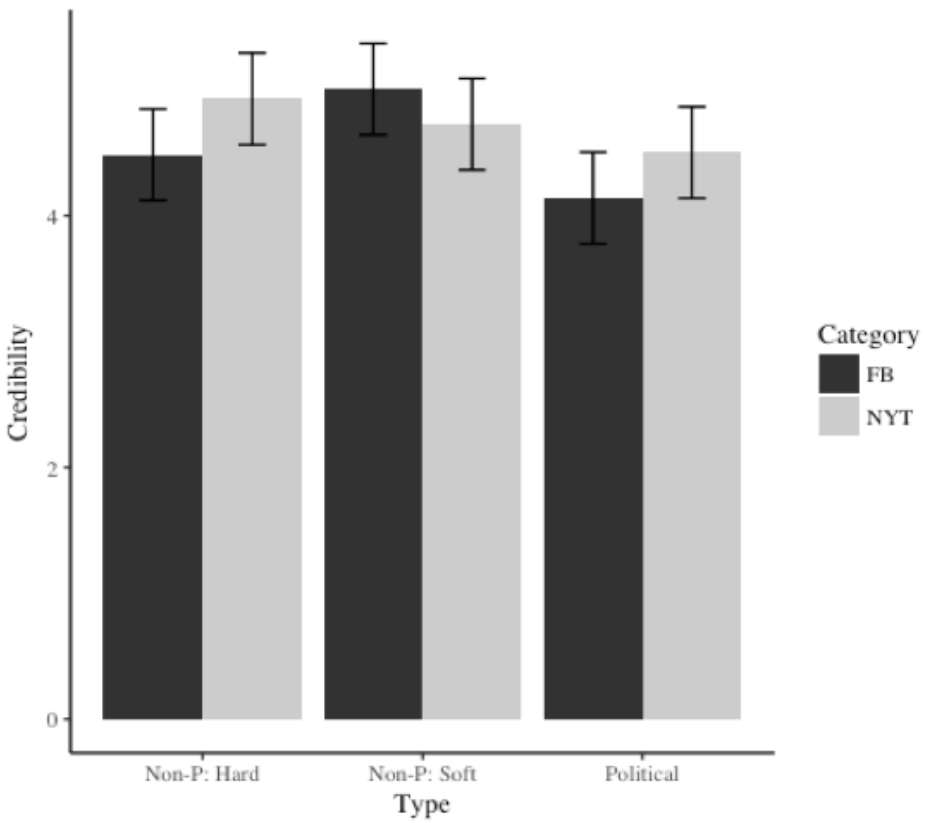

Figure 12 Credibility difference per category

Table 9. Mean difference of credibility level per category

	Non-political (Hard)	Non-political (Soft)	Political
NYT	4.930	4.728	4.504
FB	4.486	5.007	4.142

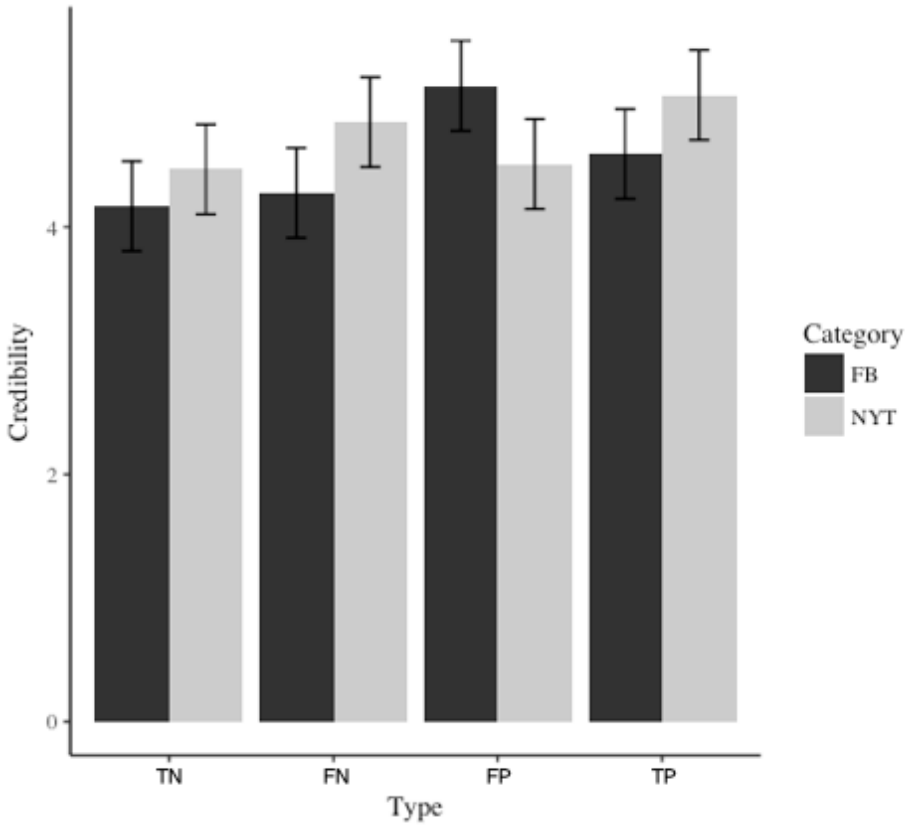

Figure 13 Credibility difference among story types

Table 10. Mean difference of credibility level among story types

	TN	FN	FP	TP
NYT	4.467	4.848	4.510	5.068
FB	4.170	4.276	5.144	4.590

The post-hoc test results using Tukey HSD showed that for 'Facebook' condition, all categories showed significant mean differences, while only political and non-political hard news category had significant difference for 'The New York Times' condition. It means that for 'Facebook' condition, the category of the story is significant for its credibility and some features of social network services would have affected this significant difference. Since the different motives of the site and user brings different aspect of credibility of media and its content (Metzger et al., 2003; Dochterman & Stamp, 2010), different motives of 'Facebook' and 'The New York Times' perceived by users might have caused the difference in credibility and significance.

As for story types, 'Facebook' also had more significant mean differences in pairwise comparison than 'The New York Times.' For 'Facebook' condition, all pairwise for except 2 pairs had all significant difference. On the other hand, 'The New York Times', half of the pairs had significant mean differences. The impressive result is that for 'Facebook', the mean difference between False-Positive and True-Negative stories was the biggest and significant but for 'The New York Times' it was the smallest and

not significant. It was because of the difference between two conditions for False-Positive story types were biggest and 'Facebook' condition received the highest credibility level. It could be inferred that people are likely to believe the story that opposes to certain statements on 'Facebook.' Also, it could be because for 'The New York Times', the positive stories that delivering truthful story received higher credibility level compared to 'Facebook' condition.

Table 11. Post-hoc test for credibility level differences among categories of 'Facebook' condition

Pair-wise Comparison	Difference	95% Confidence Interval		p-value
		Lower Bound	Upper Bound	
Non-political (Soft) -Non-political (Hard)	0.5216	0.2894	0.7538	< 0.0001
Political- Non-political (Hard)	-0.3438	-0.5759	-0.1116	0.0015
Political -Non-political (Soft)	-0.8654	-1.0976	-0.6332	0

Table 12. Post-hoc test for credibility level differences among categories of 'The New York Times' condition

Pair-wise Comparison	Difference	95% Confidence Interval		p-value
		Lower Bound	Upper Bound	
Non-political (Soft) -Non-political (Hard)	-0.2025	-0.4448	0.03975	0.1223
Political- Non-political (Hard)	-0.4275	-0.6698	-0.1852	0.0001
Political -Non-political (Soft)	-0.2250	-0.4673	0.0173	0.0751

Table 13. Post-hoc test for credibility level differences among story types of 'Facebook' condition

Pair-wise Comparison	Difference	95% Confidence Interval		p-value
		Lower Bound	Upper Bound	
FN-TN	0.1058	-0.1870	0.3986	0.7892
FP-TN	0.9744	0.6815	1.2672	0
TP-TN	0.4199	0.1271	0.7127	0.0013
FP-FN	0.8686	0.5758	1.1614	0
TP-FN	0.3141	0.0213	0.6069	0.0299
TP-FP	-0.5545	-0.8473	-0.2617	< 0.0001

Table 14. Post-hoc test for credibility level differences among story types of 'The New York Times' condition

Pair-wise Comparison	Difference	95% Confidence Interval		p-value
		Lower Bound	Upper Bound	
FN-TN	0.3810	0.0763	0.6855	0.0073
FP-TN	0.0428	-0.2598	0.3453	0.9836
TP-TN	0.6005	0.2959	0.9051	< 0.0001
FP-FN	-0.3381	-0.6427	-0.0335	0.0226
TP-FN	0.2196	-0.0870	0.5262	0.2540
TP-FP	0.5577	0.2531	0.8623	< 0.0001

4.2.3. Designing Online Platform Elements

Total 1,301 answers were collected by using Amazon's Mechanical Turk. The average time for the HIT was 2 minute 41 seconds. The time taken is comparably short compared to the first and second experiment as it includes only a few sentences

to read and assess credibility. The posts were randomly presented to users when they started working on the HIT and in order to receive their reward the workers had to answer all the questions and submit the HIT.

The credibility was evaluated with a 7-point Likert scale and the average credibility of all posts was 3.589. The online post that was written by a person with nickname and no picture but with a high level received the lowest credibility level which was 2.750. The location for this post was open and this post received high number of comments but 95% of other users opposed to this post's content. The highest credibility level received was 4.337 and this post was written by a user who opens her picture and location but has la ow level. Also, the user did not use real name to write post, but received a high number of comments and agreements from other users. The 3 posts that received the lowest credibility levels and 3 posts with the highest credibility levels are listed in Table 15.

In order to observe the relationship between credibility level and variables that were used in the online post, linear regression

was conducted. The number of yes or no was the only significant variable (ß = 0.3, p <0.001). This means that agreement from other users are most important to increase credibility level. Reputation and past behavior of the user is more significant than social information disclosure for crowdsourced fact-checking results to obtain usefulness. Considering the R-squared value of the regression result, it can be concluded that the user components do not affect the credibility level of crowdsourced fact-checking results. Nonetheless, the consensus from other users (Yes/No ratio) showed significance to the credibility level and this result could be further developed into future studies.

Table 15. Highest 3 and lowest 3 credibility level variables

Rank	Name	Picture	Location	Level	Yes/No	Comment	Credibility
High 1	Nickname	Yes	Open	Low	Yes	High	4.337
High 2	Nickname	Yes	Closed	High	Yes	Low	4.250
High 3	Real name	No	Closed	High	Yes	Low	4.211
Low 1	Nickname	No	Open	High	No	High	2.750
Low 2	Nickname	Yes	Open	High	No	Low	2.833
Low 3	Nickname	No	Open	High	No	Low	2.913

Table 16. The statistical value for linear regression

	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	3.55712	0.103727	34.293 <	2.00E-16
Name	-0.013109	0.078688	-0.167	0.868
Picture	-0.006521	0.07881	-0.083	0.934
Location	-0.084645	0.078866	-1.073	0.283
Level	-0.087142	0.078694	-1.107	0.268
Yes/No	0.327616	0.078766	4.159	3.40E-05
Comment	-0.064832	0.079139	-0.819	0.413

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.418 on 1294 degrees of freedom
Multiple R-squared: 0.01517,
F-statistic: 3.322 on 6 and 1294 DF, p-value: 0.002993

5. Discussion

5.1 Possibility of Public Discussion Through Fact-checking Process

The correct answer rate of crowdsourced fact-checking is not impressively high but when the statements that were partially true were excluded, the correct answer rate increases to 75.34%. When the statement about 'Twitter', which was "Twitter will increase its per-tweet character count from 140 to 10,000", was excluded, the correct answer rate jumps to 80.41%. This statement confused many people because Twitter recently changed the method for counting per-tweet characters by excluding links and names for counting. The result of the first experiment implies that except for ambiguous claims, people are capable of searching for information to support their answer. For partially true sentences, many people raised issues on binary answer choices and provided reasoning for their answers. They tried to discuss the claims and explain why the statements were only partly correct. Though it is important to pull up the correct answer rate for crowdsourced fact-checking results, it has a

significant meaning that people invested their time to search for information.

Considering the average time the workers spent for answering these questions was 23 minutes and the highest reward the participants received was $0.60, they would have earned $1.50 for 1 hour. This is a very cheap labor and encouraging results since fact-checking was an expensive task. However, the motivational factor of crowdsourced task is not examined in this research. If a crowdsourced fact-checking platform is to be designed, it is most likely the workers would not be paid for their work. Therefore, comparison of the correct answer rate between paid and unpaid workers would be interesting. Previous studies on relationship between incentives and performance of crowds reveal that financial incentives do not necessarily increase the quality of the work (Bonner, et al., 2000; Mason & Watts, 2010; Marge, Banerjee, & Rudnicky, 2010). As workers in Amazon's Mechanical Turk are not primarily motivated by financial incentives (Mason & Suri, 2012), the amount of reward only increases the quantity of the work but not quality (Mason & Watts, 2010). However, low compensations may lead to longer delay of gathering workers

(Mason & Suri, 2012) and their completion time (Merge et al., 2010). For motivational aspect, the financial incentive would not be the right solution and other approaches should be explored in order for crowdsourced fact-checking platforms to last.

The correct answer rate for political category was lower compared to the other categories. In line with the concern of crowdsourced fact-checking, some people simply criticized particular politicians. Regarding the sentences about Donald Trump, some users attacked Trump without any logic. Their reasons included insults such as "*Of course it's true! Trump is an idiot*" or "*I know this because I did a lot of research on that psychopath for the election. But here is a link as well.*" These hateful comments were the same for Hillary Clinton. Some people called her 'liar' or pointed out the election results with comments like "*She's lying Hillary*" or "*She lost elections.*" However, as mentioned above, many participants provided their answer with specific reasoning and detailed explanation, especially for partially true statements. They provided some evidence together with profound explanations. The quotes below are the actual explanation from the participants which include profound thoughts

and discussions.

> *"Partially true, there is a disease that's causing the potential risk of losing one breed of banana, but not all of them."*

> *"Putin only congratulated Donald Trump. And there is no additional message, according to Google search results and many newspaper articles."*

> *"This statement is too broad- where will it become illegal? When will it become illegal? Are we talking some day in the future, then yeah, probably. Are we talking next year? Maybe in some places, and it is likely illegal in many localities already.*
> *http://www.snopes.com/smoking-in-cars-with-children-illegal/"*

Their serious and sincere explanations imply the possibility of the public discussions. If their deliberate opinions and explanation for their answer are accumulated and open to other people, public opinion would go closer to the truth. If the given statements were ambiguous or too broad, participants raised an issue and elaborated the statements together to judge the truthfulness. For

example, if the users were to judge the sentence “Smoking in a car with children will become illegal”, they narrowed it down the sentence to clarify it.

5.2 Design Implications

When perceived credibility levels of fact-checking results of ‘The New York Times’ and ‘Facebook’ are compared, ‘The New York Times’ had a generally higher level. However, for non-political soft news and False-Positive stories, “Facebook” condition had a higher credibility level. For False-Positive stories, “Facebook”, crowdsourced fact-checking results condition, had the highest credibility level. This might be because of the characteristics of new media. New media journalism, due to its independence of financial issues and objectivity controversies, usually opposes to traditional media reports. Therefore, new media news articles denying the true stories that are reported by traditional journalism might have earned high credibility.

In this experiment, the political category generally received the lowest credibility level and this might be due to the political bias of the participants and other users. Also, during the recent

presidential election of the United States of America, fake political news articles rapidly spread through social network services were troublesome. These factors may have declined the credibility level of political news. “Facebook” condition had a higher credibility level compared to other condition for non-political soft news content. This could be because the general news category and many soft news articles are read in portal sites and social network sites. Moreover, the soft news article that received high credibility for “Facebook” was about Google’s autonomous car. Since new technology related news articles are shared by many people on social networks, those articles may have received high credibility levels. The motives of “Facebook” site, and the intentions of the users accessing the social network services might have affected the credibility of its content (Metzger et al., 2003; Dochterman & Stamp, 2010). Since many users of new media and social network services are young and more interested in technological issues, some specific categories might have received higher credibility compared to other stories.

The last experiment showed the possibility of crowdsourced fact-checking since the consent from other users were the most

influential and significant to the credibility of the results. This implies that rather than social information disclosure level, managing reputation of the participants in fact-checking would increase credibility of the fact-check results and its usefulness. By using more stories to collect data, this experiment could have revealed more about variables that affect credibility of fact-checking results, however, there are some implications we could use when designing a crowdsourced fact-checking platform. If users feel uncomfortable about publicizing their social information, they could disclose either their name or picture of themselves. Also, instead of assigning levels to participants, indicating the percentage of consent that they have received would be useful.

5.3 Limitations and Future Studies

As mentioned above, the correct answer rate is not remarkably high. In this study, the factors that could increase the correct answer rate are not examined. In future studies, supplement from existing fact-checking algorithms could be considered in the experiment to observe the synergy effect. Since, fact-checking

algorithms are being developed and real-time fact-checking is being implemented, observing the possibility of real-time fact-checking by using algorithms and crowdsourced labor would have been given an impressive result. Collaboration with the professional fact-checkers and crowd also could be examined to identify the method to increase the correct answer rate.

When observing credibility difference between 'The New York Times' and 'Facebook', 2 stories were used in each category of news. However, since some participants have heard about those topics before the experiment, using more stories in the experiment would have been helpful in examining the credibility difference. Also, if the variables that affected the credibility difference was observed together, it would have uncovered more about factors that affect credibility of fact-checked results which could be used in the last experiment. This study only observed the credibility variables of crowdsourced fact-checking results. The future study may focus on how those variables affect credibility difference of professional and crowdsourced fact-checking results to elaborate the platform design.

6. Conclusion

This study has explored the possibility of crowdsourced fact-checking. Except for too broad statements, the crowd did reasonable research for the answer and provided accurate answers with the evidence. More, many people provided critical links for the answers and even raised questions on ambiguous sentences. The result showed that the crowd is capable of extracting and narrowing down verifiable statements from broad sentences and checking veracity of those claims.

'The New York Times' news articles, which are the traditional journalism fact-checking results, received generally high credibility level compared to 'Facebook' postings, as expected. However, for some conditions 'Facebook', the crowdsourced fact-checking results had a higher credibility level. The reasons behind this result need to be further investigated but the characteristics of social media might have brought this result. The features of social media should be taken into account when building a platform for crowdsourced fact-checking. The result of the last experiment implies that consensus of other people can

bring high credibility of fact-check results. Social disclosure level did not influence the credibility level of the result but reputation or assessment of other users did.

The results of this study confirm that the crowdsourced fact-checking has potential possibility and found some feature that can raise the credibility level of crowdsourced fact-checking. Using the cheap labor of the crowd, fact-checking can be done rapidly compared to the traditional method. Since several new projects on fact-checking are planned in the near future, this exploratory study could be helpful in designing fact-checking system that needs help from the crowd. When collaborating with the fact-checking algorithms that extract the check-worthy claim and sentences, the labor of the crowd would generate synergy and rapidly "self-purify" the pollution from false information on the Internet and media.

References

1. Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. Groups, leadership, and men, 222-236.

2. Amazeen, M. A. (2013). Making a difference: A critical assessment of fact-checking in 2012. New America Foundation Media Policy Initiative Research Paper.

3. Andersen, M. (2009). Four crowdsourcing lessons from the Guardian's (spectacular) expenses-scandal experiment. Nieman Journalism Lab, 23, 2009.

4. Ban Hyun (2003). Internet news media eui silledo-e gwanhan shilhum yeongu [Experimental study of user credibility on Internet newsmedia]. Korean Journal of Broadcasting and Telecommunication Studies, 17(2),207-231.

5. Ban Hyun (2007). Online, offline news media eui simin journalism yeokhal-e deahan sooyongja pyungga-wa silledo yeongu [Audience perception and trust toward civic journalism in online and offline news media]. The Journal of Political Science and Communication, 10(2), 297-316.

6. Bantz, C. R., McCorkle, S., & Baade, R. C. (1980). The news factory. Communication Research, 7(1), 45-68.

7. Beckett, C., & Mansell, R. (2008). Crossing boundaries: New media and networked journalism. Communication, Culture and Critique, 1(1), 92-104.

8. Bonner, S. E., Hastie, R., Sprinkle, G. B., & Young, S. M. (2000). A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. Journal of Management Accounting Research, 12(1), 19-64.

9. Chang, Jeong-Heon J,C,, Ha, Ju-Yong, & Kim, Son-Ho (2014). Jeongchi jeongbowon-euirosu Internet media silledo yeongu [Credibility of the Internet as a Source of Political Information: A Survey Analysis the 2012 Korean Presidential Election Campaign Period]. Korean Journal of Journalism &Communication Studies,58(4), 96-128.

10. Chen, W., Zeng, Q., Wenyin, L., & Hao, T. (2007). A user reputation model for a user-interactive question answering system. Concurrency and Computation: Practice and Experience, 19(15),2091-2103.

11. Chen, Y., Conroy, N. J., & Rubin, V. L. (2015, November). News in an online world: the need for an automatic crap detector. In Proceedings of the 78th ASIS & T Annual Meeting: Information Science with Impact: Research in and for the Community (p. 81). American Society for Information Science.

12. Cho, J., Kwon, K., & Park, Y. (2009). Q-rater: A collaborative reputation system based on source credibility theory. Expert Systems with Applications, 36(2), 3751-3760.

13. Chong, S. K., & Abawajy, J. H. (2007, April). Feedback

credibility issues in trust management systems. In Multimedia and Ubiquitous Engineering, 2007. MUE' 07. International Conference on (pp.387-394). IEEE.

14. Chung, C. J., Nam, Y., & Stefanone, M. A. (2012). Exploring online news credibility: The relative influence of traditional and technological factors. Journal of Computer-Mediated Communication, 17(2), 171-186.

15. Ciampaglia, G. L., Shiralkar, P., Rocha, L. M.,Bollen, J., Menczer, F., & Flammini, A. (2015). Computational fact checking from knowledge networks. PloS one, 10(6), e0128193.

16. Clerwall, C. (2014). Enter the Robot Journalist: Users' perceptions of automated content. Journalism Practice, 8(5), 519-531.

17. Coddington, M., Molyneux, L., & Lawrence, R. G.(2014). Fact checking the campaign how political reporters use Twitter to set the record straight (or not). The International Journal of Press/Politics,19(4), 391-409.

18. Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1), 1-4.

19. Cunningham, B. (2003). Rethinking Objectivity. Columbia Journalism Review. July/August 2003.
http://www.cjr.org/feature/rethinking_objectivity.php.

20. Dellarocas, C. (2010). Designing reputation systems for the social web. Boston U. School of Management Research Paper, (2010-18).

21. Dobbs, M. (2012). The Rise of Political Fact-checking How Reagan Inspired a Journalistic Movement: A Reporter's Eye View.

22. Dochterman, M. A., & Stamp, G. H. (2010). Part 1: The Determination of Web Credibility: A Thematic Analysis of Web User's Judgments. Qualitative Research Reports in Communication, 11(1), 37-43.

23. Donath, J. S. (1999). Identity and deception in the virtual community. Communities in cyberspace, 1996, 29-59.

24. Dunwoody, S. (1982). A question of accuracy. IEEE Trans Prof Commun, 25, 196-9.

25. Ennals, R., Trushkowsky, B., & Agosta, J. M. (2010, April). Highlighting disputed claims on the web. In Proceedings of the19th international conference on World wide web (pp. 341-350). ACM.

26. Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. Journalism & Mass Communication Quarterly, 77(3), 515-540.

27. Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors

on the perceived credibility of web-based information. New Media & Society, 9(2), 319-342.

28. Flanagin, A. J., Metzger, M. J., Pure, R., & Markov, A. (2011, January). User-generated ratings and the evaluation of credibility and product quality in ecommerce transactions. In System Sciences (HICSS), 2011 44th Hawaii International Conference on (pp. 1-10). IEEE.

29. Florin, F. (November 16, 2010). Crowdsourced Fact-Checking? What We Learned from Truthsquad. http://mediashift.org/2010/11/crowdsourced-fact-checking-what-we-learned-from-truthsquad320/

30. Fogg, B. J., & Tseng, H. (1999, May). The elements of computer credibility. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems (pp. 80-87). ACM.

31. Fogg, B.J., J. Marshall,O. Laraki, A. Osipovich, C.Varma, N. Fang, J. Paul, A. Rangnekar, J. Shon, P. Swani and M.Treinen (2001b). 'What Makes Web Sites Credible? A Report on a Large Quantitative Study',Proceedings of ACM CHI 2001 Conference on Human Factors in Computing Systems. Seattle,WA.

32. Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. Information Systems Research, 19(3), 291-313.

33. Freeman, K. S., & Spyridakis, J. H. (2004). An examination of

factors that affect the credibility of online health information. Technical Communication, 51(2), 239-263.

34. Giudice, K. D. (2010). Crowdsourcing credibility: The impact of audience feedback on Web page credibility. Proceedings of the American Society for Information Science and Technology, 47(1), 1-9.

35. Graefe, A. (2016). Guide to automated journalism. Tow Center for Digital Journalism. Janeiro.

36. Graves, L., & Glaisyer, T. (2012). The fact-checking universe in Spring 2012. New America.

37. Graves, L., Nyhan, B., & Reifler, J. (2016). Understanding Innovations in Journalistic Practice: A Field Experiment Examining Motivations for Fact-Checking. Journal of Communication.

38. Hanitzsch, T. (2007). Deconstructing journalism culture: Toward a universal theory. Communication theory, 17(4), 367-385.

39. Hansen, D. L., Schone, P. J., Corey, D., Reid, M., & Gehring, J. (2013, February). Quality control mechanisms for crowdsourcing: peer review, arbitration, & expertise at family search indexing. In Proceedings of the 2013 conference on Computer supported cooperative work (pp. 649-660). ACM.

40. Hassan, N., Li, C., & Tremayne, M. (2015, October). Detecting check-worthy factual claims in presidential debates. In

Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (pp. 1835-1838). ACM.

41. Hill, K. A., & Hughes, J. E. (1999). Cyberpolitics: Citizen activism in the age of the Internet. Rowman& Littlefield Publishers, Inc..

42. Horrigan, J. B., Garrett, K., & Resnick, P. (2004). The Internet and democratic debate. Pew Internet &American Life Project.

43. Jensen C, Davis J, & Farnham S. (2002). Finding others online: reputation systems for social online spaces. In: Proceedings of CHI2002 Conference on Human Factors in Computer Systems. New York:

44. Johnson, T. J., & Kaye, B. K. (2000). Using is believing: The influence of reliance on the credibility of online political information among politically interested Internet users. Journalism &Mass Communication Quarterly, 77(4), 865-879.

45. Johnson, T. J., & Kaye, B. K. (2004). Wag the blog: How reliance on traditional media and the Internet influence credibility perceptions of weblogs among blog users. Journalism & Mass Communication Quarterly, 81(3), 622-642.

46. Johnson, T. J., Kaye, B. K., Bichard, S. L., & Wong, W. J. (2007). Every Blog Has Its Day: Politically-interested Internet Users' Perceptions of Blog Credibility. Journal of Computer-Mediated Communication, 13(1), 100-122.

47. Johnson, T. J., & Kaye, B. K. (2010). Still cruising and believing? An analysis of online credibility across three presidential campaigns. American Behavioral Scientist, 54(1), 57-77.

48. Kim, Y., & Srivastava, J. (2007, August). Impact of social influence in e-commerce decision making. In Proceedings of the ninth international conference on Electronic commerce (pp. 293-302). ACM.

49. Kiousis, S. (2001). Public trust or mistrust? Perceptions of media credibility in the information age. Mass Communication &Society, 4(4), 381-403.

50. Kriplean, T., Bonnar, C., Borning, A., Kinney, B.,& Gill, B. (2014, February). Integrating on-demand fact-checking with public dialogue. In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (pp. 1188-1199). ACM.

51. Kusumasondjaja, S., Shanka, T., & Marchegiani, C. (2012). Credibility of online reviews and initial trust The roles of reviewer's identity and review valence. Journal of Vacation Marketing, 18(3), 185-195.

52. Lease, M. (2011). On Quality Control and Machine Learning in Crowdsourcing. Human Computation, 11, 11.

53. Lowry, P. B., Roberts, T. L., & Higbee, T. (2007, July). First impressions with websites: the effect of the familiarity and credibility of corporate logos on perceived consumer swift

trust of websites. In International Conference on Human-Computer Interaction (pp. 77-85). Springer Berlin Heidelberg.

54. Ma, M., & Agarwal, R. (2007). Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. Information systems research, 18(1), 42-67.

55. Magdy, A., & Wanas, N. (2010, October). Web-based statistical fact checking of textual documents. In Proceedings of the 2nd international workshop on Search and mining user-generated contents (pp.103-110). ACM.

56. Marge, M., Banerjee, S., & Rudnicky, A. I. (2010, March). Using the amazon mechanical turk for transcription of spoken language. In 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 5270-5273). IEEE.

57. Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's MechanicalTurk. Behavior research methods, 44(1), 1-23.

58. Mason, W., & Watts, D. J. (2010). Financial incentives and the performance of crowds. ACM SigKDD Explorations Newsletter, 11(2), 100-108.

59. McKenna, K. Y., & Bargh, J. A. (1999). Causes and consequences of social interaction on the Internet: A conceptual framework. Media Psychology, 1(3), 249-269.

60. Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & McCann, R. M. (2003). Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. Annals of the International Communication Association, 27(1), 293-335.

61. Mikkelson, D. (October 16, 2014). Ebola in the AirTonight. http://www.snopes.com/medical/disease/ebolaairborne.asp

62. National Conference of Editorial Writers (Oct. 18,1996). Public journalism. Http://www.ncew.org/civic.html

63. Nyhan, B., & Reifler, J. (2014). Misinformation and fact-checking: research findings from social science. New America Foundation Media Policy Initiative Research Paper; 2012.

64. Nyhan, B., & Reifler, J. (2015). The Effect of Fact-Checking on Elites: A Field Experiment on US State Legislators. American Journal of Political Science, 59(3), 628-640.

65. Pavlik, J. (2000). The impact of technology on journalism. Journalism Studies, 1(2), 229-237.

66. Petty, R. E., & Brinol, P. (2010). Attitude change. Advanced social psychology: The state of the science, 217-259.

67. Quinn, A. J., & Bederson, B. B. (2011, May). Human computation: a survey and taxonomy of a growing field. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1403-1412). ACM.

68. Reinhard, M. A., Scharmach, M., & Sporer, S. L. (2012). Situational familiarity, efficacy expectations, and the process of credibility attribution. Basic and Applied Social Psychology, 34(2), 107-127.

69. Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. Organization studies, 28(3), 377-408.

70. Schudson, M. (2001). The objectivity norm in American journalism*. Journalism, 2(2), 149-170.

71. Shapiro, I., Brin, C., Bédard-Brûlé, I., & Mychajlowycz, K. (2013). Verification as a strategic ritual: How journalists retrospectively describe processes for ensuring accuracy. Journalism Practice, 7(6), 657-673.

72. Spivak, C. (2011). The fact-checking explosion. American Journalism Review, 32(4), 38-43.

73. Sundar, S. S., & Nass, C. (2001). Conceptualizing sources in online news. Journal of Communication, 51(1), 52-72.

74. Sundar, S. S. (1998). Effect of source attribution on perception of online news stories. Journalism &Mass Communication Quarterly, 75(1), 55-68.

75. Sussman, S. W., & Siegal, W. S. (2003).Informational influence in organizations: An integrated approach to knowledge adoption. Information systems research, 14(1), 47-65.

76. The Guardian. (June 23, 2009). MPs' expenses: The Guardian launches major crowdsourcing experiment. http://www.theguardian.com/gnm-press-office/crowdsourcing-mps-expenses

77. Van der Haak, B., Parks, M., & Castells, M.(2012). The future of journalism: Networked journalism. InternationalJournal of Communication, 6, 16.

78. Van der Kaa, H., & Krahmer, E. (2014, October). Journalist versus news consumer: The perceived credibility of machine written news. In Research Paper presented at the.

79. Vehkoo, J. (2013). Crowdsourcing in investigative journalism. Reuters Institute for the Study of Journalism Report. https://reutersinstitute.politics. ox. ac. uk/fileadmin/documents/Publications/fellows_papers/2009-2010/Crowdsourcing_in_Investigative_Journalism.pdf.

80. Wang, Z., Walther, J.B., Pingree, S., & Hawkins,R.P. (2008). Health information, credibility, homophily, and influence via the internet: Web sites versus discussion groups. Health Communication, 23, 358-368.

81. Whitehead Jr, J. L. (1968). Factors of source credibility. Quarterly Journal of Speech, 54(1), 59-63.

82. Wiesenfeld, B. M., Raghuram, S., & Garud, R. (1999). Communication patterns as determinants of organizational identification in a virtual organization. Organization Science,

10(6), 777-790.

83. Woodly, D. (2008). New competencies in democratic communication? Blogs, agenda setting and political participation. Public Choice, 134(1-2), 109-123.

84. Wu, F., Li, H. H., & Kuo, Y. H. (2011). Reputation evaluation for choosing a trustworthy counterparty in C2C e-commerce. Electronic Commerce Research and Applications, 10(4), 428-436.

85. Wu, Y., Agarwal, P. K., Li, C., & Yang, J. (2014). C. Yu. Toward computational fact-checking.

국문초록

크라우드소싱을 이용한 팩트체킹의 유효성과 신뢰도

홍나희
서울대학교
사회과학대학원
언론정보학과

팩트체킹은 '정확성'을 주요 가치로 여기는 저널리즘에서 중요한 과정이다. 디지털 기술의 발전으로 정보의 양이 증가하면서 팩트체킹의 수요도 함께 증가했다. 하지만 팩트체킹을 위해서는 시간과 비용이 적지 않게 필요하고 전문가의 노동력 또한 정보의 양에 비해 모자라기 때문에 빠르게 정보가 확산되는 디지털 미디어 환경에서 팩트체킹 과정이 누락되는 일이 다수다. 따라서 거짓 정보 및 오보가 증가했으며 이를 방지하기 위해 많은 기업과 연구진들이 자동 팩트체킹 알고리즘 개발을 위해 노력하고 있다. 하지만 해당 알고리즘은 아직 상용화 되기에는 부족하며

대중의 참여를 통해 이를 보완하고자 하는 시도들도 있었다. 독자들이 직접 데이터를 수집하고 분류하여 분석 작업에 참여하는 것부터 독자의 의견을 수용하여 보도에 활용하는 경우도 찾아볼 수 있다.

본 연구는 크라우드소싱의 대표적인 플랫폼인 아마존의 메커니컬 터크를 이용하여 집단지성을 이용한 팩트체킹의 가능성에 대해 알아보고 실험을 통해 전통 저널리즘과 집단지성을 통한 팩트체킹의 신뢰도를 측정한다. 마지막으로 집단지성을 이용한 팩트체킹 결과의 신뢰도에 영향을 미치는 요소를 찾기 위한 실험을 진행한다. 결과를 통해 너무 넓은 주제를 다루거나 애매한 내용을 담고 있는 문장 외에는 사람들의 팩트체킹 정답률이 높은 것을 확인할 수 있었다. 특히 사람들이 자신들의 의견을 전달하거나 중요한 증거 자료를 제출하기도 했으며 대부분의 사람들이 답변에 대한 합리적인 이유를 적었다. 사람들은 팩트체킹에 참여하면서 토론이 가능하도록 자신의 의견과 답변에 대한 이유를 제공했으며 특히 주어진 문장에 대해 사실 확인이 가능한 문장으로 좁혀 나가기도 했다. 전통 저널리즘의 팩트체킹 결과가 집단지성을 통한 결과에 비해 전반적으로 높았지만 몇 개의 조건에서는 집단지성을 이용한 팩트체킹의 신뢰도가 더 높은 것을 확인할 수 있었다. 이는 소셜 미디어의 특성에 기인한 것을 유추해볼 수 있다. 사용자의 사회적 정보 노출 정도보다 다른 사용자의 평가 내용이 결과의 신뢰도에 영향을 준다는 것을 알 수 있었으며 이러한 결과를 종합하여 팩트체킹 시스템의 디자인 요소를 제안

할 수 있다. 또 현재 완벽하지 않은 팩트체킹 알고리즘을 집단지성을 이용하여 보완할 수 있을 것이다.

주요어 : 팩트체킹, 크라우드소싱, 저널리즘, 온라인 미디어, 뉴스 신뢰도

학번 : 2015-20247