Online participation and crowdsourcing as a solution to mitigate news bias

Himesha Wijekoon*

Faculty of Economics and Management,
Department of Information Engineering,
Czech University of Life Sciences Prague,
Kamýcká 129, 16500, Prague, Czech Republic

and

Faculty of Science,
Department of Industrial Management,
University of Kelaniya,
Kelaniya, 11600, Sri Lanka
Email: wijekoon@pef.czu.cz
*Corresponding author

Boris Schegolev

Faculty of Economics and Management,
Department of Information Engineering,
Czech University of Life Sciences Prague,
Kamýcká 129, 16500, Prague, Czech Republic
Email: schegolev@pef.czu.cz

Vojtěch Merunka

Faculty of Economics and Management,
Department of Information Engineering,
Czech University of Life Sciences Prague,
Kamýcká 129, 16500, Prague, Czech Republic

and

Faculty of Nuclear Sciences and Engineering,
Department of Software Engineering,
Czech Technical University in Prague,
Jugoslávských partyzánů 1580/3,
16000, Prague, Czech Republic
Email: merunka@pef.czu.cz

Abstract: It is apparent that certain news outlets occasionally report biased news. Forcefully planted biased viewpoints can cause harmful effects on society thus democracy. Since online news is gradually becoming the main source of news particularly for the new generation, new mechanisms to tackle online news bias will benefit the society as a whole in future. In this study we investigate how online participation and crowdsourcing techniques can be

utilised to fight against biased news to leverage the power of internet users who are now becoming active participants rather than merely being passive readers. Even though there have been some previous attempts to mitigate online news bias, the aforementioned methods have been overlooked and underutilised. Based on the findings of this study a novel news platform to mitigate online news bias is proposed and discussed in the end.

Keywords: online participation; crowdsourcing; news bias; online news; news aggregation; news platform; bias mitigation.

Biographical notes: Himesha Wijekoon is a Lecturer in Information Technology at the Department of Industrial Management in University of Kelaniya, Sri Lanka. He holds a MSc in Systems Engineering and Informatics from Czech University of Life Sciences Prague, Czech Republic. At present, he is a PhD candidate at the Department of Information Engineering in the Czech University of Life Sciences Prague, Czech Republic. He also has more than 10 years of experience as a software engineer in multiple companies. His main research interests include crowdsourcing, software engineering, software modelling and model driven engineering.

Boris Schegolev is a post-graduate student at Czech University of Life Sciences in the field of Systems Engineering and Informatics. As an experienced Software Engineer, in his research he mostly focuses on patterns and antipatterns in various engineering environments. He has published his research in conference proceedings and journals.

Vojtěch Merunka is an Associate Professor in Information Management since 2005 at the Czech University of Life Sciences in Prague and the Czech University of Technology in Prague. He teaches software engineering and information management. He holds a Master's in Computer and Software Engineering and has obtained his PhD in Data Processing and Mathematical Modelling in 1998. He also has 15 years of experience in the international management and consulting company Deloitte. He is a founding member of the MOBA (Modelling of Business Agility) workshop and the SIGMAS (Special Interest Group on Modelling and Simulation) at the AIS.

1 Introduction

Nowadays there exists a huge number of news providers. If an incident happens, they report this news with different headlines, words, pictures etc. This is an unavoidable situation. However, for an example the usage of certain words and pictures can also express the viewpoints of the author of a particular news story to the readers. The

professional journalists should avoid these occurrences as much as possible by only reporting the news incident without mentioning their own perspectives of it. However, it is apparent that some news sources unethically use such practises to artificially plant viewpoints in the audience. This phenomenon can be identified as *media bias* which is formally defined as below.

> "By definition, the word bias refers to showing an unjustified favouritism toward something or someone. Thus, on a very simplistic level, media bias refers to the media exhibiting an unjustifiable favouritism as they cover the news. When the media transmit biased news reports, those reports present viewers with an inaccurate, unbalanced, and/or unfair view of the world around them." (Levasseur, 2008)

Bias can be implanted in various stages of news production (Park et al., 2009). However, it can be mainly classified as *Selection Bias* and *Presentation Bias* (Groeling, 2013). Selection bias occurs when the news provider chooses an unrepresentative sample of news stories from all the possible news stories. For an example one news provider may deliberately not cover positive news events related to the government. Presentation Bias occurs when a news provider employs various ways to influence the reader with the content and the placement of the news items. The presentation bias can be incorporated into news stories in following ways (Park et al., 2009; Groeling, 2013).

- Bias by labelling
- Bias by word choice
- Bias by tone
- Bias by spin
- Bias by placement
- Bias by ordering

- Bias by photo and media
- Bias by selection of sources
- Bias by news framing
- Bias by the visual dimension
- Bias by length

As there are numerous ways to inject bias in news, a common reader will not be able to easily understand the biasness of a particular news story by himself alone as it will require critical analysis and time (Park et al., 2011). The news providers use this opportunity to manipulate the general opinion of the reader, consequently the public altogether. This is happening in media industry due to different reasons (Herman, 2003). Political views of the media owners and the demand for biased news are the foremost reasons for media bias according to the researchers (Gentzkow et al., 2015; Tully et al., 2020).

Thanks to the modern communication technologies, an online news article can be spread to millions of readers within few minutes. But if this news is biased or wrong, it can also bring great harm to people. Especially if this news article makes people angry it can lead to riots, clashes and other destructive consequences. Furthermore, biased news can change peoples' opinions and this can be deliberately used to change their viewpoints. For an example there is a claim by researchers that *Fox News* has affected the behaviour of voters during an election in the USA using this approach (DellaVigna and Kaplan, 2007). Thus, biased news can directly influence democracy. Hence, discovering methods to mitigate the media bias is valuable to the entire society.

Online news is becoming a main source of news for the people along with the imminent abundance of the internet usage (Kohut, 2013; Kennedy and Prat, 2019). For example, online news has become as popular as television news while the usage of

printed media was declining in UK (Oxford University, 2017). Further, online media has become the number one source for news for people younger than 35 years.

There have been various methods proposed to fight against bias in online news. In this paper, we wish to propose a solution in this regard based on online participation and crowdsourcing techniques. *Merriam-Webster* dictionary defines crowdsourcing as "*the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers*" (Merriam-Webster, 2021). With the advent of Web 2.0 internet users have become more active and participative in comparison with the passive usage in the past. Therefore, crowdsourcing platforms and techniques can be advantageous and yet a feasible solution to mitigate biased online news.

The structure of this paper is as follows. First, previous work related to mitigating biased online news are reviewed. Then, the online participation methods related to news sector are discussed. Next, the lessons from crowdsourcing platforms are evaluated. Subsequently the findings of this review are discussed from the viewpoint of what will bring benefits to address news bias. Based on the discussion, a novel framework is proposed to mitigate biased online news. Then the implementation and evaluation of the proposed system are discussed. Finally, some conclusions are drawn from the experience of this research study.

2   Previous work

Researchers have employed diverse methods to mitigate online news bias. This section reviews most related work.

Aggregating news from multiple sources and presenting them to the readers is a popular technique to fight against biased news. *DiversiNews* (Trampuš et al., 2015) and *NewsBird* (Hamborg et al., 2020) use automated news aggregation for this purpose. The aggregated news is presented to the readers with interactive controls like grouping, filtering and sorting. Further *DiversiNews* utilises automatic sentiment analysis which calculates the polarity (positive, neutral or negative) of each news article. The polarity value is displayed in the user interface as an indication about news bias. *NewsBird* fetches news articles from *Europe Media Monitor* (Atkinson and der Goot, 2009) and automatically group them into topics using Latent Dirichlet Allocation (LDA) method.

Social news aggregation is a non-automated technique which depends on users to manually register news in a news aggregation system. *NewsCube2.0* exploits this method to gather news on the topics created in their system along with the framing metrics (Park et al., 2011). This facilitates future incoming readers to the website to get a diverse view about topics. However user participation and management of framing spectrum were difficult and problematic according to the researchers.

There are also some attempts of using crowdsourcing for tackling news bias. Budak et al. have used *Mechanical Turk* to classify and identify biased news. The task doers (crowd) in this case have mainly done their job for the compensation they received. Therefore, they were not regular news readers who did not know much about the context of news stories. This might have affected the quality of the bias identification. In a similar vein *UnbiasedCrowd* has used not a crowd but selected activists to identify and report visual bias in news (Narwal et al., 2017). However, in this case, staffing of

activists for every single news story has been an issue and excessive manual effort is involved.

Researchers have also attempted to automate the bias recognition using text processing techniques. *Blews* system uses text processing along with link analysis to find the emotional sentiment towards liberal or conservative stance (Gamon et al., 2008). *DisputeFinder* which is a web browser extension utilises text processing to highlight disputed text in a webpage (Ennals et al., 2010).

There are some researches focusing on combining crowdsourcing with machine learning techniques to automate bias identification of news items. Vincent and Mestre have used the crowdsourcing platform *CrowdFlower* to utilise a crowd for annotating a set of selected news stories (Vincent and Mestre, 2018). Further they have also used domain experts to leverage the individual biasness of the inexperienced crowd workers. The annotated dataset was then used to train the machine learning algorithms that can be used to automate the bias detection. However, a similar research study concludes that human labelling is more suitable than automatic labelling for this task (Lazaridou et al., 2020).

Table 1 summarises the techniques used by other researchers regarding mitigating bias in news. It can be observed that most of the scholars think that providing multiple news items for a given story will help the readers to overcome the news bias. Only two examples could be found about trying to automate the detection of bias in news. Given the complexity of bias, automating seems to be an infeasible option. Even though crowdsourcing seems a very viable solution for this problem, yet few have tried to employ it. Some researchers have only utilised crowdsourcing to produce a labelled sample to train the machine learning algorithms.

Table 1   Summary of the previous work

Name of the project	Used techniques
DiversiNews	• Automatic news aggregation • Automatic sentiment analysis
NewsBird	• Automatic news aggregation
NewsCube2.0	• Social news aggregation • User framing
Budak et al.	• Crowdsourcing
UnbiasedCrowd	• Manual bias reporting
Blews	• Text processing • Link analysis
DisputeFinder	• Text processing
Vincent and Mestre	• Crowdsourcing • Expert knowledge • Machine learning
Lazaridou et al.	• Crowdsourcing • Expert knowledge • Machine learning

3   Methods of online participation

Nowadays users of websites actively interact in different ways rather than just referring the content. Online news consumption has brought forward techniques to involve the news readers as active participants such as news sharing, news aggregation, sharing opinion about news items, uploading user generated content and commenting (Fletcher and Park, 2017). Such online participation approaches are studied extensively and discussed in this section.

3.1   Wiki journalism (Opensource journalism)

Wiki journalism is mainly about collaborative editing and sharing of news articles. It is analogical to opensource software, where once a news article is published in the platform by a writer, other users can edit it. The quality of the articles is driven by the contributions of the users of the platform. The articles can be submitted by anyone without limiting to trained journalists.

Wikipedia which has been founded in 2004 can be named as the initiator of wiki journalism. However, *Wikipedia* itself is focused as an online collaborative encyclopaedia rather than merely focused on news. Therefore, *Wikipedia Foundation* has launched *WikiNews* in 2004 with the main objective of promoting citizen journalism (Bradshaw, 2007). However, *WikiNews* has not gained enough popularity even though it is still an active website. Therefore, Jimmy Wales who is the founder of *Wikipedia* created *WikiTribune* in 2017 as a platform to fight against fake news (Hern, 2017). In contrast to *WikiNews*, few reputed journalists are hired for *WikiTribune* to maintain the quality of the submitted news articles. The articles which are created by users must be approved or edited by a member of the hired staff before making them publicly accessible.

However major challenge of wiki journalism is the inability of laymen to quickly compose a quality news article (Seward, 2010). In news sector where timeliness of news is critical, writers do not get enough time to fine tune the writing style which requires coherent narrative which contrasts with factual style practised in *Wikipedia*. This can be attributed to the low attractiveness of news wikis. Otherwise, wiki journalism provides a strong basis to fight news bias. Apart from this, vandalism and authority are other main issues which exist in wiki journalism (Lih, 2004). Wiki platforms have devised techniques to fight vandalism such as limiting features for newly registered users until they pass a set time duration.

Amidst the aforementioned challenges, some of the wiki journalism projects such as *Ushahidi* (Okolloh, 2009) and *OhMyNews* (Bentley, 2008) have become successful. However, these projects have employed paid staff members to verify and finalise the news items before publishing. Another observation is that these wiki news sites focus on selected contemporary events or topics so that they could successfully attract necessary contributors from interest groups.

3.2   Social news aggregation

Common news aggregators automatically crawl news from different news providers and utilise algorithms to group, summarise and visualise news for the readers (Hamborg et al., 2020). *Google News*, *Yahoo News* and *Huffington Post* are popular examples for common news aggregation servers. On the other hand, social news aggregators employ their user

community to manually fetch and submit news items from different sources. Multiple successful social news aggregator sites have been emerged as a result of Web 2.0 (i.e., Participative and Social Web). Users of these sites can post stories (with links to source) together with their comments and tags. These posts are then ranked by the help of users and/or a set algorithm and then displayed in the front page. Stories are usually classified under subtopics (e.g., political, business, technology, sports etc.) as well (Steinbaur, 2011).

Reddit, *Digg*, *Fark* and *Slashdot* are leading social news websites at the moment. *Reddit* being the most popular website had around 330 million users in 2018 while *Digg* had 12 million monthly users in 2015 (McCormick, 2017; Engel, 2016). The substantial base of users enables these platforms to deal with large amount of news stories and to engage the labour of the crowd to filter and rank posts to be summarised and presented in their websites (Bruns, 2005).

News aggregation helps the readers to realise bias in news by providing multiple viewpoints of a news story from different sources (Trampuš et al., 2015; Hamborg et al., 2020). Social news aggregation further empowers users by actively engaging them to tag, rank and comment on news items.

3.3 User generated content (UGC)

Another way to involve the online audience in news sites is by facilitating them to contribute with the content such as images, videos, text and audio (Berthon et al., 2015). This content is defined as user generated content or user created content. Functionality to upload user generated content has been already implemented in major news providers like *BBC*, *CNN* and *TIMES Magazine*. UGC is a very good tool to showcase first-hand ground level information from the general public from places where journalists cannot access. This is quite useful during dangerous situations such as conflicts, accidents, natural disasters etc. However, due to the difficulty in the verification UGC should be used with care even though it is advantageous. There exist also legal concerns over the responsibility as well as copyrights. Therefore, usually the news provider does not take responsibility of UGC.

3.4 User comments

In some news websites, readers are given facility to write their own comments regarding the news articles. But recently some of the major news websites have cut off this feature (Green, 2018). Main reason behind this is uncivil and impolite commenting practises of the users which has made it harder to moderate (Naab and Kalch, 2017). Commenting enables the readers to communicate with the author and other interested readers. This can certainly help to resolve bias in news as it creates a platform for expressing opinions thus good to be encouraged than been restricted. Identifying new ways to help the moderation of user comments should be beneficial in this regard.

4   Lessons from crowdsourcing platforms

There are multiple knowledge production crowdsourcing platforms such as *Wikipedia*, *Everipedia*, *Scholarpedia*, *Citizendium*, *Stack Overflow*, *Yahoo Answers* and *Quora*. Yet

Wikipedia and *Stack Overflow* have been identified as major knowledge production sites as they are the most prominent crowdsourcing platforms (Matei et al., 2018). *Wikipedia* is a web-based, multilingual, free-content encyclopaedia which operates on a model of openly editable content (Wikipedia, 2021a). Currently *English Wikipedia* has about 6.2 million content articles with around 52.8 million total pages (Wikipedia, 2021a). There are about 41 million registered users and 1109 administrators. It is currently listed as the 13th most popular website in the *Alexa* rankings (Alexa Internet Inc., 2021). *Stack Overflow* is the number one question and answer site which caters a wide range of topics in computer programming. As of now, *Stack Overflow* has about 14.3 million registered users with around 265 million of monthly visits (Stack Exchange Inc., 2021a).

Based on the above facts there is no doubt about the success and popularity of above-mentioned crowdsourcing platforms. Therefore, *Wikipedia* and *Stack Overflow* platforms are further studied to learn aspects which could be beneficial for the news sector to specifically support with bias mitigation.

4.1   Lessons from Wikipedia

- *Underlying software framework*: *Wikipedia* is backed by its very powerful software called *MediaWiki*. The *Wikimedia Foundation* offers it as a free and opensource wiki engine (MediaWiki, 2021). Lot of other people use this engine to create their own wiki type websites for various purposes as well. Customisation is also possible as per the requirements of the context of use as this software is purely opensource.

- *System of moderation and conflict resolution*: In principle, an article can be submitted or modified by any user in *Wikipedia*. However sometimes user actions are restricted to confront vandalism and content disputes. *Wikipedia* has adopted a strong protection policy which specifies a system of moderation and conflict resolution in this regard. Furthermore, *Wikipedia* also has a comprehensive guide to address bias (Wikipedia, 2021b). Readers can notify administrators about biased articles through the system for them to take necessary action for resolving bias. Administrators even have the power to temporarily flag an article as *biased* until the issues are completely resolved.

- *Multilanguage support*: The *Wikimedia* Foundation has offered *Wikipedias* in different languages. As of June 2021, its articles have been created in 321 languages (WikiMedia, 2021) while having 310 active *Wikipedias* in different languages. This achievement can be also contributed to its underlying software *MediaWiki*. Therefore, *Wikipedia* model can be combined to overcome language barrier.

- *Motivation to contribute*: *Wikipedia* does not have a reputation system or provide financial benefits for its contributors. Yet it has a good set of contributing users and audience. This phenomenon can be exclusively attributed to the personal values of the users. Figure 1 displays April 2011 Wikipedia Editor Survey with top reasons for continuing to contribute (Wikipedia, 2021c). It is surprising fact to observe that 53% of its users contribute to find articles that are biased or incomplete.

Figure 1   Wikipedia editor survey – April 2011 (see online version for colours)
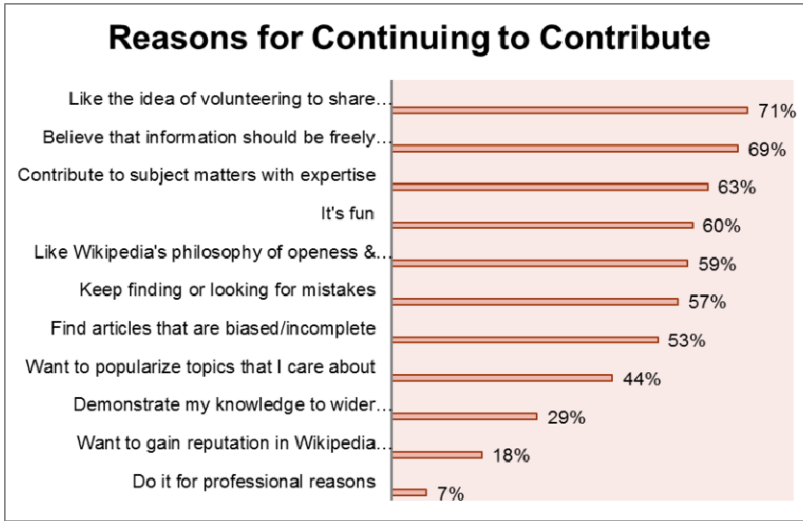
Source: Wikipedia (2021c)

4.2   Lessons from Stack Overflow

- *User reputation system*: Registered users in *Stack Overflow* mainly post questions and provide answers to them. Users can also tag questions with topics and vote/comment for both questions and answers. Users can earn points or merits for these activities they perform in *Stack Overflow* which will be considered as the *reputation*. The reputed users will get more control over the activities they can be involved in. For an example, new users cannot comment on posts until they earn the required minimum level of reputation (Stack Exchange Inc., 2021b).

- *Tagging, commenting and replying*: A question posted by a user in *Stack Overflow* analogues to a news item posted by a user in a social news website. Tagging, voting, commenting and replying can be performed for each question posted. Crowd's help can be used to tag news stories into topics or subtopics. Similarly, users can be asked to mark the biasness of a news article they read along with the reasoning through comments if necessary. However unlike *Stack Overflow*, the news articles have a time value. Therefore, the ranking algorithms should consider that as well. Replies will enrich the reader's understanding as they facilitate a dialog between concerned parties and they also provide a platform for people to share their first-hand experiences. Members can comment on both questions and answers in *Stack Overflow*. Likewise, commenting on news article itself or on each reply can be enabled. This will aid to validate each activity (news article or reply) separately by the help of crowd and thus work as crowd-based moderation.

- *Voting*: *Stack Overflow* has a voting system which is crucial for the governance of the whole site according to the principles of representative democracy. It is possible to vote up (upvote) or vote down (downvote) towards a question or answer posted in the site. Upvoting will indicate the community that a post is exciting, well-researched, and valuable, while downvoting indicates the opposite. These metrics

help the site to reward the good contributors with reputation. Reducing reputation points from users who receive downvotes sends them a message for improvement or correction. Voting is also used to rank the answers for a question and display them in a way which is easier for a reader to quickly find the best answer to a question.

• *Motivation to contribute*: In *Stack Overflow*, there is no financial incentive to the users for their contribution. However, it provides intrinsic motivation such as reputation, satisfaction and values for their contributors. For this purpose, *Stack Overflow* has devised a comprehensive user reputation system where users are assigned with points/badges to value their contribution level. This provides the necessary motivation along with additional powers of control over the platform.

5   Discussion

As per the study done in Sections 3 and 4, following mechanisms from online participation and crowdsourcing were chosen which can be incorporated when devising solutions to fight against biased online news.

• social news aggregation

• tagging

• commenting

• voting

• user reputation system.

Following sections describe these mechanisms in detail.

5.1   Social news aggregation

Many researchers believe and have already tested providing several news articles from multiple news providers to the readers as a method to minimise the effects of news bias (Park et al., 2011; Trampuš et al., 2015; Hamborg et al., 2020). Almost half of online news readers directly visit websites of broadcasters or news articles rather than searching or using social media (Kennedy and Prat, 2019). Therefore, there is a huge chance that these readers are affected by the biasedness of the limited set of news providers they follow. Hence this approach may help them to get an unbiased viewpoint of the news stories they read.

Even though wiki journalism is an option to provide different viewpoints on news stories it has not been successful so far due to various reasons. Hence, social news aggregation is a viable mechanism for this purpose as it is a well proven technique. But the crowd itself will not have enough power to handle the huge amount of common news stories emerging real-time. To overcome this situation, automated news aggregation can be also incorporated to fetch news. *NewsBird* has successfully used *NewsFeed* crawling software to automatically aggregate news (Trampuš et al., 2015; Trampuš and Novak, 2012). However social news aggregation itself will be enough in situations where only specific news stories or topics are considered.

5.2   Tagging

Crowdsourcing can be used to identify and tag biasness of news articles. The readers can be asked to rate the biasness of articles along with their comments or labels. These inputs can be utilised to generate a metric which displays the biasness of an article so that future readers will be aware of such bias. The reported biasness values can be used to devise metrics for news providers or journalists in general as well. These indicators will help the audience to get a good picture about the news providers.

The automatic text processing techniques are not yet capable of recognising the complex scenarios of different types of media bias (Hamborg et al., 2019; Lazaridou et al., 2020). Therefore, those techniques have a very limited use in bias detection and evaluation. In contrast, the crowd will be able to perform this task effectively as a collective force. However, the crowd will need some amount of time to react upon a newly popped up news item. Thus, when compared with the automated approach of bias detection, there will be a time lag.

5.3   Commenting

Empowering readers to comment will provide a good environment to mitigate bias in the news. This can be further enhanced if readers can attach user generated content in the comments section. However, there should be a proper mechanism to control vandalism and edit-wars. The already proven moderation systems which are used in crowdsourcing platforms like *Wikipedia* and *Stack Overflow* can be employed for this purpose.

5.4   Voting

Voting helps to moderate user activities with the help of the community. For an example unnecessary or unethical comments can be downvoted by the community and subsequently discouraged while good feedback is upvoted. The number of votes received by users can help to take necessary action against vandalism and unethical commenting. The voting system will encourage the good contributors while degrading or limiting bad users. Voting system can be used together with the reputation system to limit user behaviour as per the feedback they receive from the community.

Further, voting will facilitate presenting the popular news of the community at a given moment. The presentation of news items can be sorted according to the votes they receive. But sum of votes itself will not be sufficient to sort the news items because the news stories have a time-value. Therefore, there should be a mechanism to consider time factor when showing top news in the site.

5.5   User reputation system

Community driven websites entirely depend on the contribution of their users (Zhao and Zhu, 2014). Therefore, motivating the users to contribute and ensuring the quality of their submissions are main critical factors to become a successful crowdsourcing platform. The user reputation system facilitates to combat vandalism as well. The rewarding criteria must be devised skilfully to sustain the level of user participation. The mechanism which is used in *Stack Overflow* can be studied as a good example to begin with.

6   Proposed solution

Based on the analysis a high-level architecture of a web-based news platform is proposed as depicted in Figure 2. This solution uses advantageous mechanisms from both online participation and crowdsourcing to help with mitigating news bias. This platform consists of five main components named Aggregation, Presentation, Crowd Activities, User Management and Administration in a functional point of view. These components are discussed in detail below.

6.1   Aggregation

This component is responsible for fetching news items for the site. Automated news aggregation and/or social news aggregation can be used for this purpose and the aggregated news items should be stored in the database. Most of the popular websites offer news through RSS (Really Simple Syndication) feeds where third-party applications can automatically fetch in a standardised machine-readable form. The proposed platform can use RSS feeds of news providers to aggregate news from them. There are already developed free and open-source libraries like *NewsFeed* that can be used for this purpose as well (Trampuš and Novak, 2012). *NewsFeed* periodically crawls a specified set of RSS feeds and acquire links to news articles. The set of RSS feeds and the interval to refresh can be set as per the requirement. This should be done carefully not to flood readers with huge number of news articles, but to provide a sufficient combination of sources to minimise the effect of bias.

Figure 2   High-level architecture of the proposed solution
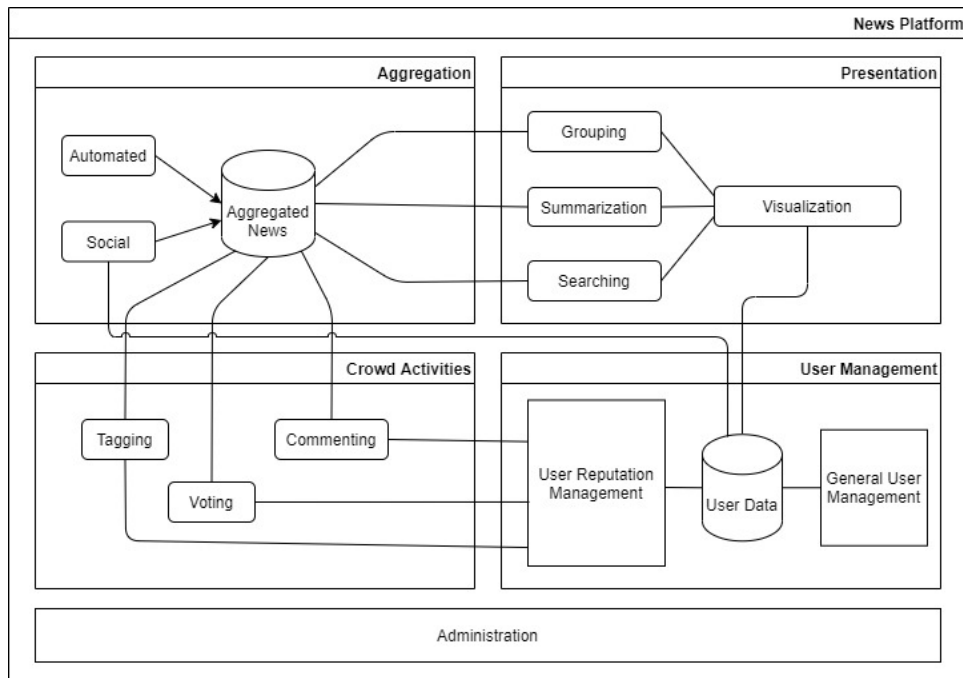
Web browser plugins (extensions) can be implemented for the social news aggregation. The users of the proposed news platform should be asked to install these plugins in their browsers. When they read a particular news article in a news website, they can use the plugin to post it to the proposed news platform. Necessary information from the news item can be automatically fetched by the browser plugin to make the process convenient and efficient. The RSS specification can be used as a standard to decide which information to collect. Example schema for a news article is displayed in Figure 3. There should be also a mechanism to avoid the duplicate entries as multiple users may submit the same news article. The unique *guid* element in RSS schema can be used to check for duplicates. In addition, when submitting articles, users can be asked to provide the bias rankings, tags and their comments. Therefore, social news aggregation component should be able to access user data.

6.2   *Presentation*

This component handles the presentation of the site. The way news stories are displayed on the website is very important as the usability of the platform will be the key factor for attracting and retaining the viewers. Grouping into topics or subtopics (e.g., Politics, Financial, Sports, Science, etc.) should be done according to the tags. There is also possibility to use topic analysis to automatically categorise news articles. So that in the main page of the site, main topics can be displayed for a user to select. The news articles inside these topics can be ranked in reverse chronological order. However, the popular articles which have more viewership should be displayed for at least a limited time for the daily users to read.

Figure 3   Example schema for a news article (see online version for colours)

```
▼<rss version="2.0">
  ▼<channel>
      <title>XYZ News Company</title>
      <link>http://xyznews.com/</link>
      <description>News from all around the world.</description>
      <language>en-us</language>
    ▼<item>
        <title>Headline of the News Article</title>
        <link>http://xyznews.com/2020/10/10newsitem1.html</link>
        <description>Short description about the news.</description>
        <pubDate>Sat, 10 Oct 2020 09:39:21 GMT</pubDate>
        <guid>http://xyznews.com/2020/10/10newsitem1.html</guid>
      </item>
    </channel>
  </rss>
```

The full news article will not be available for the site to display. Therefore, only the title, link to original article, source, author, published date and a small description will be shown. The reader should open the link to read the full article. The crowd activities must also dynamically affect the presentation of gathered news items to the users. The metrics like biasness, number of votes, views, comments, and replies also should be displayed along with the news item.

When the site is up and running, lots of data will be gathered about the news sources, topics, biasness rankings, view counts, votes etc. These collected data will provide a good

source of information about the news sector in general. Therefore, providing a sub section to display these statistics will help the audience to get a good picture about the industry. For an example, a metric like *reported biasness over the number of views* for each news source/journalist will provide a message to the audience about their biasness. So, the readers can avoid biased sources or use unbiased sources as much as possible.

Apart from the above-mentioned functionalities, a search facility should be also implemented to provide a comprehensive news searching as per the user inputs. The visualisation component, which is integrated with grouping, summarisation, searching components along with user data will create the presentation of news on the website.

6.3   Crowd activities

This component takes care of the activities of the crowd. After reading the full news article the users should be able to use the browser plugin to report bias, add comments or vote. Perceived biasness will differ for each user as it is a subjective measurement. But ultimately the cumulative figures of the whole community will act as a respectable normalised result for the biasness. This normalisation could circumvent the effect of biased users of the platform itself. A scale from 0 to 10 can be provided for the readers to submit a value for the biasness. However, when rating the biasness users should be also provided the facility to enter tags to relate with the biasness they report. For an example if the article is favoured for a certain political party, the user can enter the party name as a tag. When the biasness metric is displayed keywords can be also shown along with their occurrences. As specified in section 1, there are numerous ways bias can be injected into a news story. Therefore, a single average reader can be easily tricked. Utilising community labelling will help the readers in a collective manner to identify and mitigate bias in news items they read.

Moderation of the user comments will be handled by the community itself with the help of voting and user reputation system. Voting functionality should be available for the news article and for the individual comments with the possibility of upvoting and downvoting.

Tagging, voting and commenting components are integrated with the user reputation management module as these activities will be only available for users with different reputation levels. Further these activities will add meta data about the already aggregated news in the database. This updated information about news items will then be propagated into the presentation module.

6.4   User management

User Management contains two submodules named General User Management and User Reputation Management. General User Management must handle general maintenance of users of the news platform such as new user registration, user profile management, changing password etc. The User Reputation Management should cover the mechanism of reputation handling of the registered users. User reputation system is a key part of a crowdsourcing platform which does not provide any financial benefits to the users for their contributions. The users contribute for their intrinsic motivation like personal values and respect from the community. Therefore, rewarding good contributors with points and giving them more controls over the others motivate them to contribute more. A hierarchy of user levels should be also devised along with the badges. Criteria should also be

prepared to decide how each contributing activity should affect points of the users. Upvotes and downvotes for the comments should be also taken into consideration in this regard.

The user data should be stored in the database and shared with the social news aggregation and visualisation components.

6.5 Administration

The Administration module should support the general maintenance activities of the news platform such as configurations, basic data etc. For an example, the URL links to RSS news feeds should be entered in the system when the automatic news aggregation is used. This component should facilitate the configuration of such inputs.

7 Implementation and evaluation of the proposed system

A possible technical architecture of the proposed new platform is shown in Figure 4. The system will share a common database to store data. The News Aggregator component should automatically crawl the news from the Web and save in the database. The settings of the News Aggregator (e.g., The URLs of new sources, frequency, etc.) should be setup by an administrator using the Frontend Web Application. The Web Browser Plugin component should provide functionality for registering new users, user login and adding news articles manually to the news platform. The Web Browser Plugin will use the Backend Web API component to provide these functionalities.

Figure 4   Technical components of the proposed system
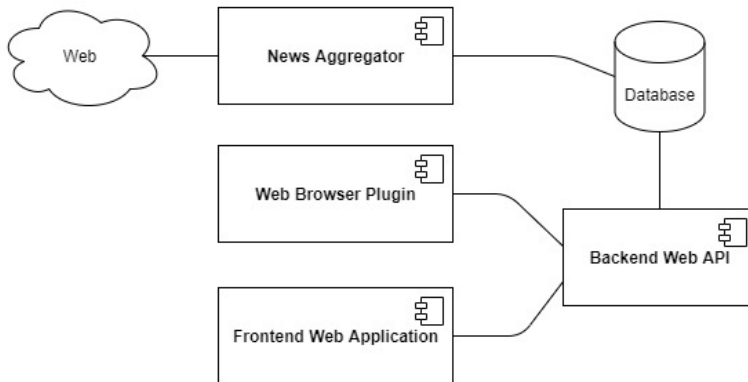
The Frontend Web Application should provide the presentation of news to the readers over the internet in the means of a public website. Unregistered users should also be able to see the news along with the metadata (i.e., biasness, tags, comments, etc.). It should also provide facility to register new members and log them in to the system. The registered users (members) should be provided additional features like reporting bias, tagging and commenting. This component should also provide administrative module to provide general administration of the news platform (e.g., basic data setup, user

management). The Backend Web API should keep the business logic and handle the communication with the database.

Scalability and the performance of the platform can be guaranteed by deploying the technical components into a cloud platform such as Amazon Web Services (AWS) or Microsoft Azure. At the user registration the potential members should be provided with a comprehensive data and privacy policy about how their personal data are saved and used which they must agree with, to become a member. This can be enriched with general terms and conditions agreement. Since this platform will become a social platform, user registration and management should be planned well to minimise anonymous or false identities (e.g., verification via mobile phone).

Next target of the authors is to provide the proof of concept in the means of a prototype of this novel news platform in a successive phase. Once the prototype is built the system should be evaluated. Volunteers must be sought to represent a good sample of the general online news readers. Then feedback of these users should be gathered to evaluate the system. Special consideration should be given to collect their experience about identifying and mitigating the biased news through the platform and how it differs from the way they read news earlier. The evaluation should assess if the platform itself would affect bias to prove the measures taken in this regard are successful. Furthermore, usability and performance of the system should be also main considerations in the evaluation.

The system can be further improved by taking the results of evaluation into consideration. As commercialisation of the product is not the intention of the authors, the source code of the final version can be shared with the public similar to Wiki software. Hence interested parties can setup the system on their own. For an example interested person/organisation can set it up for a specific country/domain. But this does not mean the news platform cannot be a global platform. Yet, that should be the ultimate goal. In the ideal case the news platform may host as a not-for-profit service like Wikipedia which depends on donations. However, it should be also possible to gain revenue through online advertisements.

8   Conclusion

News can affect the minds of individual readers and then ultimately the entire society. Even irrevocable consequences can occur when readers are fed with biased news. Therefore, mitigating biased news is important to preserve the values of democracy. This paper provides a solution in this regard from the best practices of online participation and crowdsourcing platforms.

At first previous work from scholars to fight against biased news were reviewed. Then the methods of online participation were discussed following an analysis of best practices from successful crowdsourcing platforms which can benefit news sector. All the findings of this study were then reviewed and a novel news platform has been proposed harnessing the strengths of online participation and crowdsourcing with special focus on online news bias mitigation. Main benefit of the proposed solution is the community labelling of bias utilising the news readers themselves as the crowd workers. The readers will be collectively empowered to identify the biasness of the news items they read. Additional mechanisms are proposed to support and sustain this.

The proposed solution can be used for a generic news platform or for a specific domain/topic/event. For an example, separate news platforms can be implemented for different countries which focus on local communities and news stories. Future research should be carried out to build a prototype news platform using the provided architecture as a proof of concept. Attracting a sufficient crowd is crucial as the success of a crowdsourcing platform depends on user contribution. One way to achieve this is to make the news website freely accessible for viewing news items while making additional functionality only available for registered users.

A major challenge of the proposed solution is to keep up with the time value of news. However, the previous statistics about the news source and the reporter can be displayed along with the newly emerged news items which will help the readers to get some understanding until the crowd comes up with tags and comments. Another concern is that only a link to the original news article and a small description will be shown in the news site for posted news items. The reader should open the link to go through the full news article. However, the browser extension must be developed in a user-friendly manner to overcome this usability concern.

Manipulating the system and vandalism will be controlled by the community itself with the help of user reputation system and voting. For an example, new users will not be able to comment until they earn some level of reputation, which will be powerful to avoid situations where a group of people collectively organise to make false comments overnight. Scalability will not be also an issue with the cloud-based options available nowadays. If the realised news platform is successful, it will provide a lucrative database which can be used for data mining. Therefore, as an additional benefit this will provide chances to reveal interesting patterns related to news industry.

References

Alexa Internet Inc. (2021) *The Top.500 Sites on the Web*, https://www.alexa.com/topsites (Accessed 15 June, 2021).

Atkinson, M. and der Goot, E. (2009) 'Near real time information mining in multilingual news', *Proceedings of the 18th International Conf. on World Wide Web,* Madrid, Spain, pp.1153–1154.

Bentley, C.H. (2008) 'Citizen journalism: back to the future', *Discussion Paper Presented at the Carnegie Knight Conference on the Future of Journalism,* Cambridge, MA.

Berthon, P., Pitt, L., Kietzmann, J. and McCarthy, I.P. (2015) 'CGIP: managing consumer-generated intellectual property', *California Management Review*, Vol. 57, No. 4, pp.43–62.

Bradshaw, P. (2007) *WIKI JOURNALISM Are Wikis the New Blogs?. Paper Presented at the Future of Newspapers Conference,* Cardiff, UK.

Bruns, A. (2005) Gatewatching: Collaborative Online News Production.

Budak, C., Goel, S. and Rao, J.M. (2016) 'Fair and balanced? quantifying media bias through crowdsourced content analysis', *Public Opinion Quarterly*, Vol. 80, pp.250–271.

DellaVigna, S. and Kaplan, E. (2007) 'The fox news effect: media bias and voting', *The Quarterly Journal of Economics*.

Engel, K. (2016) *Digg: Its Rise and Fall and Rebirth*, https://www.whoishostingthis.com/blog/2016/08/22/digg-rise-fall/ (Accessed 15 June, 2021).

Ennals, R., Trushkowsky, B. and Agosta, J.M. (2010) 'Highlighting disputed claims on the web', *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*, ACM, New York, NY, USA, pp.341–350.

Fletcher, R. and Park, S. (2017) 'The impact of trust in the news media on online news consumption and participation', *Digital Journalism*, Vol. 5, No. 10, pp.1281–1299.

Gamon, M., Basu, S., Belenko, D., Fisher, D., Hurst, A. and König, A. (2008) 'BLEWS: Using blogs to provide context for news articles', *Proceedings of the Second International Conference on Weblogs and Social Media, ICWSM 2008,* Seattle, WA, USA.

Gentzkow, M., Shapiro, J.M. and Stone, D.F. (2015) 'Media bias in the marketplace: theory', in Anderson, S.P., Waldfogel, J. and Stromberg, D. (Eds.): *Handbook of Media Economics*, Elsevier, North Holland, pp.623–45.

Green, M. (2018) *No Comment! Why More News Sites Are Dumping Their Comment Sections*, https://www.kqed.org/lowdown/29720/no-comment-why-a-growing-number-of-news-sites-are-dumping-their-comment-sections (Accessed 15 June, 2021).

Groeling, T. (2013) 'Media bias by the numbers: challenges and opportunities in the empirical study of partisan news', *Annual Review of Political Science*, Vol. 16, No. 1, pp.129–151.

Hamborg, F., Donnay, K. and Gipp, B. (2019) 'Automated identification of media bias in news articles: an interdisciplinary literature review', *International Journal on Digital Libraries,* Vol. 20, pp.391–415.

Hamborg, F., Meuschke, N. and Gipp, B. (2020) 'Bias-aware news analysis using matrix-based news aggregation', *International Journal on Digital Libraries, Vol. 21,* pp.129–147.

Herman, E.S. (2003) *The Propaganda Model: A Retrospective Against All Reason*, Vol. 1, pp.1–14.

Hern, A. (2017) *Wikipedia Founder to Fight Fake News with New Wikitribune Site*, https://www.theguardian.com/technology/2017/apr/25/wikipedia-founder-jimmy-wales-to-fight-fake-news-with-new-wikitribune-site (Accessed 8 March, 2021).

Kennedy, P.J. and Prat, A. (2019) 'Where do people get their news?', *Economic Policy*, Vol. 34, No. 97, January, pp.5–47.

Kohut, A.P.R. (2013) *Pew Surveys of Audience Habits Suggest Perilous Future for News*, Pew Research Center Survey Report, Pew Research Center, Washington, DC.

Lazaridou, K., Loser, A., Mestre, M. and Naumann, F. (2020) 'Discovering biased news articles leveraging multiple human annotations', *Proceedings of the 12th Conference on Language Resources and Evaluation*, May*,* Marseille, France, pp.1268–1277.

Levasseur, D.G. (2008) *Media Bias*', *Encyclopedia of Political Communication*, Lynda Lee Kaid, Sage Publications.

Lih, A. (2004) *Wikipedia as Participatory Journalism: Reliable Sources? Metrics for Evaluating Collaborative Media as a News Source*, Nature.

Matei, S.A., Abu Jabal, A. and Bertino, E. (2018) 'Social-collaborative determinants of content quality in online knowledge production systems: comparing Wikipedia and stack overflow', *Soc. Netw. Anal. Min.*, Vol. 8, p.36.

McCormick, R. (2017) *Reddit Overhauls Its Front Page for New Users and Lurkers*, https://www.theverge.com/2017/2/15/14632390/reddit-front-page-popular-change-new-users (Accessed 15 June, 2021).

MediaWiki (2021) *MediaWiki*, https://www.mediawiki.org/wiki/MediaWiki (Accessed 15 June, 2021).

Merriam-Webster (2021) *Crowdsourcing*, https://www.merriam-webster.com/dictionary/crowd sourcing (Accessed 15 June, 2021).

Naab, T. and Kalch, A. (2017) *Replying, Evaluating, Flagging: How Users Engage With Uncivil and Impolite Comments on News Sites, Paper Presented at the Annual Conference of the International Communication Association,* San Diego, USA.

Narwal, V., Salih, M.H., Lopez, J.A., Ortega, A., O'Donovan, J., Höllerer, T. and Savage, S. (2017) 'Automated assistants to identify and prompt action on visual news bias', *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*, ACM, New York, NY, USA, pp.2796–2801.

Okolloh, O. (2009) 'Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information', *Participatory Learning and Action*, Vol. 59, pp.65–70.

Oxford University (2017) *Where Do People Get Their News?*, https://medium.com/oxford-university/where-do-people-get-their-news-8e850a0dea03 (Accessed 15 June, 2021).

Park, S., Kang, S., Chung, S. and Song, J. (2009) 'NewsCube: delivering multiple aspects of news to mitigate media bias', *Proceedings of CHI '09: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Boston, MA, USA April, pp.443–452.

Park, S., Ko, M., Kim, J., Choi, H. and Song, J. (2011) 'NewsCube2.0: an exploratory design of a social news website for media bias mitigation', *Paper Presented at the 2nd International Workshop on Social Recommender Systems*, 19–23 March, Hangzhou, China.

Seward, Z.M. (2010) *Why Wikipedia Beats Wikinews as a Collaborative Journalism Project*, https://www.niemanlab.org/2010/02/why-wikipedia-beats-wikinews-as-a-collaborative-journalism-project/ (Accessed 15 June, 2021).

Stack Exchange Inc. (2021a) *About – Stack Exchange*, https://stackexchange.com/about (Accessed 15 June, 2021).

Stack Exchange Inc. (2021b) *Help Center – Stack Overflow*, https://stackoverflow.com/help (Accessed 15 June, 2021).

Steinbaur, T. (2011) *Information and Social Analysis of Reddit*, http://snap.stanford.edu/class/cs224w-2011/proj/tbower_Finalwriteup_v1.pdf (Accessed 15 June, 2021).

Trampuš, M. and Novak, B. (2012) 'Internals of an aggregated web news feed', *Paper Presented at the Fifteenth International Multiconference on Information Society*, 8–12 October, Ljubljana, Slovenia.

Trampuš, M., Fuart, F., Pighin, D., Štajner, T., Berčič, J., Novak, B., Rusu, D., Stopar, L. and Grobelnik, M. (2015) 'DiversiNews: surfacing diversity in online news', *AI Magazine*, Vol. 36, No. 4, December, pp.87–104.

Tully, M., Vraga, E. and Smithson, A. (2020) *News Media Literacy, Perceptions of Bias, and Interpretation of News Journalism*. Vol. 21, pp.209–226.

Vincent, E. and Mestre, M. (2018) 'Crowdsourced measure of news articles bias: assessing contributors' reliability, *Proceedings of the 1st Workshop on Subjectivity, Ambiguity and Disagreement in Crowdsourcing, and Short Paper Proceedings of the 1st Workshop on Disentangling the Relation Between Crowdsourcing and Bias Management*, pp.1–10.

WikiMedia (2021) *List of Wikipedias*, https://meta.wikimedia.org/wiki/List_of_Wikipedias (Accessed 15 June, 2021).

Wikipedia (2021a) *Wikipedia: About*, https://en.wikipedia.org/wiki/Wikipedia: About (Accessed 15 June, 2021).

Wikipedia (2021b) Wikipedia: Guide to Addressing Bias, https://en.wikipedia.org/wiki/Wikipedia: Guide_to_addressing_bias (Accessed 15 June, 2021).

Wikipedia (2021c) Wikipedia Community, https://en.wikipedia.org/wiki/Wikipedia_community (Accessed 15 June, 2021).

Zhao, Y. and Zhu, Q. (2014) *Evaluation on Crowdsourcing Research: Current Status and Future Direction, Information Systems Frontiers*, Vol. 16, pp.417–434.