Google Snippets and Twitter Posts; Examining Similarities to Identify Misinformation

Saud Althabiti[1,2], Mohammad Ammar Alsalka[3] and Eric Atwell [4]

[1] *University of Leeds, scssal@leeds.ac.uk, Leeds, United Kingdom*

[2] *King Abdulaziz University, salthabiti@kau.edu.sa, Jeddah, Saudi Arabia*

[3] *University of Leeds, m.a.alsalka@leeds.ac.uk, Leeds, United Kingdom*

[4] *University of Leeds, e.s.atwell@leeds.ac.uk, Leeds, United Kingdom*

Abstract

Despite numerous efforts to address the persistent issue of fake news, its proliferation continues due to the vast volume of information circulating on social media platforms. This poses a significant challenge to manual fact-checking processes. To explore a potential solution, this study investigates the applicability of Google search and its results as a practical tool for detecting fake news on platforms like Twitter. The research focuses explicitly on comparing Google search result snippets with tweets to assess their similarity and determine if such similarity can serve as an indicator of misinformation. However, the study reveals that the observed similarity between tweets and snippets does not necessarily correlate with news credibility. Consequently, alternative techniques, such as retrieving complete news articles and assessing sources, may be necessary to effectively tackle the challenge of fake news detection on social media. This research spots light on the limitations of relying solely on snippet similarity. In addition, it suggests the importance of considering comprehensive content analysis and source credibility in future works to combat misinformation.

Keywords

Misinformation detection, Google snippets, Automatic fact checking, Cosine similarity, Sentence similarity, sBERT.

1. Introduction

Many individuals use social networks site to obtain news and information. These platforms, like Twitter, Facebook, and others, have become popular news and information sources for many people [1]–[4]. This is because these media allow users to easily access and share information and often offer real-time updates on events and issues [5]. Further, many news organizations and journalists use social media to disseminate their stories and updates, which makes it easy for people to access news and information on these platforms [6]. However, it is important to note that not all information on these media is trustworthy [7].

Misleading information, also known as "fake news," is a type of information presented as factual, but is actually false or intentionally deceptive. It is often circulated on social media platforms and can be challenging to differentiate from legitimate news sources [8]. This can be detrimental, as it can lead individuals to make decisions based on inaccurate information [5]. Therefore, it is vital for individuals to evaluate the information they come across on social media critically and to verify its accuracy from multiple sources before acknowledging it or sharing it with others.

There have been considerable efforts by social media platforms, researchers, and other organizations to enhance the quality of information on the internet and reduce the spread of misleading information

[2]. For instance, many networking sites have implemented policies and algorithms to detect and remove false content. In addition, they often provide users with tools to flag or report this type of content [1]. Besides, researchers and fact-checking organizations have evolved techniques for detecting and debunking incorrect information. Furthermore, they often work with social media platforms to assist them in identifying and removing such content [9]–[12]. Also, many initiatives and organizations focus on educating people on how to evaluate the information they encounter online critically [13], so there are considerable efforts underway to improve the online information environment and reduce the spread of false or misleading information.

Nevertheless, plenty of deliberately fabricated or manipulated rumors have been propagated on the internet, raising concern that many naive users believe them without checking their authenticity [14]. Individuals and organizations use this common way to spread propaganda, promote political agendas, or simply cause confusion [15], [16]. Unfortunately, in various cases, this type of information can be challenging to detect because it is often designed to look like legitimate news or information, and it may be shared by seemingly reputable sources [2]. This is why it's essential for users to be critical of the information they encounter online and to verify its accuracy from multiple sources before accepting it as accurate and making a decision or drawing a conclusion.

One way to manually fact-check information is to use a search engine, such as Google, to look for information or sources that support or contradict the posted information to evaluate the credibility of a source or piece of information, see Figure 1.
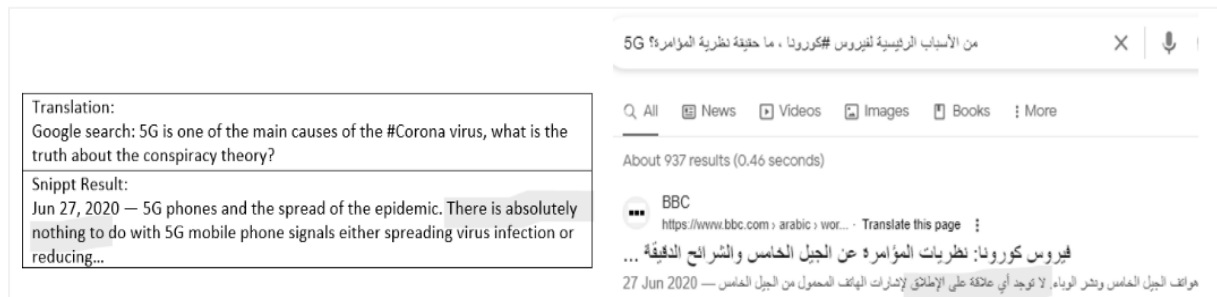

Figure 1. A tweet that is manually fact-checked using Google results.

This experiment's main objective is to ascertain the effectiveness of Google's search for detecting misinformation published on social networking sites and to develop a feature-based approach for comparing Google's results (snippets) to tweets to determine their similarity.

Our contributions comprise two main components. First, we present the first Arabic snippets dataset, which was created by collecting snippets from Google search results based on a previously published tweets dataset. Second, we introduce a new methodology for evaluating the similarity between tweets and snippets extracted from Google.

The remainder of the paper is organized as follows. First, we present related works in Section 2 and explain our methodology in Section 3. While Section 4 discussed and analyzed observations. Finally, we conclude our work in Section 5.

2. Related work

Several techniques have been used to detect fake news on social media platforms. These include utilizing machine learning algorithms to classify patterns in the language and style of the content and checking the information source to verify its credibility. Other methods are to check fact-checking organizations to verify the accuracy of the information or use community feedback and reporting tools to flag false content [2], [17]–[21]. Additionally, some platforms, such as Facebook, have developed measures like alerting labels or reduced visibility for content that was identified as misleading or potentially untrusted [4].

An essential step in the process of detecting fake news is extracting features. It includes determining the data's most relevant and informative characteristics that can be used to differentiate between real and fake news. These features can then be fed to a classification model for training and prediction [22],

[23]. For instance, some studies have used NLP techniques to extract features such as the existence of particular words or phrases and the sentiment or emotion expressed in the text [24]–[26]. These features can be combined with other content-based and source-based features to improve the accuracy of fake news detection models. Also, some researchers have used machine learning and transformer-based algorithms to examine the responses or reactions to a particular post to assess its credibility [27], [28]. For example, a posted tweet that has many negative comments may be more likely to be fake news. Similarly, a widely conveyed and discussed tweet is more likely genuine.

Some experimenters have turned to approaches such as comparing the news to other sources of information. One way to do this is to search for a given story or news on Google and see if multiple reputable sources have reported it. Based on the study [29], it is suggested that using the cosine similarity score, which is calculated after conducting topic modeling on a collection of texts, can improve the accuracy of classification. In one such process, the cosine similarity was calculated between headlines and contents as a new feature by comparing normalized TF-IDF vectors in Investigation [30]. In another experiment by [31], they developed a method for translating news titles into multiple languages, searching for related articles using the Google Search API, and evaluating the similarity between the initial news and the search results. Likewise, a study [32] proposed a similar approach involving a comparison of user-submitted articles with those from reliable sources. Also, in a paper published by [33], they tested their system on a set of 100 news and found that a matching value of at least 70 percent for more than three articles indicated credibility. While study [34], for example, calculated and used the highest similarity score and its associated title to be presented to users so they could compare and check credibility with the most related source. On the other hand, we aim to assess the effectiveness of Google's search results, which contains sources, titles, and snippets, in detecting misinformation on social media by comparing extracted snippets to tweets and evaluating their similarity. Table 1 briefly compares these studies to our proposed methodology. The news articles column represents both the title and the contents of the used dataset, while snippets are the results from Google searches. To the best of our knowledge, this is the first study investigating the efficacy of similarity between Google snippets and tweets to detect fake news.

Table 1.

Comparing studies used cosine similarity and Google search to detect fake news.

Study	Social media	News articles Titles	News articles Contents	Snippets	Extracted from	Vector representation
[29]		✓	✓		Google search	TF-IDF
[30]		✓	✓		--	TF-IDF
[31]		✓	✓		Google search	mBERT
[32]			✓		Reliable sources.	TF-IDF
[33]			✓		Google search	N gram & TF*IDF
[34]		✓			News website	TF-IDF
Our work	✓			✓	**Google search**	**sBERT**

3. Methodology

3.1. Tweets dataset

Since our primary goal in this study is to compare news from social media with Google snippets, we used a published dataset from Twitter. Specifically, we chose the ArCOV19-Rumors dataset, which contains Arabic Twitter posts about COVID-19 misinformation [35]. It includes 138 verified claims, mostly from reliable fact-checking online sources, and 9.4K tweets that relate to those claims. The tweets have been annotated with information about their truthfulness in order to support research on detecting false information. The collection covers the period from January 27 to April 2020. We used this dataset to collect similar news from Google as described in Subsection 3.3.

3.2. Sentence and cosine similarity

A Sentence similarity is a degree to which two texts are semantically alike or equal in meaning [36]. In contrast, cosine similarity is a computation of similarity between two non-zero vectors of an inner product space that calculates the cosine of the angle between them [37]. In natural language processing, cosine similarity is often used to measure the similarity between a pair of texts by treating each piece of text as a vector [38].

sBERT is a variant of BERT (Bidirectional Encoder Representations from Transformers) [39]. It is specifically developed to process and understand the meaning of individual sentences rather than entire documents or paragraphs. This makes it particularly useful for experiments that need a deep understanding of the meaning and context of individual sentences. The sBERT model can be fine-tuned for specific NLP tasks by training it on labelled data. It has been shown to perform well on a variety of NLP tasks, including sentiment analysis, text classification, and language translation [39], [40].

3.3. System description and collecting snippets

We utilized the ArCOV19-Rumors and followed the following steps to conduct our experiment:

Pre-processing: Due to the existence of URLs embedded within tweets, a google search does not provide any results when searching for a particular tweet. This necessitates the removal of unwanted data, such as URLs, usernames, and hashtags. However, it yielded in the existence of many repeated tweets after preprocessing, so we also removed duplicates. The reason is that the dataset was collected based on 138 claims from Twitter, so many similar tweets were collected with different hashtags and usernames. As a result, only 2821 tweets are included in the experiment after this step.

Google search: The second step is to harvest additional data. We automatically query each tweet in the Google search engine to scrape results using requests-HTML. Each extracted output contains the website title, the link or source, and a snippet, which is the text that appears in Google's search results to describe a website's content quickly and briefly. This query returned 2267 responses, so we only experimented with tweets that have responses retrieved.

Machine translation: We aim to investigate various open-source libraries, some of which do not support the Arabic language, such as spaCy. Therefore, we automatically translated all tweets and the retrieved titles and snippets of each query using the Google translator library. In this paper, we refer to the translated tweet from Arabic to English as *ET* and the translated Google snippets as *ES*.

Calculating similarities: We utilized the pre-trained model sBERT to compute the embedding of an English tweet (*ET*) and English snippets (*ES*). Then, calculate the cosine similarity between two embeddings (*ETi* and *ESij*), where *i* represents a particular tweet and *j* represents a retrieved snippet queried by tweet *i*. In addition, we also calculate the sentence similarity between *ETi* and *ESij* using spaCy. As a result, the calculated mean value of the computed similarities between each tweet and their corresponding snippets , as in (1), are considered new features.

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^{n} \cos(ET_i, ES_{ij}) \tag{1}$$

The final dataset named 'TweetsWithSnippets' can be found on GitHub[2], which includes the following columns:
- Tweet ID
- Label: True or False tweet.
- Tweet text: Original tweet from the ArCOV19-Rumors.
- Cleaned text: Tweet text after pre-processing.
- Snpt_titles: A list of titles retrieved from Google results in both languages.
- Snpt_links: A list of URLs retrieved from Google results.
- Snpt: A list of snippets retrieved from Google results in both languages.

---

[2] https://github.com/althabiti/TweetsWithSnippets

- SenSi_ET-ES: A list of the calculated sentence similarity between *ET* and *ES*.
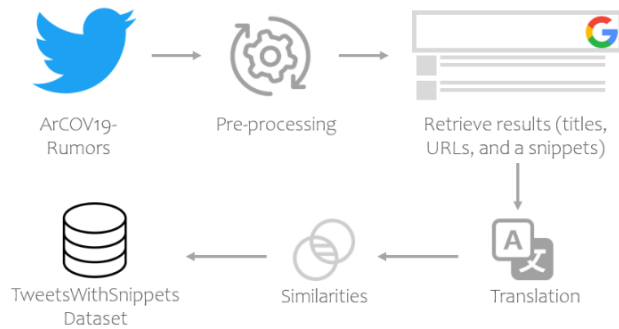- CoSi_ET-ESTxt: A list of the calculated cosine similarity between *ET* and *ES*

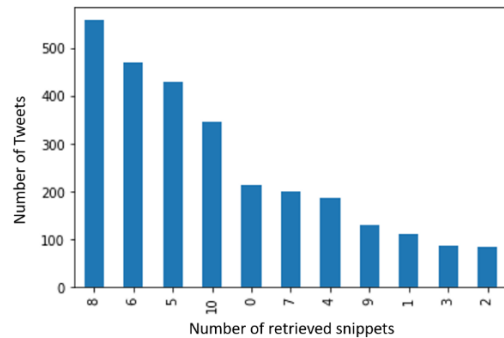**Figure 2.** Dataset collection process

Figure 3. Retrieved snippets per tweet

4. Results and discussion

After calculating the average similarity between *ET* and *ES*, we analyzed whether it was possible to take advantage of these extracted features to verify the credibility of the news. As shown to us in the Figure below, which shows the average cosine similarities and sentence similarities between the True news and their snippets in blue line, as well as in red for false news. However, upon examining Figure 4and Figure 5, we observed mixed fluctuating lines in both the true and fake news samples. There is no clear indication that one group consistently displayed higher or lower similarities than the other. Therefore, it cannot be concluded that the similarities between a tweet and related snippets can be used to predict the credibility of news. Further research is required to explore alternative approaches to detect fake news.
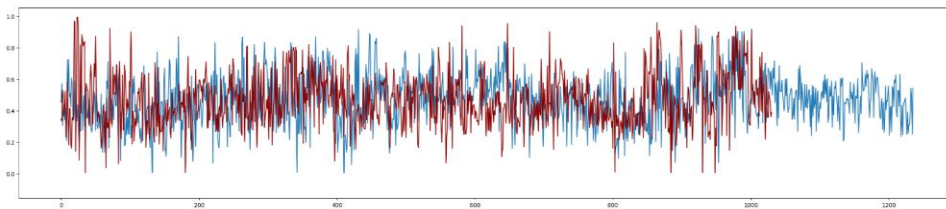

Figure 4. Cosine similarities averages. The x-axis represents samples (True in blue; False in red)
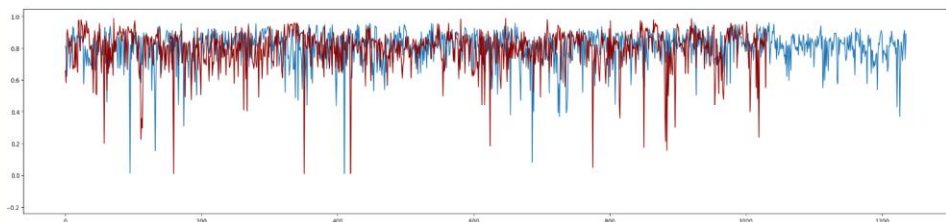

Figure 5. Sentence similarities averages. The x-axis represents samples (True in blue; False in red)

Following that, we manually analyzed several examples to ascertain the reasons for the dissimilarity in many cases by categorizing whether the snippet conveys the same information as the searched tweet. We found that cosine helps to detect if it is talking about the same matter but cannot determine the semantic similarity or the entire meaning of a sentence. Therefore, based on the randomly selected examples, 50% and above may indicate they are discussing the same topic. It is possible to benefit from those snippets to proceed with the experiment. So, we eliminated all snippets with less than 50% cosine. In addition, snippets with high similarity percentages indicate that the tweet is a copy of elsewhere content or vice versa, as shown in Table 2. The following Tables provide examples of true and false

tweeted information, searched Google snippets, sentence similarity using the spaCy (SS), cosine similarity (CS), and whether each snippet conveys and relates to the searched tweet.

Table 2.

Three examples of the samples we manually analyzed.

---

label: False

Arabic Tweet: هل يُساعد تناول الثوم في منع الإصابة بـ كورونا ؟ الدواء بلا حدود نعم يساعد تناول الثوم، في منع الأصابة بالفيروس كورونا، لتقوية المناعة في جسم الأنسان

English Translation: Does eating garlic help prevent corona infection? Medicine without limits Yes, eating garlic helps prevent infection with the Corona virus, to strengthen immunity in the human body

Snippets from Google	SS	CS	Our analyses
مكافحة كورونا: كشفت دراسات علمية عديدة أن تناول الثوم بشكل يومي يمكنه مكافحة مختلف أعراض نزلات البرد ومكافحة فيروس كورونا المستجد بنسبة 63%.'	0.78	**0.69**	**Related, agrees**
'هل يمكن أن يساعد تناول الثوم في منع الإصابة بكورونا الجديد؟ الثوم غذاء صحي قد يحتوي على بعض الخصائص المضادة للميكروبات. ومع ذلك، لا يوجد أي دليل …	0.9	**0.89**	**Related, disagree**
'تقديم الخدمات المساندة للمؤسسة. الدولة، تجسيداً للتلاحم بين أفراد المجتمع والجهات. الرسمية في مواجهة أزمة فيروس كورونا (كوفيد-١٩).'	0.83	0.3	Not related

label: True

Tweet text: وزارة الصحة تنفي مطالبتها بإغلاق المنافذ الحدودية مع ايران

English Translation: The Ministry of Health denies its demand to close the border crossings with Iran

Snippets from Google	SS	CS	Our analyses
نفت وزارة الصحة، اليوم الخميس، مطالبتها بإغلاق المنافذ الحدودية مع إيران. Feb وقال المكتب الإعلامي في الوزارة لوكالة الأنباء العراقية (واع): …	0.89	**0.68**	**Related**, agree
يذكر ان وزارة الصحة الايرانية سجلت اصابات بفايروس كورونا ادت الى وفاة شخصين لغاية الان، فيما اتخذت السلطات الايرانية الاجراءات الفورية لمنع انتشار الفايروس.	0.91	0.43	Not related
نفت وزارة الصحة، اليوم الخميس، مطالبتها بإغلاق المنافذ الحدودية مع إيران. وقالت الوزارة في بيان تابعه لـ ديجتال ميديا ": ان "الانباء...	0.92	**0.69**	**Related**, agree
ودعت الوزارة الى "توخي الدقة في نقل المعلومة واستسقائها من المصادر الرسمية". … يذكر ان وزارة الصحة الايرانية سجلت اصابات بفيروس	0.87	0.38	Not related
نفت وزارة الصحة، اليوم الخميس، مطالبتها بإغلاق المنافذ الحدودية مع إيران. وذكر المكتب الإعلامي في الوزارة لوكالة (واع) شبه ...	0.91	**0.69**	**Related**, agree

label: False

Tweet text: إلتقطت هذه الصورة صباح اليوم الجمعة لطائرة الخطوط الاماراتية مقلعه من دبي إلى بكين الصين وعلى متنها راكب واحد فقط يحمل الجنسية اليمنيه رايح يشتري بضاعه عرطه

English translation: This picture was taken this Friday morning of an Emirates Airlines plane taking off from Dubai to Beijing, China, with only one passenger on board of Yemeni nationality, going to buy Arta goods.

Snippets from Google	SS	CS	Our analyses
إلتقطت هذه الصورة صباح اليوم الجمعة لطائرة الخطوط الاماراتية مقلعه من دبي إلى بكين ... وعلى متنها راكب واحد فقط يحمل الجنسية اليمنيه رايح يشتري بضاعه عرطه.	0.98	**0.98**	**Copied tweet**
إلتقطت هذه الصورة صباح اليوم الاحد لطائرة الخطوط الاماراتية مقلعه من دبي إلى بكين (الصين) وعلى متنها راكب واحد فقط يحمل الجنسية اليمنيه رايح يشتري بضاعه ...	0.98	**0.94**	**Copied tweet**

Additional samples have been considered and manually explored. The following cases we found in most examples seem to be the main challenges preventing this experiment from taking advantage of the extracted features.

- Not all tweets convey a story, such as "The fact that an Egyptian young man died of Corona in China via news".
- Many tweets are talking about the same claim. Because the dataset is built based on 18 claims, they searched for tweets that talked about the same or negated the claim.

● An unclear tweet that is challenging to be compared with Google snippets. For example, the tweet is a question, which is meant to negate specific news, such as " Did you believe the rumor of the evacuation of Oman for Yemeni students?"
● The result is a copied tweet.
● A tweet might convey more than one claim. In some cases, one claim is correct and the other one is false.
● Snippets agree with Fake once. Although we hypothesized that snippets should agree with correct information and disagree with rumors, it was not always the case in many examples.

5. Conclusion and future work

Although there are many studies to combat fake news, the problem still exists, as the large volume of information on social media makes it challenging to fact-check such news manually. Our aim in this research is to study Google search and results if it can serve as a valuable tool in detecting misinformation on Twitter and similar platforms. The experiment involved comparing snippets, brief summaries of a web page that appear below its URL and title in Google Search, with tweets to check their similarity, and our research question was whether the similarity between a tweet and the retrieved snippets could be used as a feature to help detect fake news. The study found that the similarity between tweets and snippets does not indicate a clear difference. Thus, this approach is not enough to predict news credibility. We also manually examined random samples to find possible reasons discussed in the results and discussion section. Additional research is required to investigate alternative methods for extracting more features. One possibility is to retrieve complete news articles and their sources, instead of relying only on brief descriptions or snippets. Moreover, we aim to explore different approaches rather than relying exclusively on similarities. For instance, one potential method is to extract abstractive summaries from the newly gathered news articles and assess their impact as supplementary information for detecting misinformation.

Acknowledgements

References

[1] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of economic perspectives*, vol. 31, no. 2, pp. 211–236, 2017.
[2] X. Zhou and R. Zafarani, "Fake News: a survey of research, Detection Methods, and Opportunities," *ACM Comput Surv*, pp. 1–40, 2018.
[3] S. Sommariva, C. Vamos, A. Mantzarlis, L. U.-L. Đào, and D. Martinez Tyson, "Spreading the (fake) news: exploring health messages on social media and the implications for health professionals using a case study," *Am J Health Educ*, vol. 49, no. 4, pp. 246–255, 2018.
[4] N. Krishnan, J. Gu, R. Tromble, and L. C. Abroms, "Research note: Examining how various social media platforms have responded to COVID-19 misinformation," *Harvard Kennedy School Misinformation Review*, vol. 2, no. 6, pp. 1–25, 2021.
[5] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD explorations newsletter*, vol. 19, no. 1, pp. 22–36, 2017.
[6] A. Hermida, "Social media and journalism," *The Sage handbook of social media*, pp. 497–511, 2018.

[7]   M. Viviani and G. Pasi, "Credibility in social media: opinions, news, and health information—a survey," *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 7, no. 5, p. e1209, 2017.

[8]   X. Zhou and R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," *ACM Comput Surv*, 2020, doi: 10.1145/3395046.

[9]   R. Oshikawa, J. Qian, and W. Y. Wang, "A survey on natural language processing for fake news detection," in *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*, 2020.

[10]   X. Zeng, A. S. Abumansour, and A. Zubiaga, "Automated fact-checking: A survey," *Lang Linguist Compass*, vol. 15, no. 10, p. e12438, 2021.

[11]   S. Kumar, S. Kumar, P. Yadav, and M. Bagri, "A Survey on Analysis of Fake News Detection Techniques," in *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021*, 2021. doi: 10.1109/ICAIS50930.2021.9395978.

[12]   M. K. Elhadad, K. Fun Li, and F. Gebali, "Fake News Detection on Social Media: A Systematic Survey," in *2019 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, PACRIM 2019 - Proceedings*, 2019. doi: 10.1109/PACRIM47961.2019.8985062.

[13]   N. M. Lee, "Fake news, phishing, and fraud: a call for research on digital media literacy education beyond the classroom," *Commun Educ*, vol. 67, no. 4, pp. 460–466, 2018.

[14]   S. K. Uppada, K. Manasa, B. Vidhathri, R. Harini, and B. Sivaselvan, "Novel approaches to fake news and fake account detection in OSNs: user social engagement and visual content centric model," *Soc Netw Anal Min*, vol. 12, no. 1, p. 52, 2022.

[15]   Y. M. Rocha, G. A. de Moura, G. A. Desidério, C. H. de Oliveira, F. D. Lourenço, and L. D. de Figueiredo Nicolete, "The impact of fake news on social media and its influence on health during the COVID-19 pandemic: A systematic review," *J Public Health (Bangkok)*, pp. 1–10, 2021.

[16]   S. A. Khan, M. H. Alkawaz, and H. M. Zangana, "The use and abuse of social media for spreading fake news," in *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, IEEE, 2019, pp. 145–148.

[17]   "Fake News Detection on Social Media A Data Mining Perspective," *International Journal of Innovative Technology and Exploring Engineering*, 2020, doi: 10.35940/ijitee.i7098.079920.

[18]   Z. Guo, M. Schlichtkrull, and A. Vlachos, "A survey on automated fact-checking," *Trans Assoc Comput Linguist*, vol. 10, pp. 178–206, 2022.

[19]   X. Zhou and R. Zafarani, "Fake News: a survey of research, Detection Methods, and Opportunities," *ACM Comput Surv*, 2018.

[20]   S. Althabiti, M. Alsalka, and E. Atwell, "SCUoL at CheckThat! 2021: An AraBERT model for check-worthiness of Arabic tweets," in *CEUR Workshop Proceedings*, 2021.

[21]   S. Althabiti, M. A. Alsalka, and E. Atwell, "SCUoL at CheckThat! 2022: fake news detection using transformer-based models," in *CEUR Workshop Proceedings*, CEUR Workshop Proceedings, 2022, pp. 428–433.

[22]   G. Jardaneh, H. Abdelhaq, M. Buzz, and D. Johnson, "Classifying Arabic tweets based on credibility using content and user features," in *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology, JEEIT 2019 - Proceedings*, 2019. doi: 10.1109/JEEIT.2019.8717386.

[23]   E. A. Hassan and F. Meziane, "A survey on automatic fake news identification techniques for online and socially produced data," in *Proceedings of the International Conference on Computer, Control, Electrical, and Electronics Engineering 2019, ICCCEEE 2019*, 2019. doi: 10.1109/ICCCEEE46830.2019.9070857.

[24]   X. Zhang, J. Cao, X. Li, Q. Sheng, L. Zhong, and K. Shu, "Mining dual emotion for fake news detection," in *Proceedings of the Web Conference 2021*, 2021, pp. 3465–3476.

[25]   M. A. Alonso, D. Vilares, C. Gómez-Rodríguez, and J. Vilares, "Sentiment analysis for fake news detection," *Electronics (Switzerland)*. 2021. doi: 10.3390/electronics10111348.

[26]   B. Bhutani, N. Rastogi, P. Sehgal, and A. Purwar, "Fake News Detection Using Sentiment Analysis," in *2019 12th International Conference on Contemporary Computing, IC3 2019*, 2019. doi: 10.1109/IC3.2019.8844880.

[27]   S. S. Alanazi and M. B. Khan, "Arabic Fake News Detection In Social Media Using Readers' Comments: Text Mining Techniques In Action," *INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND NETWORK SECURITY*, 2020.

[28]   S. Althabiti, M. A. Alsalka, and E. Atwell, "Detecting Arabic Fake News on Social Media using Sarcasm and Hate Speech in Comments".

[29]   S. Mhatre and A. Masurkar, "A hybrid method for fake news detection using cosine similarity scores," in *2021 International Conference on Communication information and Computing Technology (ICCICT)*, IEEE, 2021, pp. 1–6.

[30]   A. P. S. Bali, M. Fernandes, S. Choubey, and M. Goel, "Comparative performance of machine learning algorithms for fake news detection," in *International conference on advances in computing and data sciences*, Springer, 2019, pp. 420–430.

[31]   D. Dementieva and A. Panchenko, "Fake news detection using multilingual evidence," in *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, 2020, pp. 775–776.

[32]   S. H. Long and M. P. Bin Hamzah, "Fake news detection," in *Computational Science and Technology*, Springer, 2021, pp. 295–303.

[33]   S. D. Samantaray and G. Jodhani, "Fake news detection using text similarity approach," *International Journal of Science and Research (IJSR)*, vol. 8, no. 1, pp. 1126–1132, 2019.

[34]   B. Al Asaad and M. Erascu, "A tool for fake news detection," in *2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, IEEE, 2018, pp. 379–386.

[35]   F. Haouari, M. Hasanain, R. Suwaileh, and T. Elsayed, "ArCOV19-rumors: Arabic COVID-19 twitter dataset for misinformation detection," *arXiv preprint arXiv:2010.08768*, 2020.

[36]   M. T. R. Laskar, X. Huang, and E. Hoque, "Contextualized embeddings based transformer encoder for sentence similarity modeling in answer selection task," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 2020, pp. 5505–5514.

[37]   P. Sunilkumar and A. P. Shaji, "A survey on semantic similarity," in *2019 International Conference on Advances in Computing, Communication and Control (ICAC3)*, IEEE, 2019, pp. 1–8.

[38]   V. Zhelezniak, A. Savkov, A. Shen, and N. Y. Hammerla, "Correlation coefficients and semantic textual similarity," *arXiv preprint arXiv:1905.07790*, 2019.

[39]   N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," *arXiv preprint arXiv:1908.10084*, 2019.

[40]   F. Feng, Y. Yang, D. Cer, N. Arivazhagan, and W. Wang, "Language-agnostic bert sentence embedding," *arXiv preprint arXiv:2007.01852*, 2020.