The CLEF-2023 CheckThat! Lab: Checkworthiness, Subjectivity, Political Bias, Factuality, and Authority

Alberto Barrón-Cedeño[1(✉)], Firoj Alam[2], Tommaso Caselli[3], Giovanni Da San Martino[4], Tamer Elsayed[5], Andrea Galassi[1], Fatima Haouari[5], Federico Ruggeri[1], Julia Maria Struß[6], Rabindra Nath Nandi[7], Gullal S. Cheema[8], Dilshod Azizov[9], and Preslav Nakov[9]

[1] Università di Bologna, Bologna, Italy
a.barron@unibo.it
[2] Qatar Computing Research Institute, HBKU, Ar-Rayyan, Qatar
[3] University of Groningen, Groningen, The Netherlands
[4] University of Padova, Padua, Italy
[5] Qatar University, Doha, Qatar
[6] University of Applied Sciences Potsdam, Potsdam, Germany
[7] BJIT Limited, Dhaka, Bangladesh
[8] TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany
[9] Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates

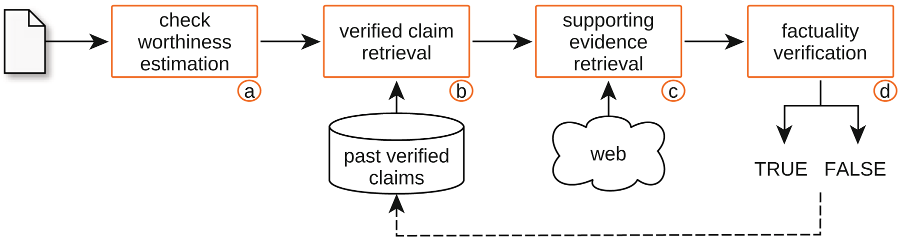
Abstract. The five editions of the CheckThat! lab so far have focused on the main tasks of the information verification pipeline: check-worthiness, evidence retrieval and pairing, and verification. The 2023 edition of the lab zooms into some of the problems and—for the first time—it offers five tasks in seven languages (Arabic, Dutch, English, German, Italian, Spanish, and Turkish): Task 1 asks to determine whether an item, text or a text plus an image, is check-worthy; Task 2 requires to assess whether a text snippet is subjective or not; Task 3 looks for estimating the political bias of a document or a news outlet; Task 4 requires to determine the level of factuality of a document or a news outlet; and Task 5 is about identifying authorities that should be trusted to verify a contended claim.

Keywords: Disinformation · Fact-checking · Check-worthiness · Subjectivity · Political bias · Factuality · Authority finding

1 Introduction

During its first five editions, the CheckThat! lab has focused on developing technology to assist the *journalist fact-checker* during the main steps of verification [7,8,18,19,47–49,51,52]. Figure 1 (top) shows the pipeline. First, a document (or a claim) is assessed for check-worthiness, i.e., whether a journalist should check its veracity. If this is so, the system needs to retrieve claims verified in the past that could be useful to fact-check the current one.

J. Kamps et al. (Eds.): ECIR 2023, LNCS 13982, pp. 506–517, 2023.
https://doi.org/10.1007/978-3-031-28241-6_59



	ar	bg	de	du	en	es	tr
a)	[3,28,58,46]	[58,46]		[46]	[3,4,59,58,46]	[58,46]	[58,46]
b)	[57,50]				[59,57,50]		
c)	[29,28]						
d)	[9,29,28]		[38]		[9,60,38]		

	ar	de	du	en	it	es	tr
T1	■			□		■	
T2	■	■	■	■	■		■
T3				■			
T4				■			
T5	■						

Fig. 1. Overview of the `CheckThat!` verification pipeline. The left table shows the core tasks addressed between the 2018 and the 2022 editions of the lab, including pointers to the relevant papers. The right table overviews the languages we target for the five tasks of the 2023 edition. Task 1 this year is the only one that belongs to the core tasks (a in the diagram), including multimodal data in English; a premier for the `CheckThat!` lab.

Further evidence to verify the claim is retrieved from the Web, if necessary. Finally, with the evidence gathered from the diverse sources, a decision can be made: whether the claim is factually true or not. The bottom-left table in Fig. 1 is the key to the technology developed for the tasks of the pipeline for all languages over the five editions of the `CheckThat!` lab so far.

Expert journalists consider that the most impactful technology in the verification process is check-worthiness, and that there are other aspects of news and social media that are relevant during analysis and verification, which have been overlooked.[1] With this in mind, the 2023 edition of the `CheckThat!` lab is organized around five tasks, four of which are run for the first time:

Task 1 Check-worthiness in tweets and political debates; the only task that has been organized during all the editions of the lab. It allows to reduce the workload of *listening* to social media for tweets that could be interesting. We introduce for the first time this year a multimodal track; cf. Sect. 2.

Task 2 Subjectivity in news articles to spot text that should be processed with specific strategies [56] (e.g., opinions may be filtered out and not checked, sarcasm and hyperboles might need further processing to extract the message they aim to convey); benefiting the fact-checking pipeline [33,35,64]; cf. Sect. 3.

Task 3 Political bias of news articles and news media to identify the political leaning of an article or media source, since a biased ones are more likely to make statements that are false or should be checked when they pursue the agenda and align with the bias of the author or the publisher; cf. Sect. 4.

[1] Private communication with organisations in various countries.

The Malaysian Ministry of Education has introduced E-Skrol, an application built on the NEM blockchain to deal with the issue of certificate fraud through the use of blockchain technology.

(a) Checkworthy

Turns out I've been doing airborne precautions wrong my whole life. #coronavirus

(b) Not-checkworthy

الان عودة الاشتباكات العنيفه في نهم والجيش الوطني يسحق الحوثيين في وادي حريب

Translation: Now the violent clashes are back in Nehm, and the national army is crushing the Houthis in Wadi Harib

(c) Checkworthy

واقعياً مليشيا الإصلاح بـ #مأرب هي خط الدفاع الأول عن الحوثيين

Translation: Realistically, the Islah militia in Marib is the first line of defense for the Houthis

(d) Not checkworthy

Fig. 2. Examples of tweets with their checkworthiness labels for Task 1.

Task 4 Factuality of reporting of news media is critical for media profiling; cf. Sect. 5.

Task 5 Authority finding in Twitter to help fact-checkers who aim to verify rumors propagating in social media find a trusted source (an authority that has "real knowledge" on the matter) that might help to confirm or to debunk a specific rumor. This task can be seen as a sub-problem of topical expert finding in Twitter [22,39,65]; cf. Sect. 6.

The bottom-right table in Fig. 1 gives an overview of the language coverage that we target for the five tasks this year.

2 Task 1: Check-Worthiness in Tweets

Task Definition. The aim of this task is to determine whether a claim is worth fact-checking. This year, we offer two kinds of data, which translate to the following two subtasks:

Subtask 1A (Multimodal – Tweets): The tweets to be judged include both a text snippet and an image.

Subtask 1B: Check-worthiness estimation from Multigenre (Unimodal) (text: A text snippet alone—from a tweet or a debate/speech transcription—has to be assessed for check-worthiness.

Subtask 1A is offered in Arabic and English, Subtask 1B is offered in Arabic, English and Spanish.

Data. For Task 1A, we use the annotation schema of [13]. Each tweet is annotated based on both the image and the text it contains for *(i)* the presence of a factual claim, *(ii)* check-worthiness, and *(iii)* visual relevance. The latter holds for two aspects: there is a piece of evidence (e.g., an event, an action, a situation, a person's identity, etc.) or illustration of certain aspects from the textual claim, or the image contains overlayed text that contains a claim in a textual form. The English data consists of 3k tweets. We also provide 82k unlabeled tweets that consist of text–image pairs and can be used for semi-supervised learning. The Arabic data consists of 3k tweets on topics such as COVID-19 and politics [1,47].

Table 1. Instances of subjective and objective sentences for Task 2.

	Instance	Class
1.	While it's misguided to put all focus or hope onto one section of the working class, we can't ignore this immense latent power that logistics workers possess	subj
2.	Taking refuge in public credit will cause that same infection to attack business, banking, industry, agriculture, the entire body of private enterprise	subj
3.	Workers would have a 24 percent wage increase by 2024, including an immediate 14 percent raise	obj
4.	University of Washington epidemiologist Ali Mokdad predicted a rise in reported COVID-19 cases	obj

The dataset for Subtask 1B consists of tweets in Arabic and Spanish. The Spanish tweets are collected from Twitter accounts and transcriptions from Spanish politicians and are manually annotated by professional journalists who are experts in fact-checking. The Arabic tweets for subtask 1B are collected using keywords related to COVID-19 and vaccines, using the annotation schema in [1]. The dataset for Subtask 1B (English) consists of political debates collected from U.S. general election presidential debates and annotated by human coders. Figure 2 shows examples of checkworthy and non-checkworthy tweets.

Evaluation. This is a binary classification task. The official evaluation measure is F_1 score for the positive class.

3 Task 2: Subjectivity in News Articles

Task Definition. The systems are challenged to distinguish whether a sentence from a news article expresses the subjective view of its author or presents an objective view of the covered topic. Given a list of sentences taken from a news article, the task asks to classify each of the sentences as subjective or objective. The task is offered in Arabic, Dutch, English, Italian, German, and Turkish.

Data. The focus is on sentences from newspaper articles. The data for Italian is partially derived from SubjectivITA [2] and consists of 2.2k examples, 25% of which are subjective. For English, we release a new dataset containing 1.2k sentences. The annotation process involved multiple annotators that labeled instances individually. Later on, annotators discussed and resolved the disagreements. We measured the Inter-Annotators Agreement (IAA) for the Italian dataset using Fleiss' kappa, and obtained a score of 0.61, which corresponds to substantial agreement. For the English dataset, we computed Krippendorf's alpha of 0.83. For the other languages, we plan to follow the same methodology to release datasets of comparable size. Table 1 shows examples of the English part of the dataset for Task 2.

Table 2. Examples of media with different biases for Task 3.

Name	URL	Bias
Loser.com	http://loser.com	Left
Die Hard Democrat	http://dieharddemocrat.com	Left
Democracy 21	http://www.democracy21.org	Center
Federal Times	http://www.federaltimes.com	Center
Gulf News	http://gulfnews.com	Center
Fox News	http://www.foxnews.com	Right

Evaluation. This is a binary classification task, and thus we use macro-averaged F_1 score as the official evaluation measure.

4 Task 3: Political Bias of News Articles and News Media

Task Definition. The goal of the task is to detect political bias of news reporting at the article and at the media level. This is an ordinal classification task and it is offered in English. It includes two subtasks:

Subtask 3A: Given an article, classify its leaning as left, center, or right.

Subtask 3B: Given the URL to a news outlet (e.g., www.cnn.com), predict the overall political bias of that news outlet as left, center, or right.

Data. We release a collection of 95k articles from 900 media sources annotated for bias at the article and at the media level, respectively. We used a subset of this data in previous research [6], but we have now crawled additional articles and sources for training and testing purposes.[2] Table 2 shows examples of news media with their political leaning. Note that we map the bias from a 7-point scale (Extreme-Left, Left, Center-Left, Center, Center-Right, Right, and Extreme-Right) to 3-point scale: left, center, and right.

Evaluation. This is an ordinal classification task, and thus we use mean absolute error as the official measure for both subtasks.

5 Task 4: Factuality of Reporting of News Media

Task Definition. We ask to predict the factuality of reporting at the media level, given the URL to a news outlet (e.g., www.cnn.com): low, mixed, and high. We offer the task in English.

Data. We use the same kind of data as for task 3, but with labels for factuality (again on an ordinal scale). We obtain the annotations and the analysis of the factuality of reporting and/or bias from mediabiasfactcheck.org, which are manually labeled by fact-checkers. The dataset consists of over 2k news media. Table 3 shows examples of news media and their factuality labels.

[2] The annotated labels for the articles are obtained from http://www.allsides.com/ and http://mediabiasfactcheck.org/.

Table 3. Examples of news media with different factuality labels for Task 4.

Name	URL	Factuality
Associated Press	http://apnews.com	High
NBC News	http://www.nbcnews.com/	High
Russia Insider	http://russia-insider.com	Mixed
Patriots Voice	https://www.patriotvoices.com	Low

Evaluation. This is an ordinal classification task, and we use mean absolute error as the official evaluation measure.

6 Task 5: Authority Finding in Twitter

Task Definition. The task asks systems to retrieve authority Twitter accounts for a given rumor that propagates in Twitter. Given a tweet spreading a rumor, the participating systems need to retrieve a ranked list of authority Twitter accounts that can help verify that rumor, as such accounts may tweet evidence that supports or denies the rumor [26]. This task is offered in Arabic.

Data. The training set comprises 150 rumors expressed in tweets associated with 1k authority Twitter accounts, a set of 400k Twitter accounts, 1.2M unique Twitter lists, and 878M timeline tweets. To construct the data, we selected rumors from Misbar[3], an Arabic fact-checking platform adopted by recent studies to construct datasets for Arabic rumor verification [27] and fake news detection [37]. For each rumor, two annotators were individually asked to find all possible authority Twitter accounts who can help confirm or deny that rumor following our detailed annotation guidelines. As part of the annotation process, the annotators were required to assign a grade for each authority to determine whether she is *highly relevant* or *relevant* to the rumor, i.e., having a higher priority to be contacted for verification or not. Finally, the annotators discussed their agreement on each others' selected authorities and their grades, and a third annotator helped resolve the disagreements. To evaluate the quality of the annotations, we considered the agreement both on whether the target Twitter account is labeled as authority with respect to the considered rumor as well as the graded relevance. The Cohen's Kappa inter-annotator agreement [14] was 0.78 and 0.71 for the former and for the latter, respectively, both scores corresponding to *substantial* agreement [40]. Table 4 shows an example rumor with authorities ranked according to relevance.

Evaluation. As this is a ranking task, we adopt P@5 as the official evaluation measure to evaluate how well the participating systems retrieve Twitter authorities at the top of a short retrieved list. We further report NDCG@5 to measure the ability of systems to retrieve highly relevant authority Twitter accounts higher up in that list.

[3] https://misbar.com/.

Table 4. An example of a rumor with corresponding authorities for Task 5.

Rumor Tweet: The Saudi Federation decided to ban the 'Al-Alamy' title from all league clubs and prohibit clubs from using it, whether banners inside stadiums or through clubs' websites		
	Authority	Relevance
1	Saudi Arabian Football Federation	Highly relevant
2	President of the Saudi Arabian Football Federation	Highly relevant
3	Saudi Arabian Football Federation Media and Communications	Highly relevant
4	Ministry of Sport in Saudi Arabia	Relevant
5	Minister of Sports and President of the Saudi Olympic and Paralympic Committee	Relevant

7 Related Work

There has been a lot of research on checking the factuality of a claim, of a news article, or of an information source [5,6,34,41,45,67]. Given that misleading content is causing harm across different dimensions, a lot of attention has been paid to identifying disinformation and misinformation in social media [24,36,43, 61,66]. Check-worthiness estimation is still an understudied problem, especially in social media [21,30–32,63], and fake news detection for news articles is mostly approached as a binary classification problem [53].

`CheckThat!` is related to several tasks at SemEval: on determining rumor veracity [16,23], on stance detection [44], on fact-checking in community question answering forums [42], and on propaganda detection [15,17]. It is also related to the FEVER task [62] on fact extraction and verification, to the Fake News Challenge [25,55] and to the FakeNews task at MediaEval [54].

8 Conclusion

We presented the 2023 edition of the `CheckThat!` lab, which features complementary tasks to assist in the full fact-checking pipeline: from spotting check-worthy claims to identifying an authority that could help verify a rumor in social media. In line with one of the main missions of CLEF, we promote multi-linguality by offering tasks in seven languages: Arabic, Dutch, English, German, Italian, Spanish, and Turkish. Moreover, for the first time, we also promote a multimodal task.

Acknowledgments. The work of Tamer Elsayed was made possible by NPRP grant #NPRP-11S-1204-170060 from the Qatar National Research Fund (a member of Qatar Foundation). The work of Fatima Haouari is supported by GSRA grant #GSRA6-1-0611-19074 from the Qatar National Research Fund. The statements made herein are solely the responsibility of the authors.

References

1. Alam, F., et al.: Fighting the COVID-19 infodemic: modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society. In: Findings of EMNLP 2021, pp. 611–649 (2021)
2. Antici, F., Bolognini, L., Inajetovic, M.A., Ivasiuk, B., Galassi, A., Ruggeri, F.: SubjectivITA: an Italian corpus for subjectivity detection in newspapers. In: Candan, K.S., et al. (eds.) CLEF 2021. LNCS, vol. 12880, pp. 40–52. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-85251-1_4
3. Atanasova, P., et al.: Overview of the CLEF-2018 CheckThat! lab on automatic identification and verification of political claims. Task 1: check-worthiness. In: Cappellato et al. [12]
4. Atanasova, P., Nakov, P., Karadzhov, G., Mohtarami, M., Da San Martino, G.: Overview of the CLEF-2019 CheckThat! lab on automatic identification and verification of claims. Task 1: check-worthiness. In: Cappellato et al. [11]
5. Ba, M.L., Berti-Equille, L., Shah, K., Hammady, H.M.: VERA: a platform for veracity estimation over web data. In: Proceedings of the 25th International Conference on World Wide Web, WWW 2016, pp. 159–162 (2016)
6. Baly, R., et al.: What was written vs. who read it: news media profiling using text analysis and social media context. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, pp. 3364–3374 (2020)
7. Barrón-Cedeño, A., et al.: CheckThat! at CLEF 2020: enabling the automatic identification and verification of claims in social media. In: Advances in Information Retrieval, ECIR 2020, pp. 499–507 (2020)
8. Barrón-Cedeño, A., et al.: Overview of CheckThat! 2020: automatic identification and verification of claims in social media. In: Arampatzis, A., et al. (eds.) CLEF 2020. LNCS, vol. 12260, pp. 215–236. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58219-7_17
9. Barrón-Cedeño, A., et al.: Overview of the CLEF-2018 CheckThat! lab on automatic identification and verification of political claims. Task 2: factuality. In: Cappellato et al. [12]
10. Cappellato, L., Eickhoff, C., Ferro, N., Névéol, A. (eds.): CLEF 2020 Working Notes. CEUR Workshop Proceedings (2020)
11. Cappellato, L., Ferro, N., Losada, D., Müller, H. (eds.): Working Notes of CLEF 2019 Conference and Labs of the Evaluation Forum. CEUR Workshop Proceedings (2019)
12. Cappellato, L., Ferro, N., Nie, J.Y., Soulier, L. (eds.): Working Notes of CLEF 2018-Conference and Labs of the Evaluation Forum. CEUR Workshop Proceedings (2018)
13. Cheema, G.S., Hakimov, S., Sittar, A., Müller-Budack, E., Otto, C., Ewerth, R.: MM-claims: a dataset for multimodal claim detection in social media. In: Findings of NAACL, pp. 962–979 (2022)
14. Cohen, J.: A coefficient of agreement for nominal scales. Educ. Psychol. Measur. **20**(1), 37–46 (1960)
15. Da San Martino, G., Barrón-Cedeno, A., Wachsmuth, H., Petrov, R., Nakov, P.: SemEval-2020 task 11: detection of propaganda techniques in news articles. In: Proceedings of the 14th Workshop on Semantic Evaluation, SemEval 2020, pp. 1377–1414 (2020)

16. Derczynski, L., Bontcheva, K., Liakata, M., Procter, R., Wong Sak Hoi, G., Zubiaga, A.: SemEval-2017 task 8: RumourEval: determining rumour veracity and support for rumours. In: Proceedings of the 11th International Workshop on Semantic Evaluation, SemEval 2017, pp. 69–76 (2017)
17. Dimitrov, D., et al.: SemEval-2021 task 6: detection of persuasion techniques in texts and images. In: Proceedings of the International Workshop on Semantic Evaluation, SemEval 2021, pp. 70–98 (2021)
18. Elsayed, T., et al.: CheckThat! at CLEF 2019: automatic identification and verification of claims. In: Azzopardi, L., Stein, B., Fuhr, N., Mayr, P., Hauff, C., Hiemstra, D. (eds.) ECIR 2019. LNCS, vol. 11438, pp. 309–315. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15719-7_41
19. Elsayed, T., et al.: Overview of the CLEF-2019 CheckThat! lab: automatic identification and verification of claims. In: Crestani, F., et al. (eds.) CLEF 2019. LNCS, vol. 11696, pp. 301–321. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-28577-7_25
20. Faggioli, G., Ferro, N., Joly, A., Maistro, M., Piroi, F. (eds.): CLEF 2021 Working Notes. Working Notes of CLEF 2021-Conference and Labs of the Evaluation Forum (2021)
21. Gencheva, P., Nakov, P., Màrquez, L., Barrón-Cedeño, A., Koychev, I.: A context-aware approach for detecting worth-checking claims in political debates. In: Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, pp. 267–276 (2017)
22. Ghosh, S., Sharma, N., Benevenuto, F., Ganguly, N., Gummadi, K.: Cognos: crowdsourcing search for topic experts in microblogs. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2012, pp. 575–590 (2012)
23. Gorrell, G., et al.: SemEval-2019 task 7: RumourEval, determining rumour veracity and support for rumours. In: Proceedings of the 13th International Workshop on Semantic Evaluation, SemEval 2019, pp. 845–854 (2019)
24. Gupta, A., Kumaraguru, P., Castillo, C., Meier, P.: TweetCred: real-time credibility assessment of content on Twitter. In: Proceedings of the 6th International Social Informatics Conference, SocInfo 2014, pp. 228–243 (2014)
25. Hanselowski, A., et al.: A retrospective analysis of the fake news challenge stance-detection task. In: Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, pp. 1859–1874 (2018)
26. Haouari, F., Elsayed, T.: Detecting stance of authorities towards rumors in Arabic tweets: a preliminary study. In: Proceedings of the 45th European Conference on Information Retrieval (ECIR 2023) (2023)
27. Haouari, F., Hasanain, M., Suwaileh, R., Elsayed, T.: ArCOV19-Rumors: Arabic COVID-19 Twitter dataset for misinformation detection. In: Proceedings of the Arabic Natural Language Processing Workshop, WANLP 2021, pp. 72–81 (2021)
28. Hasanain, M., et al.: Overview of CheckThat! 2020 Arabic: automatic identification and verification of claims in social media. In: Cappellato et al. [10]
29. Hasanain, M., Suwaileh, R., Elsayed, T., Barrón-Cedeño, A., Nakov, P.: Overview of the CLEF-2019 CheckThat! lab on automatic identification and verification of claims. Task 2: evidence and factuality. In: Cappellato et al. [11]
30. Hassan, N., Li, C., Tremayne, M.: Detecting check-worthy factual claims in presidential debates. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM 2015, pp. 1835–1838 (2015)
31. Hassan, N., et al.: ClaimBuster: the first-ever end-to-end fact-checking system. Proc. VLDB Endow. **10**(12), 1945–1948 (2017)

32. Jaradat, I., Gencheva, P., Barrón-Cedeño, A., Màrquez, L., Nakov, P.: ClaimRank: detecting check-worthy claims in Arabic and English. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, NAACL-HLT 2018, pp. 26–30 (2018)
33. Jerônimo, C.L.M., Marinho, L.B., Campelo, C.E.C., Veloso, A., da Costa Melo, A.S.: Fake news classification based on subjective language. In: Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services, pp. 15–24 (2019)
34. Karadzhov, G., Nakov, P., Màrquez, L., Barrón-Cedeño, A., Koychev, I.: Fully automated fact checking using external sources. In: Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, pp. 344–353 (2017)
35. Kasnesis, P., Toumanidis, L., Patrikakis, C.Z.: Combating fake news with transformers: a comparative analysis of stance detection and subjectivity analysis. Information **12**(10), 409 (2021)
36. Kazemi, A., Garimella, K., Gaffney, D., Hale, S.: Claim matching beyond English to scale global fact-checking. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL-IJCNLP 2021, pp. 4504–4517 (2021)
37. Khalil, A., Jarrah, M., Aldwairi, M., Jararweh, Y.: Detecting Arabic fake news using machine learning. In: Proceedings of the International Conference on Intelligent Data Science Technologies and Applications, IDSTA 2021, pp. 171–177 (2021)
38. Köhler, J., et al.: Overview of the CLEF-2022 CheckThat! lab task 3 on fake news detection. In: Working Notes of CLEF 2022–Conference and Labs of the Evaluation Forum, CLEF 2022 (2022)
39. Lahoti, P., De Francisci Morales, G., Gionis, A.: Finding topical experts in Twitter via query-dependent personalized PageRank. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2017, pp. 155–162 (2017)
40. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. Biometrics 159–174 (1977)
41. Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B.J., Wong, K.F., Cha, M.: Detecting rumors from microblogs with recurrent neural networks. In: Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI 2016, pp. 3818–3824 (2016)
42. Mihaylova, T., Karadzhov, G., Atanasova, P., Baly, R., Mohtarami, M., Nakov, P.: SemEval-2019 task 8: fact checking in community question answering forums. In: Proceedings of the 13th International Workshop on Semantic Evaluation, SemEval 2019, pp. 860–869 (2019)
43. Mitra, T., Gilbert, E.: CREDBANK: a large-scale social media corpus with associated credibility annotations. In: Proceedings of the Ninth International AAAI Conference on Web and Social Media, ICWSM 2015, pp. 258–267 (2015)
44. Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., Cherry, C.: SemEval-2016 task 6: detecting stance in tweets. In: Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval 2016, pp. 31–41 (2016)
45. Mukherjee, S., Weikum, G.: Leveraging joint interactions for credibility analysis in news communities. In: Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, pp. 353–362 (2015)
46. Nakov, P., et al.: Overview of the CLEF-2022 CheckThat! lab task 1 on identifying relevant claims in tweets. In: Working Notes of CLEF 2022–Conference and Labs of the Evaluation Forum, CLEF 2022 (2022)

47. Nakov, P., et al.: Overview of the CLEF-2022 CheckThat! lab on fighting the COVID-19 infodemic and fake news detection. In: Proceedings of the 13th International Conference of the CLEF Association: Information Access Evaluation meets Multilinguality, Multimodality, and Visualization, CLEF 2022 (2022)
48. Nakov, P., et al.: The CLEF-2022 CheckThat! lab on fighting the COVID-19 infodemic and fake news detection. In: Hagen, M., et al. (eds.) ECIR 2022. LNCS, vol. 13186, pp. 416–428. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-99739-7_52
49. Nakov, P., et al.: Overview of the CLEF-2018 lab on automatic identification and verification of claims in political debates. In: Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, CLEF 2018 (2018)
50. Nakov, P., Da San Martino, G., Alam, F., Shaar, S., Mubarak, H., Babulkov, N.: Overview of the CLEF-2022 CheckThat! lab task 2 on detecting previously fact-checked claims. In: Working Notes of CLEF 2022–Conference and Labs of the Evaluation Forum, CLEF 2022 (2022)
51. Nakov, P., et al.: Overview of the CLEF–2021 CheckThat! lab on detecting check-worthy claims, previously fact-checked claims, and fake news. In: Candan, K.S., et al. (eds.) CLEF 2021. LNCS, vol. 12880, pp. 264–291. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-85251-1_19
52. Nakov, P., et al.: The CLEF-2021 CheckThat! lab on detecting check-worthy claims, previously fact-checked claims, and fake news. In: Hiemstra, D., Moens, M.-F., Mothe, J., Perego, R., Potthast, M., Sebastiani, F. (eds.) ECIR 2021. LNCS, vol. 12657, pp. 639–649. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-72240-1_75
53. Oshikawa, R., Qian, J., Wang, W.Y.: A survey on natural language processing for fake news detection. In: Proceedings of the 12th Language Resources and Evaluation Conference, LREC 2020, pp. 6086–6093 (2020)
54. Pogorelov, K., et al.: FakeNews: corona virus and 5G conspiracy task at MediaEval 2020. In: Proceedings of the MediaEval 2020 Workshop, MediaEval 2020 (2020)
55. Pomerleau, D., Rao, D.: The fake news challenge: exploring how artificial intelligence technologies could be leveraged to combat fake news (2017). http://www.fakenewschallenge
56. Riloff, E., Wiebe, J.: Learning extraction patterns for subjective expressions. In: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, EMNLP 2003, pp. 105–112 (2003)
57. Shaar, S., et al.: Overview of the CLEF-2021 CheckThat! lab task 2 on detecting previously fact-checked claims in tweets and political debates. In: Faggioli et al. [20]
58. Shaar, S., et al.: Overview of the CLEF-2021 CheckThat! lab task 1 on check-worthiness estimation in tweets and political debates. In: Faggioli et al. [20]
59. Shaar, S., et al.: Overview of CheckThat! 2020 English: automatic identification and verification of claims in social media. In: Cappellato et al. [10]
60. Shahi, G.K., Struß, J.M., Mandl, T.: Overview of the CLEF-2021 CheckThat! lab: task 3 on fake news detection. In: Faggioli et al. [20]
61. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: a data mining perspective. SIGKDD Explor. Newsl. **19**(1), 22–36 (2017)
62. Thorne, J., Vlachos, A., Christodoulopoulos, C., Mittal, A.: FEVER: a large-scale dataset for fact extraction and VERification. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, pp. 809–819 (2018)

63. Vasileva, S., Atanasova, P., Màrquez, L., Barrón-Cedeño, A., Nakov, P.: It takes nine to smell a rat: neural multi-task learning for check-worthiness prediction. In: Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP 2019, pp. 1229–1239 (2019)
64. Vieira, L.L., Jerônimo, C.L.M., Campelo, C.E.C., Marinho, L.B.: Analysis of the subjectivity level in fake news fragments. In: Proceedings of the Brazillian Symposium on Multimedia and the Web, WebMedia 2020, pp. 233–240. ACM (2020)
65. Wei, W., Cong, G., Miao, C., Zhu, F., Li, G.: Learning to find topic experts in Twitter via different relations. IEEE Trans. Knowl. Data Eng. **28**(7), 1764–1778 (2016)
66. Zhao, Z., Resnick, P., Mei, Q.: Enquiring minds: early detection of rumors in social media from enquiry posts. In: Proceedings of the 24th International Conference on World Wide Web, WWW 2015, pp. 1395–1405 (2015)
67. Zubiaga, A., Liakata, M., Procter, R., Hoi, G.W.S., Tolmie, P.: Analysing how people orient to and spread rumours in social media by looking at conversational threads. PLoS ONE **11**(3), e0150989 (2016)