The CLEF-2022 CheckThat! Lab on Fighting the COVID-19 Infodemic and Fake News Detection

Nakov, Preslav; Barrón-Cedeño, Alberto; Da San Martino, Giovanni; Alam, Firoj; Struß, Julia Maria; Mandl, Thomas; Míguez, Rubén; Caselli, Tommaso; Kutlu, Mucahid; Zaghouani, Wajdi

Published in:
Advances in Information Retrieval - 44th European Conference on IR Research, ECIR 2022, Proceedings

DOI:
10.1007/978-3-030-99739-7_52

The CLEF-2022 CheckThat! Lab on Fighting the COVID-19 Infodemic and Fake News Detection

Preslav Nakov[1(✉)], Alberto Barrón-Cedeño[2], Giovanni Da San Martino[3], Firoj Alam[1], Julia Maria Struß[4], Thomas Mandl[5], Rubén Míguez[6], Tommaso Caselli[7], Mucahid Kutlu[8], Wajdi Zaghouani[9], Chengkai Li[10], Shaden Shaar[11], Gautam Kishore Shahi[12], Hamdy Mubarak[1], Alex Nikolov[13], Nikolay Babulkov[14], Yavuz Selim Kartal[15], and Javier Beltrán[6]

[1] Qatar Computing Research Institute, HBKU, Doha, Qatar
{pnakov,fialam,hmubarak}@hbku.edu.qa
[2] DIT, Università di Bologna, Forlì, Italy
a.barron@unibo.it
[3] University of Padova, Padua, Italy
dasan@math.unipd.it
[4] University of Applied Sciences Potsdam, Potsdam, Germany
struss@fh-potsdam.de
[5] University of Hildesheim, Hildesheim, Germany
mandl@uni-hildesheim.de
[6] Newtral Media Audiovisual, Madrid, Spain
{ruben.miguez,javier.beltran}@newtral.es
[7] University of Groningen, Groningen, The Netherlands
t.caselli@rug.nl
[8] TOBB University of Economics and Technology, Ankara, Turkey
m.kutlu@etu.edu.tr
[9] Hamad Bin Khalifa University, Doha, Qatar
wzaghouani@hbku.edu.qa
[10] University of Texas at Arlington, Arlington, TX, USA
cli@cse.uta.edu
[11] Cornell University, Ithaca, NY, USA
ss2753@cornell.edu
[12] University of Duisburg-Essen, Duisburg, Germany
gautam.shahi@uni-due.de
[13] CheckStep, London, UK
alex.nikolov@checkstep.com
[14] Sofia University, Sofia, Bulgaria
[15] GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany
yavuzselim.kartal@gesis.org

Abstract. The fifth edition of the CheckThat! Lab is held as part of the 2022 Conference and Labs of the Evaluation Forum (CLEF). The lab evaluates technology supporting various factuality tasks in seven languages: Arabic, Bulgarian, Dutch, English, German, Spanish, and Turkish. Task 1 focuses on disinformation related to the ongoing COVID-19

M. Hagen et al. (Eds.): ECIR 2022, LNCS 13186, pp. 416–428, 2022.
https://doi.org/10.1007/978-3-030-99739-7_52

infodemic and politics, and asks to predict whether a tweet is worth fact-checking, contains a verifiable factual claim, is harmful to the society, or is of interest to policy makers and why. Task 2 asks to retrieve claims that have been previously fact-checked and that could be useful to verify the claim in a tweet. Task 3 is to predict the veracity of a news article. Tasks 1 and 3 are classification problems, while Task 2 is a ranking one.

Keywords: Fact-checking · Disinformation · Misinformation · Check-worthiness · Verified claim retrieval · Fake news · Factuality · COVID-19

1 Introduction

The mission of the `CheckThat!` lab is to foster the development of technology to assist in the process of fact-checking news articles, political debates, and social media posts. Four editions of the lab have been held previously, targeting various natural language processing and information retrieval tasks related to factuality.

The 2018 edition of the lab focused on check-worthiness and fact-checking of claims in political debates [54]. The 2019 edition covered the various modules necessary to verify a claim: from check-worthiness, to ranking and classification of evidence in the form of Web pages, to actual fact-checking of claims against specific text snippets [24,25]. The 2020 edition featured three main tasks: detecting previously fact-checked claims, evidence retrieval, and actual fact-checking of claims [9,11]. Similarly, the 2021 edition focused on detecting check-worthy claims, previously fact-checked claims, and fake news [56,57]. Whereas the first editions focused mostly on political debates and speeches, and eventually tweets, the 2021 edition added the verification of news articles. Notably, all editions covered one of the most important initial stages in the fact-checking process: the identification of check-worthy claims—in debates, speeches, press conferences, and tweets. Finally, over the years, `CheckThat!` has witnessed an expansion in terms of language coverage, going from two (Arabic and English) to seven languages now (Arabic, Bulgarian, Dutch, English, German, Spanish, and Turkish).

The 2022 edition of the lab features three tasks to foster the technology on three timely problems in multiple languages.[1] Task 1 asks to detect relevant tweets: check-worthy, verifiable, harmful, and attention-worthy. Task 2 aims at detecting previously fact-checking claims. Task 3 focuses on checking the factuality of news articles. Automated systems to detect and to verify such multifaceted aspects can be very useful as supportive technology for investigative journalism, as they could provide help and guidance, thus saving time [3,27,37,39,55,81]. For example, a system could automatically identify check-worthy claims, could make sure they have not been fact-checked before by a reputable fact-checking organization, and can then present them to a journalist for further analysis in a ranked list. Similarly, a system can identify harmful and attention-worthy social

[1] http://sites.google.com/view/clef2022-checkthat/.

media content to support different stakeholders in their day-to-day decision-making process.

2 Description of the Tasks

The lab is organized around three tasks, each of which in turn has several subtasks. Figure 1 shows the full `CheckThat!` verification pipeline, with the three tasks we target this year highlighted.

2.1 Task 1: Identifying Relevant Claims in Tweets

Task 1 has four subtasks, three binary and one multi-class; Table 1 shows the class labels for each task. More detail about the original dataset on which this task is based can be found in [2,3].

Subtask 1A: Check-Worthiness Estimation Given a tweet, predict whether it is worth fact-checking by professional fact-checkers.

Subtask 1B: Verifiable Factual Claims Detection. Given a tweet, predict whether it contains a verifiable factual claim.

Subtask 1C: Harmful Tweet Detection. Given a tweet, predict whether it is harmful to the society.

Subtask 1D: Attention-Worthy Tweet Detection. Given a tweet, predict whether it should get the attention of policy makers and why.

2.2 Task 2: Detecting Previously Fact-Checked Claims

Given a check-worthy claim, and a set of previously-checked claims, determine whether the claim has been previously fact-checked with respect to a collection of fact-checked claims [67,69].

Subtask 2A: Detecting Previously Fact-checked Claims From Tweets. Given a tweet, detect whether the claim the tweet makes was previously fact-checked with respect to a collection of previously fact-checked claims. This is a ranking task, where the systems are asked to produce a list of top-n candidates.

Subtask 2B: Detecting Previously Fact-Checked Claims in Political Debates/Speeches. Within the context of a political debate or a speech, detect whether a claim has been previously fact-checked with respect to a collection of previously fact-checked claims. This is a ranking task, where systems are asked to produce a list of top-n candidates. It is offered in English only.

2.3 Task 3: Fake News Detection

This task targets news articles. Given the text and the title of an article, determine whether the main claim made in the article is true, partially true, false, or other (e.g., articles in dispute and unproven articles) [75,76]. This task is offered as a monolingual task in English and an English-German cross-lingual task.

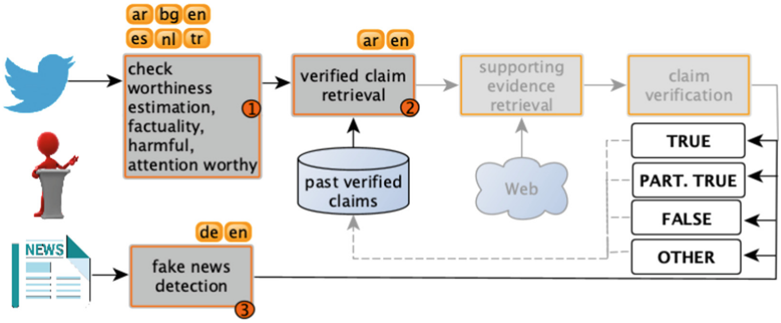

Fig. 1. The full verification pipeline. The lab covers three tasks from this pipeline: 1. identifying relevant claims in tweets, 2. verified claim retrieval, and 3. fake news detection. The languages covered are Arabic (ar), Bulgarian (bg), Dutch (nl), English (en), German (de), Spanish (es), and Turkish (tr).

Table 1. Overview of the classes for Subtasks 1A, 1B, 1C, and 1D.

Subtask 1A	**Subtask 1C**	**Subtask 1D**	
1. No	1. No	1. No	6. Yes, contains advice
2. Yes	2. Yes	2. Yes, asks question	7. Yes, discusses action taken
Subtask 1B		3. Yes, blame authorities	8. Yes, discusses cure
1. No		4. Yes, calls for action	9. Yes, other
2. Yes		5. Yes, Harmful	

3 Data

Table 2 summarizes the data available for each task and for each language.

Task 1: Identifying Relevant Claims in Tweets. We have more than 34K annotations about several topics, including COVID-19 and politics, which cover all subtasks 1A-1D [10,57].

Task 2: Detecting Previously Fact-Checked Claims. For Subtask 2A, we have 1,400 annotated examples ranging from 2016 till 2021. The test set comes from Snopes. For more details, refer to [71]. For Subtask 2B, we have 800 claims available for training [70].

Task 3: Fake News Detection We have 1,254 annotated examples in English on various topics including COVID-19, climate change, and politics from the 2021 edition of the lab [77]. We provide a new test set for English and a similarly annotated test data set in German.

Table 2. Data for all tasks. *VC*: Verified claim, *Input-VC*: Input–verified claim pair.

Task	ar	bl	en	nl	es	tr	Task	ar	en	Task	en	de
1A	3.4k	2.6k	2.9k	1.2k	15.0k	3.3k	2A: Input-claim	858	1.4k	2B: Input-claim	669	
1B	5.0k	3.7k	4.5k	2.7k			2A: Input-VC	1.0k	1.4k	2B: Input-VC	804	
1C	5.0k	3.7k	4.5k	2.7k			2A: VC	30.3k	13.8k	2B: VC	19.2k	
1D	5.0k	3.7k	4.5k	2.7k						3	1.3k	400*

*The process of crawling and annotating the data is not finished, and the final number will vary.

4 Evaluation Measures

For the classification tasks 1A and 1C, we use the F_1 measure with respect to the positive class, for Task 1B, we use accuracy, and for Task 1D, we use weighted-F_1. For the ranking problems in Tasks 2A and 2B, the official evaluation measure is *Mean Average Precision* (MAP), as in the two previous editions of these tasks. We also report reciprocal rank, and $P@k$ for $k \in \{1, 3, 5, 10, 20, 30\}$. Finally, for Task 3, we use macro-average F_1-measure as the official evaluation measure.

5 Previously in the `CheckThat!` Lab

Four editions of `CheckThat!` have been held so far, and some of the tasks in the 2022 edition are closely related to tasks from previous editions. Thus, considering the most successful approaches applied in the past is a good starting point to address the 2022 tasks. Below, we briefly discuss the tasks from previous years.

5.1 CheckThat! 2021

Task 1_{2021}. Given a topic and a set of potentially related tweets (or political debates/speeches), rank the tweets (or the sentences in the political debate/speech) according to their check-worthiness for the topic. BERT, AraBERT, and RoBERTa were by far the most popular large-scale pre-trained language models for the task [72,83]. Other approaches used WordNet [85] and LIWC [66].

Task 2_{2021}. Given a tweet, a political debate, or a speech, detect whether the claim it makes was previously fact-checked with respect to a collection of fact-checked claims. The most successful approaches were based on AraBERT, RoBERTa, and Sentence-BERT [18,47,65].

Task 3_{2021}. Given the text and the title of a news article, determine whether the main claim it makes is true, partially true, false, or other. Also, identify the domain of the article: health, crime, climate, elections, or education. The most successful pre-trained language model was RoBERTa [7,19,44]. Ensembles were also popular, with components using BERT [44] and LSTMs [19,44].

5.2 CheckThat! 2020

Task 1_{2020}. Given a topic and a stream of potentially-related tweets, rank the tweets by check-worthiness for the topic [35,73]. The most successful runs adopted state-of-the-art transformers. The top-ranked teams for the English version of this task used BERT [17] and RoBERTa [60,82]. For the Arabic version, the top systems used AraBERT [42,82] and a multilingual BERT [34].

Task 2_{2020}. Given a check-worthy claim and a collection of previously verified claims, rank these verified claims, so that those that verify the input claim (or a sub-claim in it) are ranked on top [73]. The most effective approaches fine-tuned BERT or RoBERTa [13].

Task 3_{2020}. Given a check-worthy claim in a tweet on a specific topic and a set of text snippets extracted from potentially relevant Web pages, return a ranked list of evidence snippets for the claim.

Task 4_{2020}. Given a check-worthy claim on a specific topic and a set of potentially-relevant Web pages, predict the veracity of the claim [35]. The top model used a scoring function that computes the degree of concordance and negation between a claim and all input text snippets for that claim [80].

Task 5_{2020}. Given a debate segmented into sentences, together with speaker information, prioritize sentences for fact-checking [73]. Only one out of eight runs outperformed a strong bi-LSTM baseline [46].

5.3 CheckThat! 2019

Task 1_{2019}. Given a political debate, an interview, or a speech, rank its sentences by the priority with which they should be fact-checked [5]. The most successful approaches used neural networks for the individual classification of the instances, e.g., based on domain-specific word embeddings, syntactic dependencies, and LSTMs [33].

Task 2_{2019}. Given a claim and a set of potentially relevant Web pages, identify which of these pages (and passages thereof) are useful for a human to fact-check the claim. Finally, determine the factuality of the claim [36]. The best approach used textual entailment and external data [28].

5.4 CheckThat! 2018

Task 1_{2018}. It was identical to Task 1_{2019} [4]. The best approaches used *pseudo-speeches* as a concatenation of all interventions by a debater [87], and represented the entries with embeddings, POS tags, and syntactic dependencies [32].

Task 2_{2018}. Given a check-worthy claim in the form of a (transcribed) sentence, determine whether the claim is true, half-true, or false [12]. The best approach grabbed information from the Web and fed the claim with the most similar Web text into a CNN [32].

6 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 [6,8,41,45,51,59,64,68,86]. Special attention has been paid to disinformation and misinformation in social media [30,43,49, 74,78,84], more recently with focus on fighting the COVID-19 infodemic [2,3, 52,53]. Check-worthiness estimation is still an understudied problem, especially in social media [27,37–40,81], and fake news detection for news articles is mostly approached as a binary classification problem [59,61].

`CheckThat!` is related to several tasks at SemEval: on determining rumour veracity [22,29], on stance detection [50], on fact-checking in community question answering forums [48], on propaganda detection [21,23], and on semantic textual similarity [1,58]. It is also related to the FEVER task [79] on fact extraction and verification, to the Fake News Challenge [31,63], to the FakeNews task at MediaEval [62], as well as to the NLP4IF tasks on propaganda detection [20] and on fighting the COVID-19 infodemic in social media [68].

7 Conclusion

We presented the 2022 edition of the `CheckThat!` lab, which features tasks that span the full fact-checking pipeline: from spotting check-worthy claims to checking whether an input claim has been fact-checked before. We further have a fake news detection task. Last but not least, in line with one of the main missions of CLEF, we promote multi-linguality by offering tasks in seven languages: Arabic, Bulgarian, Dutch, English, German, Spanish, and Turkish.

Acknowledgments. This research is part of the Tanbih mega-project, developed at the Qatar Computing Research Institute, HBKU, which aims to limit the impact of "fake news", propaganda, and media bias, thus promoting media literacy and critical thinking. The Arabic annotation effort was partially made possible by NPRP grant NPRP13S-0206-200281 from the Qatar National Research Fund (a member of Qatar Foundation).

References

1. Agirre, E., et al.: SemEval-2016 task 1: semantic textual similarity, monolingual and cross-lingual evaluation. In: Proceedings of the 10th International Workshop on Semantic Evaluation. SemEval 2016, pp. 497–511 (2016)
2. Alam, F., et al.: Fighting the COVID-19 infodemic in social media: a holistic perspective and a call to arms. In: Proceedings of the International AAAI Conference on Web and Social Media, ICWSM 2021, pp. 913–922 (2021)
3. 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)
4. 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. [16]

5. 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. [15]
6. 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)
7. Balouchzahi, F., Shashirekha, H., Sidorov, G.: MUCIC at CheckThat! 2021:FaDo-fake news detection and domain identification using transformersensembling. In: Faggioli et al. [26], pp. 455–464D
8. 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)
9. Barrón-Cedeño, A., et al.: `CheckThat!` at CLEF 2020: enabling the automatic identification and verification of claims in social media. In: Jose, J.M., et al. (eds.) ECIR 2020. LNCS, vol. 12036, pp. 499–507. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-45442-5_65
10. Barrón-Cedeño, A., et al.: Overview of CheckThat! 2020 – automatic identification and verification of claims in social media. In: Proceedings of the 11th International Conference of the CLEF Association: Experimental IR Meets Multilinguality, Multimodality, and Interaction. CLEF 2020, pp. 215–236 (2020)
11. 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
12. 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. [16]
13. Bouziane, M., Perrin, H., Cluzeau, A., Mardas, J., Sadeq, A.: Buster.AI at CheckThat! 2020: Insights and recommendations to improve fact-checking. In: Cappellato et al. [14]
14. Cappellato, L., Eickhoff, C., Ferro, N., Névéol, A. (eds.): CLEF 2020 Working Notes. CEUR Workshop Proceedings (2020)
15. 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)
16. 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)
17. Cheema, G.S., Hakimov, S., Ewerth, R.: Check_square at CheckThat! 2020: claim detection in social media via fusion of transformer and syntacticfeatures. In: Cappellato et al. [14]
18. Chernyavskiy, A., Ilvovsky, D., Nakov, P.: Aschern at CLEF CheckThat! 2021: lambda-calculus of fact-checked claims. In: Faggioli et al. [26]
19. Cusmuliuc, C.G., Amarandei, M.A., Pelin, I., Cociorva, V.I., Iftene, A.: UAICS at CheckThat! 2021: fake news detection. In: Faggioli et al. [26]
20. Da San Martino, G., Barrón-Cedeño, A., Nakov, P.: Findings of the NLP4IF-2019 shared task on fine-grained propaganda detection. In: Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda. NLP4IF 2019, pp. 162–170 (2019)

21. 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)
22. 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)
23. 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)
24. 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.) Advances in Information Retrieval, pp. 309–315. Springer International Publishing, Cham (2019)
25. 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
26. 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)
27. 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)
28. Ghanem, B., Glavaš, G., Giachanou, A., Ponzetto, S., Rosso, P., Rangel, F.: UPV-UMA at CheckThat! lab: verifying Arabic claims using cross lingual approach. In: Cappellato et al. [15]
29. 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)
30. Gupta, A., Kumaraguru, P., Castillo, C., Meier, P.: TweetCred: real-time credibility assessment of content on Twitter. In: Aiello, L.M., McFarland, D. (eds.) SocInfo 2014. LNCS, vol. 8851, pp. 228–243. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-13734-6_16
31. 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)
32. Hansen, C., Hansen, C., Simonsen, J., Lioma, C.: The Copenhagen team participation in the check-worthiness task of the competition of automatic identification and verification of claims in political debates of the CLEF-2018 fact checking lab. In: Cappellato et al. [16]
33. Hansen, C., Hansen, C., Simonsen, J., Lioma, C.: Neural weakly supervised fact check-worthiness detection with contrastive sampling-based ranking loss. In: Cappellato et al. [15]
34. Hasanain, M., Elsayed, T.: bigIR at CheckThat! 2020: multilingual BERT for ranking Arabic tweets by check-worthiness. In: Cappellato et al. [14]
35. Hasanain, M., et al.: Overview of CheckThat! 2020 Arabic: automatic identification and verification of claims in social media. In: Cappellato et al. [14]

36. 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. [15]
37. 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)
38. Hassan, N., Tremayne, M., Arslan, F., Li, C.: Comparing automated factual claim detection against judgments of journalism organizations. In: Computation+Journalism Symposium, pp. 1–5 (2016)
39. Hassan, N., et al.: ClaimBuster: the first-ever end-to-end fact-checking system. Proc. VLDB Endowment **10**(12), 1945–1948 (2017)
40. 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)
41. 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)
42. Kartal, Y.S., Kutlu, M.: TOBB ETU at CheckThat! 2020: prioritizing English and Arabic claims based on check-worthiness. In: Cappellato et al. [14]
43. 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 202, pp. 4504–45171 (2021)
44. Kovachevich, N.: BERT fine-tuning approach to CLEF CheckThat! fake news detection. In: Faggioli et al. [26]
45. Ma, J., et al.: Detecting rumors from microblogs with recurrent neural networks. In: Proceedings of the International Joint Conference on Artificial Intelligence. IJCAI 2016, pp. 3818–3824 (2016)
46. Martinez-Rico, J., Araujo, L., Martinez-Romo, J.: NLP&IR@UNED at CheckThat! 2020: a preliminary approach for check-worthiness and claim retrieval tasks using neural networks and graphs. In: Cappellato et al. [14]
47. Mihaylova, S., Borisova, I., Chemishanov, D., Hadzhitsanev, P., Hardalov, M., Nakov, P.: DIPS at CheckThat! 2021: verified claim retrieval. In: Faggioli et al. [26]
48. 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)
49. 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)
50. 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)
51. 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)

52. Nakov, P., Alam, F., Shaar, S., Da San Martino, G., Zhang, Y.: COVID-19 in Bulgarian social media: factuality, harmfulness, propaganda, and framing. In: Proceedings of the International Conference on Recent Advances in Natural Language Processing. RANLP 2021, pp. 997–1009 (2021)
53. Nakov, P., Alam, F., Shaar, S., Da San Martino, G., Zhang, Y.: A second pandemic? Analysis of fake news about COVID-19 vaccines in Qatar. In: Proceedings of Conference on Recent Advances in Natural Language Processing, pp. 1010–1021 (2021)
54. 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)
55. Nakov, P., et al.: Automated fact-checking for assisting human fact-checkers. In: Proceedings of the 30th International Joint Conference on Artificial Intelligence. IJCAI 2021, pp. 4551–4558 (2021)
56. 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
57. 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
58. Nakov, P., et al.: SemEval-2016 Task 3: community question answering. In: Proceedings of the 10th International Workshop on Semantic Evaluation. SemEval 2015, pp. 525–545 (2016)
59. Nguyen, V.H., Sugiyama, K., Nakov, P., Kan, M.Y.: FANG: leveraging social context for fake news detection using graph representation. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM 2020, pp. 1165–1174 (2020)
60. Nikolov, A., Da San Martino, G., Koychev, I., Nakov, P.: Team_Alex at CheckThat! 2020: identifying check-worthy tweets with transformer models. In: Cappellato et al. [14]
61. 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)
62. Pogorelov, K., et al.: FakeNews: corona virus and 5G conspiracy task at MediaEval 2020. In: Proceedings of the MediaEval 2020 Workshop. MediaEval 2020 (2020)
63. Pomerleau, D., Rao, D.: The fake news challenge: exploring how artificial intelligence technologies could be leveraged to combat fake news (2017). http://www.fakenewschallenge
64. Popat, K., Mukherjee, S., Strötgen, J., Weikum, G.: Credibility assessment of textual claims on the web. In: Proceedings of the 25th ACM International Conference on Information and Knowledge Management. CIKM 2016, pp. 2173–2178 (2016)
65. Pritzkau, A.: NLytics at CheckThat! 2021: check-worthiness estimation as a regression problem on transformers. In: Faggioli et al. [26]
66. Sepúlveda-Torres, R., Saquete, E.: GPLSI team at CLEF CheckThat! 2021: fine-tuning BETO and RoBERTa. In: Faggioli et al. [26]
67. Shaar, S., Alam, F., Da San Martino, G., Nakov, P.: The role of context in detecting previously fact-checked claims. arXiv:2104.07423 (2021)

68. Shaar, S., et al.: Findings of the NLP4IF-2021 shared tasks on fighting the COVID-19 infodemic and censorship detection. In: Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda. NLP4IF 2021, pp. 82–92 (2021)
69. Shaar, S., Alam, F., Martino, G.D.S., Nakov, P.: Assisting the human fact-checkers: detecting all previously fact-checked claims in a document. arXiv preprint arXiv:2109.07410 (2021)
70. Shaar, S., Babulkov, N., Da San Martino, G., Nakov, P.: That is a known lie: Detecting previously fact-checked claims. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. ACL 2020, pp. 3607–3618 (2020)
71. 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. [26]
72. 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. [26]
73. Shaar, S., et al.: Overview of CheckThat! 2020 English: automatic identification and verification of claims in social media. In: Cappellato et al. [14]
74. Shahi, G.K.: AMUSED: an annotation framework of multi-modal social media data. arXiv:2010.00502 (2020)
75. Shahi, G.K., Dirkson, A., Majchrzak, T.A.: An exploratory study of COVID-19 misinformation on Twitter. Online Social Networks Media **22**, 100104 (2021)
76. Shahi, G.K., Nandini, D.: FakeCovid - a multilingual cross-domain fact check news dataset for COVID-19. In: Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media (2020)
77. 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. [26]
78. 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)
79. 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 2018, pp. 809–819 (2018)
80. Touahri, I., Mazroui, A.: EvolutionTeam at CheckThat! 2020: integration of linguistic and sentimental features in a fake news detection approach. In: Cappellato et al. [14]
81. 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)
82. Williams, E., Rodrigues, P., Tran, S.: Accenture at CheckThat! 2021: interesting claim identification and ranking with contextually sensitive lexical training data augmentation. In: Faggioli et al. [14]
83. Williams, E., Rodrigues, P., Tran, S.: Accenture at CheckThat! 2021: interesting claim identification and ranking with contextually sensitive lexical training data augmentation. In: Faggioli et al. [26]
84. 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)

85. Zhou, X., Wu, B., Fung, P.: Fight for 4230 at CLEF CheckThat! 2021: domain-specific preprocessing and pretrained model for ranking claims by check-worthiness. In: Faggioli et al. [26]
86. 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)
87. Zuo, C., Karakas, A., Banerjee, R.: A hybrid recognition system for check-worthy claims using heuristics and supervised learning. In: Cappellato et al. [16]