Overview of the CLEF-2023 CheckThat! Lab Task 4 on Factuality of Reporting of News Media

Preslav Nakov[1,*], Firoj Alam[2,*], Giovanni Da San Martino[3], Maram Hasanain[1], Dilshod Azizov[1], Rabindra Nath Nandi[4] and Panayotov Panayot[5]

[1]*Mohamed bin Zayed University of Artificial Intelligence, UAE*
[2]*Qatar Computing Research Institute, HBKU, Qatar*
[3]*University of Padova, Italy*
[4]*Hishab Singapore Pte. Ltd, Singapore*
[5]*Sofia University, Bulgaria*

Abstract

We present an overview of the CLEF-2023 `CheckThat!` lab Task 4, which focused on predicting the factuality of reporting of entire news outlets. This is a different level of granularity compared to previous efforts, which focused on fact-checking, where the target is a claim, or fake news detection, where the target is an article. We briefly summarize the participating systems and discuss the dataset, the task, and the evaluation setup. The task attracted a large number of registrations, and eventually five teams made submissions. All participants improved over the baseline by a margin using both deep learning and traditional machine learning approaches. We make the dataset and the associate code freely available to the research community with the aim to promote further research on this problem.

Keywords

factuality, news media, veracity

1. Introduction

The study of factuality of news media reporting is of paramount importance for society at a time where anybody can create a website and become a "news producer." This has given rise to various initiatives to promote the dissemination of accurate, impartial, and truthful information to the public [1]. This includes manual fact-checking efforts, as well as automatic systems. These include the SemEval-2017 task on Rumor Detection [2], the FEVER challenge at EMNLP'2018 [3], the CLEF'2018 CheckThat! Lab on Automatic Identification and Verification of Claims in Political Debates [4, 5, 6], the SemEval-2019 task on Fact-Checking in Community Question Answering Forums [7, 8], and the CLEF 2021-2022 CheckThat! lab tasks on fake news detection [8, 9]. All these previous initiatives focused on fact-checking a claim or an article. In contrast, the Task 4 of the CheckThat! Lab 2023 [10] specifically targets the factuality of reporting of entire news outlets.

CLEF 2023: Conference and Labs of the Evaluation Forum, September 18–21, 2023, Thessaloniki, Greece
*Corresponding author.
preslav.nakov@mbzuai.ac.ae (P. Nakov); fialam@hbku.edu.qa (F. Alam); dasan@math.unipd.it (G. Da San Martino); mhasanain@hbku.edu.qa (M. Hasanain); dilshod.azizov@mbzuai.ac.ae (D. Azizov); rabindro.rath@gmail.com (R. N. Nandi); p.d.panayotov09@gmail.com (P. Panayot)

The CheckThat! 2023 Lab offered five tasks in its sixth edition: Task 1 focuses on check-worthiness in multimodal and multigenre content [11], Task 2 investigates the subjectivity of news articles [12], Task 3 studies the political bias of news articles and news media [13], Task 4 (this paper) examines the factuality of reporting of news media, and Task 5 delves into the process of authority finding in Twitter [14].

Below, we focus on Task 4, which focuses on the factuality of reporting of news media. We describe the dataset that we created for the task, which consists of news outlets annotated for factuality of reporting on a 3-point scale (left/center/right), and a set of articles from these outlets, which the participants had to use to determine the factuality of the news outlet.

The participants adopted a variety of approaches, ranging from transformer-based models to traditional machine learning. The most successful team used RoBERTa, incorporating a robustness-improving strategy using adversarial training components, a common method in deep learning, to enhance the performance of their system. They also capitalized on the extensive training data available. Other participants combined stylometry features with a more traditional Random Forest classifier.

The rest of this paper is organized as follows: Section 2 provides an overview of the task and explains the dataset used for it. Section 3 discusses the evaluation setup. Section 4 presents the results and provides details about the systems submitted by the participants. Section 5 highlights related work. Finally, Section 6 concludes the paper with key findings and outlines potential avenues for future research.

2. Task and Dataset

In this section, we first formulate the task, and then we discuss the dataset we developed for the task.

2.1. Task definition

The goal of the task is to determine the factuality of reporting at an entire news outlet. We offer the task in English, and we formally define it as follows:

Task 4: *Given a set of news articles from a news outlet, predict the factuality of that news outlet's reporting as low, mixed, or high factuality.*

2.2. Dataset

For this task, we developed and released a new dataset consisting of 1,189 media sources, along with their annotations for factuality in English. We further release a set of articles for each news outlet, a total of 10k news articles, which are to be used to predict the factuality of their source. To assess the factuality of news media, we used a 3-point scale: *low*, *mixed*, and *high*. We derived these labels from the Media Bias/Fact Check.[1] Examples of news media and their corresponding factuality assessments are provided in Table 1. Table 2 gives some statistics about the sources and the articles in the training, the development, and the test sets, respectively.

[1] http://mediabiasfactcheck.org

Table 1
Examples of media outlets with different level of factuality.

Name	URL	Bias
BBC	https://www.bbc.com	High
Reuters	https://www.reuters.com	High
Fox News	https://foxnews.com	Mixed
Breitbart	https://www.breitbart.com	Mixed
WhiteHouse.News	https://whitehouse.news	Low
Infowars-Alex Jones	https://apnews.com	Low

Table 2
Statistics about the dataset: training, development, and test partitions.

Class	Train	Dev	Test	Total
High	593	72	72	**737**
Mixed	233	32	31	**296**
Low	121	16	19	**156**
	947	**120**	**122**	**1,189**

3. Evaluation Settings

The evaluation process is divided into two phases: development and testing. In the development phase, we provided training and development datasets, which allowed the participants to fine-tune their systems and to adjust the system parameters based on the results on the development dataset. Throughout the testing phase, the participants were asked to submit the predictions generated by their systems using the provided test set that did not include any gold labels. Although they had the freedom to submit multiple runs of their system's output, only the last submitted run was counted as their official entry on the leaderboard.

This is an ordinal classification task, and thus we used *Mean Absolute Error* (MAE) as the official evaluation measure.

4. Results and Overview of the Systems

4.1. Results

Table 3 shows the results for four baselines on the development set. We can see that the *Ngram* baseline works best, while the random baseline is the worst, with middle and majority class baselines falling in between.

Five teams participated in this task, and their results are shown in Table 4. The *CUCPLUS [15]* team was ranked first with a MAE of 0.295, exhibiting the highest predictive accuracy. They are followed by the *NLPIR-UNED* and *Accenture* teams, with MAEs of 0.344 and 0.467, respectively. *UBCS* and *Awakened* ranked fourth and fifth with MAEs of 0.541 and 0.705, respectively. Notably, all teams improved over the baseline by a margin.

Table 3
Results for the baseline models on the development set.

Team	MAE
Ngram	**0.392**
Majority class (i.e., *high*)	0.533
Middle class (i.e., *center*)	0.733
Random	0.800

Table 4
Official evaluation results on the leaderboard.

Rank	Team	MAE
1	CUCPLUS [15]	0.295
2	NLPIR-UNED	0.344
3	Accenture [16]	0.467
4	UBCS [17]	0.541
5	Awakened	0.705
6	Ngram baseline	0.943

4.2. Overview of the Systems

CUCPLUS[15] tried to reduce the influence of redundant data and to enhance the model resilience using RoBERTa coupled with regularized adversarial training.
Accenture[16] aimed to maximize the amount of training data and developed a RoBERTa model that learns the factual reporting patterns of news articles and news sources.
UBCS [17] explored the effectiveness of stylometric features combined with a Random Forest classifier to evaluate the factuality of news media reporting. They proposed to leverage writing styles as a distinctive marker to differentiate between accurate and less factual news sources.
Ngram baseline uses TF.IDF to transform the text data into a numerical form, and an SVM to generate label predictions.

5. Related Work

Journalists, policymakers and researchers have shown significant interest in studying factuality across different levels: claim-level, user-level, article-level, and medium-level [18, 19, 20, 21, 22, 23, 24, 25]. Claim-level fact-checking has often been analyzing user interactions with these assertions on social media platforms, as explored in [26, 27]. In the realm of user-level reliability assessment, some studies, such as [28], have focused on the automatic detection of Twitter trolls during the COVID-19 pandemic, leveraging Test Time Evasion (TTE) in combination with a Markov chain-based mechanism. Additional studies, e.g.,!by [29, 30], focused on identifying opinion-manipulating trolls and sockpuppets in social media. At the article level, factuality has sometimes been examined alongside political bias, as the two are closely intertwined [31, 32].

In terms of online resources, the source's reliability usually pertains to the source's credibility, like the *URL* domain or the media outlet. Conversely, when it comes to social media and Internet forums, reliability is often linked to analyzing user behavior, including efforts to spot trolls who are suspected of or are confirmed to be swaying opinions, according to [29, 33].

The focus on factuality has been a constant across previous editions of the CLEF `CheckThat!` lab in 2018-2022. In the inaugural 2018 edition [34], Task 2 [6], offered in English and Arabic, asked the participants to evaluate the factual accuracy of claims made by politicians as part of debates or speeches. The data, mainly drawn from the 2016 US Presidential Campaign, was used to classify these claims as true, half-true, or false. Subsequently, in CLEF 2019 [7], Subtask 2D delved into the concept of factuality by predicting the trustworthiness of a claim using reliable websites. A claim was deemed true if it was accurate as stated or backed by sufficient credible evidence; otherwise, it was considered false.

Maintaining the `CheckThat!` tradition [35, 36, 37], one task was explicitly dedicated to the assessment of veracity. Compared to previous editions, which used binary classification for claim reliability, that edition introduced a 3-point scale. Given a check-worthy claim as a transcribed sentence, it could now be classified as true, half-true, or false. In the `CheckThat!` lab CLEF 2021 [38, 39], a pilot task was introduced with the aim to predict the veracity of a news article along with its topical domain. The subtask 3A, primarily focused on classifying the factuality of news articles on a 4-point scale: true, partially true, false, or other. Lastly, in the most recent `CheckThat!` 2022 lab, Nakov et al. [9] expanded Task 3 to include the prediction of the veracity of the main claim in a news article.

More directly related to the current task is work on detecting the factuality of reporting at the level of news outlets. Baly et al. [31] collected gold labels from Media Bias/Fact Check using diverse information for media bias and factuality analysis. Baly et al. [23] proposed a multi-task ordinal regression framework to simultaneously model the factuality of reporting and the political bias of entire news outlets [18]. Further enhancing their methodology, Baly et al. [19] incorporated Facebook followers and speech signals from the news outlet's YouTube channel, where available, as part of their information sources. Hounsel et al. [40] made predictions based on the domain, the certificate, and the hosting information from the website infrastructure as potential indicators of source reliability [18]. Bozhanova et al. [41] predicted the factuality of news outlets using observations about user attention in their YouTube channels. Finally, Panayotov et al. [42] modeled the inter-media similarity based on audience overlap.

6. Conclusion and Future Work

We presented an overview of Task 4 of the CLEF 2023 `CheckThat!` lab, which focused on predicting the factuality of reporting of news media. The task was offered in English and attracted a diverse range of approaches from transformer-based such as RoBERTa to traditional machine learning methods such as Random Forests and using stylometry features, which achieved sizable improvements over the baseline.

In future work, we plan to extend our dataset to more languages. We further aim to move beyond 3-way classification, towards a more finer-grained ordinal scale.

Acknowledgments

The work of F. Alam and M. Hasanain is partially supported by NPRP 13S-0206-200281 and NPRP 14C-0916-210015 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors.

References

[1] C. Gaziano, K. McGrath, Measuring the concept of credibility, Journalism quarterly 63 (1986) 451–462.

[2] L. Derczynski, K. Bontcheva, M. Liakata, R. Procter, G. Wong Sak Hoi, A. Zubiaga, SemEval-2017 task 8: RumourEval: Determining rumour veracity and support for rumours, in: Proceedings of the 11th International Workshop on Semantic Evaluation, SemEval '17, 2017, pp. 69–76.

[3] J. Thorne, A. Vlachos, C. Christodoulopoulos, A. Mittal, FEVER: a large-scale dataset for fact extraction and VERification (2018) 809–819.

[4] P. Nakov, A. Barrón-Cedeno, T. Elsayed, R. Suwaileh, L. Màrquez, W. Zaghouani, P. Atanasova, S. Kyuchukov, G. Da San Martino, Overview of the CLEF-2018 CheckThat! lab on automatic identification and verification of political claims, in: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2018, pp. 372–387.

[5] P. Atanasova, L. Marquez, A. Barrón-Cedeño, T. Elsayed, R. Suwaileh, W. Zaghouani, S. Kyuchukov, G. Da San Martino, P. Nakov, Overview of the CLEF-2018 CheckThat! lab on automatic identification and verification of political claims. Task 1: Check-worthiness, in: [43], 2018.

[6] A. Barrón-Cedeño, T. Elsayed, R. Suwaileh, L. Marquez, P. Atanasova, W. Zaghouani, S. Kyuchukov, G. Da San Martino, P. Nakov, Overview of the CLEF-2018 CheckThat! lab on automatic identification and verification of political claims. Task 2: Factuality, in: [43], 2018.

[7] T. Elsayed, P. Nakov, A. Barrón-Cedeño, M. Hasanain, R. Suwaileh, G. Da San Martino, P. Atanasova, Overview of the CLEF-2019 CheckThat!: Automatic identification and verification of claims, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, LNCS, 2019, pp. 301–321.

[8] P. Nakov, G. Da San Martino, T. Elsayed, A. Barrón-Cedeño, R. Míguez, S. Shaar, F. Alam, F. Haouari, M. Hasanain, W. Mansour, B. Hamdan, Z. S. Ali, N. Babulkov, A. Nikolov, G. K. Shahi, J. M. Struß, T. Mandl, M. Kutlu, Y. S. Kartal, Overview of the CLEF-2021 CheckThat! lab on detecting check-worthy claims, previously fact-checked claims, and fake news, LNCS (12880), 2021.

[9] P. Nakov, A. Barrón-Cedeño, G. Da San Martino, F. Alam, J. M. Struß, T. Mandl, R. Míguez, T. Caselli, M. Kutlu, W. Zaghouani, C. Li, S. Shaar, G. K. Shahi, H. Mubarak, A. Nikolov, N. Babulkov, Y. S. Kartal, J. Beltrán, M. Wiegand, M. Siegel, J. Köhler, Overview of the CLEF-2022 CheckThat! lab on fighting the COVID-19 infodemic and fake news detection, in: A. Barrón-Cedeño, G. Da San Martino, M. Degli Esposti, F. Sebastiani, C. Macdonald,

G. Pasi, A. Hanbury, M. Potthast, G. Faggioli, F. Nicola (Eds.), Proceedings of the 13th International Conference of the CLEF Association: Information Access Evaluation meets Multilinguality, Multimodality, and Visualization, CLEF '2022, Bologna, Italy, 2022.

[10] A. Barrón-Cedeño, F. Alam, T. Caselli, G. Da San Martino, T. Elsayed, A. Galassi, F. Haouari, F. Ruggeri, J. M. Struß, R. N. Nandi, G. S. Cheema, D. Azizov, P. Nakov, The CLEF-2023 CheckThat! Lab: Checkworthiness, subjectivity, political bias, factuality, and authority, in: J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, A. Caputo (Eds.), Advances in Information Retrieval, Springer Nature Switzerland, Cham, 2023, pp. 506–517.

[11] F. Alam, A. Barrón-Cedeño, G. S. Cheema, S. Hakimov, M. Hasanain, C. Li, R. Míguez, H. Mubarak, G. K. Shahi, W. Zaghouani, P. Nakov, Overview of the CLEF-2023 CheckThat! lab task 1 on check-worthiness in multimodal and multigenre content, in: Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF '2023, Thessaloniki, Greece, 2023.

[12] A. Galassi, F. Ruggeri, A. Barrón-Cedeño, F. Alam, T. Caselli, M. Kutlu, J. Struss, F. Antici, M. Hasanain, J. Köhler, K. Korre, F. Leistra, A. Muti, M. Siegel, T. Mehmet Deniz, M. Wiegand, W. Zaghouani, Overview of the CLEF-2023 CheckThat! lab task 2 on subjectivity in news articles, in: Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF '2023, Thessaloniki, Greece, 2023.

[13] G. Da San Martino, F. Alam, M. Hasanain, R. N. Nandi, D. Azizov, P. Nakov, Overview of the CLEF-2023 CheckThat! lab task 3 on political bias of news articles and news media, in: Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF '2023, Thessaloniki, Greece, 2023.

[14] F. Haouari, Z. Sheikh Ali, T. Elsayed, Overview of the CLEF-2023 CheckThat! lab task 5 on authority finding in twitter, in: Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF '2023, Thessaloniki, Greece, 2023.

[15] C. Li, R. Xue, C. Lin, W. Fan, CUCPLUS at CheckThat! 2023: Text Combination and Regularized Adversarial Training for News Media Factuality Evaluation., in: Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF '2023, Thessaloniki, Greece, 2023.

[16] S. Tran, P. Rodrigues, B. Strauss, E. Williams, Accenture at CheckThat! 2023: Learning to Detect Factuality Levels of News Sources, in: Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF '2023, Thessaloniki, Greece, 2023.

[17] T. Leburu-Dingalo, E. Thuma, N. P. Motlogelwa, M. Mudongo, G. Mosweunyane, UBCS at CheckThat! 2023: Stylometric features in detecting factuality of reporting of news media, in: M. Aliannejadi, G. Faggioli, N. Ferro, Vlachos, Michalis (Eds.), Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF 2023, Thessaloniki, Greece, 2023.

[18] P. Nakov, H. T. Sencar, J. An, H. Kwak, A survey on predicting the factuality and the bias of news media, arXiv preprint arXiv:2103.12506 (2021).

[19] R. Baly, G. Karadzhov, J. An, H. Kwak, Y. Dinkov, A. Ali, J. Glass, P. Nakov, 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, Online, 2020, pp. 3364–3374.

[20] R. Baly, M. Mohtarami, J. Glass, L. Màrquez, A. Moschitti, P. Nakov, Integrating stance detection and fact checking in a unified corpus, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), New Orleans, Louisiana, 2018, pp. 21–27.

[21] J. Ma, W. Gao, Z. Wei, Y. Lu, K. Wong, Detect rumors using time series of social context information on microblogging websites, in: J. Bailey, A. Moffat, C. C. Aggarwal, M. de Rijke, R. Kumar, V. Murdock, T. K. Sellis, J. X. Yu (Eds.), Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015, 2015, pp. 1751–1754.

[22] A. Zubiaga, M. Liakata, R. Procter, G. W. S. Hoi, P. Tolmie, Analysing how people orient to and spread rumours in social media by looking at conversational threads, PloS one 11 (2016) e0150989.

[23] R. Baly, G. Karadzhov, A. Saleh, J. Glass, P. Nakov, Multi-task ordinal regression for jointly predicting the trustworthiness and the leading political ideology of news media, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019, pp. 2109–2116.

[24] F. Alam, S. Shaar, F. Dalvi, H. Sajjad, A. Nikolov, H. Mubarak, G. D. S. Martino, A. Abdelali, N. Durrani, K. Darwish, A. Al-Homaid, W. Zaghouani, T. Caselli, G. Danoe, F. Stolk, B. Bruntink, P. Nakov, 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, 2021, pp. 611–649.

[25] F. Alam, F. Dalvi, S. Shaar, N. Durrani, H. Mubarak, A. Nikolov, G. Da San Martino, A. Abdelali, H. Sajjad, K. Darwish, P. Nakov, 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 '21, 2021, pp. 913–922.

[26] E. Kochkina, M. Liakata, A. Zubiaga, All-in-one: Multi-task learning for rumour verification, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 3402–3413.

[27] G. Karadzhov, P. Nakov, L. Màrquez, A. Barrón-Cedeño, I. Koychev, Fully automated fact checking using external sources, in: Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, 2017, pp. 344–353.

[28] P. Jachim, F. Sharevski, P. Treebridge, Trollhunter [evader]: Automated detection [evasion] of Twitter trolls during the COVID-19 pandemic, in: New Security Paradigms Workshop 2020, 2020, pp. 59–75.

[29] T. Mihaylov, G. Georgiev, P. Nakov, Finding opinion manipulation trolls in news community forums, in: Proceedings of the Nineteenth Conference on Computational Natural Language Learning, 2015, pp. 310–314.

[30] S. K. Maity, A. Chakraborty, P. Goyal, A. Mukherjee, Detection of sockpuppets in social media, in: Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17, 2017, pp. 243–246.

[31] R. Baly, G. Karadzhov, D. Alexandrov, J. Glass, P. Nakov, Predicting factuality of reporting and bias of news media sources, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 3528–3539.

[32] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K. Wong, M. Cha, Detecting rumors from microblogs with recurrent neural networks, in: S. Kambhampati (Ed.), Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016, 2016, pp. 3818–3824.
[33] T. Mihaylov, I. Koychev, G. Georgiev, P. Nakov, Exposing paid opinion manipulation trolls, in: Proceedings of the International Conference Recent Advances in Natural Language Processing, 2015, pp. 443–450.
[34] P. Nakov, A. Barrón-Cedeño, T. Elsayed, R. Suwaileh, L. Màrquez, W. Zaghouani, P. Atanasova, S. Kyuchukov, G. Da San Martino, Overview of the CLEF-2018 Check-That! lab on automatic identification and verification of political claims, in: Proceedings of the Ninth International Conference of the CLEF Association: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Lecture Notes in Computer Science, 2018, pp. 372–387.
[35] A. Barrón-Cedeño, T. Elsayed, P. Nakov, G. Da San Martino, M. Hasanain, R. Suwaileh, F. Haouari, N. Babulkov, B. Hamdan, A. Nikolov, S. Shaar, Z. Ali, 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, 2020, pp. 215–236.
[36] M. Hasanain, F. Haouari, R. Suwaileh, Z. Ali, B. Hamdan, T. Elsayed, A. Barrón-Cedeño, G. Da San Martino, P. Nakov, Overview of CheckThat! 2020 Arabic: Automatic identification and verification of claims in social media, in: [44], 2020.
[37] S. Shaar, A. Nikolov, N. Babulkov, F. Alam, A. Barrón-Cedeño, T. Elsayed, M. Hasanain, R. Suwaileh, F. Haouari, G. Da San Martino, P. Nakov, Overview of CheckThat! 2020 English: Automatic identification and verification of claims in social media, in: [44], 2020.
[38] G. K. Shahi, J. M. Struß, T. Mandl, Overview of the CLEF-2021 CheckThat! lab: Task 3 on fake news detection, 2021.
[39] P. Nakov, G. Da San Martino, T. Elsayed, A. Barrón-Cedeño, R. Míguez, S. Shaar, F. Alam, F. Haouari, M. Hasanain, N. Babulkov, et al., The CLEF-2021 CheckThat! lab on detecting check-worthy claims, previously fact-checked claims, and fake news, in: ECIR (2), 2021.
[40] A. Hounsel, J. Holland, B. Kaiser, K. Borgolte, N. Feamster, J. Mayer, Identifying disinformation websites using infrastructure features, in: 10th USENIX Workshop on Free and Open Communications on the Internet (FOCI 20), 2020.
[41] K. Bozhanova, Y. Dinkov, I. Koychev, M. Castaldo, T. Venturini, P. Nakov, Predicting the factuality of reporting of news media using observations about user attention in their YouTube channels, in: Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP 21, 2021, pp. 182–189.
[42] P. Panayotov, U. Shukla, H. T. Sencar, M. Nabeel, P. Nakov, GREENER: Graph neural networks for news media profiling, in: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022, pp. 7470–7480.
[43] L. Cappellato, N. Ferro, J.-Y. Nie, L. Soulier (Eds.), Working Notes of CLEF 2018–Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, 2018.
[44] L. Cappellato, C. Eickhoff, N. Ferro, A. Névéol (Eds.), CLEF 2020 Working Notes, CEUR Workshop Proceedings, 2020.