

ROMCIR 2023: Overview of the 3rd Workshop on Reducing Online Misinformation through Credible Information Retrieval

Marinella Petrocchi¹[0000-0003-0591-877X] and
Marco Viviani²[0000-0002-2274-9050]

¹ Institute of Informatics and Telematics – National Research Council (IIT – CNR),
Pisa, Italy

marinella.petrocchi@iit.cnr.it

<https://www.iit.cnr.it/marinella.petrocchi/>

² University of Milano-Bicocca – Department of Informatics, Systems, and
Communication (UNIMIB – DISCo), Milan, Italy

marco.viviani@unimib.it

<https://ikr3.disco.unimib.it/people/marco-viviani/>

Abstract. With the advent of the Social Web, we are constantly and more than ever assaulted by different kinds of information pollution, which may lead to severe issues for both individuals and society as a whole. In this context, it becomes essential to guarantee users access to genuine information that does not distort their perception of reality. For this reason, in recent years, numerous approaches have been proposed for the identification of misinformation, in different contexts and for different purposes. However, the problem has not yet been sufficiently addressed in the field of Information Retrieval, because it has been treated primarily as a classification task to identify information versus misinformation. Hence, the purpose of this Workshop is to address the IR community for solutions in which, among other issues, the genuineness of information is considered as one of the dimensions of relevance within search engines or recommender systems, early detection of misinformation can be achieved, the results obtained are explainable with respect to the users of Information Retrieval Systems, user’s privacy is taken into consideration.

Keywords: Information Retrieval · Information Access · Information Disorder · Information Genuineness · Misinformation.

1 Motivation and relevance to ECIR

*“Technology is so much fun but we can drown in our technology.
The fog of information can drive out knowledge”.*

This quote by American historian Daniel J. Boorstin, about the computerization of libraries, appeared in July 1983 in the New York Times. Some 40 years later, it is probably still more relevant than it was then. In fact, the lack of mediation that Web 2.0 technologies brought in the generation and dissemination

of online content within the Social Web has led to the well-known problems of *information overload* [23] and the spread of *misinformation* [6], which make it difficult for users to find information that is truly useful for their purposes, leading to possible severe issues for both individuals and society as a whole [10, 20, 26, 3]. Fake news can, for example, guide public opinion in political and financial choices [24, 30]; false reviews can promote or discredit economic activities [5, 14]; unverified medical information can lead people to follow behaviors that can be harmful to their own health and to that of society as a whole (let the reader think, for example, to the set of unverified news stories that have been disseminated about Covid-19) [4, 18, 8, 25].

Hence, the central topic of the third edition of the ROMCIR Workshop concerns providing access to users to both topically relevant and genuine information (without neglecting other relevant dimensions), to mitigate the *information disorder* phenomenon with respect to distinct domains. By “information disorder” we mean all forms of communication pollution, from misinformation made out of ignorance, to the intentional sharing of false content [27, 28]. In this context, all those approaches that can serve to assess the genuineness of information circulating online and in social media find their place. This topic is very broad, as it concerns different contents (e.g., Web pages, news, reviews, medical information, online accounts, etc.), different Web and social media platforms (e.g., microblogging platforms, social networking services, social question-answering systems, etc.), and different purposes (e.g., identifying false information, accessing information based on its genuineness, retrieving genuine information, etc.). Each of these aspects is relevant to the European Conference on Information Retrieval.

2 Aim and Topics of Interest

The aim of the Workshop is to generate a discussion on, and possibly provide countermeasures to, the problem of online information disorder, by providing access to information that is genuine. Given that the problem in recent years has been addressed from various points of view (e.g., fake news detection, bot detection, information genuineness assessment, and news source reputability, to cite a few), the purpose of this Workshop is to consider these issues in the context of Information Access and Retrieval, also considering related Artificial Intelligence fields such as Natural Language Processing (NLP), Natural Language Understanding (NLU), Computer Vision, Machine and Deep Learning. Hence, the topics of interest of ROMCIR 2023 include, but are not limited to, the following:

- Access to genuine information
- Bias detection
- Bot/spam/troll detection
- Computational fact-checking
- Crowd-sourcing for information genuineness assessment

- Deep fakes
- Disinformation/misinformation detection
- Evaluation strategies to assess information genuineness
- Fake news/review detection
- Harassment/bullying/hate speech detection
- Information polarization in online communities, echo chambers
- Propaganda identification/analysis
- Retrieval of genuine information
- Security, privacy and information genuineness
- Sentiment/emotional analysis
- Stance detection
- Trust and reputation
- Societal reaction to misinformation

3 Past Editions

The first two editions of the Workshop, both co-located with the ECIR conference, led to fervent discussion and presentation of innovative work with respect to a variety of open issues related to information genuineness and IR. The first edition, namely ROMCIR 2021, took place in online mode on April 1, 2021;³ the second edition, namely ROMCIR 2022 took place in hybrid mode on April 10, 2022.⁴ Further details are provided here below.

- ROMCIR 2021 received 15 submissions, of which 6 were accepted, with an acceptance rate of 40%. The accepted articles, collected in CEUR Proceedings [22], proposed distinct solutions for distinct open issues. There were those tangentially related to Credible Information Retrieval, such as those of *authorship verification* [29] and *bias detection* in science evaluation [2]. Furthermore, the problems of opinion mining and misinformation identification were tackled, such as those of *hate speech detection* [11] and *claim verification* [13]. Finally, the problem of access to genuine information was considered, by proposing the definition of Information Retrieval Systems to support users in *retrieving genuine news* [12], and the study of new IR methods able to consider the *genuineness* of the data collected in the retrieval process [7].
- ROMCIR 2022 received 10 submissions, of which 6 were accepted, so with an acceptance rate of 60%. The articles, collected in CEUR Proceedings [21] and summarized in the June 2022 SIGIR Forum volume,⁵ have primarily considered two issues from distinct and new points of view compared to those considered the previous year, reflecting the interest and new challenges related to the aim of the Workshop. The first theme concerned *genuine health IR* [9, 15, 16, 19], an application domain that is generating much discussion

³ <https://romcir.disco.unimib.it/2021-edition/2021-workshop/>

⁴ <https://romcir.disco.unimib.it/2022-edition/2022-workshop/>

⁵ <https://sigir.org/forum/issues/june-2022/>

among researchers; the second theme concerned *multi-modal genuine* IR [1, 17], and papers discussed the challenges that multimedia content analysis were raising with respect to those of text content analysis in information genuineness assessment.

4 List of Organizers

The following people contributed in different capacities to the organization of the Workshop and to the verification of the quality of the submitted work.

4.1 Workshop Chairs



Marinella Petrocchi is a Senior Researcher at the Institute of Informatics and Telematics of the National Research Council (IIT-CNR) in Pisa, Italy, under the Trust, Security, and Privacy research unit. She also collaborates with the Sysma unit at IMT School for Advanced Studies, in Lucca, Italy. Her field of research lies between Cybersecurity, Artificial Intelligence, and Data Science. Specifically, she studies novel techniques for online fake

news/fake accounts detection and automated methods to rank the reputability of online news media. She is the author of several international publications on these themes and she usually gives talks and lectures on the topic. She is in the core team of the TOFFEE project (TOols for Fighting FakEs), funded by IMT. *Website:* <https://www.iit.cnr.it/marinella.petrocchi/>



Marco Viviani is an Associate Professor at the University of Milano-Bicocca, Department of Informatics, Systems, and Communication (DISCo). He works in the Information and Knowledge Representation, Retrieval and Reasoning (IKR3) Lab. He has been co-chair of several special tracks and workshops at international conferences, and general co-chair of MDAI 2019. He is an Associate Editor of Social Network Analysis and Mining, an Editorial Board Member of Online Social Networks and Media, and a Guest Editor of several Special Issues in International Journals related to information disorder detection. His main research activities include Social Computing, Information Retrieval, Text Mining, Natural Language Processing, Trust and Reputation Management, and User Modeling. On these topics, he has written several international publications. *Website:* <https://ikr3.disco.unimib.it/people/marco-viviani/>

4.2 Proceedings Chair

Rishabh Upadhyay is a Research Fellow at the University of Milano-Bicocca, Department of Informatics, Systems, and Communication (DISCo). His research

interests are related to Machine and Deep Learning, Information Retrieval, and Social Computing. He is currently working within the EU Horizon 2020 ITN/ETN DoSSIER project on Domain-Specific Systems for Information Extraction and Retrieval, in particular on the project: “Assessing Credibility, Value, and Relevance”. He was one of the co-organizers of Task 2: Consumer Health Search, at CLEF 2021 eHealth Lab Series. He has recently published papers at International Conferences on the topic of health misinformation detection. *Website:* <https://www.unimib.it/rishabh-gyanendra-upadhyay/>

4.3 Program Committee Members

- Rino Falcone, ISTC – National Research Council (CNR), Italy;
- Carlos A. Iglesias, Universidad Politécnica de Madrid, Spain
- Petr Knoth, Research Studio Data Science (DSc), Austria
- Udo Kruschwitz, University of Regensburg, Germany
- Yelena Mejova, ISI Foundation, Italy
- Preslav Nakov, Hamad Bin Khalifa University, Qatar
- Symeon Papadopoulos, Information Technologies Institute (ITI), Greece
- Gabriella Pasi, University of Milano-Bicocca, Italy
- Marinella Petrocchi, IIT – National Research Council (CNR), Italy
- Francesco Pierri, Politecnico di Milano, Italy
- Manuel Pratelli, IMT School for Advanced Studies Lucca, Italy
- Fabio Saracco, Centro Ricerche Enrico Fermi, Italy
- Marco Viviani, University of Milano-Bicocca, Italy
- Arkait Zubiaga, Queen Mary University, UK

Acknowledgements

This work was supported by re-DESIRE: *DissEmination of ScIentific REsults* 2.0, funded by IIT-CNR, TOFFEE: *TOols for Fighting FakES*, funded by IMT Scuola Alti Studi Lucca, and DoSSIER: *Domain Specific Systems for Information Extraction and Retrieval*, EU Horizon 2020 ITN/ETN (H2020-EU.1.3.1., ID: 860721).

References

1. Aghada, K.A.: An alternative approach to ranking videos and measuring dissimilarity between video content and titles. In: Petrocchi, M., Viviani, M. (eds.) ROMCIR 2022 CEUR Workshop Proc. vol. 3138, pp. 89–99. CEUR-WS.org (2022)
2. Bethencourt, A.M., Luo, J., Feliciani, T.: Bias and truth in science evaluation: a simulation model of grant review panel discussions. In: ROMCIR 2021 CEUR Workshop Proc. vol. 2838, pp. 16–24 (2021)
3. Caldarelli, G., De Nicola, R., Petrocchi, M., Del Vigna, F., Saracco, F.: The role of bot squads in the political propaganda on Twitter. *Commun Phys.* **3**(81) (2020)

4. Caldarelli, G., De Nicola, R., Petrocchi, M., Pratelli, M., Saracco, F.: Flow of online misinformation during the peak of the covid-19 pandemic in italy. *EPJ data science* **10**(1), 34 (2021)
5. Crawford, M., Khoshgoftaar, T.M., Prusa, J.D., Richter, A.N., Al Najada, H.: Survey of review spam detection using machine learning techniques. *Journal of Big Data* **2**(1), 1–24 (2015)
6. Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H.E., Quattrociocchi, W.: The spreading of misinformation online. *Proceedings of the National Academy of Sciences* **113**(3), 554–559 (2016)
7. Denaux, R., Gomez-Perez, J.M.: Sharing retrieved information using Linked Credibility Reviews. In: *ROMCIR 2021 CEUR Workshop Proc.* vol. 2838, pp. 59–65 (2021)
8. Di Sotto, S., Viviani, M.: Health misinformation detection in the social web: An overview and a data science approach. *International Journal of Environmental Research and Public Health* **19**(4), 2173 (2022)
9. Fröbe, M., Günther, S., Bondarenko, A., Huck, J., Hagen, M.: Using keyqueries to reduce misinformation in health-related search results. In: Petrocchi, M., Viviani, M. (eds.) *ROMCIR 2022 CEUR Workshop Proc.* vol. 3138, pp. 1–10. CEUR-WS.org (2022)
10. Ginsca, A.L., Popescu, A., Lupu, M., et al.: Credibility in information retrieval. *foundations and trends® in information retrieval* **9**(5), 355–475 (2015)
11. Gupta, S., Nagar, S., Nanavati, A.A., Dey, K., Barbhuiya, F.A., Mukherjea, S.: Consumption of Hate Speech on Twitter: A Topical Approach to Capture Networks of Hateful Users. In: *ROMCIR 2021 CEUR Workshop Proc.* vol. 2838, pp. 25–34 (2021)
12. Gupta, V., Beckh, K., Giesselbach, S., Wegener, D., Wi, T.: Supporting verification of news articles with automated search for semantically similar articles. In: *ROMCIR 2021 CEUR Workshop Proc.* vol. 2838, pp. 47–58 (2021)
13. Hatua, A., Mukherjee, A., Verma, R.M.: On the Feasibility of Using GANs for Claim Verification- Experiments and Analysis. In: *ROMCIR 2021 CEUR Workshop Proc.* vol. 2838, pp. 35–46 (2021)
14. Heydari, A., ali Tavakoli, M., Salim, N., Heydari, Z.: Detection of review spam: A survey. *Expert Systems with Applications* **42**(7), 3634–3642 (2015)
15. Huang, Y., Xu, Q., Wu, S., Nugent, C., Moore, A.: Fight against COVID-19 misinformation via clustering-based subset selection fusion methods. In: Petrocchi, M., Viviani, M. (eds.) *ROMCIR 2022 CEUR Workshop Proc.* vol. 3138, pp. 11–26. CEUR-WS.org (2022)
16. Janzen, S., Orr, C., Terp, S.: Cognitive security and resilience: A social ecological model of disinformation and other harms with applications to COVID-19 vaccine information behaviors. In: Petrocchi, M., Viviani, M. (eds.) *ROMCIR 2022 CEUR Workshop Proc.* vol. 3138, pp. 48–88. CEUR-WS.org (2022)
17. Kirdemir, B., Adeliyi, O., Agarwal, N.: Towards characterizing coordinated inauthentic behaviors on YouTube. In: Petrocchi, M., Viviani, M. (eds.) *ROMCIR 2022 CEUR Workshop Proc.* vol. 3138, pp. 100–116. CEUR-WS.org (2022)
18. Mattei, M., Pratelli, M., Caldarelli, G., Petrocchi, M., Saracco, F.: Bow-tie structures of Twitter discursive communities. *Scientific Reports* **12**, 12944 (2022)
19. Pankovska, E., Schulz, K., Rehm, G.: Suspicious sentence detection and claim verification in the COVID-19 domain. In: Petrocchi, M., Viviani, M. (eds.) *ROMCIR 2022 CEUR Workshop Proc.* vol. 3138, pp. 27–47. CEUR-WS.org (2022)
20. Pasi, G., Viviani, M.: Information credibility in the social web: Contexts, approaches, and open issues. *arXiv preprint arXiv:2001.09473* (2020)

21. Petrocchi, M., Viviani, M.: Overview of ROMCIR 2022: The 2nd Workshop on Reducing Online Misinformation through Credible Information Retrieval. In: ROMCIR 2022 CEUR Workshop Proc. vol. 3138, pp. i–vii (2022)
22. Saracco, F., Viviani, M.: Overview of ROMCIR 2021: Workshop on Reducing Online Misinformation through Credible Information Retrieval. In: ROMCIR 2021 CEUR Workshop Proc. vol. 2838, pp. i–vii (2021)
23. Schmitt, J.B., Debbelt, C.A., Schneider, F.M.: Too much information? Predictors of information overload in the context of online news exposure. *Information, Communication & Society* **21**(8), 1151–1167 (2018)
24. Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., Liu, Y.: Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology (TIST)* **10**(3), 1–42 (2019)
25. Upadhyay, R., Pasi, G., Viviani, M.: Vec4cred: a model for health misinformation detection in web pages. *Multimedia Tools and Applications* pp. 1–20 (2022)
26. Viviani, M., Pasi, G.: Credibility in social media: opinions, news, and health information—a survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery* **7**(5), e1209 (2017)
27. Wardle, C., Derakhshan, H.: Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe* **27** (2017)
28. Wardle, C., Derakhshan, H., et al.: Thinking about ‘information disorder’: Formats of misinformation, disinformation, and mal-information. Ireton, Cheryl; Posetti, Julie. *Journalism, ‘fake news’ & disinformation*. Paris: Unesco pp. 43–54 (2018)
29. Zhang, Y., Bumber, D., Hosseinia, M., Yang, Mukherjee, A.: Improving Authorship Verification using Linguistic Divergence. In: ROMCIR 2021 CEUR Workshop Proc. vol. 2838, pp. 1–15 (2021)
30. Zhou, X., Zafarani, R.: A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)* **53**(5), 1–40 (2020)