

Overview of ROMCIR 2022: The 2nd Workshop on Reducing Online Misinformation through Credible Information Retrieval

Marinella Petrocchi^{a,b}, Marco Viviani^c

^aIIT-CNR, Via G. Moruzzi, 1 – 56124 Pisa, Italy

^bIMT School for Advanced Studies, Piazza San Francesco, 19 – 55100 Lucca, Italy

^cUniversity of Milano-Bicocca (DISCo – IKR3 Lab), Edificio U14, Viale Sarca, 336 – 20126 Milan, Italy

Abstract

The 2022 Workshop on Reducing Online Misinformation through Credible Information Retrieval (ROMCIR 2022), at its Second Edition as part of the Satellite Events of the 44th European Conference on Information Retrieval (ECIR 2022), is concerned with providing users with access to genuine information, to mitigate the information disorder phenomenon characterizing the current online environment. This problem is very broad, as it concerns different information objects (e.g., Web pages, online accounts, social media posts, etc.) on different platforms, and different domains and purposes (e.g., detecting fake news, retrieving genuine health-related information, reducing propaganda and hate-speech, etc.). In this context, all those approaches that can serve, from different perspectives, to tackle the genuine information access problem, find their place. In particular, this year articles have been submitted that addressed the problem of preventing access to health misinformation and assessing the genuineness of multi-modal information.

Keywords

Information Disorder, Information Genuineness, Information Retrieval, Health Misinformation, Multi-modal Information

1. Motivations

Coming into contact with different types of misleading content that is propagated online, especially through social media platforms [1, 2], can lead to several problems for individuals and society as a whole, as we have experienced especially in recent years. False news can, for example, influence public opinion in political and financial choices [3, 4]; false reviews can promote substandard products or, on the contrary, damage florid economic activities by means on discredit campaigns [5, 6]; unverified health information can lead people to follow behaviors that can be harmful both individually and globally [7, 8, 9, 10] (let us think, for example, of the risk of following negationism hypotheses in the context of the recent COVID-19 pandemic [11]).

This scenario is due to the so-called *information disorder* phenomenon [12], which indicates the proliferation of different forms of (online) communication pollution, encompassing *dis-*, *mis-*, and *mal-information*. Specifically, *misinformation* is the spread of false content resulting from the spreader's

ROMCIR 2022: The 2nd Workshop on Reducing Online Misinformation through Credible Information Retrieval, held as part of ECIR 2022: the 44th European Conference on Information Retrieval


✉ marinella.petrocchi@iit.cnr.it (M. Petrocchi); marco.viviani@unimib.it (M. Viviani)

🌐 <https://www.iit.cnr.it/en/marinella.petrocchi> (M. Petrocchi); <http://www.ir.disco.unimib.it/people/marco-viviani/> (M. Viviani)

📞 0000-0003-0591-877X (M. Petrocchi); 0000-0002-2274-9050 (M. Viviani)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

ignorance; *disinformation* is a form of intentional sharing of false content to produce harm; *malinformation* indicates the spread of (private) information that is based on reality, having the same harmful intent (e.g., the despicable act of revenge porn).

Access to this non-genuine information is made easier and easier due to the fact that, from a technological point of view, information is produced at a speed and volume never seen before, almost without any trusted traditional intermediary [13, 14]. Faced with this huge amount of information, and the uncertainty associated with its degree of genuineness, human cognitive abilities are not always sufficient to take well-informed decisions [15].

In this context, it is clear that the problem of guaranteeing access to genuine information online needs to find effective solutions, despite (and precisely because of) it is very broad, as it concerns different information objects (e.g., Web pages, online accounts, social media posts, etc.), different online platforms (e.g., Web portals, social networking services, question-answering systems, etc.), and different domains and purposes (e.g., detecting fake news, retrieving genuine health-related information, reducing propaganda and hate-speech, etc.).

2. Scientific Objectives

Within the ECIR conference, the key goal of the Workshop is to encourage a discussion between researchers, also belonging to different disciplines, and propose innovative approaches, about the problem of guarantee to users access to genuine information that does not distort their perception of reality, through Information Retrieval solutions [16]. In recent years, despite numerous approaches have been proposed to tackle the considered issue in different contexts, and for different purposes, we are still a long way from having found completely effective and domain-independent solutions.

The problem is still of great interest with respect to many research directions, such as the access to and retrieval of genuine information, the early detection of dis-/mis-/mal-information, the development of solutions that can be understood by final users (explainable AI), the study of the problem in the health domain, the study of the relationship between security, privacy and genuineness in information access and dissemination, the consideration of multi-modal information in assessing genuineness.

In this scenario, the role of researchers working in the fields of Information Retrieval, Social Computing, Social Sciences, Data and Web Science and other related research areas, is crucial to investigate the above-mentioned research directions.

3. Topics of Interest

All those approaches that can serve, from different perspectives, to tackle the genuine information access problem, find their place in ROMCIR 2022. Specifically, the topics of interest include, but are not limited to:

- Access to/retrieval of genuine information
- Bias detection
- Bot/spam/troll detection
- Computational fact-checking
- Crowdsourcing for information genuineness assessment

- Deep fakes and multi-modal misinformation detection
- Dis/misinformation detection
- Evaluation strategies in assessing dis/misinformation
- Fake news/reviews detection
- Filter bubble/echo chamber/polarization detection
- Harassment/bullying/hate-speech/propaganda detection
- Security, privacy and information genuineness
- Sentiment/emotional analysis and stance detection
- Trust and reputation systems
- Understanding and guiding the societal reaction in the presence of dis/misinformation

Both theoretical studies, model-driven, and data-driven approaches, supported by publicly available datasets, are more than welcome.

4. Submissions

The ROMCIR 2022 Workshop received 10 submissions, of which 6 were accepted, so with an acceptance rate of 60%. Articles have been submitted from four different countries, i.e., Germany, India, UK, and USA. The accepted articles, collected in these Proceedings, have primarily considered two issues from distinct points of view. The first issue concerns access to genuine information in the health-related domain; the second issue concerns access to multi-modal genuine information.

With respect to the first issue, in the article by Fröbe *et al.* entitled: “Using Keyqueries to Reduce Misinformation in Health-Related Search Results”, the authors investigate whether explicit relevance feedback provided by experts can guide query expansion methods to formulate queries that return fewer misleading or wrong results. In contrast to standard query expansion methods that pay no attention to feedback document ranks in expanded query results, the authors experiment with a keyquery-based approach to identify expanded queries for which feedback documents are ranked as high as possible. In the article by Huang *et al.* entitled: “Fight Against COVID-19 Misinformation via Clustering-Based Subset Selection Fusion Methods”, the authors try to improve the retrieval quality of search engines w.r.t. different relevance dimensions via a data fusion technique. In particular, a clustering-based approach is proposed for selecting a subset of IR systems from all the available ones, for finding the most relevant, credible, and correct documents talking about COVID-19. Still in the health domain, the article by Pankovska *et al.*, entitled: “Suspicious Sentence Detection and Claim Verification in the COVID-19 Domain”, addresses the claim verification and fact-check-worthiness issues, by proposing an approach that involves a two-step claim verification procedure consisting of a fake news detection task in the form of binary sequence classification and fact-checking using the Google Fact Check Tools. The aim of the authors is to alert the reader that a document contains suspicious statements, even if no already validated similar claims exist. The last article that deals with health misinformation, this time from a more high-level perspective, is the one by Janzen *et al.* entitled “Cognitive security and resilience: A social-ecological model of disinformation and other harms with applications to COVID-19 vaccine information behaviors”. In this paper, the authors propose a

novel application of Brofenbrenner’s social ecological model to the study of cognitive security and resilience in the context of information disorder. First, they describe the refitting of the model from public health and human development to cognitive security. Using existing literature in the field, they identify the key factors at each level of influence that shape susceptibility and resilience to information disorder. They also consider the dynamic interactions between individuals, groups, societies, and characteristics of the technological environment, including how algorithms interact with individual behaviors, policies, and organizational decision-making to shape access to and discoverability of genuine information. Finally, they describe an application of the model to a use case involving COVID-19-related information behaviors.

Coming to the second problem addressed by the articles submitted to ROMCIR 2022, namely that of multi-modal genuine information access, the article by Aghada entitled: “An Alternative Approach to Ranking Videos and Measuring Dissimilarity Between Video Content and Titles”, proposes a statistical approach to video retrieval and ranking by introducing a novel dissimilarity measure acting on a video’s audio-visual content and its title, hence aiding in video click-bait detection. Finally, in the paper by Kirdemir *et al.* entitled: “Towards Detecting Coordinated Inauthentic Behaviors on YouTube”, the authors aim to explore new approaches to assess latent and implicit characteristics of coordination among users in YouTube channels that can indicate manipulation of information and communication. In particular, they propose computational models leveraging multi-step time-series analysis of engagement trends, network structural feature-based analysis, and a combination of unsupervised and supervised machine learning techniques.

5. Keynote Speech

As part of the Workshop, a Keynote Speech entitled “Fake News Detection: Limited Ground Truth, Limited Text, No Understanding of Spreading Intent” was given by Reza Zafarani, on current open issues that still concern fake news detection.

Abstract: “Fake news” is now viewed as one of the greatest threats to democracies and journalism. The massive spread of fake news has weakened public trust in governments and its potential impact on various political outcomes such as the Brexit is yet to be realized. We will briefly review fake news detection techniques, along with some of the current challenges that these methods face. We will discuss some recent advancements to tackle these challenges, particularly focusing on multi-modal fake news analysis and assessing the intent of fake news spreaders.



Reza Zafarani is an Assistant Professor of Electrical Engineering and Computer Science at Syracuse University. His research interests are in Data Mining, Machine Learning, Social Media Mining, and Social Network Analysis. His research has been published in major academic venues and highlighted in various scientific and news outlets. He is the principal author of “Social Media Mining: An Introduction” a textbook by Cambridge University Press and the associate editor for SIGKDD Explorations and Frontiers in communication. He is the winner of the NSF CAREER award, President’s Award for Innovation, and outstanding teaching award at Arizona State University. *Web site:* <http://reza.zafarani.net/>

Web site: <http://reza.zafarani.net/>

6. Organizing Team

The ROMCIR 2022 Organizing Team was composed of the following people with respect to their distinct roles.

6.1. Co-chairs



Marinella Petrocchi is a Senior Researcher at the Institute of Informatics and Telematics, part of the National Research Council (IIT-CNR), Pisa, Italy, under the Trust, Security and Privacy research unit. She collaborates with the Sysma unit at IMT School for Advanced Studies, in Lucca, Italy. Her field of research lies between Cybersecurity and Data Science. She studies novel techniques for online fake news/fake accounts detection. She is in the core team of the *TOols for Fighting FakeEs* (TOFFEE) project, funded by IMT, and WP leader in H2020 Medina, where she studies how to automatically translate NL cloud security requirements to machine-readable, enforceable policies. *Web site:* <https://www.iit.cnr.it/en/marinella.petrocchi/>



Marco Viviani is an Associate Professor at the University of Milano-Bicocca, Department of Informatics, Systems, and Communication (DISCo), Milan, Italy. He works in the Information and Knowledge Representation, Retrieval and Reasoning (IKR3) Lab. He is involved in numerous research initiatives that involve accessing and retrieving information, especially genuine information. He has been Co-chair of several Special Tracks and Workshops at International Conferences, General Co-chair of MDAI 2019, and Co-organizer of the First Edition of the ROMCIR Workshop. He is Associate Editor of *Social Network Analysis and Mining* (SNAM), Springer-Verlag, and Editorial Board Member of *Online Social Networks and Media* (OSNEM), Elsevier. His main research activities include Social Computing, Information Retrieval, Natural Language Processing, Text Mining, and User Modeling. On these topics, he has published more than 80 research works in International Journals, at International Conferences, as Monographs, and Book Chapters. *Web site:* <https://ikr3.disco.unimib.it/people/marco-viviani/>

6.2. Publicity and Publication Chair

Rishabh Upadhyay, PhD Student. University of Milano-Bicocca, Department of Informatics, Systems, and Communication (DISCo), Milan, Italy.

Web site: <https://ikr3.disco.unimib.it/people/rishabh-upadhyay/>

6.3. Program Committee

- **Rino Falcone**, Inst. of Cognitive Sciences and Technologies (ISTC) – CNR, Rome, Italy
- **Carlos A. Iglesias**, Universidad Politécnica de Madrid, Madrid, Spain
- **Petr Knoth**, The Open University, London, UK
- **Udo Kruschwitz**, University of Regensburg, Regensburg, Germany

- **Preslav Nakov**, Qatar Computing Research Institute, HBKU, Doha, Qatar
- **Symeon Papadopoulos**, Inf. Tech. Inst. (ITI), Thessaloniki, Greece
- **Marinella Petrocchi**, Inst. of Informatics and Telematics (IIT) – CNR, Pisa, Italy
- **Francesco Pierri**, Politecnico di Milano, Milan, Italy
- **Adrian Popescu**, CEA LIST, Gif-sur-Yvette, France
- **Paolo Rosso**, Universitat Politècnica de València, València, Spain
- **Fabio Saracco**, Centro Ricerche Enrico Fermi (CREF), Florence, Italy
- **Marco Viviani**, University of Milano-Bicocca, Milan, Italy
- **Xinyi Zhou**, Syracuse University, Syracuse, NY, USA
- **Arkaitz Zubiaga**, Queen Mary University of London, London, UK

Acknowledgments

We would like to thank the authors of the submitted articles for their interest in the considered problem, the Keynote Speaker for the interest aroused in new research directions, and the members of the Program Committee for their valuable contribution to the success of the ROMCIR 2022 Workshop.

More information about the 2022 Edition of the ROMCIR Workshop can be found at the following URL: <https://romcir2022.disco.unimib.it/>

References

- [1] G. Pasi, M. Viviani, Information credibility in the social web: Contexts, approaches, and open issues, arXiv preprint arXiv:2001.09473 (2020).
- [2] M. Viviani, G. Pasi, Credibility in social media: opinions, news, and health information—a survey, WIREs: Data Mining and Knowl. Discovery 7 (2017) e1209.
- [3] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, et al., The science of fake news, Science 359 (2018) 1094–1096.
- [4] G. Caldarelli, R. De Nicola, F. Del Vigna, et al., The role of bot squads in the political propaganda on Twitter, Commun Phys 3 (2020).
- [5] R. Barbado, O. Araque, C. A. Iglesias, A framework for fake review detection in online consumer electronics retailers, Information Processing & Management 56 (2019) 1234–1244.
- [6] Y. Wu, E. W. Ngai, P. Wu, C. Wu, Fake online reviews: Literature review, synthesis, and directions for future research, Decision Support Systems (2020) 113280.
- [7] F. Tagliabue, L. Galassi, P. Mariani, The “pandemic” of disinformation in covid-19, SN Comprehensive Clinical Medicine 2 (2020) 1287–1289.
- [8] R. Upadhyay, G. Pasi, M. Viviani, Health misinformation detection in web content: A structural-, content-based, and context-aware approach based on web2vec, in: Proceedings of the Conference on Information Technology for Social Good, 2021, pp. 19–24.

- [9] S. Di Sotto, M. Viviani, Health misinformation detection in the social web: An overview and a data science approach, *International Journal of Environmental Research and Public Health* 19 (2022) 2173.
- [10] G. Caldarelli, R. De Nicola, M. Petrocchi, M. Pratelli, F. Saracco, Flow of online misinformation during the peak of the COVID-19 pandemic in Italy, *EPJ Data Sci.* 10 (2021) 34. doi:10.1140/epjds/s13688-021-00289-4.
- [11] V. E. Valenti, A. P. d. Silva, The effect of negationism on public health, *Journal of Human Growth and Development* 31 (2021) 189–191.
- [12] C. Wardle, H. Derakhshan, Information disorder: Toward an interdisciplinary framework for research and policy making, *Council of Europe Report* 27 (2017).
- [13] B. Carminati, E. Ferrari, M. Viviani, Security and trust in online social networks, *Synthesis Lectures on Information Security, Privacy, & Trust* 4 (2013) 1–120.
- [14] G. Eysenbach, Credibility of health information and digital media: New perspectives and implications for youth, in: M. M. Metzger, A. J. Flanagin (Eds.), *Digital Media, Youth, and Credibility*, The MIT Press, 2008, pp. 123–154.
- [15] M. J. Metzger, A. J. Flanagin, Credibility and trust of information in online environments: The use of cognitive heuristics, *Journal of pragmatics* 59 (2013) 210–220.
- [16] A. L. Ginsca, A. Popescu, M. Lupu, Credibility in information retrieval, *Foundations and Trends in Information Retrieval* 9 (2015) 355–475.