BeyondFacts'22: 2nd International Workshop on Knowledge Graphs for Online Discourse Analysis

Konstantin Todorov
todorov@lirmm.fr
LIRMM, University of Montpellier, CNRS
Montpellier, France

Pavlos Fafalios
fafalios@ics.forth.gr
Information Systems Laboratory, ICS-FORTH
Heraklion, Greece

Stefan Dietze
stefan.dietze@gesis.org
GESIS & Heinrich-Heine-University Düsseldorf
Cologne, Germany

ABSTRACT

Expressing opinions and interacting with others on the Web has led to an abundance of online discourse: claims and viewpoints on controversial topics, their sources and contexts. This constitutes a valuable source of insights for studies into mis- / disinformation spread, bias reinforcement, echo chambers or political agenda setting. While knowledge graphs promise to provide the key to a Web of structured information, they are mainly focused on facts without keeping track of the diversity, connection or temporal evolution of online discourse. As opposed to facts, claims and viewpoints are inherently more complex. Their interpretation strongly depends on the context and a variety of intentional or unintended meanings, where terminology and conceptual understandings strongly diverge across communities from computational social science, to argumentation mining, fact-checking, or viewpoint/stance detection. The 2nd International Workshop on Knowledge Graphs for Online Discourse Analysis (BeyondFacts'22, equivalently abbreviated as KnOD'22) aims at strengthening the relations between these communities, providing a forum for shared works on the modeling, extraction and analysis of discourse on the Web. It addresses the need for a shared understanding and structured knowledge about discourse in order to enable machine-interpretation, discoverability and reuse, in support of studies into the analysis of societal debates.

CCS CONCEPTS

• **Information systems** → **World Wide Web**; **Information retrieval**; **Data management systems**; • **Computing methodologies** → **Machine learning**; **Artificial intelligence**.

KEYWORDS

Online Discourse Analysis, Knowledge Graphs, Social Web Mining, Computational Fact-checking, Mis-/Dis-information Spread, Stance/Viewpoint Detection

BEYONDFACTS'22: A CROSS-DISCIPLINARY COMMUNITY

The Web is evolving into a ubiquitous platform giving the opportunity to everyone to publish content, express opinions and interact with others. Understanding online discourse hence becomes an increasingly important issue. We define online discourse as any kind of narrative, debate or conversation that happens on the Web, including social networks or news media, involving *claims* and *stances* on *controversial topics*, their *sources* and *contexts*, such as related events or entities.

Recently, a wide range of interdisciplinary research investigates discourse analysis by involving a variety of scientific disciplines. Such works are either focused on gaining new scientific insights, for instance, by studying the spreading patterns of false claims on Twitter [16], or they aim at computational methods, like pipelines for detecting the stance of claim-relevant Web documents [17], understanding/quantifying hidden biases [11], classifying sources of news [13], or research into *fake news detection* [14], and automatic *fact-checking* [2].

One crucial requirement to facilitate the aforementioned research is the availability of reliable structured knowledge about the key notions of discourse (e.g. claims, truth ratings, evidence, sources, arguments and their relations). On the one hand, initiatives such as the schema.org Claim Review vocabulary[1] aim at encouraging website providers to offer such data through Web markup. On the other hand, knowledge graphs (KGs) such as MultiFC [3], ClaimsKG [15][2], TweetsKB [10][3], or TweetsCOV19 [8][4] have been proposed aimed at consolidating Web-mined data about online discourse. While KGs promise to provide the key to a Web of structured information, they are mainly focused on facts without keeping track of the diversity, connection or temporal evolution of discourse. As opposed to facts, claims are inherently more complex. Their interpretation strongly depends on the context and a variety of intentional or unintended meanings, where terminology and conceptual understandings strongly diverge across communities from computational social science, to argumentation mining, fact-checking, or stance detection [5, 6].

Initial efforts have been made to gather communities working in those areas, for instance through dedicated challenges, such

---

[1] http://schema.org/ClaimReview
[2] https://data.gesis.org/claimskg/
[3] https://data.gesis.org/tweetskb/
[4] https://data.gesis.org/tweetscov19/

WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France.

Konstantin Todorov, Pavlos Fafalios, and Stefan Dietze

as the Fake News Challenge,[5] or sessions at major conferences, such as the Journalism, Misinformation and Fact Checking track at TheWebConf 2018.[6] The BeyondFacts'22 Workshop brings together the various disciplines involved in or benefiting from (a) approaches for representing online discourse and involved notions, (b) methods for mining these notions, like for instance, claims, stances, sources and their relations from the Web, and (c) inter-disciplinary research investigating online discourse.

Beyond research into information and knowledge extraction, data modeling and consolidation for KG building, the workshop targets communities focusing on the analysis of online discourse, relying on methods from *Machine Learning (ML), Natural Language Processing (NLP)* and *Data Mining (DM)*. These include communities centered around:

- discourse analysis
- social web mining
- argumentation mining
- computational fact-checking
- mis- and dis-information spread
- bias and controversy detection and analysis
- stance/viewpoint detection and representation
- opinion mining
- rumour, propaganda and hate-speech detection
- computational social science
- computational journalism

Hence, BeyondFacts'22 provides a meeting point for these related but distinct communities that address similar or closely related questions from different perspectives and in different fields, using different models and definitions of the main notions of interest. The workshop aims at strengthening the relations between these communities, providing a forum for shared works on the modeling, extraction and analysis of discourse on the Web. It addresses the need for a shared understanding and structured knowledge about discourse data in order to enable machine-interpretation, discoverability and reuse, in support of scientific or journalistic studies into the analysis of societal debates on the Web. Often the aforementioned communities apply their research in particular domains, such as scientific publishing, medicine, journalism or social science. Therefore, the workshop is particularly interested in works that apply an *interdisciplinary* approach, such as works on *computational social sciences* or *computational journalism*.

The BeyondFacts'22 workshop continues its successful first edition (BeyondFacts'21 / KnOD 2021)[7] which took place as a virtual event jointly with the 30th Web Conference (WWW 2021).[8]

WORKSHOP OVERVIEW

The BeyondFacts'22 workshop[9] takes place as a virtual event jointly with the 31st Web Conference (WWW 2022)[10], as it closely relates to the topics of the venue in terms of the nature of the analysed data and the targeted communities. In particular, it complements and *bridges* a number of research tracks of the conference, such as

"Semantics and Knowledge", "Web and Society", "Social Web", "Web Mining and Content Analysis" and "Social Network Analysis and Graph Algorithms". BeyondFacts'22 also fits into and continues a line of former WebConf forums such as the Fact Checking track in 2018 or the workshops (and a special track in 2019) on Data Science for Social Good.

ACCEPTED PAPERS AND KEYNOTE SPEAKERS OVERVIEW

The second edition of the BeyondFacts workshop brings together a diverse community of researchers from different fields such as stance detection, fact verification, bias analysis and knowledge graph building. Five papers were accepted for publication (two long papers and three short ones) after a peer-review process[11], spanning a palette of topics including: fact verification [4], bias analysis in recommendation systems based on stance detection [1], enriching COVID19 knowledge base with geographic information [9], twitter reaction trends and the identification of disinformation communities [7], and building knowledge graphs for science [12].

In addition, we are very happy to host three excellent keynotes. Serena Villata (Laboratoire I3S, CNRS, Sophia Antipolis, France) gives a talk on argument-based explanatory dialogues: from argument mining to argument generation; Harith Alani (Knowledge Media Institute, The Open University, UK) talks about the misinformation tipping point – a computational look into the dynamics of false and corrective information; finally, José Manuel Gómez Pérez (Expert.AI, Spain) looks into the problem and limitations of an accurate and explainable misinformation detection.

The current volume contains the five accepted papers as well as the abstracts of the three keynote talks.

REFERENCES

[1] Mehwish Alam, Andreea Iana, Alexander Grote, Katharina Ludwig, Philipp Müller, and Heiko Paulheim. 2022. Towards Analyzing the Bias of News Recommender Systems Using Sentiment and Stance Detection. In *Companion Proceedings of the Web Conference 2022*. ACM.

[2] Pepa Atanasova, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, Georgi Karadzhov, Tsvetomila Mihaylova, Mitra Mohtarami, and James Glass. 2019. Automatic fact-checking using context and discourse information. *Journal of Data and Information Quality (JDIQ)* 11, 3 (2019), 1–27.

[3] Isabelle Augenstein, Christina Lioma, Dongsheng Wang, Lucas Chaves Lima, Casper Hansen, Christian Hansen, and Jakob Grue Simonsen. 2019. MultiFC: A Real-World Multi-Domain Dataset for Evidence-Based Fact Checking of Claims. In *Proceedings of the 2019 EMNLP and the 9th IJCNLP*. 4677–4691.

[4] Anab Maulana Barik, Wynne Hsu, and Mong Li Lee. 2022. Incorporating External Knowledge for Evidence-based Fact Verification. In *Companion Proceedings of the Web Conference 2022*. ACM.

[5] Katarina Boland, Pavlos Fafalios, Andon Tchechmedjiev, Stefan Dietze, and Konstantin Todorov. 2022. Beyond Facts–A Survey and Conceptualisation of Claims in Online Discourse Analysis. *Semantic Web* (2022).

[6] Katarina Boland, Pavlos Fafalios, Andon Tchechmedjiev, Konstantin Todorov, and Stefan Dietze. 2019. Modeling and contextualizing claims. In *2nd International Workshop on Contextualized Knowledge Graphs (CKG 2019 @ISWC)*.

[7] Debraj De, Gautam Thakur, Drahomira Herrmannova, and Christopher Carter. 2022. Methodology to Compare Twitter Reaction Trends between Disinformation Communities, to COVID related Campaign Events at Different Geospatial Granularities. In *Companion Proceedings of the Web Conference 2022*. ACM.

[8] Dimitar Dimitrov, Erdal Baran, Pavlos Fafalios, Ran Yu, Xiaofei Zhu, Matthäus Zloch, and Stefan Dietze. 2020. TweetsCOV19-A Knowledge Base of Semantically Annotated Tweets about the COVID-19 Pandemic. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2991–2998.

[9] Dimitar Dimitrov, Dennis Segeth, and Stefan Dietze. 2022. Geotagging TweetsCOV19: Enriching a COVID-19 Twitter Discourse Knowledge Base with

---

[5]http://www.fakenewschallenge.org/

[6]https://www2018.thewebconf.org/call-for-papers/misinformation-cfp/

[7]https://knod2021.wordpress.com/

[8]https://www2021.thewebconf.org/

[9]https://knod22.wordpress.com/

[10]https://www2022.thewebconf.org/

---

[11]Each submitted paper was reviewed by three programme committee members.

Geographic Information. In *Companion Proceedings of the Web Conference 2022*. ACM.

[10] Pavlos Fafalios, Vasileios Iosifidis, Eirini Ntoutsi, and Stefan Dietze. 2018. Tweet-skb: A public and large-scale rdf corpus of annotated tweets. In *European Semantic Web Conference*. Springer, 177–190.

[11] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Quantifying controversy on social media. *ACM Trans. on Soc. Comp.* 1, 1 (2018), 3.

[12] Anh Le-Tuan, Carlos Franzreb, Danh Le Phuoc, Sonja Schimmler, and Manfred Hauswirth. 2022. Towards Building Live Open Scientific Knowledge Graphs. In *Companion Proceedings of the Web Conference 2022*. ACM.

[13] Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2017. Where the truth lies: Explaining the credibility of emerging claims on the web and social media. In *Proceedings of the 26th International Conference on World Wide Web Companion*. 1003–1012.

[14] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter* 19, 1 (2017), 22–36.

[15] Andon Tchechmedjiev, Pavlos Fafalios, Katarina Boland, Malo Gasquet, Matthäus Zloch, Benjamin Zapilko, Stefan Dietze, and Konstantin Todorov. 2019. Claim-sKG: A knowledge graph of fact-checked claims. In *International Semantic Web Conference*. Springer, 309–324.

[16] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.

[17] Xuezhi Wang, Cong Yu, Simon Baumgartner, and Flip Korn. 2018. Relevant Document Discovery for Fact-Checking Articles. In *WWW*. 525–533.

APPENDIX

Programme Committee of BeyondFacts'22

- Katarina Boland, GESIS, Germany
- Alexander Brand, University of Hildesheim, Germany
- Sandra Bringay, Paul Valéry University of Montpellier, France
- Ronald Denaux, Amazon.com, Spain
- Gianluca Demartini, University of Queensland, Australia
- Vasilis Efthymiou, FORTH, Greece
- Michael Färber, Karlsruhe Institute of Technology, Germany
- Mario Haim, University of Leipzig, Germany
- Kyle Hamilton, Technological University Dublin, Ireland
- Daniel Hardt, Copenhagen Business School, Denmark
- Julio Amador Diaz Lopez, Imperial College London, UK
- Petr Motlicek, Idiap Research Institute, Switzerland
- Panagiotis Papadakos, FORTH, Greece
- José Manuel Gómez Pérez, Expert.AI, Spain
- Achim Rettinger, University of Trier, Germany
- Kostas Stefanidis, Tampere University, Finland
- Daniel Schwabe, Pontificia Universidade Católica, Brazil
- Pedro Szekely, University of South California, USA
- Andon Tchechmedjiev, Ecoles des Mines d'Alès, France
- Yannis Tzitzikas, FORTH, Greece
- Xiaofei Zhu, Chongqing University of Technology, China