

# Beyond Facts: Online Discourse and Knowledge Graphs

## A preface to the proceedings of the 1st International Workshop on Knowledge Graphs for Online Discourse Analysis (KnOD 2021, co-located with TheWebConf'21)

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@geis.org  
GEIS & Heinrich-Heine-University  
Düsseldorf, Cologne, Germany

### ABSTRACT

Expressing opinions and interacting with others on the Web has led to the production of an abundance of online discourse data, such as claims and viewpoints on controversial topics, their sources and contexts. This data constitutes a valuable source of insights for studies into misinformation 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 data. 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 viewpoint/stance detection. The 1st International Workshop on Knowledge Graphs for Online Discourse Analysis (KnOD 2021) 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.

### KEYWORDS

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

### BEYOND FACTS: A CROSS-DISCIPLINARY COMMUNITY

With the Web evolving into a ubiquitous platform giving the opportunity to everyone to publish content, express opinions and interact with others, understanding online discourse has become 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 directions are being explored involving a variety of scientific disciplines. Such works either are focused on gaining new scientific insights, for instance, by investigating the spreading pattern of false claims on Twitter [11], or they aim at computational methods, for instance, pipelines for detecting the stance of claim-relevant Web documents

[12], understanding/quantifying hidden biases [7], approaches for classifying sources of news, such as Web pages, pay-level domains, users or posts [8], or research into *fake news detection* [9], and automatic *fact-checking* [1].

One crucial requirement to facilitate the aforementioned research areas is the availability of reliable structured knowledge about key notions such as claims, truth ratings, evidence, sources, arguments and their relations. On the one hand, initiatives such as the schema.org Claim Review vocabulary<sup>1</sup> aim at encouraging website providers to offer such data through embedded Web markup. On the other hand, initial knowledge graphs (KG) such as MultiFC [2], ClaimsKG [10]<sup>2</sup>, TweetsCOV19 [5]<sup>3</sup> or TweetsKB [6]<sup>4</sup> have been proposed aimed at consolidating Web-mined data about the aforementioned notions. While knowledge graphs (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 online 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 viewpoint/stance detection [3, 4].

Initial efforts have been made to gather communities working in those areas, for instance through dedicated challenges, such as the Fake News Challenge,<sup>5</sup> or sessions at major conferences, such as the Journalism, Misinformation and Fact Checking track at The Web Conf 2018.<sup>6</sup> The KnOD Workshop brings together the various disciplines involved in or benefitting from (a) approaches for representing online discourse and involved notions, (b) methods for mining such notions (for instance, claims, stances, sources, etc.) and their relations from the Web, and (c) inter-disciplinary research investigating online discourse.

Beyond research into information and knowledge extraction, and 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 on:

- discourse analysis
- social web mining

<sup>1</sup><http://schema.org/ClaimReview>

<sup>2</sup><https://data.geis.org/claimskg/site>

<sup>3</sup><https://data.geis.org/tweetscov19/>

<sup>4</sup><https://data.geis.org/tweetskb/>

<sup>5</sup><http://www.fakenewschallenge.org/>

<sup>6</sup><https://www.2018.thewebconf.org/call-for-papers/misinformation-cfp/>

- 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

Hence, KnOD 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*.

## WORKSHOP OVERVIEW

The **KnOD 2021 workshop**<sup>7</sup> took place as a virtual event (due to COVID-19 outbreak) jointly with the **30th The Web Conference (WWW 2021)**<sup>8</sup>, 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", "Web Mining and Content Analysis" and in part "Social Network Analysis and Graph Algorithms". KnOD 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.

This first edition of the KnOD workshop brought together a diverse community of researchers from different fields such as argument mining, knowledge graphs and neural language models or databases, but also social and political science. Seven papers were accepted for publication (2 long papers and 5 short ones) after a peer-review process,<sup>9</sup> spanning a palette of topics such as claim detection, relation extraction for online discourse modeling, interpretable graph embeddings for misinformation detection, disinformation on social networks, fact-checking in relation to argumentation schemes and false narratives, political and social scientific perspectives on propaganda chains and discourse mapping. The current volume contains the seven accepted papers.

In addition, we were very happy to host three excellent keynotes: **Preslav Nakov** talked about detecting 'Fake News' before it was even written, media literacy and flattening the curve of the COVID-19 infodemic; **Daniel Schwabe** proposed his take on trust and

information disorders seen as disputes of narratives, while **Ioana Manolescu** gave an overview and lessons learned from the ANR ContentCheck project, focusing on content management approaches for assisting journalists in their day-to-day fact-checking efforts.

We consider the first edition of the workshop a successful first step towards fostering a community on discourse analysis via structured knowledge in the context of the Web. We would like to warmly thank all authors and keynote speakers for their contributions, participation and exciting discussions during the workshop day. We also thank the members of the programme committee (see Appendix) for their constructive reviews, as well as the WebConf 2021 organizers and workshop chairs for their cooperation and support.

## REFERENCES

- [1] 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.
- [2] 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.
- [3] Katarina Boland, Pavlos Fafalios, Andon Techechmedjiev, Stefan Dietze, and Konstantin Todorov. 2021. Beyond Facts—A Survey and Conceptualisation of Claims in Online Discourse Analysis. *Under review for the Semantic Web Journal* (2021).
- [4] Katarina Boland, Pavlos Fafalios, Andon Techechmedjiev, Konstantin Todorov, and Stefan Dietze. 2019. Modeling and contextualizing claims. In *2nd International Workshop on Contextualized Knowledge Graphs (CKG 2019 @ISWC)*.
- [5] 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.
- [6] Pavlos Fafalios, Vasileios Iosifidis, Eirini Ntoutsis, and Stefan Dietze. 2018. Tweet-skb: A public and large-scale rdf corpus of annotated tweets. In *European Semantic Web Conference*. Springer, 177–190.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] Andon Techechmedjiev, 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.
- [11] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.
- [12] 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 KnOD 2021

- **Harith Alani**, KMI, The Open University, UK
- **Katarina Boland**, GESIS, Germany
- **Sandra Bringay**, Paul Valéry University of Montpellier, France
- **Gianluca Demartini**, University of Queensland, Australia
- **Ronald Denaux**, Expert.AI, Spain
- **Vasilios Efthymiou**, FORTH, Greece
- **Michael Färber**, Karlsruhe Institute of Technology, Germany
- **Jose Manuel Gomez-Perez**, Expert.AI, Spain

<sup>7</sup><https://knod2021.wordpress.com/>

<sup>8</sup><https://www2021.thewebconf.org/>

<sup>9</sup>Each submitted paper was reviewed by 3 programme committee members.

- **Daniel Hardt**, Copenhagen Business School, Denmark
- **Ioana Manolescu**, INRIA Saclay and LIX/Ecole Polytechnique, France
- **Preslav Nakov**, Qatar Computing Research Institute, Qatar
- **Panagiotis Papadakos**, FORTH, Greece
- **Rajesh Piryani**, South Asian University, New Delhi, India
- **Daniel Schwabe**, Pontificia Universidade Católica do Rio de Janeiro, Brazil
- **Kostas Stefanidis**, Tampere University, Finland
- **Pedro Szekely**, University of Southern California, USA
- **Andon Tchechmedjiev**, Ecoles des Mines d'Alès, France
- **Yannis Tzitzikas**, FORTH and University of Crete, Greece
- **Xiaofei Zhu**, Chongqing University of Technology, China