Beyond Facts: 4th International Workshop on Computational Methods for Online Discourse Analysis

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

Pavlos Fafalios
pfafalios@tuc.gr
Technical University of Crete, FORTH-ICS,
Chania, Greece

Stefan Dietze
stefan.dietze@gesis.org
GESIS, Cologne, Germany

Dimitar Dimitrov
dimitar.dimitrov@gesis.org
GESIS, Cologne, Germany

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; **Natural language processing**; • **Information systems** → **World Wide Web**.

KEYWORDS

Online discourse analysis, Language models, Knowledge graphs, Social web mining, Computational fact-checking, Mis- and disinformation spread and detection, Stance / viewpoint discovery, intent detection, computational journalism

ACM Reference Format:
Konstantin Todorov, Pavlos Fafalios, Stefan Dietze, and Dimitar Dimitrov. 2024. Beyond Facts: 4th International Workshop on Computational Methods for Online Discourse Analysis. In *Companion Proceedings of the ACM Web Conference 2024 (WWW '24 Companion), May 13–17, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3589335.3641296

1 SCOPE AND OBJECTIVES OF BEYONDFACTS

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 (events, entities). This data constitutes a valuable source of insights for studies into misinformation spread, bias reinforcement, echo chambers or political agenda setting. Computational methods, mostly from the field of NLP, have emerged that tackle a wide range of tasks in this context, including argument and opinion mining, claim detection, checkworthiness detection, stance detection or fact verification. However, computational models require robust definitions of classes and concepts under investigation. Thus, these computational tasks require a strong interdisciplinary and epistemological foundation, specifically with respect to the underlying definitions of key concepts such as claims, arguments, stances, check-worthiness or veracity. This requires a highly interdisciplinary approach combining expertise from fields such as communication studies, computational linguistics and computer science. 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 a computational perspective, in order to address this complexity, the synergy of multiple approaches, coming both from symbolic (knowledge representation) and statistical AI seem to be promising to tackle such challenges.

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, or, they aim at computational methods, for instance, pipelines for detecting the stance of claim-relevant Web documents, approaches for classifying sources of news, such as Web pages, pay-level domains (PLDs), users or posts, or research into **fake news detection** and automatic **fact-checking** and mining of facts and claims from the Web for knowledge base augmentation.

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[1] aim at encouraging website providers to offer such data through embedded Web markup. On the other hand, knowledge graphs (KG) such as ClaimsKG[2] have been proposed aimed at consolidating Web-mined data about the aforementioned notions.

Initial efforts have been made to gather communities working in those areas, for instance through dedicated challenges (such as the Fake News Challenge[3] or sessions at major conferences, such as the Journalism, Misinformation and Fact Checking track at The Web Conf 2018.[4] The BeyondFacts Workshop brings together the various disciplines involved in or benefitting from (a) approaches for representing online discourse and involved notions, (b) methods

WWW '24 Companion, May 13–17, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-0172-6/24/05.
https://doi.org/10.1145/3589335.3641296

[1] http://schema.org/ClaimReview
[2] https://data.gesis.org/claimskg/site
[3] http://www.fakenewschallenge.org/
[4] https://archives.iw3c2.org/www2018/call-for-papers/misinformation-cfp/

for mining such notions (for instance, claims, stances, sources, etc) and their relations from the Web, or (c) inter-disciplinary research investigating online discourse.

Beyond research into information and knowledge extraction, data consolidation and modeling, the workshop targets communities focusing on the analysis of online discourse, relying on methods from **machine learning (ML), natural language processing (NLP) and Web data mining (DM)**. These include communities on:

- 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 journalism

BeyondFacts 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. Often these 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**.

BeyondFacts fits the Web Conference 2024 in terms of both the nature of the analysed data and the targeted communities. In particular, it relates closely to, 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". It also complements the special track "Web4Good", while it fits into and continues a line of former WWW forums such as the Fact Checking track in 2018 or the workshops (and a special track in 2019) on Data Science for Social Good.

2 TOPICS OF INTEREST

The workshop focuses on the following main topics of interest:

- Large language models for online discourse
- Computational fact-checking / truth discovery
- Computational journalism
- Social, ethical and legal aspects of online discourse
- Bias and controversy detection and analysis
- Stance and viewpoint discovery
- Interpretability / explainability of computational methods for discourse analysis
- Rumour, propaganda and hate-speech detection
- Intent discovery for claims
- Integration, aggregation, linking and enrichment of discourse data
- Multilingual analysis of online discourse data
- Ontologies and data models for online discourse data
- Reuse and extension of existing models such as schema.org and Wikidata
- KGs and knowledge extraction techniques in the context of online discourse
- Semantic and exploratory search of online discourse data
- Argumentation and reasoning over online discourse
- Recommender systems for discourse data
- Quality, uncertainty, provenance, and trust of discourse data
- Dealing with online audiovisual content
- Benchmarks and training data for extraction, verification or linking of discourse data
- Use-cases, applications and cross-community interfaces

3 FORMAT

The workshop includes a mix of research papers, position papers, demo/system papers, resource papers, keynotes and invited talks. Note that we will accept position papers or papers with intermediate/incomplete results, the idea being to foster discussion as much as presenting finished results. The workshop will contain:

- Paper presentations (full, short, resource, position, posters, demos)
- Two one-hour poster sessions for all accepted contributions providing an opportunity for discussions and networking
- Two keynotes / invited talks from senior researchers on relevant topics (with one of them being highly interdisciplinary)
- Closing session and discussion on future directions and challenges, with a best paper award ceremony.

All contributions are eligible for the "Best BeyondFacts Paper" award, which will be awarded to the best contribution during the closing session. In addition, the workshop organisation team is planning to invite selected workshop submissions to a dedicated journal special issue (to be specified).

The 2024 edition of BeyondFacts received 11 submissions, out of which 7 were accepted (1 short paper and 6 full papers).

In addition, the workshop features two keynote talks.

- Isabelle Augenstein (Head Copenhagen NLU group and Natural Language Processing section Department of Computer Science University of Copenhagen)
- Abraham Bernstein (Head Dynamic and Distributed Information Systems Group and Director UZH Digital Society Initiative)

4 ORGANISATION AND PROGRAM COMMITTEE

The workshop is co-organized by:

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

Dimitar Dimitrov, GESIS, Cologne, Germany, dimitar.dimitrov@gesis.org

Pavlos Fafalios, Technical University of Crete and FORTH-ICS, Chania, Greece, pfafalios@tuc.gr

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

Programme Committee.

- Julio Amador Diaz Lopez, Imperial College London, UK
- Katarina Boland, Heinrich Heine University, Germany

- Sandra Bringay, Paul Valéry University of Montpellier, France
- John Culbert, GESIS, Germany
- Gianluca Demartini, University of Queensland, Australia
- Ronald Denaux, Amazon.com, Spain
- Vasilis Efthymiou, Harokopio University of Athens and FORTH-ICS, Greece
- José Manuel Gómez-Pérez, Expert.AI, Spain
- Salim Hafid, LIRMM, Université Montpellier, France
- Kyle Hamilton, Technological University Dublin, Ireland
- Daniel Hardt, Copenhagen Business School, Denmark
- Andreea Iana, University of Mannheim, Germany
- Stephan Linzbach, GESIS, Germany
- Ioana Manolescu, INRIA Saclay and LIX/Ecole Polytechnique, France
- Yannis Marketakis, University of Crete and FORTH-ICS, Greece
- Michalis Mountantonakis, University of Crete and FORTH-ICS, Greece
- Panagiotis Papadakos, FORTH-ICS, Greece
- Katerina Papantoniou, FORTH-ICS, Greece
- Ensiyeh Raoufi, LIRMM, Université Montpellier, France
- Achim Rettinger, University of Trier, Germany
- Francois Scharffe, LIRMM, Université Montpellier, France
- Sebastian Schellhammer, GESIS, Germany
- Daniel Schwabe, Pontificia Universidade Católica, Brazil
- Andon Tchechmedjiev, Ecoles des Mines d'Alès, France
- Yannis Tzitzikas, University of Crete and FORTH-ICS, Greece
- Ran Yu, University of Bonn, Germany

5 PROGRAMME

After rigorous peer review we have accepted seven papers on topics covering LLM-based methods to detect propaganda and measure influence in social networks, approaches towards summarisation of online discourse, claim retrieval or to understand emotional dynamics in online discourse. Next to that, we would like to express our deepest gratitude to our keynote speakers, namely **Isabelle Augenstein** (Head of Copenhagen NLU Group and Natural Language Processing Section, Department of Computer Science University of Copenhagen) and **Abraham Bernstein** (Head Dynamic and Distributed Information Systems Group and Director UZH Digital Society Initiative).

6 PAST EDITIONS OF THE WORKSHOP

The **BeyondFacts**[5] **2021 workshop**[6] and the **BeyondFacts 2022 workshop**[7] took place as virtual events (due to COVID-19 outbreak) jointly with the **29th and 30th Web Conferences (WWW 2021 and 2022)**, respectively. The third edition, **BeyondFacts 2023**[8] took place as a hybrid event jointly with the 31st Web Conference (WWW 2023).

All three editions brought together diverse communities of researchers from different fields such as argument mining, knowledge graphs and neural language models or databases, but also social, political science and journalism scholars. The accepted and presented papers at the past workshops span a palette of topics such as claim detection, relation extraction for online discourse modeling, interpretable graph embeddings for misinformation detection, disinformation on social networks, explainability, stance detection, fact-checking in relation to argumentation schemes and false narratives, political and social scientific perspectives on propaganda chains and discourse mapping.

In the first edition, seven papers were accepted for publication, out of nine submissions, (2 long papers and 5 short ones) after a peer-review process (each submitted paper was reviewed by three PC members). The **proceedings** are available at: http://ceur-ws.org/Vol-2877/. The **best paper award** was given to the paper by Giovanni Luca Ciampaglia and John Licato: *Fact-checking, False Narratives, and Argumentation Schemes*. In addition, the first edition of the workshop hosted 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.

In the second edition, five papers were accepted for publication, out of seven submissions, (2 full papers, 2 short papers and one position paper), after a peer-review process (each submitted paper was reviewed by three PC members). The **proceedings** are available at: https://dl.acm.org/doi/proceedings/10.1145/3487553. The **best paper award** was given to the paper by Anab Maulana Barik, Wynne Hsu and Mong Li Lee: *Incorporating External Knowledge for Evidence-based Fact Verification*. The second edition of the workshop hosted three brilliant keynote talks: **Serena Villata** talked about argument-based explanatory dialogues, **Harith Alani** discussed computational tools and analysis of our interactions with false and corrective information while **José Manuel Gómez Pérez** gave an industry keynote on accurate and explainable misinformation detection.

In the third edition, three papers were accepted for publication, out of four submissions, after a peer-review process (each submitted paper was reviewed by three PC members). The **proceedings** are available at: https://dblp.org/db/conf/www/www2023c.html#TodorovF0D23. The **best paper award** was given to the paper by Youri Peskine, Raphaël Troncy and Paolo Papotti, "Analyzing COVID-Related Social Discourse on Twitter using Emotion, Sentiment, Political Bias, Stance, Veracity and Conspiracy Theories". In addition, the third edition of the workshop hosted two excellent interdisciplinary keynotes: **Kai Shu** talked about combating disinformation on social media and its challenges, while **Colin Porlezza**, a journalism scholar, discussed the responsible future of AI in journalism.

Based on the high interest and participation (around 80 participants in all three events), we consider the first three editions of the workshop successful steps towards fostering a community on discourse analysis via structured knowledge in the context of the Web. The fourth edition continues this effort, by opening up towards **novel communities** and strengthening the interdisciplinary nature of the forum. To do so, we have invited as members of the

[5]The workshop used to carry the acronym KnOD, which was later on changed to BeyondFacts.
[6]https://knod2021.wordpress.com/
[7]https://knod22.wordpress.com/
[8]https://beyondfacts2023.wordpress.com/

programme committee authors from the previous editions that come from fields such as *social sciences* and *political sciences*, as well as confirmed experts in *computational journalism* and experts in analysing *audiovisual content*, enabling to go beyond textual data (e.g. for the analysis of deep fakes). In addition, we have reached out to renowned experts from the *industry*. We also plan to attract scholars in ethics and law. We have updated the new edition's topics of interest and relevant communities accordingly. The considerably higher number of submissions that the fourth edition received (11 in total) shows that the efforts have gone in the right direction.

As a main difference in this year's edition, we plan (a) the inclusion of a poster session in addition to the presentations (for strengthening interactions among the workshop participants); (b) a round table discussing interactions between academia and industry.

7 A SHORT BIO OF THE WORKSHOP ORGANIZERS

Stefan Dietze is full professor for Data and Knowledge Engineering at the Institute for Computer Science at Heinrich-Heine-University Düsseldorf (UDUS), Scientific Director of the department Knowledge Technologies for the Social Sciences at GESIS – Leibniz Institute for the Social Sciences and affiliated member at the L3S Research Center of the Leibniz University Hanover, Germany. His research interests are at the intersection of information retrieval, semantic technologies and artificial intelligence, and in particular, the extraction, fusion and search of knowledge and data on the Web in various application domains. Stefan's work has been published at major scientific venues, such as WWW/The Web Conf, SIGIR, or ISWC, where he also frequently serves as PC and OC member. Stefan has been involved in the organisation of major conferences (ESWC, ACM WebSci, ISWC, co-chair of CIKM2020), and he has been involved in the organisation of several succesful workshop series (for instance, KnOD/BeyondFacts, PROFILES, LILE and LDOW).

Dimitar Dimitrov is a postdoctoral researcher at the department Knowledge Technologies for the Social Sciences at GESIS – Leibniz Institute for the Social Sciences (Cologne, Germany), leading the team Information Extraction and Linking. Dimitar was a visiting researcher at the Graz University of Technology, Graz, Austria, and at the eXascale Infolab at the University of Fribourg, Fribourg, Switzerland. His research focuses on discourse data analysis from social media platforms. He co-organized the Doctoral Consortium at ECIR 2020 and co-chaired the CIKM2020 AnalytiCup for which he also prepared the COVID-19 Retweet Prediction Challenge. He regularly serves as a programme committee member at computer science conferences, such as ISWC, ESWC, HT, and CIKM. Dimitar was also a co-organizer of the 3rd KnOD/BeyondFacts workshops (collocated with The WebConf 2023).

Pavlos Fafalios is assistant professor at the School of Production Engineering and Management of the Technical University of Crete (Chania, Greece) and Affiliated Researcher at the Institute of Computer Science of the Foundation for Research and Technology - Hellas (FORTH-ICS). His research interests fall in the areas of information systems, knowledge engineering, information retrieval, and Semantic Web, with a special interest on interdisciplinary research in these areas and the application of his research findings in different disciplines, including Cultural Heritage and the Humanities. He regularly participates to the programme committee of international conferences and workshops (including ISWC, ESWC and HT), and reviews papers for international journals such as the ACM Transactions on the Web, the ACM Journal on Computing and Cultural Heritage, and the Journal of Web Semantics (Elsevier). Pavlos has co-organised the 1st, 2nd, 3rd and 4th KnOD/BeyondFacts workshops (collocated with The WebConf 2021, 2022, 2023 and 2024).

Konstantin Todorov is associate professor for Computer Science at the University of Montpellier since 2012. His research lies in the fields of Web data science, knowledge graphs and machine learning with applications in the cultural heritage, social science, agro-bio and medical domains. He is scientific coordinator of the AISci French-German project and the French national ANR DACE-DL project. K. Todorov serves as a a senior PC member of ESWC since 2022, a PC member of a number of international conferences (e.g. ESWC, ISWC, the WebConf, WebScience) and has (co-)organised the following events: (i) 1st/2nd/3d KnOD/BeyondFacts workshops jointly with the WebConf 2021, 2022 and 2023; (ii) The Ontology Alignment Evaluation Initiative in 2016 and 2017, jointly with ISWC; (iii) The Linked Data tutorial at J-DEV 2017 in Marseille; (iv) The DOREMUS tutorial on structuring and publishing music-related metadata at ESWC 2016; (v) The Special Session "Uncertainty and imprecision on the Web of Data" at IPMU 2014, the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems; (vi) The International Workshop on Ontologies for Multimedia Interpretation and Retrieval at SAMT 2010, the International Conference on Semantic and Digital Media Technologies.