SemEval 2023

The 17th International Workshop on Semantic Evaluation (SemEval-2023)

Proceedings of the Workshop

July 13-14, 2023

The SemEval organizers gratefully acknowledge the support from the following organizations.

Gold

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-959429-99-9

Introduction

The Semantic Evaluation (SemEval) workshops focus on the evaluation and comparison of systems that analyze diverse semantic phenomena in text, with the aim of extending the current state of the art in semantic analysis and creating high quality annotated datasets in a range of increasingly challenging problems in natural language semantics. SemEval provides an exciting forum for researchers to propose challenging research problems in semantics and to build systems/techniques to address such research problems.

SemEval-2023 is the seventeenth workshop in the series of International Workshops on Semantic Evaluation. The first three workshops, SensEval-1 (1998), SensEval-2 (2001) and SensEval-3 (2004) focused on the word sense disambiguation (expanding in the number of languages offered, the number of tasks and the number of participating teams each year). In 2007, the workshop was renamed SemEval, and evolved to include semantic tasks beyond word sense disambiguation. Starting 2012, SemEval has been organized every year. The tasks for the next iteration of the workshop, SemEval-2024, have been selected and are underway.

SemEval-2023 is co-located (hybrid) with The 61st Annual Meeting of the Association for Computational Linguistics (ACL'2023) to be held between July 13 - 14, 2023 in Toronto and it includes the following 12 tasks:

- Semantic Structure
  - Task 1: V-WSD: Visual Word Sense Disambiguation
  - Task 2: Multilingual Complex Named Entity Recognition
- Discourse and Argumentation
  - Task 3: Detecting the Category, the Framing, and the Persuasion Techniques in Online News in a Multi-lingual Setup
  - Task 4: ValueEval: Identification of Human Values behind Arguments
  - Task 5: Clickbait Spoiling
  - Task 6: LegalEval: Understanding Legal Texts
- Medical Applications
  - Task 7: Multi-Evidence Natural Language Inference for Clinical Trial Data
  - Task 8: Causal medical claim identification and related PICO frame extraction from social media posts
- Social Attitudes
  - Task 9: Multilingual Tweet Intimacy Analysis
  - Task 10: Towards Explainable Detection of Online Sexism
  - Task 11: Learning with Disagreements (Le-Wi-Di), 2nd edition
  - Task 12: AfriSenti-SemEval: Sentiment Analysis for Low-resource African Languages using Twitter Dataset

This volume contains both the task description papers (12), that describe each of the above tasks, and the system description papers (306) that present the systems that participated in the tasks.
In addition, SemEval-2023 features two awards (one for the organizers of a task and one for a team participating in a task). The Best Task award recognizes a task that stands out for making an important intellectual contribution to empirical computational semantics, as demonstrated by a creative, interesting, and scientifically rigorous dataset and evaluation design, and a well-written task overview paper.

The Best System Description Paper award recognizes a system description paper (written by a team participating in one of the tasks) that advances our understanding of a problem and available solutions with respect to a task. It does not need not be the highest scoring system in the task, but it should have a strong analysis component in the evaluation, as well as a clear and reproducible description of the problem, algorithms, and methodology.

We are grateful to the task organizers for their dedication in carrying out twelve very successful tasks and to the large number of participants whose enthusiastic participation has made SemEval 2023 a successful event. We also appreciated the efforts of the task organizers and participants who reviewed the paper submissions. These proceedings have greatly benefited from their detailed and thoughtful feedback. We thank and acknowledge our assistant organizer Elisa Sartori (University of Padova) for her careful work during the production of the proceedings. Finally, we also thank the members of the program committee who reviewed the submitted task proposals and helped us to select this exciting set of tasks, and the ACL 2023 conference organizers for their support and the ACL Special Interest Group on the Lexicon (SIGLEX) for sponsoring and supporting this event.

Atul Kr. Ojha, A. Seza Doğruöz, Giovanni Da San Martino, Harish Tayyar Madabushi, and Ritesh Kumar (SemEval-2023 Organizers and Co-Chairs)

Organizing Committee

SemEval Chairs

Atul Kr. Ojha, Insight SFI Research Centre for Data Analytics, DSI, University of Galway & Panlingua Language Processing LLP
A. Seza Doğruöz, Universiteit Gent
Giovanni Da San Martino, University of Padova
Harish Tayyar Madabushi, University of Bath
Ritesh Kumar, Dept. of Linguistics, Dr. Bhimrao Ambedkar University, Agra

Assistant Organizer

Elisa Sartori, University of Padova

Program Committee

Alberto Barrón-Cedeño, Università di Bologna
Alessandro Raganato, University of Milano-Bicocca
Alexis Palmer, University of Colorado Boulder
Alyssa Lees, Google
Andrey Kutuzov, University of Oslo
Anjie Fang, Amazon
Anna Sauer, University of Stuttgart
Asahi Ushio, Cardiff University
Ashutosh Modi, Indian Institute of Technology Kanpur
Besnik Fetahu, Amazon
Boyuan Zheng, Johns Hopkins University
Carla Perez-Almendros, Cardiff University
Corey A. Harper, Elsevier Labs
Cristiano Chesi, Istituto Universitario di Studi Superiori
Cunxiang Wang, Westlake University
David Jurgens, University of Michigan
Eben Holderness, Brandeis University
Edwin Simpson, University of Bristol
Egoitz Laparra, University of Arizona
Elisabetta Fersini, University of Milan-Bicocca
Elisa Leonardelli, Foundation Bruno Kessler
Enrica Troiano, Vrije Universiteit Amsterdam
Fabrizio Silvestri, Sapienza University of Rome
Federico Martelli, University of Roma "La Sapienza"
Felice Dell'Orletta, Istituto di Linguistica Computazionale "A. Zampolli" (ILC)
Giovanni Da San Martino, University of Padova
Hannah Kirk, University of Oxford
Harish Tayyar Madabushi, University of Bath
Iacer Calixto, University of Amsterdam
Ibrahim Abu Farha, University of Sheffield
Jakub Piskorski, Polish Academy of Sciences
Jeffrey S. Sorensen, Google

Jennifer D'Souza, TIB Hannover
Jeremy Barnes, University of the Basque Country
Jiaxin Pei, University of Michigan
Jingxuan Tu, Brandeis University
Johannes Kiesel, Bauhaus-Universität Weimar
John Pavlopoulos, Stockholm University
Jose Camacho-collados, Cardiff University
Kyeongmin Rim, Brandeis University
Lilja Øvrelid, Dept. of Informatics, University of Oslo
Luis Chiruzzo, Facultad de Ingeniería - Universidad de la República – Uruguay
Maël Jullien, university of Manchester
Maik Fröbe, Friedrich-Schiller-Universität Jena
Marcos Zampieri, George Mason University
Marina Danilevsky, International Business Machines
Marko Robnik-Šikonja, University of Ljubljana
Massimo Poesio, Queen Mary University of London
Matthew Purver, Queen Mary University of London
Mattia Samory, GESIS-Leibniz Institute for the Social Sciences
Michael Gamon, Microsoft Research
Mohammad Taher Pilehvar, Tehran Institute for Advanced Studies
Nathan Schneider, Georgetown University
Nicolas Stefanovitch, Joint Research Centre
Paolo Rosso, Universitat Politècnica de València
Paul Groth, University of Amsterdam
Paul Röttger, University of Oxford
Preslav Nakov, Mohamed bin Zayed University of Artificial Intelligence
Richard James Evans, University of Wolverhampton
Roberto Zamparelli, University of Trento
Rodrigo Agerri, University of the Basque Country
Ron Daniel, Elsevier
Sasha Spala, Adobe
Senja Pollak, Jozef Stefan Institute
Shamsuddeen Hassan Muhammad, Bayero University, Kano
Shervin Malmasi, Amazon
Simone Conia, Sapienza University of Rome
Simone Paolo Ponzetto, University of Mannheim
Somin Wadhwa, Northeastern University
Steven Bethard, University of Arizona
Steven Schockaert, Cardiff University
Sudipta Kar, Amazon Alexa AI
Talita Anthonio, University of Stuttgart
Thamar Solorio, University of Houston
Thien Huu Nguyen, University of Oregon
Timothee Mickus, University of Helsinki
Timothy A Miller, Harvard University
Tommaso Fornaciari, Italian National Police
Viet Dac Lai, University of Oregon
Vivek Khetan, Accenture Labs
Walid Magdy, University of Edinburgh
Wenjie Yin, Queen Mary University of London
Xiaodan Zhu, Queen's University

Zhen-Hua Ling, Queen's University
Zhiyu Chen, Amazon

Keynote Talk: Holistic Evaluation of Language Models for Information Seeking

Siva Reddy
McGill University

Abstract: Conversational models for information seeking such as question answering and dialogue systems are traditionally evaluated using exact match or word overlap between the gold answer and predicted answer. With recent advances such as instruction-tuned language models, model generated answers are often verbose and naturalistic, making word overlap unreliable for measuring correctness of the answer. Moreover, the correctness of the answer is only one dimension of the problem. A generated response may contain the gold answer along with additional hallucinated information which has to be quantified as well. In this talk, I will present a holistic evaluation of retrieval-augmented conversational models and evaluate several models such as ChatGPT, Alpaca, T5, both using automatic metrics and humans. This holistic evaluation considers aspects like correctness with respect to ground truth, hallucinated information with respect to retrieved/supplied knowledge, reasoning and generalization ability. This work is informed by the author's work on TopiOCQA [1], FaithDial [2,3], [4] and NoPE [5].

[1] TopiOCQA: Open-domain Conversational Question Answering with Topic Switching. Vaibhav Adlakha, Shehzaad Dhuliawala, Kaheer Suleman, Harm de Vries, and Siva Reddy. 2022. TACL 2022
[2] FaithDial: A Faithful Benchmark for Information-Seeking Dialogue Nouha Dziri, Ehsan Kamalloo, Sivan Milton, O. Zaiane, Mo Yu, E. Ponti, Siva Reddy. TACL 2022
[3] On the Origin of Hallucinations in Conversational Models: Is it the Datasets or the Models? Nouha Dziri, Sivan Milton, Mo Yu, Osmar R Zaiane, Siva Reddy. NAACL 2022
[4] Can Retriever-Augmented Language Models Reason? The Blame Game Between the Retriever and the Language Model Parishad BehnamGhader, Santiago Miret, Siva Reddy. arxiv 2022
[5] The Impact of Positional Encoding on Length Generalization in Transformers. Amirhossein Kazemnejad and Inkit Padhi and Karthikeyan Natesan Ramamurthy and Payel Das and Siva Reddy. arXiv 2023

Bio: Siva Reddy is an Assistant Professor in the School of Computer Science and Linguistics at McGill University. He is also a Facebook CIFAR AI Chair, a core faculty member of Mila Quebec AI Institute and a research scientist at ServiceNow Research. Before McGill, he was a postdoctoral researcher at Stanford University. He received his PhD from the University of Edinburgh in 2017, where he was a Google PhD Fellow. His research focuses on representation learning for language that facilitates systematic generalization, reasoning and conversational modeling. He received the 2020 VentureBeat AI Innovation Award in NLP, and the best paper award at EMNLP 2021.

Table of Contents

Program

Thursday, July 13, 2023

09:10 - 09:25 *Welcome and Introduction to SemEval*

09:30 - 09:25 *Invited Talk 1, Shared with *SEM*

10:30 - 11:00 *Coffee Break*

11:00 - 12:30 *Oral Session I: Tasks 1-5*

SemEval-2023 Task 1: Visual Word Sense Disambiguation
Alessandro Raganato, Iacer Calixto, Asahi Ushio, Jose Camacho-Collados and Mohammad Taher Pilehvar

SemEval-2023 Task 2: Fine-grained Multilingual Named Entity Recognition (MultiCoNER 2)
Besnik Fetahu, Sudipta Kar, Zhiyu Chen, Oleg Rokhlenko and Shervin Malmasi

DAMO-NLP at SemEval-2023 Task 2: A Unified Retrieval-augmented System for Multilingual Named Entity Recognition
Zeqi Tan, Shen Huang, Zixia Jia, Jiong Cai, Yinghui Li, Weiming Lu, Yueting Zhuang, Kewei Tu, Pengjun Xie and Fei Huang

SemEval-2023 Task 3: Detecting the Category, the Framing, and the Persuasion Techniques in Online News in a Multi-lingual Setup
Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino and Preslav Nakov

SheffieldVeraAI at SemEval-2023 Task 3: Mono and Multilingual Approaches for News Genre, Topic and Persuasion Technique Classification
Ben Wu, Olesya Razuvayevskaya, Freddy Heppell, João A. Leite, Carolina Scarton, Kalina Bontcheva and Xingyi Song

SemEval-2023 Task 4: ValueEval: Identification of Human Values Behind Arguments
Johannes Kiesel, Milad Alshomary, Nailia Mirzakhmedova, Maximilian Heinrich, Nicolas Handke, Henning Wachsmuth and Benno Stein

Adam-Smith at SemEval-2023 Task 4: Discovering Human Values in Arguments with Ensembles of Transformer-based Models
Daniel Schroter, Daryna Dementieva and Georg Groh

SemEval-2023 Task 5: Clickbait Spoiling
Maik Fröbe, Benno Stein, Tim Gollub, Matthias Hagen and Martin Potthast

Thursday, July 13, 2023 (continued)

12:30 - 14:00 *Lunch*

14:00 - 15:00 *Oral Session II: Tasks 5-10*

TohokuNLP at SemEval-2023 Task 5: Clickbait Spoiling via Simple Seq2Seq Generation and Ensembling
Hiroto Kurita, Ikumi Ito, Hiroaki Funayama, Shota Sasaki, Shoji Moriya, Ye Mengyu, Kazuma Kokuta, Ryujin Hatakeyama, Shusaku Sone and Kentaro Inui

SemEval-2023 Task 6: LegalEval - Understanding Legal Texts
Ashutosh Modi, Prathamesh Kalamkar, Saurabh Karn, Aman Tiwari, Abhinav Joshi, Sai Kiran Tanikella, Shouvik Kumar Guha, Sachin Malhan and Vivek Raghavan

SemEval-2023 Task 7: Multi-Evidence Natural Language Inference for Clinical Trial Data
Maël Jullien, Marco Valentino, Hannah Frost, Paul O'regan, Donal Landers and André Freitas

Saama AI Research at SemEval-2023 Task 7: Exploring the Capabilities of Flan-T5 for Multi-evidence Natural Language Inference in Clinical Trial Data
Kamal Raj Kanakarajan and Malaikannan Sankarasubbu

SemEval-2023 Task 8: Causal Medical Claim Identification and Related PIO Frame Extraction from Social Media Posts
Vivek Khetan, Somin Wadhwa, Byron Wallace and Silvio Amir

15:00 - 15:30 *Coffee Break*

11:00 - 12:30 *Oral Session III: Tasks 9-12*

SemEval-2023 Task 9: Multilingual Tweet Intimacy Analysis
Jiaxin Pei, Vítor Silva, Maarten Bos, Yozen Liu, Leonardo Neves, David Jurgens and Francesco Barbieri

SemEval-2023 Task 10: Explainable Detection of Online Sexism
Hannah Kirk, Wenjie Yin, Bertie Vidgen and Paul Röttger

DH-FBK at SemEval-2023 Task 10: Multi-Task Learning with Classifier Ensemble Agreement for Sexism Detection
Elisa Leonardelli and Camilla Casula

SemEval-2023 Task 11: Learning with Disagreements (LeWiDi)
Elisa Leonardelli, Gavin Abercrombie, Dina Almanea, Valerio Basile, Tommaso Fornaciari, Barbara Plank, Verena Rieser, Alexandra Uma and Massimo Poesio

University at Buffalo at SemEval-2023 Task 11: MASDA–Modelling Annotator Sensibilities through DisAggregation
Michael Sullivan, Mohammed Yasin and Cassandra L. Jacobs

SemEval-2023 Task 12: Sentiment Analysis for African Languages (AfriSenti-SemEval)
Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Seid Muhie Yimam, David Ifeoluwa Adelani, Ibrahim Said Ahmad, Nedjma Ousidhoum, Abinew Ali Ayele, Saif Mohammad, Meriem Beloucif and Sebastian Ruder

NLNDE at SemEval-2023 Task 12: Adaptive Pretraining and Source Language Selection for Low-Resource Multilingual Sentiment Analysis
Mingyang Wang, Heike Adel, Lukas Lange, Jannik Strötgen and Hinrich Schütze

16:45 - 18:00 *Poster Session I: System Description Papers (local and online)*

Friday, July 14, 2023

09:30 - 10:30 *Invited Talk: Holistic Evaluation of Language Models for Information Seeking (Siva Reddy)*

10:30 - 11:00 *Coffee Break*

11:00 - 12:30 *Poster Session II: System Description Papers (online)*

12:30 - 14:00 *Lunch*

14:00 - 15:00 *Poster Session III: System Description Papers (in presence)*

15:00 - 15:30 *Coffee Break*

15:30 - 16:00 *Best Paper Awards and Concluding Remarks*