VarDial 2024

VarDial 2024 - The Eleventh Workshop on NLP for Similar Languages, Varieties and Dialects

Proceedings of the Workshop

June 20, 2024

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-104-9

Preface

These proceedings include the 22 papers presented at the Eleventh Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2024), co-located with the 2024 annual conference of the North American Chapter of the Association for Computational Linguistics (NAACL). VarDial was held in Mexico City, Mexico, in a hybrid format, allowing participants to attend on-site or remotely.

Now at its eleventh edition, we are pleased to see that VarDial continues to serve the community as the main venue for researchers interested in the computational processing of diatopic language variation. The papers accepted this year address a wide range of topics, such as language variety identification, corpus creation, and machine translation. These proceedings once again illustrate the great linguistic diversity that VarDial embodies. They include work on Germanic (historical Dutch, Limburgish, Norwegian, Swiss German), Romance (Portuguese, Spanish, Occitan), and Slavic languages (South Slavic, Slovak), as well as Arabic and Nahuatl.

As in previous editions, VarDial 2024 features an evaluation campaign with two shared tasks: The DIALECT-COPA shared task on dialectal causal commonsense reasoning, and the DSL-ML shared task on multi-label classification of similar languages. Both tasks were organized for the first time this year, although DSL-ML relies on datasets built for earlier tasks. This volume includes the system description papers prepared by the participating teams, as well as a report written by the task organizers summarizing the results and findings of the evaluation campaign.

Finally, we would like to take this opportunity to thank all the shared task organizers and the participants for their hard work. We further thank the VarDial program committee members, and particularly the 14 PC members who newly joined this year, for being an important part of the workshop's success.

The VarDial workshop organizers:

Yves Scherrer, Tommi Jauhiainen, Nikola Ljubešić, Preslav Nakov, Jörg Tiedemann, and Marcos Zampieri

https://sites.google.com/view/vardial-2024

Organizing Committee

Organizers

Tommi Jauhiainen, University of Helsinki
Nikola Ljubešić, Jožef Stefan Institute
Preslav Nakov, Mohamed bin Zayed University of Artificial Intelligence
Yves Scherrer, University of Oslo
Jörg Tiedemann, University of Helsinki
Marcos Zampieri, George Mason University

Program Committee

Program Committee

Noëmi Aepli, University of Zurich
César Aguilar, Universidad Veracruzana
Sina Ahmadi, University of Zurich
Laura Alonso Alemany, Universidad Nacional de Cordoba
Delphine Bernhard, Lilpa, Université de Strasbourg
Gabriel Bernier-Colborne, National Research Council Canada
Verena Blaschke, LMU Munich
Aoife Cahill, Dataminr
David Chiang, University of Notre Dame
Adrian-Gabriel Chifu, Aix-Marseille Universite, Universite de Toulon, CNRS, LIS, Marseille, France
Steven Coats, University of Oulu
Jon Dehdari, Fidelity Investments
Stefanie Dipper, Ruhr University Bochum
Mark Dras, Macquarie University
Jonathan Dunn, University of Illinois Urbana-Champaign
Pablo Gamallo, CITIUS, University of Santiago de Compostela
Cyril Goutte, National Research Council Canada
Nizar Habash, New York University Abu Dhabi
Radu Tudor Ionescu, University of Bucharest
Anjali Kantharuban, University of California, Berkeley
Ekaterina Lapshinova-Koltunski, University of Hildesheim
Lung-Hao Lee, National Yang Ming Chiao Tung University
John P. McCrae, Insight Center for Data Analytics, National University of Ireland Galway
Aleksandra Miletić, Department of Digital Humanities, University of Helsinki
Filip Miletić, University of Stuttgart
Maciej Ogrodniczuk, Institute of Computer Science, Polish Academy of Sciences
Petya Osenova, Sofia University St. Kl. Ohridskiand IICT-BAS
Jelena Prokic, Leiden University
Christoph Purschke, University of Luxembourg
Francisco Manuel Rangel Pardo, Universitat Politècnica de València
Reinhard Rapp, University of Mainz
Tanja Samardžić, University of Zurich
Serge Sharoff, University of Leeds
Milena Slavcheva, Institute of Information and Communication Technologies, Bulgarian Academy of Sciences
Aarohi Srivastava, University of Notre Dame
Joel Tetreault, Dataminr
Rob Van Der Goot, IT University of Copenhagen
Pidong Wang, Google
Taro Watanabe, Nara Institute of Science and Technology
Çağrı Çöltekin, University of Tübingen

Table of Contents