WIT 2022

2nd WIT - Workshop On Deriving Insights From User-Generated Text

Proceedings of the Workshop

May 27, 2022

The WIT organizers gratefully acknowledge the support from the following sponsors.

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 978-1-955917-53-7

Introduction

Welcome to the 2nd WIT (Workshop On Deriving Insights From User-Generated Text)!
Recent advances in Conversational AI, Natural Language Processing, Natural Language Understanding, Language Generation, Machine Learning, Deep Learning, Knowledge Bases, and others, have demonstrated promising results and far-reaching uses of text. Such results can be seen in many different tasks including, but not limited to better extractions from user-generated content, better language models, new approaches related to (commonsense) knowledge-bases, knowledge graphs, better information seeking QA (or Dialogue) systems, etc. Classical data management problems such as data cleaning/integration and search may also benefit from these new approaches.
The WIT workshop series was started to provide a venue to exploit and explore the use of advanced AI/ML/NLP techniques on user-generated text, which is rich in user insights and experiences. Therefore, the goal of this workshop series is to bring together researchers interested in the development and the application of novel approaches/models/systems to address challenges around harnessing text-heavy user-generated data that is available to organizations and over the Web.
For this 2nd edition, the workshop will have a great line-up of invited speakers (Mirella Lapata - University of Edinburgh, Rada Mihalcea - University of Michigan, Ann Arbor, Nina Balcan - Carnegie Mellon University, Carlos Guestrin - Stanford University) as well oral (and poster) presentations of contributed research papers. Following the tradition started in the 1st WIT, the 2nd WIT will host a panel of experts from the academia and industry to discuss and share their experiences and challenges faced in deriving insights from user-generated text. The panel is tentatively titled "User generated content and deep learning: Sorting out 'the good, the bad, and the ugly"' and is intended to highlight and surface the effects of training data on downstream applications and whether or not organizations prepare efforts around removing biases in data that they use for training or other purposes.
We would like to congratulate the authors of accepted papers, as well as to thank all the authors of submitted papers, members of the Program Committee and all the ACL main conference organization team.
2nd WIT Organizing Committee

Organizing Committee

General Chairs and Program Chairs

Estevam Hruschka, Megagon Labs Inc.
Tom Mitchell, Carnegie Mellon University
Dunja Mladenic, Jozef Stefan Institute
Marko Grobelnik, Jozef Stefan Institute
Nikita Bhutani, Megagon Labs Inc.

Program Committee

Program Committee

Sara Abdali, Georgia Institute of Technology
Shabnam Behzad, Georgetown University
Arthur Brazinskas, University of Edinburgh
Brett Zhiyuan Chen, Google
Maisa Duarte, Bradesco Bank – Brazil
Nelson Ebecken, COPPE/UFRJ Federal University of Rio de Janeiro – Brazil
Jacob Eisenstein, Google
Joao Gama, University of Porto – Portugal
Tianyu Jiang, University of Utah
Hannah Kim, Megagon Labs
Aljaz Kosmerlj, Viaduct.ai
Thom Lake, Indeed.com
Yutong Li, Apple
Jun Ma, Amazon
Vagelis Papalexakis, UC Riverside
Jing Qian, University of California Santa Barbara
Sajjadur Rahman, Megagon Labs
Yutong Shao, UC San Diego
Evan Shie, Amazon
Nedelina Teneva, Amazon
Xiaolan Wang, Megagon Labs
Xinyi (Cindy) Wang, Carnegie Mellon University
Yusuke Watanabe, Amazon
Chris Welty, Google
Natasha Zhang Foutz, University of Virginia

Invited Speakers

Mirella Lapata, University of Edinburgh
Rada Mihalcea, University of Michigan, Ann Arbor
Nina Balcan, Carnegie Mellon University
Carlos Guestrin, Stanford University

Keynote Talk: Invited Talk 1

Mirella Lapata
School of Informatics, University of Edinburgh

Abstract: Invited Talk at the 2nd WIT: Workshop On Deriving Insights From User-Generated Text at ACL2022

Bio: Mirella Lapata is a professor in the School of Informatics at the University of Edinburgh. I'm affiliated with the Institute for Communicating and Collaborative Systems and the Edinburgh Natural Language Processing Group.
Her research focuses on computational models for the representation, extraction, and generation of semantic information from structured and unstructured data, involving text and other modalities such as images, video, and large scale knowledge bases. I have worked on a variety of applied NLP tasks such as semantic parsing and semantic role labeling, discourse coherence, summarization, text simplification, concept-to-text generation, and question answering. I have also used computational models (drawing mainly on probabilistic generative models) to explore aspects of human cognition such as learning concepts, judging similarity, forming perceptual representations, and learning word meanings. The overarching goal of my research is to enable computers to understand requests and act on them, process and aggregate large amounts of data, and convey information based on them. Critical for all these tasks are models for extracting and representing meaning from natural language text, storing meanings internally, and working with stored meanings to derive further consequences.

Keynote Talk: Invited 2

Rada Mihalcea
University of Michigan, Ann Arbor

Abstract: Invited Talk at the 2nd WIT: Workshop On Deriving Insights From User-Generated Text at ACL2022

Bio: Rada Mihalcea is the Janice M. Jenkins Collegiate Professor of Computer Science and Engineering at the University of Michigan and the Director of the Michigan Artificial Intelligence Lab. Her research interests are in computational linguistics, with a focus on lexical semantics, multilingual natural language processing, and computational social sciences. She serves or has served on the editorial boards of the Journals of Computational Linguistics, Language Resources and Evaluations, Natural Language Engineering, Journal of Artificial Intelligence Research, IEEE Transactions on Affective Computing, and Transactions of the Association for Computational Linguistics. She was a program co-chair for EMNLP 2009 and ACL 2011, and a general chair for NAACL 2015 and *SEM 2019. She currently serves as ACL President. She is the recipient of a Presidential Early Career Award for Scientists and Engineers awarded by President Obama (2009), an ACM Fellow (2019) and a AAAI Fellow (2021). In 2013, she was made an honorary citizen of her hometown of Cluj-Napoca, Romania.

Keynote Talk: Invited 3

Nina Balcan
Carnegie Mellon University

Abstract: Invited Talk at the 2nd WIT: Workshop On Deriving Insights From User-Generated Text at ACL2022

Bio: Maria-Florina (Nina) Balcan is the Cadence Design Systems Professor of Computer Science at the School of Computer Science (MLD and CSD) at Carnegie Mellon University, she is also Sloan Fellow and Microsoft Faculty Fellow. Nina's main research interests are in machine learning, artificial intelligence, and theoretical computer science. Current research focus includes developing foundations and principled, practical algorithms for important modern learning paradigms. These include interactive learning, distributed learning, learning representations, life-long learning, and metalearning. Her research addresses important challenges of these settings, including statistical efficiency, computational efficiency, noise tolerance, limited supervision or interaction, privacy, low communication, and incentives. Other research topics are i) Foundations and applications of data driven algorithm design. Design and analysis of algorithms on realistic instances (a.k.a. beyond worst case).; ii) Computational and data-driven approaches in game theory and economics; iii) computational, learning theoretic, and game theoretic aspects of multi-agent systems, and iv) Analyzing the overall behavior of complex systems in which multiple agents with limited information are adapting their behavior based on past experience, both in social and engineered systems contexts.

Keynote Talk: Invited 4

Carlos Guestrin
Stanford University

Abstract: Invited Talk at the 2nd WIT: Workshop On Deriving Insights From User-Generated Text at ACL2022

Bio: Carlos Guestrin is a Professor in the Computer Science Department at Stanford University. His previous positions include the Amazon Professor of Machine Learning at the Computer Science and Engineering Department of the University of Washington, the Finmeccanica Associate Professor at Carnegie Mellon University, and the Senior Director of Machine Learning and AI at Apple, after the acquisition of Turi, Inc. (formerly GraphLab and Dato) — Carlos co-founded Turi, which developed a platform for developers and data scientist to build and deploy intelligent applications. He is a technical advisor for OctoML.ai. His team also released a number of popular open-source projects, including XGBoost, LIME, Apache TVM, MXNet, Turi Create, GraphLab/PowerGraph, SFrame, and GraphChi.
Carlos received the IJCAI Computers and Thought Award and the Presidential Early Career Award for Scientists and Engineers (PECASE). He is also a recipient of the ONR Young Investigator Award, NSF Career Award, Alfred P. Sloan Fellowship, and IBM Faculty Fellowship, and was named one of the 2008 'Brilliant 10' by Popular Science Magazine. Carlos' work received awards at a number of conferences and journals, including ACL, AISTATS, ICML, IPSN, JAIR, JWRPM, KDD, NeurIPS, UAI, and VLDB. He is a former member of the Information Sciences and Technology (ISAT) advisory group for DARPA.

Table of Contents

Program

Friday, May 27, 2022

09:20 - 09:30 *Opening Remarks*

09:30 - 10:30 *Invited Talk I*

10:30 - 11:00 *Coffee Break*

11:00 - 11:30 *Session 1*

An Interactive Analysis of User-reported Long COVID Symptoms using Twitter Data
Lin Miao, Mark Last and Marina Litvak

Bi-Directional Recurrent Neural Ordinary Differential Equations for Social Media Text Classification
Maunika Tamire, Srinivas Anumasa and P. K. Srijith

12:30 - 11:30 *Invited Talk II*

12:30 - 14:00 *Lunch Break*

15:00 - 14:00 *Invited Talk III*

15:00 - 15:30 *Coffee Break*

15:30 - 15:45 *Session 2*

Unsupervised Abstractive Dialogue Summarization with Word Graphs and POV Conversion
Seongmin Park and Jihwa Lee

15:45 - 16:45 *Invited Talk IV*

16:45 - 17:45 *Panel*