

MATCHING 2023

**The First Workshop on Matching From Unstructured and
Structured Data (MATCHING 2023)**

Proceedings of the Workshop

July 13, 2023

©2023 Association for Computational Linguistics

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 978-1-959429-82-1

Introduction

Welcome to the inaugural Workshop on Matching from Unstructured and Structured Data - MATCHING 2023!

Matching is a fundamental task with wide-ranging applications, encompassing search, recommendation systems, and data integration, among others. Given the proliferation of social media and e-commerce platforms, the ability to match information from both structured and unstructured sources has become increasingly crucial. At its core, matching aims to identify pairs of entries in two collections that share common properties. For instance, in the realm of HR platforms/services, matching resumes to job descriptions plays a pivotal role. Similarly, online booking platforms/services strive to match customer preferences with suitable businesses, such as hotels, restaurants, and real estate establishments. Beyond these entity matching examples, matching techniques find frequent application in various domains, such as matching customer reviews about a product to customer queries, aligning snippets of web documents with search queries, and matching user responses in Q&A platforms to new questions, etc. Consequently, matching tasks can take diverse forms based on the type of input source (structured vs. unstructured), the downstream application (e.g., search, conversation, recommendation), and ethical considerations (e.g., bias and transparency).

The primary objective of this workshop is to facilitate collaboration between research communities in academia and industries related to several domains, including natural language processing, language generation, deep learning, conversational AI, information extraction, data integration, knowledge graphs, and human-centered computing. For this inaugural edition of the Matching Workshop, we received 15 original research submissions and 12 application of papers that were accepted as Findings of ACL2023. Thanks to the diligent efforts of our Program Chairs and Program Committee members, we were able to curate a collection of 9 original research contributions (with an acceptance rate of 60%) and also 5 findings papers to be presented during the Workshop's technical sessions.

In addition to the research papers, the Matching Workshop will host an exciting lineup of four distinguished invited speakers and a panel discussion, providing participants with an opportunity to engage with leading researchers from academia and prominent industries.

Keynote Speakers. We are pleased to have the following invited speakers: William W. Cohen (Principal Scientist at Google Research), Ndapa Nakashole (Assistant Professor of Computer Science at the University of California, San Diego), Alan Ritter (Associate Professor in the School of Interactive Computing at Georgia Tech) and Sameer Singh (Associate Professor of Computer Science at the University of California, Irvine).

Panel Discussion. In recent years, large language models have emerged as influential forces driving both research and development. However, the impact of these models on downstream tasks remains uncertain and inadequately understood. As exemplified by numerous events publicized in the media, the embedded bias within these models can lead to catastrophic consequences. Consequently, we aim to foster a robust discussion on this topic by inviting researchers from academia and relevant industries. The panel discussion will count on six panelists and will revolve around the theme *Matching in the Era of Large Language Models: Assessing the Good, the Bad, and the Ugly*. The panelists are Renée Miller (University Distinguished Professor of Computer Science at Northeastern University), Niket Tandon (Senior Research Scientist at the Allen Institute for AI), Barbara Plank (Full Professor and Chair for AI and Computational Linguistic at the Ludwig-Maximilians-Universität München and Full Professor at the IT University of Copenhagen), AnHai Doen (Vilas Distinguished Achievement Professor in the Department of Computer Science at the University of Wisconsin) and Lei Li (Assistant Professor in Computer Science Department at University of California Santa Barbara).

We extend our congratulations to the authors of the accepted papers and express our gratitude to all authors who submitted their work, members of the Program Committee, Mentors and Mentees who participate in the Matching Workshop Mentoring Program and the entire ACL main conference organizing team. We would like to specially thank Megagon Labs (<http://www.megagon.ai>) for supporting our workshop and hosting the website of this event: <https://megagon.ai/matching-2023/>

First Matching Workshop Organizing Committee.

Organizing Committee

General Chairs and Program Chairs

Estevam Hruschka, Megagon Labs
Tom Mitchell, Carnegie Mellon University, USA
Sajjadur Rahman, Megagon Labs
Dunja Mladenic, Jozef Stefan Institute (JSI)
Marko Grobelnik, Jozef Stefan Institute (JSI)

Program Committee

Program Committee

Abdus Azad, UC Berkeley
Bosung Kim, UC San Diego
Chen Shen, Megagon Labs
Evangelos Papalexakis, UC Riverside
Farig Sadeque, BRAC University
Grace Fan, Northeastern University
Hannah Kim, Megagon Labs
Jeniya Tabassum, Amazon
Jun Ma, Amazon
Nedelina Teneva, Megagon Labs
Nelson Ebecken, Federal University of Rio de Janeiro
Rahat Rafiq, Grand Valley State University
Sara Abdali, Georgia Institute of Technology
Sara Tonelli, Fondazione Bruno Kessler
Seiji Maekawa, Megagon Labs
Thom Lake, Indeed.com
Tianyu Jiang, University of Utah
Vishwas Mruthyunjaya, Megagon Labs
Wasi Ahmad, UCLA
Yusuke Watanabe, Amazon
Yutong Li, Apple Inc.

Mentorship Program - Mentors

Andrii Krutsylo, Polish Academy of Sciences
Pouya Pezeshkpour, Megagon Labs
Sairam Gurajada, Megagon Labs

Keynote Talk: Invited 1

William Cohen

Google

2023-07-13 – Room: Room 1

Abstract: Invited Talk I -

Bio: William W. Cohen is a Principal Scientist at Google. He received his bachelor's degree in Computer Science from Duke University in 1984, and a PhD in Computer Science from Rutgers University in 1990. From 1990 to 2000 he worked at AT&T Bell Labs and later AT&T Labs-Research, and from April 2000 to May 2002 he worked at Whizbang Labs, a company specializing in extracting information from the web. From 2002 to 2018, he worked at Carnegie Mellon University in the Machine Learning Department, with a joint appointment in the Language Technology Institute, as an Associate Research Professor, a Research Professor, and a Professor. He is a past president of the International Machine Learning Society, and was General Chair for the 2008 International Machine Learning Conference, held July 6-9 at the University of Helsinki, in Finland; Program Co-Chair of the 2006 International Machine Learning Conference; and Co-Chair of the 1994 International Machine Learning Conference. He is a AAAI Fellow, and was a winner of the 2008 SIGMOD Test of Time Award for the most influential SIGMOD paper of 1998, and the 2014 SIGIR Test of Time Award for the most influential SIGIR paper of 2002-2004.

Keynote Talk: Invited 2

Ndapa Nakashole

University of California San Diego

2023-07-13 – Room: Room 1

Abstract: Invited Talk II -

Bio: Ndapa Nakashole is an Assistant Professor of Computer Science at the University of California, San Diego. Prior to this, she was a postdoctoral fellow in the Machine Learning department at Carnegie Mellon University. She obtained her PhD from the Max Planck Institute for Informatics, and Saarland University. She is a recipient of an NSF CAREER award, and an Otto Hahn Medal by the Max Planck Society for her dissertation.

Keynote Talk: Invited 3

Alan Ritter

Georgia Institute of Technology

2023-07-13 – Room: Room 1

Abstract: Invited Talk III -

Bio: Alan Ritter is an associate professor in the School of Interactive Computing at Georgia Tech. His research interests include natural language processing, with a focus on information extraction and data driven chatbots. He completed his Ph.D. at the University of Washington and was a postdoctoral fellow in the Machine Learning Department at Carnegie Mellon University. His research aims to solve challenging technical problems that can help machines learn to read vast quantities of text with minimal supervision. In a recent project, covered by WIRED, his group built a system that reads millions of tweets for mentions of new software vulnerabilities. Alan is the recipient of an NSF CAREER award and an Amazon Research Award.

Keynote Talk: Invited 4

Sameer Singh

University of California Irvine

2023-07-13 – Room: Room 1

Abstract: Invited Talk IV -

Bio: Dr. Sameer Singh is an Associate Professor of Computer Science at the University of California, Irvine (UCI). He is working primarily on the robustness and interpretability of machine learning algorithms and models that reason with text and structure for natural language processing. Sameer was a postdoctoral researcher at the University of Washington and received his Ph.D. from the University of Massachusetts, Amherst. He has received the NSF CAREER award, UCI Distinguished Early Career Faculty award, the Hellman Faculty Fellowship, and was selected as a DARPA Riser. His group has received funding from Allen Institute for AI, Amazon, NSF, DARPA, Adobe Research, Hasso Plattner Institute, NEC, Base 11, and FICO. Sameer has published extensively at machine learning and natural language processing venues and received conference paper awards at KDD 2016, ACL 2018, EMNLP 2019, AKBC 2020, ACL 2020, and NAACL 2022. (<https://sameersingh.org/>)

Table of Contents

<i>Text-To-KG Alignment: Comparing Current Methods on Classification Tasks</i> Sondre Wold, Lilja Øvrelid and Erik Velldal	1
<i>Identifying Quantifiably Verifiable Statements from Text</i> Pegah Jandaghi and Jay Pujara	14
<i>Toward Consistent and Informative Event-Event Temporal Relation Extraction</i> Xiaomeng Jin, Haoyang Wen, Xinya Du and Heng Ji	23
<i>COFFEE: A Contrastive Oracle-Free Framework for Event Extraction</i> Meiru Zhang, Yixuan Su, Zaiqiao Meng, Zihao Fu and Nigel Collier	33
<i>Corpus-Based Task-Specific Relation Discovery</i> Karthik Ramanan	45
<i>On the Surprising Effectiveness of Name Matching Alone in Autoregressive Entity Linking</i> Elliot Schumacher, James Mayfield and Mark Dredze	58
<i>Knowledge-Augmented Language Model Prompting for Zero-Shot Knowledge Graph Question Answering</i> Jinheon Baek, Alham Aji and Amir Saffari	70
<i>Knowledge Base Completion for Long-Tail Entities</i> Lihu Chen, Simon Razniewski and Gerhard Weikum	99
<i>CoSiNES: Contrastive Siamese Network for Entity Standardization</i> Jiaqing Yuan, Michele Merler, Mihir Choudhury, Raju Pavuluri, Munindar Singh and Maja Vukovic	109

Program

Thursday, July 13, 2023

08:30 - 08:40 *Welcome Message and Opening Remarks*

08:40 - 09:30 *Invited Talk I*

09:30 - 10:30 *Research Papers - Session I*

Text-To-KG Alignment: Comparing Current Methods on Classification Tasks
Sondre Wold, Lilja Øvrelid and Erik Velldal

Identifying Quantifiably Verifiable Statements from Text
Pegah Jandaghi and Jay Pujara

Toward Consistent and Informative Event-Event Temporal Relation Extraction
Xiaomeng Jin, Haoyang Wen, Xinya Du and Heng Ji

COFFEE: A Contrastive Oracle-Free Framework for Event Extraction
Meiru Zhang, Yixuan Su, Zaiqiao Meng, Zihao Fu and Nigel Collier

Knowledge-Augmented Language Model Prompting for Zero-Shot Knowledge Graph Question Answering
Jinheon Baek, Alham Aji and Amir Saffari

CoSiNES: Contrastive Siamese Network for Entity Standardization
Jiaqing Yuan, Michele Merler, Mihir Choudhury, Raju Pavuluri, Munindar Singh and Maja Vukovic

10:30 - 11:00 *Break*

11:00 - 12:00 *Invited Talk II*

12:00 - 13:00 *Lunch + Posters Session*

Text-To-KG Alignment: Comparing Current Methods on Classification Tasks
Sondre Wold, Lilja Øvrelid and Erik Velldal

Thursday, July 13, 2023 (continued)

Identifying Quantifiably Verifiable Statements from Text

Pegah Jandaghi and Jay Pujara

Toward Consistent and Informative Event-Event Temporal Relation Extraction

Xiaomeng Jin, Haoyang Wen, Xinya Du and Heng Ji

COFFEE: A Contrastive Oracle-Free Framework for Event Extraction

Meiru Zhang, Yixuan Su, Zaiqiao Meng, Zihao Fu and Nigel Collier

Knowledge-Augmented Language Model Prompting for Zero-Shot Knowledge Graph Question Answering

Jinheon Baek, Alham Aji and Amir Saffari

CoSiNES: Contrastive Siamese Network for Entity Standardization

Jiaqing Yuan, Michele Merler, Mihir Choudhury, Raju Pavuluri, Munindar Singh and Maja Vukovic

Corpus-Based Task-Specific Relation Discovery

Karthik Ramanan

On the Surprising Effectiveness of Name Matching Alone in Autoregressive Entity Linking

Elliot Schumacher, James Mayfield and Mark Dredze

Knowledge Base Completion for Long-Tail Entities

Lihu Chen, Simon Razniewski and Gerhard Weikum

ECOLA: Enhancing Temporal Knowledge Embeddings with Contextualized Language Representations

Zhen Han, Ruotong Liao, Jindong Gu, Yao Zhang, Zifeng Ding, Yujia Gu, Heinz Koepl, Hinrich Schütze and Volker Tresp

Easy-to-Hard Learning for Information Extraction

Chang Gao, Wenxuan Zhang, Wai Lam and Lidong Bing

Silver Syntax Pre-training for Cross-Domain Relation Extraction

Elisa Bassignana, Filip Ginter, Sampo Pyysalo, Rob van der Goot and Barbara Plank

Thursday, July 13, 2023 (continued)

INFOSYNC: Information Synchronization across Multilingual Semi-structured Tables

Siddharth Khincha, Chelsi Jain, Vivek Gupta, Tushar Kataria and Shuo Zhang

Aligning Instruction Tasks Unlocks Large Language Models as Zero-Shot Relation Extractors

Kai Zhang, Bernal Gutiérrez and Yu Su

13:00 - 13:50 *Invited Talk III*

13:50 - 14:40 *Research Papers - Session II*

ECOLA: Enhancing Temporal Knowledge Embeddings with Contextualized Language Representations

Zhen Han, Ruotong Liao, Jindong Gu, Yao Zhang, Zifeng Ding, Yujia Gu, Heinz Koepl, Hinrich Schütze and Volker Tresp

Easy-to-Hard Learning for Information Extraction

Chang Gao, Wenxuan Zhang, Wai Lam and Lidong Bing

Silver Syntax Pre-training for Cross-Domain Relation Extraction

Elisa Bassignana, Filip GinterÚ, Sampo PyysaloÚ, Rob van der Goot and Barbara Plank

INFOSYNC: Information Synchronization across Multilingual Semi-structured Tables

Siddharth Khincha, Chelsi Jain, Vivek Gupta, Tushar Kataria and Shuo Zhang

Aligning Instruction Tasks Unlocks Large Language Models as Zero-Shot Relation Extractors

Kai Zhang, Bernal Gutiérrez and Yu Su

14:40 - 15:30 *Invited Talk IV*

15:30 - 16:00 *Break*

16:00 - 17:30 *Panel*

17:30 - 17:35 *Final Remarks*

Thursday, July 13, 2023 (continued)