CALD-pseudo 2024

Workshop on Computational Approaches to Language Data Pseudonymization (CALD-pseudo)

Proceedings of the Workshop

March 21, 2024

©2024 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 979-8-89176-085-1

Introduction

We are excited to offer you the first proceedings from the workshop on Computational Approaches to Language Data Pseudonymization, CALD-pseudo 2024[1]!
We have accepted 10 high-quality papers representing a wide geographic diversity, namely co-authors with affiliations in the Basque Country, Canada, Finland, France, Germany, Japan, Norway, Spain, Sweden, and the USA.
In this volume, you can read papers that deal with the topic of personal or sensitive information, and the subsequent question of accessibility of research data. Accessibility of research data is critical for advances in many research fields but textual data often cannot be shared due to the personal and sensitive information it contains, e.g. names, political opinions, sensitive personal information, and medical data. General Data Protection Regulation, GDPR (EU Commission, 2016), suggests pseudonymization as a solution to secure open access to research data but we need to learn more about pseudonymization as an approach before adopting it for the manipulation of research data (Volodina et al., 2023). The main challenge is how to effectively pseudonymize data so that such individuals cannot be identified, while at the same time keeping the data usable for research (e.g. in computational linguistics, linguistics) and natural language processing tasks for which it was collected.
This workshop has invited a broad community of researchers in all concerned cross-disciplinary fields to jointly discuss challenges within pseudonymization, such as

- automatic approaches to detection and labelling of personal information in unstructured language data, including events and other context-dependent cues revealing a person;
- developing context-sensitive algorithms for replacement of personal information in unstructured data;
- studies into the effects of pseudonymization on unstructured data, e.g. applicability of pseudonymised data for the intended research questions, readability of pseudonymised data, or addition of unwelcome biases through pseudonymization;
- effectiveness of pseudonymization as a way of protecting writer identity;
- reidentification studies, e.g. adversarial learning techniques that attempt to breach the privacy protections of pseudonymized data;
- constructing datasets for automatic pseudonymization, including methodological and ethical aspects of those;
- approaches to the evaluation of automatic pseudonymization both in concealing the private information and preserving the semantics of the non-personal data;
- pseudonymization tools and software: evaluating the available tools and software for pseudonymization in different languages, and their ease of use, scalability, and performance;
- and numerous other open questions.

The workshop was one full day and included two invited talks - by Ildikó Pilán and Anders Søgaard.
We would kindly like to thank our program committee for their valuable (and enthusiastic!) contribution to the success of the workshop:

- Lars Ahrenberg, Linköping University, Sweden
- Terhi Ainiala, University of Helsinki, Finland

[1] https://mormor-karl.github.io/events/CALD-pseudo/

- Emilia Aldrin, Halmstad University, Sweden
- Špela Arhar Holdt, University of Ljubljana, Slovenia
- Andrew Caines, University of Cambridge, United Kingdom
- Hercules Dalianis, Stockholm University, Sweden
- Dana Dannélls, University of Gothenburg, Sweden
- Simon Dobnik, University of Gothenburg, Sweden
- Cyril Grouin, LIMSI, CNRS, Université Paris-Saclay, France
- Lasse Hämäläinen, University of Helsinki, Finland
- Aron Henriksson, Stockholm University, Sweden
- Dimitrios Kokkinakis, University of Gothenburg, Sweden
- Jannika Lassus, University of Helsinki, Finland
- Therese Lindström Tiedemann, University of Helsinki, Finland
- Pierre Lison, Norwegian Computing Center, Norway
- Krister Lindén, University of Helsinki, Finland
- Peter Ljunglöf, Chalmers University of Technology / University of Gothenburg, Sweden
- Ricardo Muñoz Sánchez, University of Gothenburg, Sweden
- Boel Nelson, Aarhus University, Denmark
- Lieselott Nordman, University of Helsinki, Finland
- Ildikó Pilán, Norwegian Computing Center, Norway
- Vipul Raheja, Grammarly, USA
- Tatjana Scheffler, Ruhr University Bochum, Germany
- Vicenc Torra, Umeå University, Sweden
- Thomas Vakili, Stockholm University, Sweden
- VG Vinod Vydiswaran, University of Michigan, USA
- Elena Volodina, University of Gothenburg, Sweden
- Xuan-Son Vu, Umeå University, Sweden

Our further thanks go to the generous support from the *Swedish Research Council*[2] through its funding to the research environment project *Grandma Karl is 27 years old*[3].

[2] https://www.vr.se/english/swecris.html#/project/2022-02311_VR
[3] https://mormor-karl.github.io/

Organizing Committee

General Chair

Elena Volodina, University of Gothenburg, Sweden

General Co-chairs

Simon Dobnik, University of Gothenburg, Sweden
Therese Lindström Tiedemann, University of Helsinki, Finland
Xuan-Son Vu, Umeå University, Sweden

Organizing Co-chairs

David Alfter, University of Gothenburg, Sweden
Ricardo Muñoz Sánchez, University of Gothenburg, Sweden
Maria Irena Szawerna, University of Gothenburg, Sweden

Program Committee

Program Committee

Lars Ahrenberg, Linköping University, Sweden
Terhi Ainiala, University of Helsinki, Finland
Emilia Aldrin, Halmstad University, Sweden
Špela Arhar Holdt, University of Ljubljana, Slovenia
Andrew Caines, University of Cambridge, United Kingdom
Hercules Dalianis, Stockholm University, Sweden
Dana Dannélls, University of Gothenburg, Sweden
Simon Dobnik, University of Gothenburg, Sweden
Cyril Grouin, LIMSI, CNRS, Université Paris-Saclay, France
Lasse Hämäläinen, University of Helsinki, Finland
Aron Henriksson, Stockholm University, Sweden
Dimitrios Kokkinakis, University of Gothenburg, Sweden
Jannika Lassus, University of Helsinki, Finland
Therese Lindström Tiedemann, University of Helsinki, Finland
Krister Lindén, University of Helsinki, Finland
Peter Ljunglöf, Chalmers University of Technology / University of Gothenburg, Sweden
Ricardo Muñoz Sánchez, University of Gothenburg, Sweden
Boel Nelson, Aarhus University, Denmark
Lieselott Nordman, University of Helsinki, Finland
Ildikó Pilán, Norwegian Computing Center, Norway
Vipul Raheja, Grammarly, USA
Tatjana Scheffler, Ruhr University Bochum, Germany
Vicenc Torra, Umeå University, Sweden
Thomas Vakili, Stockholm University, Sweden
VG Vinod Vydiswaran, University of Michigan, USA
Elena Volodina, University of Gothenburg, Sweden
Xuan-Son Vu, Umeå University, Sweden

Keynote Talk: NLP is Dead - Now What?

Anders Søgaard
University of Copenhagen, Denmark
2024-03-21 09:10:00 – Room: **Corinthia hotel, Gardjola 3 (virtual talk)**

Abstract: For decades, the NLP community was on a mission to get computers to understand language. To the extent the goal of the mission was defined, our mission is complete. Now what? There are still a ton of open problems, of course. Pseudonymization is one of them. Others include bias mitigation, performance parity, or getting things to run on-device. None of these problems are NLP problems, but they are all inter-dependent. Does their locus leave room for a raison d'être for the remnants of NLP?

Bio: Anders Søgaard is Full Professor in Natural Language Processing and Machine Learning, Dpt. of Computer Science, University of Copenhagen. He is also affiliated with the Pioneer Centre for Artificial Intelligence, Dpt. of Philosophy, and Center for Social Data Science. He was previously at University of Potsdam, Amazon and Google Research. He has won eight best paper awards and several prestigious grants.

Keynote Talk: Pseudonymisation and related techniques: a quest for determining what personal information to rewrite and how

Ildikó Pilán
The Norwegian Computing Center, Norway
2024-03-21 13:00:00 – Room: **Corinthia hotel, Gardjola 3**

Abstract: In this talk, we will walk through the different steps involved in the process of concealing personal information. We will start by looking at methods for which pieces of personal information to detect and how. We will then discuss strategies for rewriting these and, finally, we will look at approaches proposed for evaluating the resulting redacted text in terms of privacy protection and utility preservation. We will discuss previous work inspired by Named Entity Recognition as well as more recent approaches employing Large Language Models. We will also explore the differences between pseudonymization and anonymization highlighting the remaining challenges in performing these automatically.

Bio: Ildikó Pilán is a Senior Research Scientist at the Norwegian Computing Center, Norway. Her most impactful research comes from linguistic complexity studies within the domain of language learning, and recently from the area of anonymization and pseudonymization where she has been actively working on preparing datasets, benchmarks and models for automatic anonymization and pseudonymization of Norwegian and English data in the project Cleanup (e.g. Lison et al., 2021; Pilán et al., 2022). Her fields of expertise include Natural Language Processing, Machine Learning, privacy protection, data privacy, medical text processing and Intelligent Computer-Assisted Language Learning.

Table of Contents

Handling Name Errors of a BERT-Based De-Identification System: Insights from Stratified Sampling and Markov-based Pseudonymization
Dalton Simancek and VG Vinod Vydiswaran 1

Assessing Authenticity and Anonymity of Synthetic User-generated Content in the Medical Domain
Tomohiro Nishiyama, Lisa Raithel, Roland Roller, Pierre Zweigenbaum and Eiji Aramaki 8

Automatic Detection and Labelling of Personal Data in Case Reports from the ECHR in Spanish: Evaluation of Two Different Annotation Approaches
Maria Sierro, Begoña Altuna and Itziar Gonzalez-Dios 18

PSILENCE: A Pseudonymization Tool for International Law
Luis Adrián Cabrera-Diego and Akshita Gheewala 25

Deidentifying a Norwegian Clinical Corpus - an Effort to Create a Privacy-preserving Norwegian Large Clinical Language Model
Phuong Ngo, Miguel Tejedor, Therese Olsen Svenning, Taridzo Chomutare, Andrius Budrionis and Hercules Dalianis 37

Extending Off-the-shelf NER Systems to Personal Information Detection in Dialogues with a Virtual Agent: Findings from a Real-Life Use Case
Mario Mina, Carlos Rodríguez, Aitor Gonzalez-Agirre and Marta Villegas 44

Detecting Personal Identifiable Information in Swedish Learner Essays
Maria Irena Szawerna, Simon Dobnik, Ricardo Muñoz Sánchez, Therese Lindström Tiedemann and Elena Volodina 54

Data Anonymization for Privacy-Preserving Large Language Model Fine-Tuning on Call Transcripts
Shayna Gardiner, Tania Habib, Kevin Humphreys, Masha Azizi, Frederic Mailhot, Anne Paling, Preston Thomas and Nathan Zhang 64

When Is a Name Sensitive? Eponyms in Clinical Text and Implications for De-Identification
Thomas Vakili, Tyr Hullmann, Aron Henriksson and Hercules Dalianis 76

Did the Names I Used within My Essay Affect My Score? Diagnosing Name Biases in Automated Essay Scoring
Ricardo Muñoz Sánchez, Simon Dobnik, Maria Irena Szawerna, Therese Lindström Tiedemann and Elena Volodina 81

Program

Thursday, March 21, 2024

09:00 - 09:10 *Opening Remarks by Elena Volodina*

09:10 - 10:00 *Invited talk 1. Anders Søgaard. Title: NLP is Dead - Now What?, chair: Elena Volodina*

10:00 - 10:30 *Session 1, chair: Maria Irena Szawerna*

Handling Name Errors of a BERT-Based De-Identification System: Insights from Stratified Sampling and Markov-based Pseudonymization
Dalton Simancek and VG Vinod Vydiswaran

Automatic Detection and Labelling of Personal Data in Case Reports from the ECHR in Spanish: Evaluation of Two Different Annotation Approaches
Maria Sierro, Begoña Altuna and Itziar Gonzalez-Dios

10:30 - 11:00 *COFFEE break*

11:00 - 12:00 *Session 2, chair: Hercules Dalianis*

PSILENCE: A Pseudonymization Tool for International Law
Luis Adrián Cabrera-Diego and Akshita Gheewala

Extending Off-the-shelf NER Systems to Personal Information Detection in Dialogues with a Virtual Agent: Findings from a Real-Life Use Case
Mario Mina, Carlos Rodríguez, Aitor Gonzalez-Agirre and Marta Villegas

Data Anonymization for Privacy-Preserving Large Language Model Fine-Tuning on Call Transcripts
Shayna Gardiner, Tania Habib, Kevin Humphreys, Masha Azizi, Frederic Mailhot, Anne Paling, Preston Thomas and Nathan Zhang

12:00 - 13:00 *LUNCH break*

13:00 - 13:50 *Invited talk 2. Ildikó Pilán. Title: Pseudonymisation and related techniques: a quest for determining what personal information to rewrite and how, chair: Elena Volodina*

13:50 - 14:00 *Short break*

14:00 - 14:45 *Session 3, chair: Ricardo Muñoz Sánchez*

Thursday, March 21, 2024 (continued)

Assessing Authenticity and Anonymity of Synthetic User-generated Content in the Medical Domain
Tomohiro Nishiyama, Lisa Raithel, Roland Roller, Pierre Zweigenbaum and Eiji Aramaki

Deidentifying a Norwegian Clinical Corpus - an Effort to Create a Privacy-preserving Norwegian Large Clinical Language Model
Phuong Ngo, Miguel Tejedor, Therese Olsen Svenning, Taridzo Chomutare, Andrius Budrionis and Hercules Dalianis

When Is a Name Sensitive? Eponyms in Clinical Text and Implications for De-Identification
Thomas Vakili, Tyr Hullmann, Aron Henriksson and Hercules Dalianis

14:45 - 14:50 *Short break*

14:50 - 15:30 *Session 4, chair: Ildikó Pilán*

Detecting Personal Identifiable Information in Swedish Learner Essays
Maria Irena Szawerna, Simon Dobnik, Ricardo Muñoz Sánchez, Therese Lindström Tiedemann and Elena Volodina

Did the Names I Used within My Essay Affect My Score? Diagnosing Name Biases in Automated Essay Scoring
Ricardo Muñoz Sánchez, Simon Dobnik, Maria Irena Szawerna, Therese Lindström Tiedemann and Elena Volodina

15:30 - 16:00 *COFFEE break*

16:00 - 17:00 *Panel discussion with Ildikó Pilán, Thomas Vakili, etc., moderator: Elena Volodina*