UNIVERSIDADE DA CORUÑA

MetaHate

Usage/License

MetaHate is distributed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0). Access to the complete dataset requires submission of relevant agreements for certain derived datasets. Without these agreements, only publicly available subsets are accessible. Users must adhere to the specified terms and conditions.

Download

The dataset is available for download on the Hugging Face platform:

https://huggingface.co/datasets/irlab-udc/metahate

Please note that access to the full dataset requires prior agreement to specific terms and conditions.

Researchers

- Paloma Piot
- Patricia Martín Rodilla
- Javier Parapar

These researchers are affiliated with the Information Retrieval Lab at the University of A Coruña.

Summary

MetaHate is a comprehensive meta-collection comprising 1,226,202 social media posts aggregated from 36 distinct hate speech datasets. Each entry includes the post content and a binary label indicating the presence (1) or absence (0) of hate speech. This dataset aims to unify various efforts in hate speech detection by providing a consolidated resource for researchers.

Key Features

- Diverse Data Sources: Integration of 36 datasets encompassing various forms of hate speech across different social media platforms.
- Binary Classification Labels: Each post is labeled as '0' for non-hate speech or '1' for hate speech, simplifying classification tasks.
- Research-Oriented: Designed to support and enhance computational linguistic approaches in identifying and understanding hate speech.

What We Offer

- Extensive Collection: A large-scale dataset combining multiple sources to facilitate diverse research in hate speech detection.
- Standardized Format: Data is provided in a TSV (Tab-Separated Values) format with clear labelling for ease of use.
- Access Options: Availability of both the complete meta-collection and subsets, depending on user agreements.

Collaboration Objectives

MetaHate project seeks to:

- Unify various hate speech detection efforts by providing a consolidated dataset.
- Facilitate research in computational linguistics and social media analysis.
- Encourage collaboration among researchers to develop robust models for hate speech detection.

hybrids

HYBRIDS Website

www.hybridsproject.eu

Get in Touch

paloma.piot@udc.es