

MultiClaimNet



UK Research
and Innovation



Funded by
the European Union

Usage/License



MultiClaimNet is released under the **Creative Commons Attribution 4.0 International License (CC BY 4.0)**. Under this license, users are permitted to share and adapt the material provided appropriate credit is given to the original authors. However, access to the dataset is **restricted to research purposes**. Users must comply with the research-only access conditions specified by the dataset providers.

Download



The MultiClaimNet datasets are available for download via Zenodo:
<https://zenodo.org/records/15100352>

Please refer to the repository for detailed documentation, dataset structure, and access instructions. Access to the full dataset requires prior agreement to specific terms and conditions.

Researchers



- Rubaa Panchendrarajan
- Rubén Míguez
- Arkaitz Zubiaga

These researchers are affiliated with Queen Mary University of London and Newtral Media Audiovisual (Spain), collaborating on multilingual automated fact-checking research.

Summary

MultiClaimNet is a massively multilingual claim clustering resource designed to support scalable automated fact-checking. It comprises three claim cluster datasets constructed from claim-matching pairs, covering claims written in 86 languages. The largest dataset contains 85.3K fact-checked claims grouped into 30.9K clusters. The dataset provides ground-truth claim clusters to facilitate research on multilingual claim retrieval, clustering, and misinformation tracking.

What We Offer

- **Three Complementary Datasets:** Two smaller curated datasets (ClaimCheck, ClaimMatch) and one large-scale automatically constructed dataset (MultiClaim).
- **Cluster-Level Ground Truth:** Enables evaluation using ARI, AMI, V-measure, and other clustering metrics.
- **Multilingual Representation:** Supports research on cross-lingual and multilingual fact-checking.

Key Features

- **Multilingual Coverage:** Claims across 86 languages
- **Large-Scale Clustering:** 85.3K claims grouped into 30.9K validated clusters in the largest dataset.
- **Ground-Truth Clusters:** First dedicated dataset collection for evaluating claim clustering approaches.

Collaboration Objectives

The MultiClaimNet project seeks to:

- Advance research in multilingual claim clustering for automated fact-checking.
- Improve scalability of fact-checking pipelines.
- Encourage development of robust, task-specific clustering methods for real-world fact-check databases.

