WojoodNER 2023: The First Arabic Named Entity Recognition Shared Task

Mustafa Jarrar, Muhammad Abdul-Mageed, Mohammed Khalilia, Bashar Talafha, AbdelRahim Elmadany, Nagham Hamad, Ala'a Omar

Birzeit University, Palestine  The University of British Columbia  MBZUAI

Task Description

❖ Subtask 1 - FlatNER

❖ Subtask 2 - NestedNER

Datasets

WojoodNER-2023 shared task employs the Wojood corpus as its primary dataset (Jarrar et al., 2022). The Wojood corpus encompasses approximately 550K tokens, spanning both MSA and two Arabic dialects, annotated using 21-entity types.

Evaluation metrics

The official evaluation metric for subtask1 and subtask2 is the macro-averaged F1 score. In addition to this metric, Precision Recall, and Accuracy is recorded for submissions to both subtasks.

Shared Task Teams & Results

- 45 unique team registrations.
- 12 teams submitted for FlatNER and Nested-NER.
- 35 submissions for FlatNER from eleven teams and 22 submissions for NestedNER from eight teams.

Results

- All the models submitted to the shared task adopt the transfer learning approach, leveraging pre-trained models trained on various data sources.
- The top-performing models addressed the challenge of identifying nested entities of the same type.

Shared Task Datasets and Evaluation

Teams

- 45 unique team registrations.
- 12 teams submitted for FlatNER and Nested-NER.
- 35 submissions for FlatNER from eleven teams and 22 submissions for NestedNER from eight teams.

Results of Subtask 2 - NestedNER

Results of Subtask 1 - FlatNER

Download and Demo

https://sina.birzeit.edu/wojood/