

TEAM-NAME at SemEval-2023 Task 2: Your Paper Title

Anonymous ACL submission

Abstract

This document contains the instructions for preparing a camera-ready manuscript for the proceedings of SemEval 2023. The document itself conforms to its own specifications, and is therefore an example of what your manuscript should look like. **Please write a brief *abstract* describing your system and the results you obtained. You can also briefly mention any insights or interesting highlights from your results. We recommend you to include the name of your team in the abstract.**

Practical Information

This paper represents an example of how a system description paper may be structured. A bib file with relevant references is also included.

The example paper structure included here is an example, you can use a different structure if you prefer.

Paper Title: SemEval has its own paper naming conventions. The title of your paper should conform to the title of this template, that is:

TEAM-NAME at SemEval-2023 Task 2: Your Paper Title

Team Name: IMPORTANT Your team name should be the **same** as the name you provided, and was used in the ranking tables. This is the name that appear in the ranks and in the shared task report. If you choose to change your team name now, readers will not be able to locate your entry in the report. Please note that we cannot process team name change requests at this point.

Paper Length: Papers should contain up to **5 pages** of content plus bibliography if you only participated in one task. You can use **8 pages** plus bibliography if you participated in multiple sub-tasks (i.e. multiple languages).

Acknowledgments, references, and appendices do NOT count toward page limits.

Submission: The system papers are due **February 28th, 11:59 PM (UTC-12)**. The paper submission link will be provided on our website.

Please ensure that your paper contains all relevant information and meets the workshop standards. Remember that your paper will appear in the SemEval workshop proceedings which is published by ACL at the ACL Anthology. The workshop organizers reserve the right to reject system description papers that do not meet the workshop standards.

Finally, we encourage **all** teams to submit system description papers regardless of their system's performance. We are interested not only in the approaches that perform well but also what makes them different and what we can learn from them.

1 Introduction

You could begin with a brief description of the task and an overview of your approach.

We would like to ensure that future readers of your paper can find the relevant shared task description, data and results. According to the SemEval guidelines, you should cite the shared task report paper in your introduction, and we have included the citation here (Fetahu et al., 2023b). The key for this citation is multiconer2-report.

In addition to multilinguality, there were two key challenges in this year's task: (1) a fine-grained entity taxonomy with over 30 different classes, and (2) simulated errors added to the test set to make the task more realistic and difficult.

2 Related Work

In this section you can briefly describe other work in this area. We provide you with a bib file with references to relevant papers on NER for complex entities.

You should position the work as a continuation of the multilingual NER task started in 2022 (Malmasi et al., 2022b). You can briefly mentioned that the key challenges in the 2022 datasets were dealing with complex entities with limited context (Malmasi et al., 2022a). If you published a paper at the task last year, you should cite it here.

The challenges of NER for recognizing complex entities and in low-context situations was recently outlined by Meng et al. (2021). You can cite this paper to describe the problem, and refer to the different types of challenges they describe. Other work has extended this to multilingual and code-mixed settings (Fetahu et al., 2021). You can refer to this paper to describe the expansion to multilingual settings, which was also included in MultiCoNER.

3 Data

The data collection methods used to compile the dataset used in MultiCoNER will be described in a paper to be published shortly. You should cite this paper to refer to the data, and we have included the citation in this template (Fetahu et al., 2023a). The key for this citation is `multiconer2-data`.

If you use other datasets or any external data or knowledge bases, please include this information in your paper.

4 Methodology

The description of your system and your different submissions can be included here.

You can describe how your system was trained and evaluated.

5 Results

You can describe your results in this section.

If you competed in multiple tracks, you can try having a single large table with all your results, or multiple smaller tables.

There are different metrics that you can report. You should report the fine-grained F1 metrics,

and can also discuss the per-class performance. We also provide the coarse-grained metrics that you could report. All of these detailed results are available for your submissions in CodaLab (look at the output and scoring logs).

Please note: The official ranking metric is the overall fine-grained macro-averaged F1.

Feel free to include cross-validation results, or any baseline models you evaluated as well.

It is expected that you will interpret and discuss your results here.

6 Conclusion

Here you conclude your paper. The readers are interested not only in your system performance but also in what could be learned with your submission.

You can also include ideas for future work.

References

- Besnik Fetahu, Zhiyu Chen, Sudipta Kar, Oleg Rokhlenko, and Shervin Malmasi. 2023a. MultiCoNER v2: a Large Multilingual dataset for Fine-grained and Noisy Named Entity Recognition.
- Besnik Fetahu, Anjie Fang, Oleg Rokhlenko, and Shervin Malmasi. 2021. Gazetteer Enhanced Named Entity Recognition for Code-Mixed Web Queries. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1677–1681.
- Besnik Fetahu, Sudipta Kar, Zhiyu Chen, Oleg Rokhlenko, and Shervin Malmasi. 2023b. SemEval-2023 Task 2: Fine-grained Multilingual Named Entity Recognition (MultiCoNER 2). In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*. Association for Computational Linguistics.
- Shervin Malmasi, Anjie Fang, Besnik Fetahu, Sudipta Kar, and Oleg Rokhlenko. 2022a. MultiCoNER: a Large-scale Multilingual dataset for Complex Named Entity Recognition.
- Shervin Malmasi, Anjie Fang, Besnik Fetahu, Sudipta Kar, and Oleg Rokhlenko. 2022b. SemEval-2022 Task 11: Multilingual Complex Named Entity Recognition (MultiCoNER). In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. Association for Computational Linguistics.
- Tao Meng, Anjie Fang, Oleg Rokhlenko, and Shervin Malmasi. 2021. GEMNET: Effective gated gazetteer

representations for recognizing complex entities in low-context input. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1499–1512.