Extended Multilingual Protest News Detection - Shared Task 1, CASE 2021 and 2022

Ali Hürriyetoğlu
KNAW Humanities Cluster DHLab
ali.hurriyetoglu@dh.huc.knaw.nl

Osman Mutlu
Koc University
omutlu@ku.edu.tr

Fırat Duruşan
Koc University
fdurusan@ku.edu.tr

Onur Uca
Mersin University
onuruca@mersin.edu.tr

Alaeddin Selçuk Gürel
Huawei
alaeddinselcukgurel@gmail.com

Benjamin Radford
UNC Charlotte
benjamin.radford@uncc.edu

Yaoyao Dai
UNC Charlotte
yaoyao.dai@uncc.edu

Hansi Hettiarachchi
Birmingham City University
hansi.hettiarachchi@mail.bcu.ac.uk

Niklas Stoehr
ETH Zurich
niklas.stoehr@inf.ethz.ch

Tadashi Nomoto
National Institute of Japanese Literature
nomoto@acm.org

Milena Slavcheva
IICT, Bulgarian Academy of Sciences
milena@lml.bas.bg

Francielle Vargas
University of São Paulo
francielleavargas@usp.br

Aaqib Javid
Koc University
ajavid20@ku.edu.tr

Fatih Beyhan
Sabanci University
fatihbeyhan@sabanciuniv.edu

Erdem Yörük
Koc University
eryoruk@ku.edu.tr

Abstract

We report results of the CASE 2022 Shared Task 1 on Multilingual Protest Event Detection. This task is a continuation of CASE 2021 that consists of four subtasks that are i) document classification, ii) sentence classification, iii) event sentence coreference identification, and iv) event extraction. The CASE 2022 extension consists of expanding the test data with more data in previously available languages, namely, English, Hindi, Portuguese, and Spanish, and adding new test data in Mandarin, Turkish, and Urdu for Sub-task 1, document classification. The training data from CASE 2021 in English, Portuguese and Spanish were utilized. Therefore, predicting document labels in Hindi, Mandarin, Turkish, and Urdu occurs in a zero-shot setting. The CASE 2022 workshop accepts reports on systems developed for predicting test data of CASE 2021 as well. We observe that the best systems submitted by CASE 2022 participants achieve between 79.71 and 84.06 F1-macro for new languages in a zero-shot setting. The winning approaches are mainly ensembling models and merging data in multiple languages. The best two submissions on CASE 2021 data outperform submissions from last year for Subtask 1 and Subtask 2 in all languages. Only the following scenarios were not outperformed by new submissions on CASE 2021: Subtask 3 Portuguese & Subtask 4 English.

1 Introduction

We aim at determining event trigger and its arguments in a text snippet in the scope of an event extraction task. The performance of an automated system depends on the target event type as it may be broad or potentially the event trigger(s) can be ambiguous. The context of the trigger occurrence may need to be handled as well. For instance, the 'protest' event type may be synonymous with 'demonstration' or not in a specific context. Moreover, the hypothetical cases such as future protest plans may need to be excluded from the results. Finally, the relevance of a protest depends on the actors as in a contentious political event only citizen-led events are in the scope. This challenge is even harder in a cross-lingual and zero-shot setting in case training data are not available in new languages.

We provide a benchmark that consists of four subtasks and multiple languages in the scope of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text at The 2022 Conference on Empirical Methods in Natural Language Processing (CASE @ EMNLP 2022) (Hürriyetoğlu et al., 2022).[1] bgenfrhumil: To paraphrase: The work presented

[1] https://emw.ku.edu.tr/case-2022/, accessed on November 13, 2022.

in this paper is a continuation of the work initiated in CASE 2021 Task 1 (Hürriyetoğlu et al., 2021) and consists in adding new documents in already available languages, as well as adding new languages to the evaluation data.

Task 1 consists of the following subtasks that ensure the task is tackled incrementally: i) Document classification, ii) Sentence classification, iii) event sentence coreference identification, and iv) event extraction. The training data consist of documents in English, Portuguese, and Spanish, while the evaluation texts are in English, Hindi, Mandarin, Portuguese, Spanish, Turkish, and Urdu. Subtask 1 ensures documents with relevant senses of event triggers are selected. Next, Subtask 2 focuses on identifying event sentences in a document. Discriminating sentences that are about separate events and grouping them is done in Subtask 3 (Hürriyetoğlu et al., 2020, 2022). Finally, the sentences that are about the same events are processed to identify the event trigger and its arguments in Subtask 4. In addition to accomplishing the event extraction task, the subtask division improves significantly the annotation quality, as the annotation team can focus on a specific part of the task and errors in previous levels are corrected during the preparation of the following subtask (Hürriyetoğlu et al., 2021). The significance of this specific task division is twofold: i) facilitating the work with a random sample of documents by first identifying relevant documents and sentences before annotating or processing a sample or a complete archive of documents respectively; ii) increasing the generalizability of the automated systems that may be developed using this data (Yörük et al., 2021; Mutlu, 2022).

The current report is about Task 1 in the scope of CASE 2022. Task 2 (Zavarella et al., 2022) and Task 3 (Tan et al., 2022b,a) complement Task 1 by evaluating Task 1 systems on events related to COVID-19 and detecting causality respectively.

The following section, which is Section 2 describes the data we use for the shared task. Next we describe the evaluation setting in Section 3. The results are provided in Section 4. Finally, the Section 5 conclude this report.

2 Data

We used the CASE 2021 training data as those for CASE 2022.[2] The CASE 2022 test data are the union of CASE 2021 test data and additional new documents in both available and new languages. The new languages are Mandarin, Turkish, and Urdu.

The new document level data, which are used to extend CASE 2021 data, were randomly sampled from MOT v1.2 (Palen-Michel et al., 2022) [3] and were annotated by co-authors of this report. Documents were annotated by native speakers of the respective language. A single label was attached to each document. The annotation manual followed in the annotation process (Duruşan et al., 2022) was the same as that used in CASE 2021. The total number of CASE 2022 documents with labels is 3,870 for English, 267 for Hindi, 300 for Mandarin, 670 for Portuguese, 399 for Spanish, 300 for Turkish, and 299 for Urdu. Teams that developed systems for Subtasks 2, 3, and 4 evaluated their systems on CASE 2021 test data.

3 Evaluation setting

We utilized Codalab for evaluation of Task 1 for CASE 2022.[4] The evaluation for CASE 2021 was performed on an additional scoring page[5] of the original[6] CASE 2021 Codalab page. Moreover, we launched an additional scoring page for CASE 2022 after completion of the official evaluation period.[7] Five submissions per subtask and language pair could be submitted in total for CASE 2022. The additional scoring phase of both CASE 2021 and CASE 2022 allow only one submission per subtask and language combination per day. The test data of CASE 2021 were shared with participants at the same time with the training data. But the CASE 2022 evaluation data were shared around two weeks before the deadline for submission. The same evaluation scores that are F1-macro for Subtasks 1 and 2, CoNLL-2012[8] for Subtask 3, and CoNLL-2000[9] script for Subtask 4 were utilized.

[2] https://github.com/emerging-welfare/case-2021-shared-task for CASE 2021 and https://github.com/emerging-welfare/case-2022-multilingual-event for CASE 2022.

[3] https://github.com/bltlab/mot

[4] https://codalab.lisn.upsaclay.fr/competitions/7438, accessed on November 13, 2022.

[5] https://codalab.lisn.upsaclay.fr/competitions/7126, accessed on November 13, 2022.

[6] https://competitions.codalab.org/competitions/31247, which is not accessible due to change of the servers of Codalab.

[7] https://codalab.lisn.upsaclay.fr/competitions/7768, accessed on November 13, 2022.

[8] https://github.com/LoicGrobol/scorch, accessed on November 13, 2022.

[9] https://github.com/sighsmile/conlleval, accessed on November 13, 2022.

4 Results

Eighteen teams were registered for the task and obtained the training and test data for both CASE 2022 and CASE 2021. Ten and seven teams submitted their results for CASE 2021 and CASE 2022 respectively. Seven papers were submitted as system description papers to the CASE 2022 workshop in total. The scores of the submissions are calculated on two different Codalab pages for CASE 2021[10] and CASE 2022[11]. The teams that have participated are ARC-NLP (Sahin et al., 2022), CamPros (Kumari et al., 2022), CEIA-NLP (Fernandes et al., 2022), ClassBases (Wiriyathammabhum, 2022), EventGraph (You et al., 2022), NSUT-NLP (Suri et al., 2022), SPARTA (Müller and Dafnos, 2022). We provide details of the results and submissions of the participating teams for each subtask in the following subsections.[12]

4.1 CASE 2022 Subtask 1

The results for CASE 2022 subtask 1 are provided in Table 1. ARC-NLP finetune an ensemble of transformer-based language models and use ensemble learning, varying training data for each target language. They also perform tests with automatic translation of both training and test sets. They achieve 1st place both in Turkish and Mandarin, 2nd place in Portuguese and 3rd to 5th place in other languages. CEIA-NLP finetune XLM-Roberta-base transformers model with all the training data to achieve 1st place in Portuguese, 3rd or 4th places in other languages. ClassBases achieve 1st place in Hindi test data finetuning XLM-Roberta-large model, 5th or 6th places in other languages.

CamPros finetune XLM-Roberta-base model with all training data, and NSUT-NLP finetune mBERT while augmenting the data by translating different languages into each other.

[10] https://codalab.lisn.upsaclay.fr/competitions/7126#results, accessed on Nov 14, 2022.

[11] https://codalab.lisn.upsaclay.fr/competitions/7438#results, accessed on Nov 14, 2022.

[12] The results and system descriptions from participants that did not submit a system description paper are provided as well. This shows the capacity of the state-of-the-art systems on our benchmark. These systems are provided with their codalab names that are colabhero, fine_sunny_day, gauravsingh, lapardnemihk9989, lizhuoqun2021_iscas.

4.2 CASE 2021 Subtask 1

The extended results for CASE 2021 subtask 1 are provided in Table 2. The boldness indicates CASE 2022 entries. ClassBases finetune XLM-Roberta-large transformers model to perform 1st in Hindi and 2nd in Portuguese test data. They also achieve 5th and 6th places in Spanish and English respectively. Another team that submitted their model to CASE 2021 test data is ARC-NLP, taking 5th, 8th and 9th places in Portuguese, Spanish and English.

4.3 Subtask 2

The extended results for CASE 2021 subtask 2 are provided in Table 3. The boldness indicates CASE 2022 entries. ARC-NLP train an ensemble of transformers models using all training data to achieve 4th, 5th and 7th places in Spanish, English and Portuguese respectively. ClassBases finetune mLUKE-base for Portuguese and Spanish placing 5th in both, XLM-Roberta-large for English taking 8th place.[13]

4.4 Subtask 3

The extended results for CASE 2021 subtask 3 are provided in Table 4. The boldness indicates CASE 2022 entries. ARC-NLP achieve 1st place in both English and Spanish, 2nd place in Portuguese. They use an ensemble of English transformers models for English, Portuguese and Spanish test data. They train with only English data and translating Portuguese test data into English during model prediction. For Spanish test data, they train with English, translated Portuguese and translated Spanish, and test on translated Spanish data.

4.5 Subtask 4

The extended results for CASE 2021 subtask 4 are provided in Table 5. The boldness indicates CASE 2022 entries. SPARTA employ two methods. Both of these methods build on pretrained XLM-Roberta-large and use a data augmentation technique (sentence reordering). For English and Portuguese, they gather articles that contain protest events from outside sources and use them for further pretraining. For Spanish, they use an XLM-Roberta-large model that was further pretrained on CoNLL 2002 Spanish data. They take 1st place both in Portuguese and Spanish, 3rd place in English.

[13] CamPros do not describe their model for subtask 2.

Team	English	Portuguese	Spanish	Hindi	Turkish	Urdu	Mandarin
ARC-NLP	80.74_{4}	79.85_{2}	69.44_{5}	80.08_{4}	84.06_{1}	77.99_{3}	83.39_{1}
CEIA-NLP	80.77_{3}	80.07_{1}	73.19_{3}	78.17_{6}	82.43_{4}	77.65_{4}	77.63_{4}
CamPros	76.52_{7}	77.11_{6}	69.55_{4}	80.49_{2}	74.75_{6}	73.77_{6}	75.90_{6}
ClassBases	78.50_{6}	77.11_{5}	69.25_{6}	80.78_{1}	78.57_{5}	75.72_{5}	77.16_{5}
NSUT-NLP	80.62_{5}	73.02_{7}	64.45_{7}	56.71_{7}	67.02_{7}	65.55_{7}	75.45_{7}
fine_sunny_day	82.22_{2}	79.05_{4}	73.84_{2}	80.11_{3}	82.91_{2}	79.71_{1}	80.99_{3}
lizhuoqun2021_iscas	82.49_{1}	79.22_{3}	74.96_{1}	80.01_{5}	82.89_{3}	78.67_{2}	83.06_{2}

Table 1: The performance of the submissions in terms of F1-macro and their ranks as a subscript for each language and each team participating in CASE 2022 subtask 1.

Team	English	Hindi	Portuguese	Spanish
ALEM	80.82_{10}	N/A	72.98_{11}	46.47_{13}
AMU-EuraNova	53.46_{15}	29.66_{11}	46.47_{14}	46.47_{13}
DAAI	84.55_{3}	77.07_{6}	82.43_{4}	69.31_{10}
DaDeFrTi	80.69_{11}	78.77_{3}	77.22_{10}	73.01_{7}
FKIE_itf_2021	73.90_{13}	54.24_{10}	62.39_{12}	68.20_{11}
HSAIR	77.58_{12}	59.55_{9}	81.21_{7}	69.84_{9}
IBM MNLP IE	83.93_{4}	78.53_{5}	84.00_{3}	77.27_{3}
SU-NLP	81.75_{8}	N/A	N/A	N/A
NoConflict	51.94_{16}	N/A	N/A	N/A
jitin	67.39_{14}	70.49_{8}	52.23_{13}	62.05_{12}
ARC-NLP	81.35_{9}	N/A	81.73_{5}	72.42_{8}
ClassBases	82.30_{6}	80.78_{1}	85.39_{2}	73.48_{5}
colabhero	82.34_{5}	74.21_{7}	81.73_{5}	73.27_{6}
fine_sunny_day	85.00_{2}	N/A	80.74_{8}	82.45_{1}
gauravsingh	82.28_{7}	78.60_{4}	79.41_{9}	73.86_{4}
lizhuoqun2021_iscas	85.12_{1}	80.01_{2}	85.87_{1}	81.19_{2}

Table 2: The performance of the submissions in terms of F1-macro and their ranks as a subscript for each language and each team participating in CASE 2021 subtask 1. Bold teams indicate CASE 2022 entries.

Team	English	Portuguese	Spanish
ALEM	79.67_{9}	42.79_{15}	45.30_{15}
AMU-EuraNova	75.64_{14}	81.61_{11}	76.39_{11}
DaDeFrTi	79.28_{10}	86.62_{6}	85.17_{6}
FKIE_itf_2021	64.96_{16}	75.81_{13}	70.49_{14}
HSAIR	78.50_{11}	85.06_{8}	83.25_{8}
IBM MNLP IE	84.56_{4}	88.47_{3}	88.61_{2}
IIITT	82.91_{7}	79.51_{12}	75.78_{12}
SU-NLP	83.05_{6}	N/A	N/A
NoConflict	85.32_{3}	87.00_{4}	79.97_{10}
jiawei1998	76.14_{13}	84.67_{9}	83.05_{9}
jitin	66.96_{15}	69.02_{14}	72.94_{13}
ARC-NLP	83.77_{5}	86.53_{7}	87.20_{4}
CamPros	77.94_{12}	81.63_{10}	83.69_{7}
ClassBases	81.12_{8}	86.83_{5}	87.10_{5}
fine_sunny_day	85.75_{2}	89.67_{1}	88.78_{1}
lizhuoqun2021_iscas	85.93_{1}	88.86_{2}	88.61_{2}

Table 3: The performance of the submissions in terms of F1-macro and their ranks as a subscript for each language and each team participating in subtask 2. Bold teams indicate CASE 2022 entries.

ARC-NLP finetune an ensemble of transformers models for each language. They use all training data for Portuguese and Spanish, and only English for English test data. They achieve 2nd place in all languages. EventGraph aim to solve event extraction as semantic graph parsing. They use a graph encoding method where the labels for triggers and arguments are represented as node labels, also splitting multiple triggers. They use the pretrained XLM-Roberta-large as their encoder. They achieve 4th place both in English and Portuguese, 5th place in Spanish. ClassBases take 9th place in all languages finetuning XLM-Roberta-base transformers model.

Team	English	Portuguese	Spanish
DAAI	80.40_{4}	90.23_{6}	81.83_{6}
FKIE_itf_2021	77.05_{7}	91.33_{4}	82.52_{4}
Handshakes AI Research	79.01_{5}	90.61_{5}	81.95_{5}
IBM MNLP IE	84.44_{2}	92.84_{3}	84.23_{2}
NUS-IDS	81.20_{3}	93.03_{1}	83.15_{3}
SU-NLP	78.67_{6}	N/A	N/A
ARC-NLP	85.11_{1}	93.00_{2}	85.25_{1}

Table 4: The performance of the submissions in terms of CoNLL-2012 average score Pradhan et al. (2014) and their ranks as a subscript for each language and each team participating in subtask 3. Bold teams indicate CASE 2022 entries.

	Scores		
Team	English	Portuguese	Spanish
AMU-EuraNova	69.96_{7}	61.87_{8}	56.64_{8}
Handshakes AI Research	73.53_{5}	68.15_{6}	62.21_{6}
IBM MNLP IE	78.11_{1}	73.24_{3}	66.20_{3}
SU-NLP	2.58_{10}	N/A	N/A
jitin	66.43_{8}	64.19_{7}	58.35_{7}
ARC-NLP	77.83_{2}	73.84_{2}	67.99_{2}
ClassBases	46.88_{9}	12.52_{9}	37.09_{9}
EventGraph	74.76_{4}	71.72_{4}	64.48_{5}
SPARTA	76.60_{3}	74.56_{1}	69.86_{1}
lapardnemihk9989	72.18_{6}	70.98_{5}	64.83_{4}

Table 5: The performance of the submissions in terms of F1 score based on CoNLL-2003 (Tjong Kim Sang and De Meulder, 2003) and their ranks as a subscript for each language and each team participating in subtask 4. Bold teams indicate CASE 2022 entries.

5 Conclusion

The CASE 2022 extension consists of expanding the test data with more data in previously available languages, namely, English, Hindi, Portuguese, and Spanish, and adding new test data in Mandarin, Turkish, and Urdu for Sub-task 1, document classification. The training data from CASE 2021 in English, Portuguese and Spanish were utilized. Therefore, predicting document labels in Hindi, Mandarin, Turkish, and Urdu occurs in a zero-shot setting.

The CASE 2022 workshop accepts reports on systems developed for predicting test data of CASE 2021 as well. We observe that the best systems submitted by CASE 2022 participants achieve between 79.71 and 84.06 F1-macro for new languages in a zero-shot setting. The winning approaches are mainly ensembling models and merging data in multiple languages. The best two submissions on CASE 2021 data outperform submissions from last year for Subtask 1 and Subtask 2 in all languages. Only the following scenarios were not outperformed by new submissions on CASE 2021: Subtask 3 Portuguese & Subtask 4 English.

We aim at increasing number of languages and subtasks such as event coreference resolution (Hürriyetoğlu et al., 2022) and event type classification(Hürriyetoğlu et al., 2021) in the scope of following edition of this shared task.

References

Fırat Duruşan, Ali Hürriyetoğlu, Erdem Yörük, Osman Mutlu, Çağrı Yoltar, Burak Gürel, and Alvaro Comin. 2022. Global contentious politics database (glocon) annotation manuals.

Diogo Fernandes, Adalberto Junior, Gabriel da Mata Marques, Anderson da Silva Soares, and Arlindo Rodrigues Galvao Filho. 2022. CEIA-NLP at CASE 2022 Task 1: Protest News Detection for Portuguese. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2022)*, online. Association for Computational Linguistics (ACL).

Ali Hürriyetoğlu, Osman Mutlu, Erdem Yörük, Farhana Ferdousi Liza, Ritesh Kumar, and Shyam Ratan. 2021. Multilingual protest news detection - shared task 1, CASE 2021. In *Proceedings of the 4th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2021)*, pages 79–91, Online. Association for Computational Linguistics.

Ali Hürriyetoğlu, Hristo Tanev, Vanni Zavarella, Reyyan Yeniterzi, Osman Mutlu, and Erdem Yörük. 2022. Challenges and applications of automated extraction of socio-political events from text (case 2022): Workshop and shared task report. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2022)*, online. Association for Computational Linguistics (ACL).

Ali Hürriyetoğlu, Osman Mutlu, Fatih Beyhan, Fırat Duruşan, Ali Safaya, Reyyan Yeniterzi, and Erdem Yörük. 2022. Event coreference resolution for contentious politics events.

Ali Hürriyetoğlu, Erdem Yörük, Osman Mutlu, Fırat Duruşan, Çağrı Yoltar, Deniz Yüret, and Burak Gürel. 2021. Cross-Context News Corpus for Protest Event-Related Knowledge Base Construction. *Data Intelligence*, 3(2):308–335.

Ali Hürriyetoğlu, Vanni Zavarella, Hristo Tanev, Erdem Yörük, Ali Safaya, and Osman Mutlu. 2020. Automated extraction of socio-political events from news (AESPEN): Workshop and shared task report. In *Proceedings of the Workshop on Automated Extraction of Socio-political Events from News 2020*, pages 1–6, Marseille, France. European Language Resources Association (ELRA).

Neha Kumari, Mrinal Anand, Tushar Mohan, and Ponnurangam Kumaraguru. 2022. CamPros at CASE 2022 Task 1: Transformer-based Multilingual Protest News Detection. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2022)*, online. Association for Computational Linguistics (ACL).

Osman Mutlu. 2022. Utilizing coarse-grained data in low-data settings for event extraction.

Arthur Müller and Andreas Dafnos. 2022. SPARTA at CASE 2021 Task 1: Evaluating Different Techniques to Improve Event Extraction. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2022)*, online. Association for Computational Linguistics (ACL).

Chester Palen-Michel, June Kim, and Constantine Lignos. 2022. Multilingual open text release 1: Public domain news in 44 languages. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2080–2089, Marseille, France. European Language Resources Association.

Sameer Pradhan, Xiaoqiang Luo, Marta Recasens, Eduard Hovy, Vincent Ng, and Michael Strube. 2014. Scoring coreference partitions of predicted mentions: A reference implementation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 30–35, Baltimore, Maryland. Association for Computational Linguistics.

Umitcan Sahin, Oguzhan Ozcelik, Izzet Emre Kucukkaya, and Cagri Toraman. 2022. ARC-NLP at CASE 2022 Task 1: Ensemble Learning for Multilingual Protest Event Detection. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2022)*, online. Association for Computational Linguistics (ACL).

Manan Suri, Krish Chopra, and Adwita Arora. 2022. NSUT-NLP at CASE 2022 Task 1: Multilingual Protest Event Detection using Transformer-based Models. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2022)*, online. Association for Computational Linguistics (ACL).

Fiona Anting Tan, Ali Hürriyetoğlu, Tommaso Caselli, Nelleke Oostdijk, Hansi Hettiarachchi, Tadashi Nomoto, Onur Uca, and Farhana Ferdousi Liza. 2022a. Event causality identification with causal news corpus - shared task 3, CASE 2022. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2022)*, Online. Association for Computational Linguistics.

Fiona Anting Tan, Ali Hürrriyetoğlu, Tommaso Caselli, Nelleke Oostdijk, Tadashi Nomoto, Hansi Hettiarachchi, Iqra Ameer, Onur Uca, Farhana Ferdousi Liza, and Tiancheng Hu. 2022b. The causal news corpus: Annotating causal relations in event sentences from news. In *Proceedings of the Language Resources and Evaluation Conference*, pages 2298–2310, Marseille, France. European Language Resources Association.

Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the conll-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003 - Volume 4*, CONLL '03, page 142–147, USA. Association for Computational Linguistics.

Peratham Wiriyathammabhum. 2022. ClassBases at the CASE-2022 Multilingual Protest Event Detection Task: Multilingual Protest News Detection and Automatically Replicating Manually Created Event Datasets. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2022)*, online. Association for Computational Linguistics (ACL).

Huiling You, David Samuel, Samia Touileb, and Lilja Øvrelid. 2022. EventGraph at CASE 2021 Task 1: A General Graph-based Approach to Protest Event Extraction. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2022)*, online. Association for Computational Linguistics (ACL).

Erdem Yörük, Ali Hürriyetoğlu, Fırat Duruşan, and Çağrı Yoltar. 2021. Random sampling in corpus design: Cross-context generalizability in automated multicountry protest event collection. *American Behavioral Scientist*, 0(0):00027642211021630.

Vanni Zavarella, Hristo Tanev, Ali Hürriyetoğlu, Peratham Wiriyathammabhum, and Bertrand De Longueville. 2022. Tracking COVID-19 protest events in the United States. Shared Task 2: Event Database Replication, CASE 2022. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2022)*, online. Association for Computational Linguistics (ACL).