Massively Multi-Lingual Event Understanding: Extraction, Visualization, and Search

Chris Jenkins, Shantanu Agarwal, Joel Barry, Steven Fincke, Elizabeth Boschee
University of Southern California Information Sciences Institute
{cjenkins, shantanu, joelb, sfincke, boschee}@isi.edu

Abstract

In this paper, we present ISI-CLEAR, a state-of-the-art, cross-lingual, zero-shot event extraction system and accompanying user interface for event visualization & search. Using only English training data, ISI-CLEAR makes global events available on-demand, processing user-supplied text in 100 languages ranging from Afrikaans to Yiddish. We provide multiple event-centric views of extracted events, including both a graphical representation and a document-level summary. We also integrate existing cross-lingual search algorithms with event extraction capabilities to provide cross-lingual event-centric search, allowing English-speaking users to search over events automatically extracted from a corpus of non-English documents, using either English natural language queries (e.g. *cholera outbreaks in Iran*) or structured queries (e.g. find all events of type *Disease-Outbreak* with agent *cholera* and location *Iran*).

1 Introduction

Understanding global events is critical to understanding the world around us—whether those events consist of pandemics, political unrest, natural disasters, or cyber attacks. The breadth of events of possible interest, the speed at which surrounding socio-political event contexts evolve, and the complexities involved in generating representative annotated data all contribute to this challenge. Events are also intrinsically global: many downstream use cases for event extraction involve reporting not just in a few major languages but in a much broader context. The languages of interest for even a fixed task may still shift from day to day, e.g. when a disease emerges in an unexpected location.

The ISI-CLEAR (CROSS-LINGUAL EVENT & ARGUMENT RETRIEVAL) system meets these challenges by building state-of-the-art, language-agnostic event extraction models on top of massively multi-lingual language models. These event models require only English training data (not even bitext—no machine translation required) and can identify events and the relationships between them in at least a hundred different languages. Unlike more typical benchmark tasks explored for zero-shot cross-lingual transfer—e.g. named entity detection or sentence similarity, as in (Hu et al., 2020)—event extraction is a complex, structured task involving a web of relationships between elements in text.

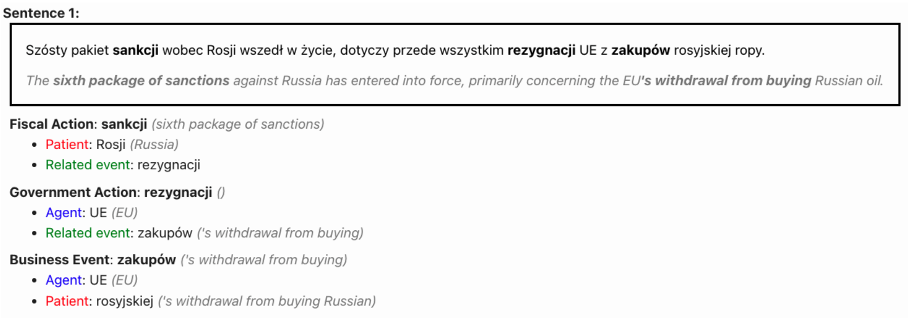
ISI-CLEAR makes these global events available to users in two complementary ways. First, users can supply their own text in a language of their choice; the system analyzes this text in that native language and provides multiple event-centric views of the data in response. Second, we provide an interface for cross-lingual event-centric search, allowing English-speaking users to search over events automatically extracted from a corpus of non-English documents. This interface allows for both natural language queries (e.g. *statements by Angela Merkel about Ukraine*) or structured queries (*event type = {Arrest, Protest}, location = Iraq*), and builds upon our existing cross-lingual search capabilities, demonstrated in (Boschee et al., 2019).

The primary contributions of this effort are threefold:

1. Strong, language-agnostic models for a complex suite of tasks, deployed in this demo on a hundred different languages and empirically tested on a representative variety of languages.
2. An event-centric user interface that presents events in intuitive text-based, graphical, or summary forms.
3. Novel integration of cross-lingual search capabilities with zero-shot cross-lingual event extraction.

We provide a video demonstrating the ISI-CLEAR user interface at `https://youtu.be/PE367pyuye8`.

Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics
Volume 3: System Demonstrations, pages 247–256
July 10-12, 2023 ©2023 Association for Computational Linguistics



Sentence 1:

Szósty pakiet **sankcji** wobec Rosji wszedł w życie, dotyczy przede wszystkim **rezygnacji** UE z **zakupów** rosyjskiej ropy.

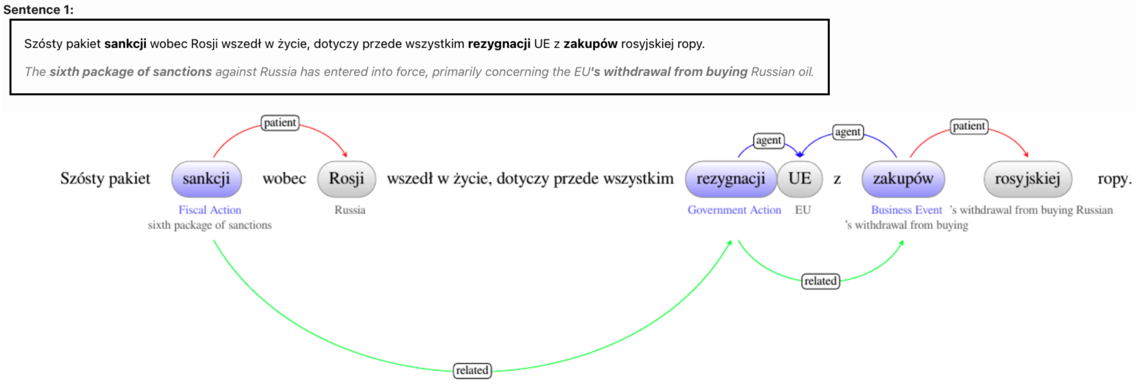
*The **sixth package of sanctions** against Russia has entered into force, primarily concerning the EU**'s withdrawal from buying** Russian oil.*

Fiscal Action: **sankcji** *(sixth package of sanctions)*
- Patient: Rosji *(Russia)*
- Related event: rezygnacji

Government Action: **rezygnacji** *()*
- Agent: UE *(EU)*
- Related event: zakupów *('s withdrawal from buying)*

Business Event: **zakupów** *('s withdrawal from buying)*
- Agent: UE *(EU)*
- Patient: rosyjskiej *('s withdrawal from buying Russian)*

Figure 1: Text-based display of Polish news. The user provides only the Polish text. To aid an English-speaking user, ISI-CLEAR displays the extracted event information not only in Polish but also in English. All processes—including anchor detection, argument extraction, machine translation and span-projection—are carried out in real time.



Sentence 1:

Szósty pakiet **sankcji** wobec Rosji wszedł w życie, dotyczy przede wszystkim **rezygnacji** UE z **zakupów** rosyjskiej ropy.

*The **sixth package of sanctions** against Russia has entered into force, primarily concerning the EU**'s withdrawal from buying** Russian oil.*

Figure 2: Graph-based display of event information extracted from user provided text in Polish.

2 User Interface

2.1 On-the-Fly Language-Agnostic Event Extraction & Display

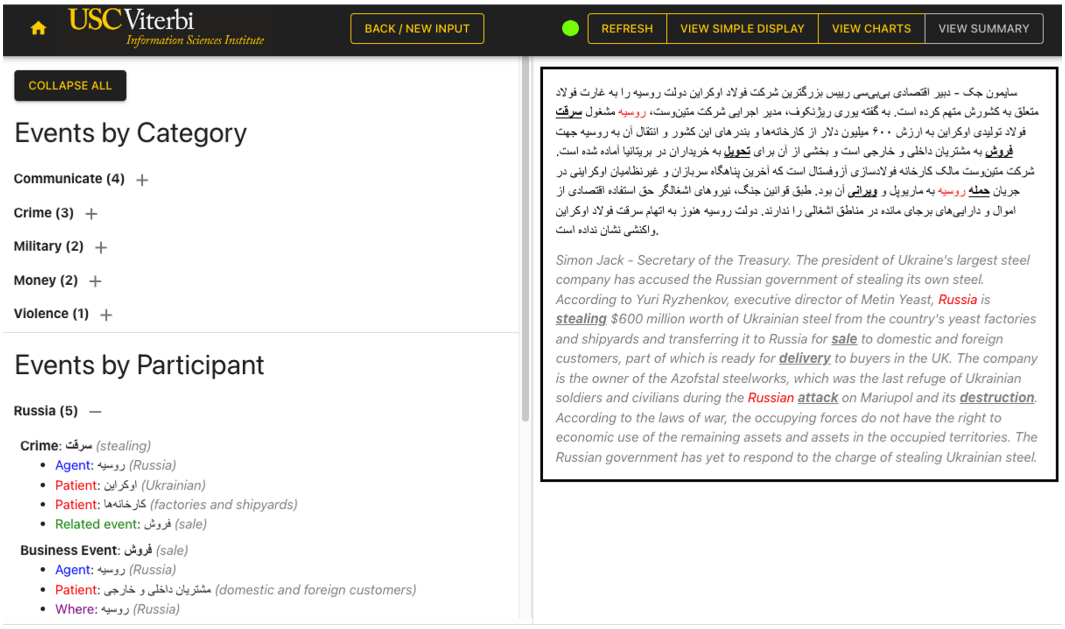
In our first mode, users are invited to supply their own text in a language of their choice. The system supports any language present in the underlying multi-lingual language model; for this demo we use XLM-RoBERTa (Conneau et al., 2020), which supports 100 languages ranging from Afrikaans to Yiddish.

After submission, the system displays the results in an initial text-based format, showing the events found in each sentence (Figure 1). For a more intuitive display of the relationships between events, users can select a graphical view (Figure 2). We can easily see from this diagram that the EU is the agent of both the *withdrawal* and the *buying* events, and that the two events are related (the EU is withdrawing from buying Russian oil).

Finally, the user can see an event-centric summary of the document, choosing to highlight either particular categories of event (e.g., *Crime*, *Military*, *Money*) or particular participants (e.g., *Ukraine*, *Putin*, *Russia*). When one or more categories or participants are selected, the system will highlight the corresponding events in both the original text and, where possible, in the machine translation. An example of a Farsi document is shown in Figure 3. Here, the system is highlighting three events in the document where Russia is either an agent or a patient of an event. For this demo, we use simple heuristics over English translations to group participant names and descriptions; in future work we plan to incorporate a zero-shot implementation of document co-reference to do this in the original language.

2.2 Cross-Lingual Event-Centric Search

The second mode of the ISI-CLEAR demo allows users to employ English queries to search over events extracted from a foreign language corpus. To enable this, we repurpose our work in cross-lingual document retrieval (Barry et al., 2020) to index and search over event arguments rather than whole documents. A query may specify target *event types* as well as *agent*, *patient*, or *location* arguments; it may also include additional words to constrain the *context*. A sample query might ask for *Communicate* events with the agent *Angela Merkel*



Events by Category

Communicate (4) +

Crime (3) +

Military (2) +

Money (2) +

Violence (1) +

Events by Participant

Russia (5) —

Crime: سرقت *(stealing)*

- Agent: روسیه *(Russia)*
- Patient: اوکراین *(Ukrainian)*
- Patient: کارخانه‌ها *(factories and shipyards)*
- Related event: فروش *(sale)*

Business Event: فروش *(sale)*

- Agent: روسیه *(Russia)*
- Patient: مشتریان داخلی و خارجی *(domestic and foreign customers)*
- Where: روسیه *(Russia)*

سایمون جک - دبیر اقتصادی بی‌بی‌سی رییس بزرگترین شرکت فولاد اوکراین دولت روسیه را به غارت فولاد متعلق به کشورش متهم کرده است. به گفته یوری ریژنکوف، مدیر اجرایی شرکت متین‌وست، روسیه مشغول **سرقت** فولاد تولیدی اوکراین به ارزش ۶۰۰ میلیون دلار از کارخانه‌ها و بندرهای این کشور و انتقال آن به روسیه جهت **فروش** به مشتریان داخلی و خارجی است و بخشی از آن برای **تحویل** به خریداران در بریتانیا آماده شده است. شرکت متین‌وست مالک کارخانه فولادسازی آزوفستال است که آخرین پناهگاه سربازان و غیرنظامیان اوکراینی در جریان **حمله** روسیه به ماریوپل و **ویرانی** آن بود. طبق قوانین جنگ، نیروهای اشغالگر حق استفاده اقتصادی از اموال و دارایی‌های برجای مانده در مناطق اشغالی را ندارند. دولت روسیه هنوز به اتهام سرقت فولاد اوکراین واکنشی نشان نداده است.

*Simon Jack - Secretary of the Treasury. The president of Ukraine's largest steel company has accused the Russian government of stealing its own steel. According to Yuri Ryzhenkov, executive director of Metin Yeast, Russia is **stealing** $600 million worth of Ukrainian steel from the country's yeast factories and shipyards and transferring it to Russia for **sale** to domestic and foreign customers, part of which is ready for **delivery** to buyers in the UK. The company is the owner of the Azofstal steelworks, which was the last refuge of Ukrainian soldiers and civilians during the Russian **attack** on Mariupol and its **destruction**. According to the laws of war, the occupying forces do not have the right to economic use of the remaining assets and assets in the occupied territories. The Russian government has yet to respond to the charge of stealing Ukrainian steel.*

Figure 3: Event-centric summary of Farsi document.

and the context *Ukraine*.

Query specification. We allow queries to be specified in two ways. The first simply asks the user to directly specify the query in structured form: using checkboxes to indicate which event types should be included and directly typing in values for each condition (*agent*, *patient*, etc.). A second and more intuitive method allows users to enter a query as natural language. The system processes the query using the ISI-CLEAR event system and populates a structured query automatically from the results. For instance, if the user enters the phrase *anti-inflation protests in Vietnam*, ISI-CLEAR will detect a *Protest* event with location *Vietnam* in that phrase. It will turn this result into a query with event type *Protest*, location *Vietnam*, and additional context word *anti-inflation*.

Display. We display corpus events in ranked order with respect to the user query. The ranking is a combination of system confidence in the underlying extractions (e.g., is this event *really* located in Vietnam?) and system confidence in the cross-lingual alignment (e.g., is *étudiants internationaux* really a good match for the query phrase *foreign students*?). To estimate the latter, we rely on our prior work in cross-lingual retrieval, where we developed state-of-the-art methods to estimate the likelihood that foreign text f conveys the same meaning as English text e (Barry et al., 2020). We note that for locations, we include containing countries (as determined via Wikidata) in the index so that a search for *Iran* will return events happening in, e.g., *Tehran*. More specific details on the ranking functions can be found in Appendix A.3.

As part of our display, we break down system confidence by query condition—that is, we separately estimate the system's confidence in the *agent* vs., say, the *location*. For each condition, we display a "traffic light" indicator that shows the system's confidence in that condition for an event. Red, yellow, and green indicate increasing levels of confidence; black indicates that there is no evidence for a match on this condition, but that other conditions matched strongly enough for the event to be returned. A sample natural language query and search results are shown in Figure 4.

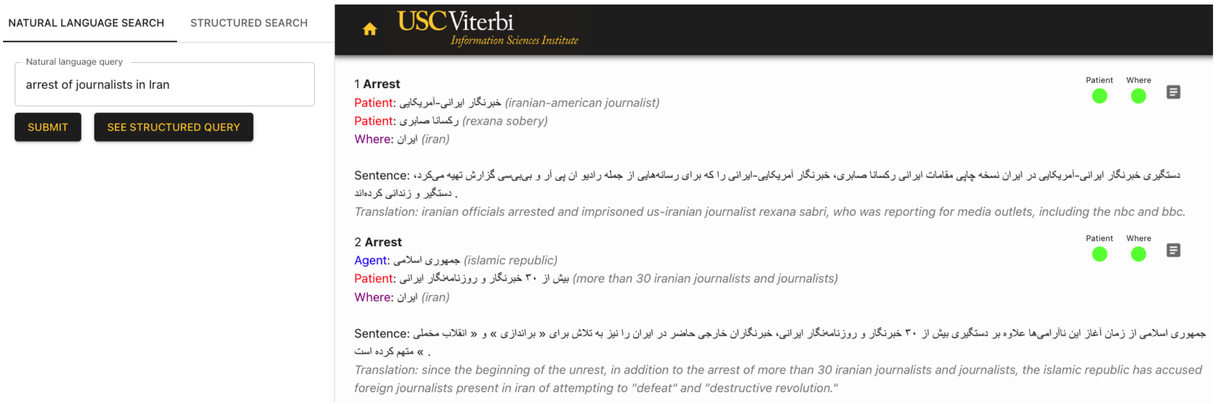
Corpora. For this demo, we support two corpora: (1) 20,000 Farsi news documents drawn from Common Crawl[1] and (2) ∼55K Weibo messages (in Chinese) on the topic of the Russo-Ukrainian crisis (Fung and Ji, 2022).

3 Ontology & Training Data

The ISI-CLEAR demo system is compatible with any event ontology that identifies a set of event types and argument roles. The system expects sentence-level English training data that identifies, for each event, one or more anchor spans and zero or more argument spans (with roles).

For this demonstration, we use the "basic event" ontology and data developed for the IARPA BET-

[1] https://commoncrawl.org/



Natural language query: arrest of journalists in Iran

SUBMIT SEE STRUCTURED QUERY

1 Arrest Patient ● Where ●

Patient: خبرنگار ایرانی-آمریکایی (iranian-american journalist)

Patient: رکسانا صابری (rexana sobery)

Where: ایران (iran)

Sentence: دستگیری خبرنگار ایرانی-آمریکایی در ایران نسخه چاپی مقامات ایرانی رکسانا صابری، خبرنگار آمریکایی-ایرانی را که برای رسانه‌هایی از جمله رادیو ان پی آر و بی‌بی‌سی گزارش تهیه می‌کرد، دستگیر و زندانی کرده‌اند .

Translation: iranian officials arrested and imprisoned us-iranian journalist rexana sabri, who was reporting for media outlets, including the nbc and bbc.

2 Arrest Patient ● Where ●

Agent: جمهوری اسلامی (islamic republic)

Patient: بیش از ۳۰ خبرنگار و روزنامه‌نگار ایرانی (more than 30 iranian journalists and journalists)

Where: ایران (iran)

Sentence: جمهوری اسلامی از زمان آغاز این ناآرامی‌ها علاوه بر دستگیری بیش از ۳۰ خبرنگار و روزنامه‌نگار ایرانی، خبرنگاران خارجی حاضر در ایران را نیز به تلاش برای « براندازی » و « انقلاب مخملی » متهم کرده است .

Translation: since the beginning of the unrest, in addition to the arrest of more than 30 iranian journalists and journalists, the islamic republic has accused foreign journalists present in iran of attempting to "defeat" and "destructive revolution."

Figure 4: Example of search results.

TER program (available at https://ir.nist.gov/better/). The ontology consists of 93 event types and a small set of argument roles (*agent*, *patient*, and *related-event*). In other settings, we have trained and tested the underlying system on the publicly available ACE event ontology[2], showing state-of-the-art zero-shot cross-lingual results in (Fincke et al., 2022). We prefer the BETTER ontology for this demo because of its broad topical coverage and its inclusion of event-event relations (in the form of *related-event* arguments). The ISI-CLEAR system is also designed to attach general-purpose *when* and *where* arguments to any event, regardless of ontology; see section 4.5.

4 System Components

We present here the highlights of our technical approach, which relies on a collection of strong, language-agnostic models to perform all aspects of event extraction and the classification of relationships between events, as well as machine translation and foreign-to-English projection of event output (for display purposes).

4.1 Ingest & Tokenization

Consistent with XLM-RoBERTa, we use Sentence Piece (Kudo and Richardson, 2018) to tokenize text, and at extraction time, our models label each input subword separately. For languages where words are typically surrounded by whitespace, our system then expands spans to the nearest whitespace (or punctuation) to improve overall performance. If the system produces a conflicting sequence of labels for a single word, we apply simple heuristics leveraging label frequency statistics to produce just one label.

4.2 Anchor Detection

ISI-CLEAR performs anchor identification and classification using a simple beginning-inside-outside (BIO) sequence-labeling architecture composed of a single linear classification layer on top of the transformer stack. For more details please see (Fincke et al., 2022).

4.3 Argument Attachment

For argument attachment, we consider one event anchor A and one role R at a time. We encourage the system to focus on A and R by modifying the input to the language model. For instance, when A=*displaced* and R=1 (*agent*), the input to the language model will be *displaced ; 1 </s> Floods < displaced > thousands last month*. This modification encourages the language model to produce representations of tokens like *thousands* that are contextualized by the anchor and role being examined. The argument attachment model concatenates the language model output vector for each input token with an embedding for event type and applies a linear classifier to generate BIO labels. For more details please see (Fincke et al., 2022).

4.4 Event-Event Relations

ISI-CLEAR can handle arbitrary event-event relations within a sentence, including the special case of event co-reference (when a given event has two or more anchor spans). We consider one event anchor A_1 at a time. Again we modify the input to the language model (by marking A_1 with special characters on either side) to encourage the model to consider all other anchors in light of A_1. We then represent each event anchor in the sentence (including A_1 itself) as a single vector, generated

[2] https://www.ldc.upenn.edu/collaborations/past-projects/ace

by feeding the language model output for its constituent tokens into a bi-LSTM and then concatenating the bi-LSTM's two final states. (This allows us to smoothly handle multi-word anchors.) To identify the relationship between A_1 and A_2, if any, we then concatenate the representations for A_1 and A_2 and pass the result to a linear classifier. The final step optimizes over the scores of all such pairwise classifications to label all relations in the sentence.

4.5 When & Where

The ontology used for this demonstration (described in Section 3) does not annotate *when* and *where* arguments. However, these event attributes are critical for downstream utility. We therefore deploy an ontology-agnostic model that can assign dates and locations to events of any type. To do this, we train a question-answering model to answer questions such as *<s> When/Where did the {anchor} happen? </s> Context </s>*. We first train the model on the SQUAD2 dataset (Rajpurkar et al., 2016) and then continue training on the event location and time annotations in the English ACE dataset.

4.6 Machine Translation & Projection

All event extraction happens in the target language; no machine translation (or bitext) is required. However, for system output to be useful to English speakers, translation is highly beneficial. Here, we rely on the 500-to-1 translation engine developed by our collaborators at ISI (Gowda et al., 2021)[3]. Translation happens after event extraction. We have not optimized this deployment of MT for speed, so we display the results without translation first and then (when the small light in the top toolbar turns green, usually after a few seconds), we can refresh the screen to show results with translations added.

To project anchor and argument spans into machine translation, we require no parallel data for training. Instead, we leverage the fact that the pre-trained XLM-RoBERTa embeddings are well aligned across languages and have been shown to be effective for word alignment tasks (Dou and Neubig, 2021). The similarity of a word in a foreign-language sentence to a word in the parallel English sentence is determined by the cosine distance between the embeddings of the two words. We leverage the Itermax algorithm (Jalili Sabet et al., 2020) to find the best phrase matches. Since we avoid making any bespoke language specific decisions, our projection technique is highly scalable and can project from any of the 100 languages on which XLM-RoBERTa was pre-trained on.

5 System Evaluation & Analysis

We evaluate our system on a variety of languages and ontologies and compare where possible to existing baselines. Following community practice, e.g. Zhang et al. (2019), we consider an anchor correct if its offsets and event type are correct, and we consider an argument correct if its offsets, event type, and role find a match in the ground truth. For event coreference (same-sentence only), we consider each anchor pair separately to produce an overall F-score.

Table 1 provides overall scores in several settings where multi-lingual event annotations are available. All models are trained on English data only. For the ACE data, we follow (Huang et al., 2022). The BETTER Basic task is described in Section 3; there are two ontologies (Basic-1 and Basic-2) from different phases of the originating program. The BETTER Abstract task is similar to BETTER Basic, but all action-like phrases are annotated as events, with no further event type specified[4]; valid roles are only *agent* and *patient* (McKinnon and Rubino, 2022). More dataset statistics are found in Appendix A.1.

It is difficult to compare system accuracy across languages; a lower score in one language may reflect a real difference in performance across languages—or just that one set of documents is harder than another. Still, we observe the following. First, performance on anchors seems most sensitive to language choice—for instance, we note that Arabic and Chinese anchor performance on ACE differs by almost 10 points. For arguments, however, non-English performance is relatively consistent given a task—but varies more widely between tasks. Second, we note that cross-lingual performance seems best on anchors, where it exceeds 80% of English performance for all but one condition. In contrast, argument performance varies more widely, with many conditions below 70% of English (though some as high as 89%).

We also compare against existing published baselines where possible. There are relatively few published results on cross-lingual event anchor detection (and none that we could find on the task of

[3] Available at `http://rtg.isi.edu/many-eng/`.

[4] Since abstract events lack event types, we also require anchor offsets to match when scoring arguments.

Task	ACE			Basic-1		Basic-2		Abstract			
Language	en	ar	zh	en	ar	en	fa	en	ar	fa	ko
Anchors	71.2	58.1	49.6	64.2	52.5	64.6	54.3	87.4	78.3	72.5	78.9
Arguments	72.1	51.5	51.7	64.5	51.5	71.6	64.0	69.8	45.0	45.7	45.0
Event coreference	–	–	–	83.4	67.9	86.5	65.9	–	–	–	–

Table 1: Component-level accuracy by language / task. Dataset statistics are available in Appendix A.1. ACE lacks same-sentence event coreference so those figures are omitted. Event coreference is peripheral to the overall Abstract task; we chose to not model it explicitly and exclude it here.

cross-lingual event co-reference as defined here). To benchmark performance on anchors, we turn to MINION (Pouran Ben Veyseh et al., 2022), a multi-lingual anchor-only dataset that uses a derivative of the ACE ontology. For a fair comparison, we retrained our model (tuned for use with XLM-RoBERTa large) with XLM-RoBERTa base; we did not adjust any hyperparameters. Table 2 shows that the ISI-Clear model performs on average 2.7 points better than the reported MINION numbers for cross-lingual settings. We also show the numbers from our actual demo models (trained with XLM-RoBERTa large) for comparison.

	base			large
	MINION	ISI-Clear	Δ	ISI-Clear
en	**79.5**	78.9	-0.6	78.0
es	**62.8**	62.3	-0.5	65.3
pt	**72.8**	71.1	-1.7	75.0
pl	**60.1**	52.6	-7.5	66.4
tr	47.2	**52.0**	+4.8	56.5
hi	58.2	**72.2**	+14.0	72.7
ko	56.8	**64.1**	+7.3	63.5
AVG	59.7	**62.4**	+2.7	66.6

Table 2: Cross-lingual anchor detection (F1) for MINION dataset, training on English only. Average is across all cross-lingual settings.

For argument detection, much more published work exists, and we show in Table 3 that ISI-Clear achieves state-of-the-art performance on all ACE datasets, comparing against the previous state-of-the-art as reported in Huang et al. (2022).

	X-GEAR	ISI-Clear
en	71.2	**72.1**
ar	44.8	**51.5**
zh	51.5	**51.7**

Table 3: Cross-lingual argument detection (F1) for ACE over gold anchors, training on English only.

6 Related Work

Several recent demos have presented multi-lingual event extraction in some form, but most assume training data in each target language (e.g. Li et al. (2019) or Li et al. (2020)) or translate foreign-language text into English before processing (e.g. Li et al. (2022)). In contrast, the focus of our demo is making events available in languages for which no training data exists. Other demos have shown the potential of zero-shot cross-lingual transfer, but on unrelated tasks, e.g. offensive content filtering (Pelicon et al., 2021). Akbik et al. (2016) uses annotation projection from English FrameNet to build target-language models for frame prediction; the focus of the demo is then on building effective queries over language-agnostic frame semantics for extraction. Finally, Xia et al. (2021) also produce FrameNet frames cross-lingually (using XLM-RoBERTa), but in contrast to our work, several of their supporting models use target-language data, and they also supply only a simpler user interface and lack the cross-lingual search-by-query capability that is a key aspect of our demo.

7 Conclusion

ISI-Clear provides a monolingual English-speaking user with effective access to global events, both on-demand (extracting events from input of a user's choice) or as a set of indexed documents accessible via cross-lingual search. The system provides a variety of visualizations and modes for engaging with system results. We look forward to future work improving the quality of the underlying components and exploring additional capabilities to cross language barriers and expand access to information around the globe.

Limitations

Our core approach is limited by the underyling multi-lingual language model it employs. For this demo, we are therefore limited to the 100 languages that make up the XLM-RoBERTa training set. Performance also varies across languages, tracking in part (though not in whole) with the volume of training data available for each language when building the multi-lingual language model. For instance, anecdotally, the performance on Yiddish (34M tokens in the CC-100 corpus used to train XLM-RoBERTa) is inferior to that of Farsi (13259M tokens). We have provided empirical results for eleven languages and five tasks, but it would be ideal to have a broader set of test conditions; unfortunately, annotated datasets for events are much less common than for simpler tasks like named entity recognition.

A second limitation of our system involves compute requirements. We employ multiple separate components for event extraction (e.g., for anchor detection vs. argument attachment), which increases memory/GPU footprint compared to a more unified system.

Finally, our system assumes an existing ontology and (English) training data set; it would be interesting to explore zero-shot ontology expansion in future work.

Ethics Statement

One important note is that our system is designed to extract information about events that are reported in text, with no judgment about their validity. This can lead a user to draw false conclusions. For instance, the system might return many results for a person X as the agent of a *Corruption* event, but this does not necessarily mean that X is actually corrupt. This should be prominently noted in any use case for this demonstration system or the underlying technologies.

Acknowledgements

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via Contract No. 2019-19051600007. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References

Alan Akbik, Laura Chiticariu, Marina Danilevsky, Yonas Kbrom, Yunyao Li, and Huaiyu Zhu. 2016. Multilingual information extraction with PolyglotIE. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pages 268–272, Osaka, Japan. The COLING 2016 Organizing Committee.

Joel Barry, Elizabeth Boschee, Marjorie Freedman, and Scott Miller. 2020. SEARCHER: Shared embedding architecture for effective retrieval. In *Proceedings of the workshop on Cross-Language Search and Summarization of Text and Speech (CLSSTS2020)*, pages 22–25, Marseille, France. European Language Resources Association.

Elizabeth Boschee, Joel Barry, Jayadev Billa, Marjorie Freedman, Thamme Gowda, Constantine Lignos, Chester Palen-Michel, Michael Pust, Banriskhem Kayang Khonglah, Srikanth Madikeri, Jonathan May, and Scott Miller. 2019. SARAL: A low-resource cross-lingual domain-focused information retrieval system for effective rapid document triage. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 19–24, Florence, Italy. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Zi-Yi Dou and Graham Neubig. 2021. Word alignment by fine-tuning embeddings on parallel corpora. *CoRR*, abs/2101.08231.

Steven Fincke, Shantanu Agarwal, Scott Miller, and Elizabeth Boschee. 2022. Language model priming for cross-lingual event extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Yi R. Fung and Heng Ji. 2022. A weibo dataset for the 2022 russo-ukrainian crisis.

Thamme Gowda, Zhao Zhang, Chris Mattmann, and Jonathan May. 2021. Many-to-English machine translation tools, data, and pretrained models. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th*

International Joint Conference on Natural Language Processing: System Demonstrations, pages 306–316, Online. Association for Computational Linguistics.

Junjie Hu, Sebastian Ruder, Aditya Siddhant, Graham Neubig, Orhan Firat, and Melvin Johnson. 2020. XTREME: A massively multilingual multi-task benchmark for evaluating cross-lingual generalisation. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 4411–4421. PMLR.

Kuan-Hao Huang, I-Hung Hsu, Prem Natarajan, Kai-Wei Chang, and Nanyun Peng. 2022. Multilingual generative language models for zero-shot cross-lingual event argument extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4633–4646, Dublin, Ireland. Association for Computational Linguistics.

Masoud Jalili Sabet, Philipp Dufter, François Yvon, and Hinrich Schütze. 2020. SimAlign: High quality word alignments without parallel training data using static and contextualized embeddings. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1627–1643, Online. Association for Computational Linguistics.

Taku Kudo and John Richardson. 2018. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 66–71, Brussels, Belgium. Association for Computational Linguistics.

Manling Li, Revanth Gangi Reddy, Ziqi Wang, Yi-shyuan Chiang, Tuan Lai, Pengfei Yu, Zixuan Zhang, and Heng Ji. 2022. COVID-19 claim radar: A structured claim extraction and tracking system. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 135–144, Dublin, Ireland. Association for Computational Linguistics.

Manling Li, Ying Lin, Joseph Hoover, Spencer Whitehead, Clare Voss, Morteza Dehghani, and Heng Ji. 2019. Multilingual entity, relation, event and human value extraction. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pages 110–115, Minneapolis, Minnesota. Association for Computational Linguistics.

Manling Li, Alireza Zareian, Ying Lin, Xiaoman Pan, Spencer Whitehead, Brian Chen, Bo Wu, Heng Ji, Shih-Fu Chang, Clare Voss, Daniel Napierski, and Marjorie Freedman. 2020. GAIA: A fine-grained multimedia knowledge extraction system. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 77–86, Online. Association for Computational Linguistics.

Timothy McKinnon and Carl Rubino. 2022. The IARPA BETTER program abstract task four new semantically annotated corpora from IARPA's BETTER program. In *Proceedings of The 14th Language Resources and Evaluation Conference*.

Andraž Pelicon, Ravi Shekhar, Matej Martinc, Blaž Škrlj, Matthew Purver, and Senja Pollak. 2021. Zero-shot cross-lingual content filtering: Offensive language and hate speech detection. In *Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation*, pages 30–34, Online. Association for Computational Linguistics.

Amir Pouran Ben Veyseh, Minh Van Nguyen, Franck Dernoncourt, and Thien Nguyen. 2022. MINION: a large-scale and diverse dataset for multilingual event detection. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2286–2299, Seattle, United States. Association for Computational Linguistics.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. SQuAD: 100,000+ Questions for Machine Comprehension of Text. *arXiv e-prints*, page arXiv:1606.05250.

Milan Straka. 2018. UDPipe 2.0 prototype at CoNLL 2018 UD shared task. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 197–207, Brussels, Belgium. Association for Computational Linguistics.

Patrick Xia, Guanghui Qin, Siddharth Vashishtha, Yunmo Chen, Tongfei Chen, Chandler May, Craig Harman, Kyle Rawlins, Aaron Steven White, and Benjamin Van Durme. 2021. LOME: Large ontology multilingual extraction. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 149–159, Online. Association for Computational Linguistics.

Tongtao Zhang, Heng Ji, and Avirup Sil. 2019. Joint entity and event extraction with generative adversarial imitation learning. *Data Intelligence*, 1:99–120.

A Appendix

A.1 Dataset Statistics

We report results for a variety of different tasks in a variety of different languages. We outline the sizes for these diverse datasets in Tables 4 and 5. The tasks use five different ontologies; we also report the number of event types for each ontology in Table 6.

A.2 Speed

Table 7 presents speed results for six representative languages, calculated as number of seconds per

	Train		Development	
	# Characters	# Events	# Characters	# Events
ACE	1,335,035	4,202	95,241	450
Basic-1	171,267	2,743	35,590	560
Basic-2	419,642	5,995	87,425	1,214
Abstract	557,343	12,390	67,266	1,499
MINION	4,388,701	14,189	544,758	1,688

Table 4: Size of English training and development sets in number of documents and number of events.

	Lang.	# Characters	# Events
ACE	en	104,609	403
	ar	44,003	198
	zh	22,452	189
Basic-1	en	33,169	569
	ar	238,133	5,172
Basic-2	en	82,296	1,139
	fa	639,6951	11,559
Abstract	en	68,863	1,527
	ar	189,174	5,339
	fa	607,429	15,005
	ko	327,811	16,704
MINION	en	554,680	1,763
	es	161,159	603
	pt	73,610	200
	pl	197,270	1,234
	tr	175,823	814
	hi	57,453	151
	ko	332,023	164

Table 5: Size of test sets in number of documents and number of events.

Ontology	# of Event Types
ACE	33
Basic-1	69
Basic-2	93
Abstract	1
MINION	16

Table 6: Number of event types in each ontology.

100 “words”. For this exercise we consider words to be the output of UDPipe’s language-specific tokenization (Straka, 2018). The primary driver of speed difference is that, given XLM-RoBERTa’s fixed vocabulary, different languages will split into more or fewer subwords on average. For instance, an average Korean word will produce at least half again as many subwords than, say, an average Farsi word; this is presumably why 100 words of Korean takes about 70% longer to process than 100 words of Farsi. On average, for a standard short news article (200 words), we expect to wait about two seconds for extraction and an additional six or seven seconds for MT and projection. We did not optimize our selection of MT package for speed (e.g., it decodes one sentence at a time instead of batching); this could easily be updated in future work to be more efficient.

	en	ar	fa	ko	ru	zh
Event	1.1	1.0	0.9	1.5	0.8	1.1
Display	n/a	2.6	2.8	4.1	3.4	3.9

Table 7: Processing speed (seconds per 100 words). Event processing includes ingest, tokenization, anchors, arguments, event-event relationships, and when/where extraction. Display processing includes components solely required for display (MT and projection). We use 11GB GTX 1080Ti GPUs for extraction/projection and use a 48GB Quadro RTX 8000 GPU for MT.

A.3 Search Ranking

ISI-CLEAR extracts a large number of events from the documents indexed from search, some of which vary in quality and some of which will match more or less confidently to an English query. The ranking function described here significantly improves the usability of our search results.

The goal of our search ranking function is to rank each extracted event E with respect to a user query Q. To calculate $score(Q, E)$, we combine two separate dimensions of system confidence:

1. *Cross-lingual alignment confidence (CAC)*: are the components of E reasonable translations of the query terms? For instance, is *étu-*

diants internationaux a good match for the query phrase *foreign students*? Here, we assume the existence of a cross-lingual retrieval method $cac(e, f)$ that estimates the likelihood that foreign text f conveys the same meaning as English text e, as in our prior work (Barry et al., 2020).

2. *Extraction confidence (EC)*: how likely is it that the elements of E were correctly extracted in the first place? Here we use confidence measures (denoted ec) produced by individual system components.

To combine these dimensions, we consider each query condition separately (summing the results). For simplicity we describe the scoring function for the *agent* condition:

$$
\begin{aligned}
&score(Q_{agent}, E_{agent}) = \\
&\quad \beta * ec(E_{agent}) * cac(Q_{agent}, E_{agent}) + \\
&\qquad\qquad (1 - \beta) * cac(Q_{agent}, E_{sentence})
\end{aligned}
$$

The first term of this equation captures the two dimensions described above. The second term allows us to account for agents missed by the system, letting us give "partial credit" when the user's search term is at least found in the nearby context (e.g., in $E_{sentence}$). Based on empirical observation, we set β to 0.75.

We follow the same formula for *patient* and *location*. For *context* we use only the final term $cac(Q_{topic}, E_{sentence})$ since *context* does not directly correspond to an event argument.

For now, event type operates as a filter with no score attached; in future work we will incorporate both the system's confidence in the event type as well as a fuzzy match over nearby event types (e.g., allowing for confusion between *Indict* and *Convict*).