

THE IARPA BETTER PROGRAM ABSTRACT TASK: FOUR NEW SEMANTICALLY ANNOTATED CORPORA FROM IARPA'S BETTER PROGRAM



Carl Rubino and Timothy McKinnon
Intelligence Advanced Research Projects Activity
Carl.Rubino,Timothy.McKinnon@iarpa.gov
LREC, June 20-25, 2022, Marseille



Abstract

IARPA's Better Extraction from Text Towards Enhanced Retrieval (BETTER) Program created multiple multilingual datasets to spawn and evaluate cross-language information extraction and information retrieval research and development in zero-shot conditions. The first set of these resources for information extraction, the "Abstract" data will be released to the public at LREC 2022 in four languages to champion further information extraction work in this area. This paper presents the event and argument annotation in the Abstract Evaluation phase of BETTER, as well as the data collection, preparation, partitioning and mark-up of the datasets.
Keywords: CLIR, CLIE, zero shot, human-in-the-loop.

The IARPA BETTER Program

The Intelligence Advanced Research Projects Activity (IARPA) kicked off the BETTER program in Boston on October 2019 to advance research in multilingual, cross-lingual and zero-shot information extraction (IE) and information retrieval (IR). BETTER was designed as a 42-month program, broken into three phases of decreasing length (18/12/12 months), each with its own evaluation language(s) and topic domain(s). The program goal was to develop enhanced methods for personalized, multilingual semantic extraction and retrieval from multilingual newswire text. Four large teams (or "Performers") were competitively selected to participate in the program based on their responses to a Broad Agency Announcement (BAA). The prime contractors for these teams were the University of Southern California Information Sciences Institute (USC-ISI), Brown University, Johns Hopkins University, and Raytheon BBN. Performers were contracted to develop systems that quickly and accurately extract complex semantic information from raw text documents in multiple languages in a way that is adaptive to the information needs of a specific monolingual English user. Systems were expected to leverage this extracted information to enable automatic IR and efficient human triage of relevant documents from massive stores of text documents. A major goal of the program was to incentivize the rapid adaptation of the technologies to new languages and domains with minimal effort.

For each of the three phases of the program IE and IR from a specific surprise domain (e.g., corporate mergers and acquisitions, protests, government corruption, epidemiology, etc.) were levied on the Performers by IARPA's Test and Evaluation Team. Within each phase, Performer systems were evaluated on three IE tasks and one human-in-the-loop (HITL) IR task. Within each phase, the initial task focused on the extraction of Abstract events, while the second and third tasks focused on the extraction of events comprising increasingly granular semantic information.

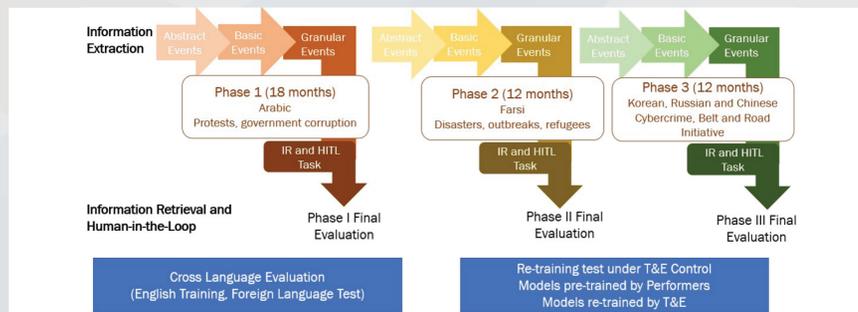


Figure 1. BETTER Program Structure

A major technical challenge posed by BETTER is that Performer systems must carry out 'zero-shot' learning: Performers are only provided training data in English, but their systems are evaluated on one or more surprise target languages, announced at the beginning of each program phase. This scheme ensures that Performers develop integrated, end-to-end IE/IR systems that are adaptive to analyst information needs and perform across differing domains and languages.

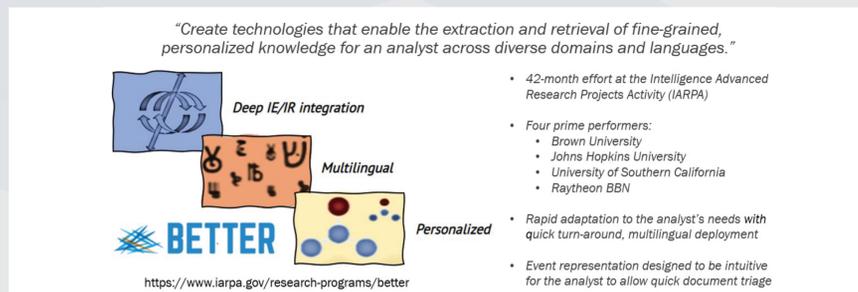


Figure 2. BETTER Program Overview

The BETTER Abstract Task

The Abstract event task is the first of a series of three IE tasks deployed in IARPA's BETTER program. True to its name, the Abstract IE task is designed to focus on high-level, domain-independent event properties, abstracting away from the construction-specific peculiarities of specific event types. The dataset was designed not only to evaluate zero-shot cross-language event and argument identification across multiple domains, but also to serve as a resource for transfer-learning: It is intended that, as a result of pretraining on the Abstract task, neural systems should require less data when fine-tuning to domain-specific IE tasks requiring extraction of finer-grained semantic information (such as the Basic and Granular IE tasks).

One hallmark of the Abstract event extraction task is that the varied elements, or arguments, that are usually distinguished in most domain-specific event representations, such as `_who_`, `_did_`, `_what_`, `_to_`, `_whom_`, `_when_`, and `_where_`, are reduced to just two: Agent and Patient.

The notions of Agent and Patient are purely semantic, and do not depend in any way on the particulars of how an event and its participants happen to be expressed syntactically within a sentence. Agents are those things that—whether sentient or not, animate or mechanical—have caused an event or in some way set it in motion. Patients, on the other hand, are those things that are most directly affected by the event.

Data Selection and Partitioning

The corpora we present for release here were used for the first IE evaluation (Abstract). Training data for this evaluation were provided in English only, and the evaluation was conducted in different languages: Arabic in Phase I, Persian in Phase II, and Korean for Phase III.

In Phase I, sentences for the Abstract Evaluation were randomly chosen from newswire texts in the target languages harvested from the Common Crawl. An Amazon Mechanical Turk pipeline was set up to confirm the suitability of the extracted sentences according to genre (news) and language, e.g. Modern Standard Arabic vs. dialectal variants. Training data for the Abstract event extraction task were presented at the sentence level on sentences that had been removed from their larger context and annotated in isolation. Each sentence included annotations for every Abstract event that occurred in the sentence. Some sentences contained no Abstract events while others contained more than one. Approximately 5,000 English sentences annotated at the Abstract event level were incrementally released to the Performers to enable them to train their multilingual extraction models.

Abstract data were partitioned by the BETTER T&E Team to optimize their use with regard to both training, system development, meaningful evaluation and error analysis.

Annotation

Annotators for the abstract task were chosen from a pool of bilingual candidates that demonstrated proficiency in the Abstract task. In Phase 1, all annotators passed a four-hour web-based qualifying exam authored by MITRE in the target language which they took after having received relevant written training in Modern Standard Arabic. In Phases 2 and 3 of the program, annotators were selected via in-person interviews to gauge their task suitability. Training for the last two phases was conducted in-person in a classroom environment. Annotators met on a weekly basis to review questions as a team, and maintained open communication as well via a slack channel hosted by the University of Maryland for any questions or concerns. Training was conducted in English across multi-language teams, to maximize information delivery and emphasize the semantic vs. syntactic nature of the annotation.

Each sentence was doubly annotated. Where discrepancies occurred in the annotation, a third annotator adjudicated to provide the official result. The adjudicators were often also the original trainers.

Annotation was performed in a tool created and hosted by MITRE. The tool is web-based, so much of the annotation was performed remotely.

Conclusion

The BETTER Abstract datasets represent a unique contribution to the resources available within the information extraction space. Unlike prior event and relation extraction tasks, which have tended to emphasize syntactic predicate-argument structure, the Abstract task adopts a semantically-based event annotation scheme that focuses on understood events and a reduced valence frame that comprises only the two most semantically salient arguments. We hope that this shift in perspective will substantially benefit research into cross-domain and multilingual event extraction.

For more information: Timothy.McKinnon@iarpa.gov

Data available at <https://www.IARPA.gov/research-programs/better>



Intelligence Advanced Research Projects Activity
I A R P A
Creating Advantage through Research and Technology

www.IARPA.gov

BETTER
BETTER EXTRACTION FROM TEXT
TOWARDS ENHANCED RETRIEVAL