



# BETTER

## BETTER EXTRACTION FROM TEXT TOWARDS ENHANCED RETRIEVAL

### INTELLIGENCE VALUE

The BETTER program aims to dramatically compress the information discovery cycle for Intelligence Community analysts by designing systems that extract personalized, mission-relevant semantic information from text and leverage this information to substantially improve search capabilities.

Massive volumes of multilingual, unstructured text present an overwhelming challenge for analysts to classify, triage, and examine. Tools and methods are needed to help analysts automatically map entities and events in text documents and use this information to perform rapid search and triage. The BETTER program is leveraging entity

and event identification, text search, and human-in-the-loop computation to create technologies that enable analysts to extract and retrieve highly-detailed, personalized knowledge across diverse domains and languages. BETTER's new tools and methods will also automatically tag incoming information, according to the analysts' pre-determined knowledge, and broaden the search criteria based on previously identified entities, events and relationships between entities, as well as human feedback. One objective is to use existing knowledge to compress the discovery cycle. Another objective is to minimize the amount of English training data needed to develop functional systems, with no training data provided in the foreign languages covered.

Information retrieval capabilities will be evaluated by the systems' ability to return a diverse set of relevant documents in an order that optimizes the user's experience, as well as identify the critical events and entities in them. Systems incorporate human feedback to drive system personalization. BETTER was launched in October 2019 and is a 42-month long program.

- Brown University
- University of Southern California Information Sciences Institute
- Johns Hopkins University

## TESTING AND EVALUATION PARTNERS

- MITRE
- National Institute of Standards and Technology

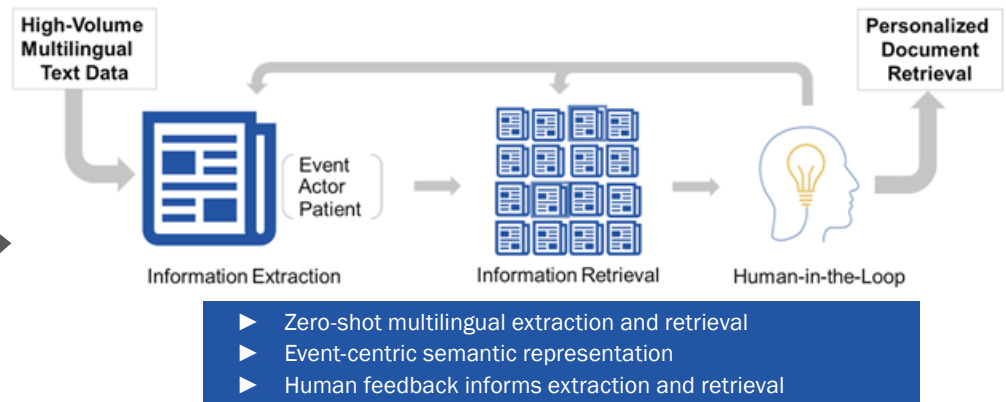
## KEYWORDS

- Information extraction
- Information retrieval
- Human language technology
- Zero shot learning
- Machine learning
- Natural language processing
- Multilingual document triage

## PRIME PERFORMERS

- Raytheon BBN Technologies

BETTER Systems extract and retrieve relevant text information for the needs of a specific user, adapting to new languages and domains with minimal effort.



### PROGRAM MANAGER

Carl Rubino, Ph.D.  
Phone: (301) 243-2081  
carl.rubino@iarpa.gov



[www.iarpa.gov](http://www.iarpa.gov)



@IARPAnews