



BETTER

BETTER EXTRACTION FROM TEXT TOWARDS ENHANCED RETRIEVAL

INTELLIGENCE VALUE

The BETTER program aims to dramatically compress the information discovery cycle for analysts in the intelligence community by developing systems that quickly extract personalized, mission-relevant semantic information from multilingual text.

Analysts face an overwhelming challenge when seeking to classify, triage, and examine the massive amounts of unstructured, multilingual data they receive. The intelligence community therefore needs tools that can rapidly sift through such data, identify entities and events of interest, and select materials that contain the most pertinent

information. Since so many analysts only speak English, these tools must be English-in, English-out.

The BETTER program leverages state-of-the-art deep learning technology to fill this need. When fed English queries and sample passages as input, BETTER systems can identify entities and events of interest, retrieve the most pertinent texts, translate those texts into English, and highlight passages of interest. Systems are designed for quick turnaround into new languages without annotated target-language training data, and a human-in-the-loop capability is provided so that analysts can tweak search criteria on the fly. This enables analysts to obtain highly detailed, personalized information across a wide range of domains and languages.

Information retrieval performance will be assessed by the systems' ability to return a set of documents that are not just relevant but also provide a diversity of pertinent information. They will also be scored on their ability to identify the critical events and entities contained therein.

BETTER is a 42-month program that launched in October 2019.

PRIME PERFORMERS

- Raytheon BBN Technologies
- Brown University
- University of Southern California Information Sciences Institute
- Johns Hopkins University

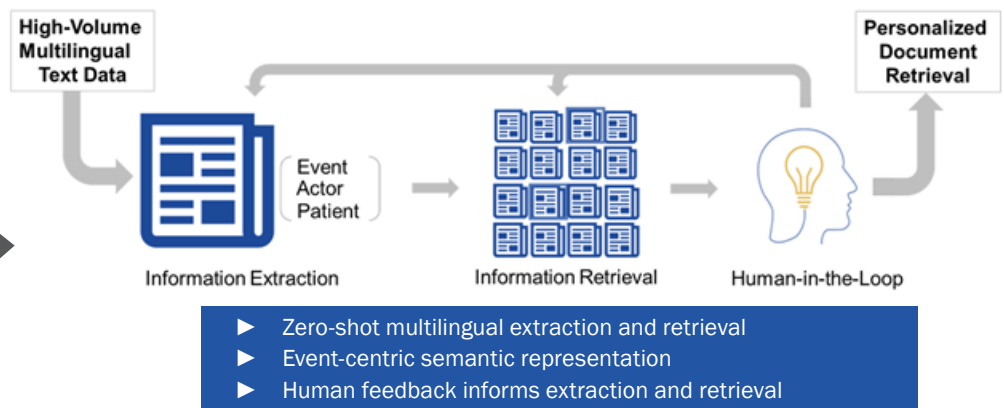
TESTING AND EVALUATION PARTNERS

- 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

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

Tim McKinnon

Phone: (301) 243-2084

timothy.mckinnon@iarpa.gov



www.iarpa.gov



@IARPAnews