



MATERIAL

MACHINE TRANSLATION FOR ENGLISH RETRIEVAL OF INFORMATION IN ANY LANGUAGE

INTELLIGENCE VALUE

The MATERIAL program aims to revolutionize the way the IC consumes foreign language information, by turning multilingual text and speech media into useable intelligence information for analysts, regardless of their language expertise.

A large portion of the ever-increasing amounts of text, audio, and video data produced in today's world is being generated by populations of emerging importance in lower-resource languages. This rich source of data is of little value if the information cannot be effectively searched. Launched in October 2017, The MATERIAL program is a 47-month venture seeking to address this challenge by building robust, automated language capabilities with limited linguistic resources, expertise, and tools.

END-TO-END MATERIAL SYSTEM



Analysts Query Foreign Language Speech and Text database using English

2 "air attack" in Pashto

- Speech and text documents
- All domains

Query in English

3 Query: *air attack*

In this regard, Paktika Police Chief Paswal Ali Must Momand told Killid that two days ago a rocket landed in Damgarkh area of Barmal district, Paktika province, during an *airstrike*, a senior Al-Qaeda network leader was said to have had close ties with Yemen's Al-Zawahiri.

airc airport, aerial, airstrike

as I found out, I said that two armed militants were killed in another similar *airstrike* in the mentioned area.

airc airport, aerial, airstrike

Defense Department has also announced that it has killed a senior Al-Qaeda network commander in an *airstrike* in Paktika province of Afghanistan.

airc airport, aerial, airstrike

This document contains information possibly relevant to the following domains:

	Military	<div style="width: 100%;"></div>	Why?
	Law & Order	<div style="width: 80%;"></div>	Why?
	Gov't & Politics	<div style="width: 60%;"></div>	Why?

Results Summarized in English

4

د پکتیکا امنیه قوماندان پاسوال علي مست مومند چې دوه ورځې وړاندې د پکتیکا ولايت د بړمېل و کې ددې پېلوته الوتکې دنيوډا هوايي برید ترڅ کې د مشر چې وېل کېږي له ايمن الظواهري سره نږدې

In this regard, Paktika Police Chief Paswal Ali Must Momand told Killid that two days ago a rocket landed in Damgarkh area of Barmal district, Paktika province, during an *airstrike*, a senior Al-Qaeda network leader was said to have had close ties with Yemen's Al-Zawahiri.

Aligned Bilingual Full Document Output

Query in English

- morphology and divergent spelling
- Text, audio, and video data crawlers operationalized for U.S. government use
- Annotated, reusable datasets in multiple languages for CLIR, ASR, and MT research
- Open-source neural MT framework Marian (co-founded)
- Bitext harvesting tool Paracrawl (funded)
- Novel cross-language query-biased summary generation technique

PRIME PERFORMERS

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

TESTING AND EVALUATION PARTNERS

- Massachusetts Institute of Technology Lincoln Laboratory
- National Institute of Standards and Technology
- University of Maryland Applied Research Laboratory for Intelligence and Security
- Tarragon Consulting Corporation



PROGRAM MANAGER

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



www.iarpa.gov



@IARPAnews