



(U) For Media Mining, the Future Is Now!

FROM: Joseph Picone and
Human Language Technology (S23)
Run Date: 08/01/2006

(TS//SI) In the first article on the Human Language Technology Program Management Office's (HLT PMO) activities and plans, we explained that we have five Strategic Thrusts. In this article, we will focus on the most active and fast-paced of the five: **Media Mining**. Its goal is to provide seamless access to information no matter what the information's source may be -- audio, image, or text. Right now over two hundred analysts have access to some Media Mining capabilities.

(S//SI) Near-Real-Time Alerts: RT-10

(S//SI) Integration of diverse information sources to produce near-real-time alerts is a major goal of a new Agency-wide program, **RT-10**. RT means REAL TIME, and 10 refers to reducing the time between collection and the generation of actionable intelligence an order of magnitude in each spin of the project.



(S//SI) The first deployment of RT-10 to the JIOC-I in Baghdad in 4th quarter 2006 will focus on integration of diverse information sources, including GSM voice intercept and geospatial coordinates, to reduce the time required to generate actionable intelligence.

(S//SI) New Voice-Services Platform: Voice RT

(S//SI) The HLT PMO is collaborating with RT-10 on the development of a new voice services platform, **Voice RT**. The first deployment of Voice RT, which is architecturally-based on an Army INSCOM* prototype known as ALICAT, will be operational in the Baghdad node of RT-10 in September 2006. This system is designed to index and tag 1 million cuts per day, and provide auxiliary HLT services such as language, dialect and speaker identification. The combination of these technologies with other RT-10 capabilities, such as geospatial coordinates, will provide a unique ability to generate actionable intelligence quickly and accurately.



(S//SI) Voice RT is a tool that allows analysts to perform keyword searching on voice content.

(S//SI) Voice Word-Search Capabilities

(TS//SI) The HLT PMO's Media Mining Thrust began as an effort to bring word-search capabilities (e.g., "Google for Voice") to Voice Language Analysts to make it easy for them to locate intercept rich in intelligence data. Voice word search technology allows analysts to find and prioritize intercept based on its intelligence content in much the same way as they now search text in PINWALE. For

SERIES: **(U) HLT**

1. [Human-Language Technology in Your Future](#)
2. For Media Mining, the Future Is Now!
3. [For Media Mining, the Future Is Now! \(conclusion\)](#)
4. ['Knowledge Discovery': Finding the Best Material](#)
5. [Human-Language Technology -- Everywhere](#)
6. [Dealing With a 'Tsunami' of Intercept](#)
7. [Building Human-Language Technology](#)
8. [Strangers in a Strange Land?](#)

example, in the Global War on Terrorism (GWOT), analysts can locate intercept dealing with explosive devices by searching for common terms such as " *operation* " or " *detonator* ," as well as more subtle terms about materials (" *hydrogen peroxide* "), place names (" *Baghdad* "), or people (" *Musharaf* ").

(S//SI) The first generation of this technology has been centered around Commercial-off-the-Shelf (COTS) software, **NEXminer** , developed by a startup company, Nexidia. The system is designed to support both **real-time searches** , in which incoming data is automatically searched by a designated set of dictionaries, and **retrospective searches** , in which analysts can repeatedly search over months of past traffic. The former capability allows the tool to function as a near real-time tipper. The latter capability allows analysts to rediscover important intelligence information and to refine their search strategies. This can be especially important in cases where pieces of a SIGINT "puzzle" become apparent and an analyst needs to go back to previous messages to see if other unnoticed pieces can be found.

(S//SI) This tool is very effective because it integrates high-performance speech processing technology with a most important agency resource, analyst knowledge of targets and missions. This technology was initially introduced to the analyst community in 2004 as a prototype, RHINEHART, which had been developed by SIGDEV Strategy and Governance (SSG).

(S//SI) RHINEHART now operates across a wide variety of missions and languages, and is used throughout the NSA/CSS Enterprise. One recent example of RHINEHART success occurred when Persian GWOT analysts searched for the words "negotiations" or "America" in their traffic, and RHINEHART located a very important call that was transcribed verbatim providing information on an important Iranian target's discussion of the formation of the new Iraqi government.

*Notes: (U) INSCOM = US Army Intelligence and Security Command

(U) Watch for the conclusion of this look at media mining, coming soon...

"(U//FOUO) SIDtoday articles may not be republished or reposted outside NSANet without the consent of S0121 ([DL sid comms](#))."