



(U//FOUO) Human-Language Technology -- Everywhere

FROM: [REDACTED] and Anita Kulman
Human Language Technology (S23)
Run Date: 08/22/2006

(U//FOUO) Looking into the future, the Human Language Technology Program Management Office's (HLT PMO) crystal ball has given us snapshots of the future for other Strategic Thrusts such as Media Mining and Knowledge Discovery. But the thrust to deliver **HLT throughout the Enterprise** needs a wide, landscape shot.

(U//FOUO) It is the broadest reaching of the five Strategic Thrusts because it will provide mature services for all of the nine HLT capabilities to users not only in SID at NSAW, but to any and all analysts who need help processing the material that floods their queues every day. This means that **more and better HLT will be deployed to analysts at the Cryptologic Centers and field sites in the US and throughout the world.**

(S//SI//REL) For some HLT services throughout the Enterprise, the future is now: many analysts are already using HLT and may not even be aware of it. At its best, HLT functions "upstream" as close as possible to collection sites, and definitely "behind the desktop." When this occurs, analysts receive queues of messages that have already been sorted and filtered by some HLT service that may have labeled the messages with information to alert the analysts to possible good intelligence or to tell them that the material is probably junk.

(S//SI//REL) When HLT is visible to analysts, they can see the services at their disposal at a touch of their fingertips by accessing a website or by hitting an option on a tool already on their desktops. There are speech, text and images services available, as follows:

- Speech Activity Detection ("*Is there speech in this cut?*");
- Language ID ("*Is this conversation in a language I'm interested in because I've found intelligence there before?*");
- Speaker Search ("*Is this the speaker who has provided good information in the past?*");
- Machine Translation in over 70 languages ("*Does this text have information that should be translated carefully by a language analyst?*"), and
- Optical Character Recognition (OCR) for most Roman, Cyrillic and Arabic scripts, which turns images of text in faxes and other electronic messages into searchable text.

(S//SI//REL) These tools are available right now, in the present, and the near future promises a lot more activity. For example, for OCR, new languages are being enabled monthly and training on how to use OCR tools will be forthcoming. The PMO has invested in an "OCR-on-demand" capability. In addition, keyword search of

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?](#)

OCR printed fax is available for Chinese and is being studied for most Roman, Cyrillic and Arabic scripts.

(S//SI//REL) As another example, currently available speech services, speech detection and language ID are being integrated into [BABBLEQUEST](#) and are being made more robust. This means that they will need less maintenance and will deliver good information consistently over time. Comparison and Speaker Watchlist services are being added to the Agency operational voice architecture as services available as on-demand pull-down menus from [HOTZONE](#).

(U) Watch out for these innovations!

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

DYNAMIC PAGE -- HIGHEST POSSIBLE CLASSIFICATION IS
TOP SECRET // SI / TK // REL TO USA AUS CAN GBR NZL
DERIVED FROM: NSA/CSSM 1-52, DATED 08 JAN 2007 DECLASSIFY ON: 20320108