Semantic Search/Research using PriArt: DoD IG Examples

Chuck Rehberg
CTO/Chief Scientist
Trigent Software/Semantic Insights
13-Jul-11
Research Using PriArt

• PriArt is a web-based application for conducting Semantic Search/Research

• By “Semantic Search” we mean:
  - “Rather than using ranking algorithms such as Google's PageRank to predict relevancy, Semantic Search uses semantics, or the science of meaning in language, to produce highly relevant search results.
  - [...] the goal is to deliver the information queried by a user rather than have a user sort through a list of loosely related keyword results.” - http://en.wikipedia.org/wiki/Semantic_search

• PriArt quickly reads through a potentially large corpus of documents and reports just the information you are looking for.
For example, let's examine a case of: “illegal export of M134”

• Research Statements:
  - “An arms dealer was convicted of illegally exporting M134. The M134 was mounted on an SUV.”

• Information Source to Search:
  - “2011 DOD IG Semiannual Report to Congress”
PriArt Results

- **Information Source:** “2011 DOD IG Semiannual Report to Congress”

- The reference found is in the caption of a photograph

---

**Investigation name:**

**M134**

is owned by chuck_r This investigation has 0 collaborators.

**Plain Text:**

An arms dealer was convicted of illegally exporting M134. The M134 was mounted on an SUV.

**Status**

Created: 10/7/2011 17:17:2
Completed: 

Reporting...

**Most relevant documents pertaining to the investigation content:**

  - An arms dealer was convicted of illegally exporting M134.
    1. A arms dealer was convicted of illegally exporting M134 mounted SUVs.
  - The M134 was mounted on an SUV.

These are the prior art findings and references for each step in the M134 investigation.

- An arms dealer was convicted of illegally exporting M134
  - A arms dealer was convicted of illegally exporting M134 mounted SUVs.
  - The M134 was mounted on an SUV.

displaying infringements
Additional Results

Information Source: a set of government websites (68 sites; 1498 documents) listed here:
http://www.semanticinsights.com/Sources/DOD_IG_Sources.htm

Items found are listed in order of the most number of concepts in common (i.e. best “semantic overlap”)

---

**An arms dealer was convicted of illegally exporting M134.**

A arms dealer was convicted of illegally exporting M134 mounted SUVs. [1]

Meng was the first defendant in the country to be convicted of exporting military source code pursuant to the Arms Export Control Act. [2]

Cervantes-Sanchez was convicted of illegally exporting munitions while Soto was convicted of lying to buy two AK-47s for Cervantes-Sanchez. [3]

Two Chinese Nationals Convicted of Illegally Exporting Electronics Used in Military Radar Overview. [4]

The M134 was mounted on an SUV.

A arms dealer was convicted of illegally exporting M134 mounted SUVs. [1]
Expand the search to include Synonyms

• “G.E.'s minigun is in use in several branches of the US military, under a number of designations. The basic fixed armament version was given the designation M134 by the U.S. Army, while exactly the same weapon was designated GAU-2/A by the U.S. Air Force.”

• So both M134 and GAU-2/A could be considered synonyms
Updated Dictionary Entry for M134

This basic fixed armament minigun version is made by G.E. The designation M134 was given by the U.S. Army, while exactly the same weapon was designated GAU-2/A by the U.S. Air Force.

Synonyms:
GAU-2/A : PRPN
Result of expanded search with Synonyms

- Essentially allows finding “GAU-2/A” in addition to “M134”
- No new results found in
  - http://www.semanticinsights.com/Sources/DOD_IG_Sources.htm
Expand the search to include “miniguns”

- M134 is a kind of “minigun”
- Therefore, any reference to illegally exporting “miniguns” could also refer to the M134.
Edit the Ontology to introduce “minigun”
Result of expanded search

- Essentially allows finding “minigun” in addition to “M134” and “GAU-2/A”
- No new results found in
  - http://www.semanticinsights.com/Sources/DOD_IG_Sources.htm
Expand the search to include all “miniguns”

• We can change the research statement to include all the miniguns we care about:
  - “An arms dealer was convicted of illegally exporting miniguns. The minigun was mounted on an SUV.”

• We simply need to change the Ontology to identify the miniguns of interest.
Edit the Ontology to add kinds of “minigun”
Result of expanded search for multiple miniguns

- Information Source: “2011 DOD IG Semiannual Report to Congress”

- It appears the only reference of illegal export in the information source to the miniguns of interest was M134.

An arms dealer was convicted of illegally exporting miniguns. The minigun was mounted on an SUV.

Most relevant documents pertaining to the investigation content:

- 2011 DOD IG Semiannual Report to Congress - SemanticInsights.info
  - An arms dealer was convicted of illegally exporting miniguns.
    - A arms dealer was convicted of illegally exporting M134 mounted SUVs.

These are the prior art findings and references for each step in the miniguns investigation.

- An arms dealer was convicted of illegally exporting miniguns.
- A arms dealer was convicted of illegally exporting M134 mounted SUVs.
Result of expanded search

- Essentially allows finding “minigun”, “GAU-17/A” and “XM196” in addition to “M134” and synonym “GAU-2/A”

- No new results found in
  - http://www.semanticinsights.com/Sources/DOD_IG_Sources.htm
Another PriArt example

• Finding a needle in a haystack:
  - Select a statement at random and use PriArt to find it again in a large corpus of documents

• First identify the statement (“needle”) we are looking for:
  - From somewhere in the set of government websites (68 sites; 1498+ documents) listed here: http://www.semanticinsights.com/Sources/DOD_IG_Sources.htm, I selected the following “research statement” from some pdf document:
    • “The rapid growth of the DoD budget since FY 2000 leaves the Department increasingly more vulnerable to fraud, waste, and abuse that undermines the Department's mission.”

• Use PriArt to find this statement and the most relevant information
Report Generated: The “needle” was found...

- Based on “reading” 1498 documents, a nine page report of the semantically relevant information was produced showing an ordered list the most relevant documents and which statements were relevant.
- The target statement (the “needle”) was found.
- A Bibliography was produced (see next slide).
Generated Bibliography for the Haystack test:

- A bibliography was generated (with hyperlinks) listing all the documents semantically relevant to the original research statement (the “needle”).
- Statements in the generated report are cross referenced to the Bibliography.
Who we are:

- Semantic Insights is the R&D division of Trigent Software, Inc. [www.trigent.com](http://www.trigent.com)
- We focus on developing semantics-based information products that produce high-value results serving the needs of general users requiring little or no training.
- Visit us at [www.semanticinsights.com](http://www.semanticinsights.com)
Chuck Rehberg

- As CTO at Trigent Software and Chief Scientist at Semantic Insights, Chuck Rehberg has developed patented high performance rules engine technology and advanced natural language processing technologies that empower a new generation of semantic research solutions.

- Chuck has more than twenty five years in the high-tech industry, developing leading-edge solutions in the areas of Artificial Intelligence, Semantic Technologies, analysis and large-scale configuration software.