

Semantic Search/Research using PriArt: *“Finding a needle in a haystack”*

Chuck Rehberg
CTO/Chief Scientist
Trigent Software/Semantic Insights

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.

Here is another PriArt example

- “*Finding a needle in a haystack*”
 - The Plan: Select a statement at random from a large corpus of documents and use PriArt to find it again
- First identify the statement (the “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_Source_s.htm, I randomly selected the following “research statement” from a randomly selected 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.”
- Next, we use PriArt to find this statement and the most relevant information

Login to PriArt



Welcome to PriArt, a powerful new way to do research.

Tired of skimming through hundreds of search results only to find a few useful facts? We're building easy-to-use research products for businesses, researchers, students, - anyone who needs in-depth info on a topic. Our tools uncover valuable information, "reading" thousands of documents - web pages... or a collection of PDF files... or a database... or your desktop...really, just about anything. The stuff you need to know is handed back to you in concise, plain-english reports - faster, better, and cheaper than an expert using a key-word search engine.

Our core tools are built on more than six years of R&D in Natural Language Processing and Semantic Technology, resulting in seven patents, granted and pending, and the worlds fastest, most scaleable Rules-Based Engine. Our tools recognize synonyms in context, identify multiple ways of saying the same thing, and consult detailed domain-specific knowledge including generalizations, specializations, relationships and instances in a given domain. And we're just getting started!

Take a quick, [guided tour](#) to see how easy it is to use.

Why should I become a registered user?

Membership allows you to save your patent research, so you can continue your research. As your line of inquiry changes, your searches can continue to grow, and as new information becomes available you can generate fresh reports.

[Join now!](#)

Registered users

username

password

remember me

[forgot password?](#)

Create the Investigation (“Needle in a Haystack”)



Home

You are using 4 of 4 investigations. [Upgrade account](#)


[Create new investigation](#)

Needle in a Haystack chuck_r owns this investigation	(0) reports completed	(1) reports in progress	Delete
Organization chuck_r owns this investigation	(0) reports completed	(0) reports in progress	Delete
Harpo chuck_r owns this investigation	(1) reports completed	(1) reports in progress	Delete
Autism chuck_r owns this investigation	(3) reports completed	(0) reports in progress	Delete

4 items found, displaying all items. 1



Enter the Research Statement (i.e. cut/past the found text)

**PRIART**
POWERED BY SEMANTIC INSIGHTS

Welcome back, [chuck_r](#)
[home](#) [profile](#) [tour](#) [help](#) [logout](#)

Investigation name

Needle in a Haystack

is owned by [chuck_r](#) This investigation has [0](#) collaborators.

Claim Description: Needle in a Haystack
Statutory Class: Method
Is a dependent of: None
Selected Domains : None is Selected
[edit details](#)

PlainText ClaimText Reports

Describe your invention

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 this tab to enter a plain english description or copy and paste a description from your favorite text editor, or any other source.

Unlike the 'Claim text' tab, this description doesn't need to use the formal language that you would use in submitting a claim.

Click the 'Validate' button to submit the text that you've entered.


After you hit the 'Validate' button, two things will happen. First, this screen will refresh, with recommendations in this right sidebar.

Second, any text you've entered or edited in this tab will automatically update the way your claim is represented in the 'Claim text' and 'Step-by-step' tabs.

[Validate](#) [Save](#) [Simplify](#) [AutoQuery](#)

Copyright (c) 2009-2011, [Semantic Insights, Inc.](#) All rights reserved. [Privacy](#) [Legalese](#) [Contact](#)

Submit the Investigation Request

**PRIART**
POWERED BY SEMANTIC INSIGHTS

Welcome back, chuck_r
[home](#) [profile](#) [tour](#) [help](#) [logout](#)

Investigation name:

Needle in a Haystack

is owned by [chuck_r](#). This investigation has '[n](#)' collaborators.

Claim Description : Needle in a Haystack
Statutory Class : Method
Is a dependent of : Needle in a Haystack
Selected Domains : None is Selected

PlainText ClaimText **Reports**

Report Type	Read	Report Status	Limit
<input type="text" value="PRIORART"/>	<input type="text" value="URL"/>	<input type="text" value="Completed..."/> View...	<input type="text" value="10"/> ▼ ReGenerate Delete
Create a new report		<input type="checkbox"/> Alert me by email when my reports are done	

Copyright (c) 2009-2011, [Semantic Insights, Inc.](#) All rights reserved. [Privacy](#) [Legalese](#) [Contact](#)

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).

1310554262581.pdf - Adobe Reader

File Edit View Window Help

1 / 10 75%

Comment Share

SEMANTIC INSIGHTS

Report generated by Semantic Insights on 13/7/2011 6:51

Report Template Used: PriorArt_Basic
SIRA Server Used: www.semanticinsights.com
WritingStyle Used: Verbatim

[The DoD IG](#)

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.

This level of growth will allow the DoD IG to increase audit and investigative capabilities to better keep pace with the growth of the DoD budget and number of contracts and the increased demand for IG oversight related to the. [1]

[Microsoft Word - 02_Addendum_B_FINAL 4.1.doc](#)

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.

To enhance accountability and transparency throughout the MoI there should be an increase in the numbers and capability of DoD IG and internal affairs staff and internal affairs investigators should have a presence in the provincial and district police headquarters. [2]

[OFFICE OF THE DEPUTY INSPECTOR GENERAL FOR AUDITING](#)

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.

The DoD IG subsequently was provided additional funding in the amount of \$24.0 million in the FY 2008 Department of Defense Appropriations Act (P.L. [3]

This level of growth 22% will adjust personnel authorizations from the current level of 1687 to 2061 and allow the DoD IG to augment oversight capabilities in order to better keep pace with the DoD budget and the number of. [3]

Growth of the DoD Budget The rapid growth of the DoD budget since FY 2000 leaves the Department increasingly more vulnerable to fraud waste and abuse that undermines the Departments mission. [3]

Growth of the DoD IG The requested increase through FY 2016 will enable us to continue to support congressional and DoD mission requirements while ensuring the integrity of DoD contracting particularly services contracting. [3]

Personnel and information technology resources of the Office of the Deputy Inspector General for Auditing

8.50 x 11.00 in

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.

Bibliography

- [1] The DoD IG. Retrieved on 07/13/2011 05:52:20, from <http://www.dodig.mil/ig.html>
- [2] Microsoft Word - 02_Addendum_B_FINAL_11_13.doc. Retrieved on 07/13/2011 05:54:05, from http://www.dodig.mil/PUBS/strategic_plan/2009_DoDIG_Summary_of_Management_Challenges.pdf
- [3] OFFICE OF THE DEPUTY INSPECTOR GENERAL FOR AUDITING. Retrieved on 07/13/2011 05:55:00, from [http://www.dodig.mil/IGInformation/IGInformationReleases/DoDIGRequirementsPlanforFY11-12\(21Aug2010\).pdf](http://www.dodig.mil/IGInformation/IGInformationReleases/DoDIGRequirementsPlanforFY11-12(21Aug2010).pdf)
- [4] Ongoing Audits. Retrieved on 07/13/2011 06:22:42, from http://www.dodig.mil/gwot_iraq/ongoing_audit.htm
- [5] Microsoft Word - FINAL_ArmoredVehicles6-27.doc. Retrieved on 07/13/2011 06:23:58, from http://www.dodig.mil/IGInformation/IGInformationReleases/FINAL_ArmoredVehicles.pdf
- [6] Department of Defense Office of Inspector General. Retrieved on 07/13/2011 06:24:18, from <http://www.dodig.mil/Audit/reports/fy09/09-032.pdf>
- [7] 1310551733644.pdf. Retrieved on 07/13/2011 06:24:21, from <http://www.dodig.mil/audit/reports/fy11/11-020.pdf>
- [8] Section I - Background. Retrieved on 07/13/2011 06:24:28, from http://www.dodig.mil/gwot_iraq/PakQTRRep033110final.pdf
- [9] DoD IG Fiscal Year 2011 Audit Plan. Retrieved on 07/13/2011 06:24:52, from http://www.dodig.mil/Audit/audit_plan.pdf
- [10] 1310551916725.pdf. Retrieved on 07/13/2011 06:25:08, from http://www.dodig.mil/sar/SAR_MAR_2011.pdf
- [11] Slide 1. Retrieved on 07/13/2011 06:25:36, from http://comptroller.defense.gov/cfs/fy2010/01_DoD_Agencywide/Fiscal_Year_2010_DoD_Agencywide_Agency%20Financial%20Report.pdf
- [12] IG Monthly Newsletter - April 2011. Retrieved on 07/13/2011 06:27:08, from <http://www.dodig.mil/eletter/2011/April2011.html>
- [13] DoD IG Growth Plan-BACKGROUND. Retrieved on 07/13/2011 06:27:38, from http://www.dodig.mil/IGInformation/dodGrowthPlan_08-15/background.html
- [14] DoD IG Fiscal Year 2011 Audit Plan. Retrieved on 07/13/2011 06:27:45, from http://www.dodig.mil/audit/audit_plan.pdf
- [15] Microsoft Word - AS5.B - D2009-D000CF-0069.000 - 10-27-2008.DOC. Retrieved on 07/13/2011 06:28:00, from <http://www.dodig.mil/audit/reports/fy11/11-021.pdf>
- [16] 1310552031948.pdf. Retrieved on 07/13/2011 06:28:10, from http://www.dodig.mil/sar/SAR_Insert_MAR_11.pdf
- [17] Interagency Coordination Group of Inspectors General for Guam Realignment A.... Retrieved on 07/13/2011 06:29:02, from http://www.dodig.mil/Audit/guam/Guam_Realignment_Annual_Report_Feb_2011.pdf
- [18] 1310551706888.pdf. Retrieved on 07/13/2011 06:29:16, from http://www.dodig.mil/sar/SAR_OCT_2010.pdf
- [19] Interagency Coordination Group of Inspectors General for Guam Realignment A.... Retrieved on 07/13/2011 06:30:07, from http://www.dodig.mil/Audit/guam/Guam_Realignment_Annual_Report_Feb_2011_compliant.pdf

PriArt gets better

- You can improve PriArt simply by adding:
 1. Synonyms to the dictionary
 2. Concepts and relationships to the Ontology
- This will be demonstrated in a different presentation



- Who we are:
 - Semantic Insights is the R&D division of Trigent Software, Inc. 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

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.