Semantic Insights Research Assistant

LandWarNet TBI [Slides](#) and [Video](#)

- **Title Slide**
- **Contents**
- **Background**
  - The Semantic Insights Research Assistant (SIRA)
  - Our Mission...with a little more Precision
  - The Semantic Advantage
- **Application: Tools and Examples**
- **Introducing the PriArt**
- "Quick Start" User’s Guide to PriArt
- What sources can SIRA read today?
- "Reading" the Internet...
- **Creating Investigations...**
- **Investigation: Autism**
- Autism: Report [source=Internet]
- Investigation: Aswan Dam
- Investigation: AdnanShukrijumah
- AdnanShukrijumah: Report Results
- **Patent Infringement and Prior Art**
- Fast Rule Selection Engine: Infringement Report [based current USPTO]
- Searching the KM Conference Track list
- Information gathering inside and outside NITRD
- Asking Questions and the “unknown thing”
- Question: “What is required for Cybersecurity?”
- **PubMedResearch**
- Auto Query Generation
- Uses live Dictionary
- Report Type=PriorArt; Information Source=PubMed
- On-line Report

http://semanticommunity.info/Binary_at_LandWarNet_2011/Semantic_Insights_Research_Assistant

Updated: Wed, 23 Sep 2015 02:15:44 GMT

Powered by [MindTouch](#)
From the PDF Report
PubMed on CYP2D6 Inhibitors
Some things to know before using SIRA
Who we are
Chuck Rehberg

Title Slide

Introducing SIRA: Semantic Insights Research Assistant

Presentation to NITRD
August 10, 2011
Chuck Rehberg
Semantic Insights/Trigent Software
http://semanticommunity.info

Contents

1. Background
   Who we are; Our mission; Mission status
2. PriArt: A Semantic Research Assistant (Web)
   Simple User Interface; Examples
3. PubMed Example
4. Knowledge Management applications
   SIRA Development Center (Desktop)
   Language Lab (Web)
5. Live demonstration

http://semanticommunity.info/Binary_at_LandWarNet_2011/Semantic_Insights_Research_Assistant
Updated: Wed, 23 Sep 2015 02:15:44 GMT
Powered by mindtouch™
Background

Who are we?

“Semantic Insights” is the R&D division of Trigent Software, Inc. [www.semanticinsights.com]

What is our Mission?

Automate research tasks (faster, better, cheaper)

Why Semantics?

Semantics allows us to operate at the “meaning level”, to “separate the know from the show”

Why us?

Bright People, Proprietary Technology (IP), Passion for Excellence and Track Record of Delivery

* [https://library.kuleuven.be/bitstream/123456789/123456789/1/TFI_a_09_Ventilkansen.pdf](https://library.kuleuven.be/bitstream/123456789/123456789/1/TFI_a_09_Ventilkansen.pdf)

The Semantic Insights Research Assistant (SIRA)

Our Mission:

- The SIRA technology was developed to automate research tasks requiring natural language, domain knowledge, understanding and reasoning.
- SIRA-based products must be easy-to-use requiring little or no training beyond what the user already understands.

Mission Status:

- We have developed PriArt, an embodiment of the SIRA technology that automates the understanding and reading of natural language text (initially English), gathers specific information of interest and produces a variety of useful reports.
- Other SIRA embodiments have been conceived and prototyped.

Today SIRA can:

1. **Semantically understand** a statement of your interest expressed in Natural Language
2. **Read through** a vast number of documents
3. **Identify** semantic relevant information of interest in Natural Language text
4. **Report** the findings in useful ways including Natural Language text
Our Mission…with a little more Precision

The SIRA technology was developed to automate research tasks requiring natural language, domain-knowledge, understanding and reasoning.

- By “research tasks” we mean, “Information gathering tasks that currently require humans with domain-knowledge, Natural Language skills or otherwise would take considerable time”
- By “domain-knowledge” we mean, “background information that is necessary to understand a given natural language text”
- By “understanding” we mean, “the ability to create a meaning map that relates experiences (e.g. natural language text) to semantic items in an Ontology”
- By “reasoning” we mean, “the ability to further process the meaning map to generate derived experiences”

The Semantic Advantage

- The Natural Language Challenge
  - In natural language there are many ways to say the same thing.
  - When reading documents you need to be able to recognize when semantically equivalent text
  - The text you read may have little (or nothing) in common with the words or structure of the original research statement.

- At SIRA core is an Ontology
  - SIRA’s Ontology acts like a dictionary of “semantic items” (concepts, relationships, instances, generalizations, specializations, etc.).
  - However, the Ontology goes beyond the expressiveness of dictionaries by explicitly stating how the semantic items are related.

- The Semantic Advantage
  - One power of the Ontology comes from providing a single logical expression for the many ways of saying the same thing.
Application: Tools and Examples

- PriArt: A Semantic Research Assistant (Web)
  - Examples presented: Autism report, high school assignment, homeland security, patent infringement

- SIRA Development Center® (Desktop)
  - Used to develop and manage World Knowledge
    - Ontologies, Dictionaries, Testing and Training
    - Live Demonstration of Ontology and Dictionary creation and curation

- Language Lab® (Web)
  - Used to define Language and Genre
    - Syntax, Grammar and Meaning Maps
  *Note: Live demonstration only

Introducing the PriArt

Welcome to PriArt,

a powerful new way to do research.

Try it: drawing through hundreds of search results only to find a few useful facts?

Here's building your own research product for businesses, researchers, students,
anyone who needs to dig deep into a topic. Our tools uncover valuable information,
"reading" thousands of documents - web pages, or a collection of PDF files, or a database - to
your disposal, ready, just about anything. The staff you need to know it bubble back to you in concrete, plain-English reports - faster, better, and cheaper than ever before using a new search engine.

Our core tech is built on decades of R&D in Natural Language Processing and Semantic Technology, resulting in several patents, granted and pending, and the world's fastest, most scalable FOLK-REX Engine. Our text recognition engines in context, identify multiple ways of saying the same thing, and consult dedicated domain-specific knowledge including generalizations, specializations, relationships and businesses for given domains and verbs just getting started!

Why should become a PriArt user?

Why should it be set up like a PriArt? Because when you see your
trusted search, in a packed schedule, or a stocked shelf, it's
quickly, you should know it and do it. Quick and easy access because
customers can take a firm lead in getting results even in minutes.

Registered users

password

* I agree to the

http://semanticommunity.info/Binary_at_LandWarNet_2011/Semantic_Insights_Research_Assistant

Updated: Wed, 23 Sep 2015 02:15:44 GMT

Powered by mindtouch™
“Quick Start” User’s Guide to PriArt

“Quick Start” User’s Guide to PriArt

- PriArt is an online Research tool that creates:
  - Gather’s information based on your research statements
  - Generates research reports with bibliography and hyperlinks containing information relevant to your research
- Here’s the quick start guide:
  1. Login
  2. Create an Investigation
     a) Enter name of your investigation
     b) Enter a description of your investigation
     c) [optionally] Select the domains to be considered
  3. Generate a Report
     a) Select the kind of report you want
     b) Select what you want PriArt to read (i.e. the source location)
     c) Submit the research
  4. Review the results

What sources can SIRA read today?

What sources can SIRA read today?

- Current document formats include:
  - HTML, PDF, DOC, RTF, TXT
- Source Location
  - The Web
    - Specific Websites to any depth
    - Results of search engines (generated manually by you or automatically by SIRA)
  - Documents on your computer
    - Specific Directories or via Google Desktop
  - Email
    - Can drive Google Desktop
“Reading” the Internet…

“Reading” the Internet…

- It isn’t necessary (or practical) to read the whole internet.
- You can direct PriArt to read specific web pages and links.
- You can also direct PriArt to use keyword search engines.
  - However, if you want PriArt to effectively read a large document corpus (like the internet), you can direct PriArt to first gather documents using a keyword search engine.
- Automated Keyword Search Queries
  - PriArt will generate and execute a search strategy for you. This includes executing a number of keyword queries, each designed to find the most relevant documents based on examining your statement of investigation.
- Balancing your time with desired accuracy
  - You still may get millions of documents to read. So, you can direct PriArt to read a specific number of most relevant documents found according to the search engine ranking.

Creating Investigations…

Creating Investigations…

![PriArt Graphical User Interface](http://semanticommunity.info/Binary_at_LandWarNet_2011/Semantic_Insights_Research_Assistant)
Investigation: Autism

Investigation: Autism

- Investigate the first paragraph of the Autism Society of America "about" page.
Investigation: Aswan Dam

- My 16 year old is a Junior in High School.
- She wrote a report on Aswan Dam and wished to generate a bibliography for the report.

Investigation: Adnan Shukrijumah

- Any recent sightings of the terrorist Adnan Shukrijumah?
AdnanShukrijumah: Report Results

Patent Infringement and Prior Art
Fast Rule Selection Engine: Infringement Report [based current USPTO]

Searching the KM Conference Track list
Information gathering inside and outside NITRD

Asking Questions and the “unknown thing”

- “Research Statements” can contain questions (who, what, which, when, where, why and how)
  - “What is required for Cybersecurity?”
- You can also turn assertions into questions using “the unknown thing” (#?#)
  - “The LCS-2 can travel at #?# knots.”
- Or “the unknown action” (also #?#)
  - “A successful #?# of a two-stage rocket launch.”
Question: “What is required for Cybersecurity?”

Corpus + Internet
Accreditation requirements
I better get cracking.

PubMedResearch

What is the relationship of CYP2D6 to breast cancer in tamoxifen?
How is CYP2D6 related to tamoxifen?
How is CYP2D6 related to breast cancer?
Auto Query Generation

Uses live Dictionary
the relationship of CYP2D6 to breast cancer and tamoxifen

1. Clinical epidemiology and pharmacology of CYP2D6 inhibition related to breast cancer outcomes. [1]
2. Clinical pharmacology and pharmacokinetics of CYP2D6: Report the PRIP. [2]
3. Tamoxifen dosage for female breast cancer patients. [3]
5. The role of tamoxifen in breast cancer prevention. [5]

1. Although there is not a clear cut directive on the subject, it is now widely recommended to treat women with aromatase inhibitors with the lower doses, while women with CYP2D6 inhibition might reduce tamoxifen via prevention of [6]
From the PDF Report

PubMed on CYP2D6 Inhibitors

CYP2D6 inhib
is needed to sta the investiga

CYP2D6 inhibitors are N:
Arylated/Aralkylated CYPs.

- Paroxetine, fluoxetine, and imipramine are strong CYP2D6 inhibitors which should be avoided in tamoxifen users. [1]
- The polymorphic CYP2D6 is the key enzyme in this biomutation and recent mechanistic pharmacological and clinical evidence suggests that genetic CYP2D6 and drug interaction by CYP2D6 inhibition influence the plasma concentrations of active tamoxifen metabolites and the outcomes of tamoxifen treated. [2]
- According to CYP2D4 (cytochrome P450 family 2 subfamily D polypeptide 4) genotyping before treatment to predict metabolizer status may open new avenues for individualized endocrine treatment with the maximum benefit being expected for extensive metabolizers. [3]
- However, strong CYP2D6 inhibitors such as the selective serotonin-norepinephrine inhibitors paroxetine and fluoxetine which are used to treat hot flashes should be avoided because they severely impair formation of the active metabolites. [3]
- CYP2D6 inhibitors such as paroxetine are associated with changes in alternative pharmacokinetics similar to those observed among poor CYP2D6 metabolizers. [3]
- A new study (the impact of polymorphisms on CYP2D6) was elucidated using model compounds. [3]
Some things to know before using SIRA

- Some things to know before using SIRA
  - Research Statements
    - Short clear sentences work so much better (e.g. “The student of the professor on the balcony is happy.” Is the student or professor “on the balcony”?)
  - Information Sources
    - Web content is made for humans to read/look-at (“eye candy”, advertisements, links to other content, etc.). This poses a challenge for automated readers to identify topical information and context boundaries.
    - For selected sites (like Wikipedia and the US Patent Office), we have readers with individual “Reading Plans” that focus on the information of interest.
  - An Automated Research Assistant
    - Automated domain-aware reading and writing is not enough for a satisfactory Research Assistant. A Research Assistant often needs to translate (expand) a “research request” into semantically equivalent and findable reading objectives (e.g. What do you mean by “similar” and what constitutes evidence that something is “similar”?). We have additional tools (not covered here) that address this need.

Who we are

- 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.