

An Introduction to Semantic Insights™ Approach to Natural Language Understanding

22-SEP-2014

Chuck Rehberg, CTO

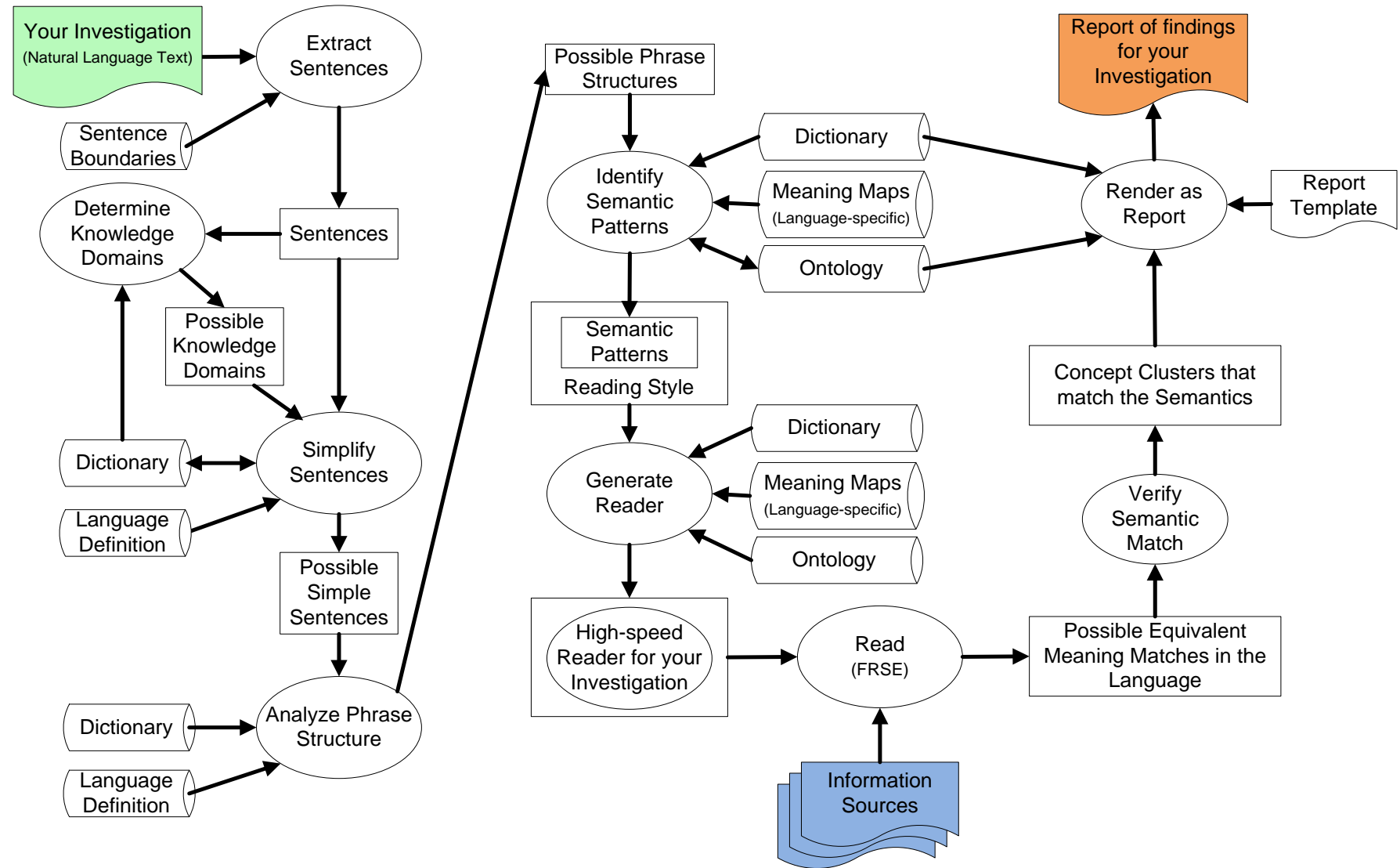
Semantic Insights™ a Division of Trigent Software, Inc.

Covered in this brief Presentation

- Presented here is a very simplified overview of the Natural Language Understanding process used in the Semantic Insights Research Assistant (SIRA) technology.
- It is intended to help potential collaborators understand where their work may fit in.
- Feel free to visit our website www.semanticinsights.com to seek further information or schedule a demo.

Simplified Overview of the NLU Processes

1. Extract Sentences
2. Determine Knowledge Domains
3. Simplify Sentences
4. Analyze Phrase Structure
5. Identify Semantic Patterns
6. Generate Reader
7. Read
8. Verify Semantic Match
9. Report Findings



Extract Sentences

- All information sources (pdf, html, txt, docx, rtf, pptx,...) containing Natural Language text are reduced to a single input stream of text.
- This stream of text is partitioned into “sentence” boundaries.

Determine Knowledge Domains

- The dictionary has terms.
- The terms have senses.
- The senses may be domain-specific or not.
- If the Investigation contains a significant number of possible senses with the same domain, those senses are will be added for consideration.
- Down-stream NLU processing is informed of this by identifying the possible domains in this Investigation.

Simplify Sentences

- The dictionary is consulted for possible part-of-speech, word class, number, gender, verb form...
- All possible simplifications of each sentence is generated.
- Duplicates are removed.
- Noun Phrases, and other clauses/phrases may be expanded:
 - For Example, “The fat man went to California.” becomes “The man is fat.” and the “The man went to California”
 - This allows the facts to be somewhat independently found in the information sources
- May grow the dictionary

Analyze Phrase Structure

- All possible “valid” phrase structures are determined.
- These structures are used to find the correct alternative sentence structures that render the same meaning.
- May grow the Ontology.
- To reduce the possible combinations, Word Sense Disambiguation may be applied here.

Identify Semantic Patterns

- Find the correct alternative sentence structures that render the same meaning.
- May grow the Ontology.

Generate Reader

- Package all the sentence variations and all the synonyms pattern matching rules.
- These are used to do “speed reading”

Read

- Apply the FRSE engine to quickly process a large amount of pattern rules.
- Return the matches for further analysis

Verify Semantic Match

- Do the sentences found match the desired semantics?

Report Findings

- Organize the results as specified in the Report Template.
- Apply the specified natural language generation/rendering algorithm.
 - Mostly now this is “verbatim”
 - You can also select “simplified English”
 - Others based on “writing style” have been contemplated

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 understanding technologies that empower a new generation of semantic research solutions.

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



SEMANTIC INSIGHTS™

A Division of Trigent Software, Inc.