

# Live Categorization using SIRA

Chuck Rehberg, CTO Semantic Insights

[chuck\\_r@semanticinsights.com](mailto:chuck_r@semanticinsights.com)

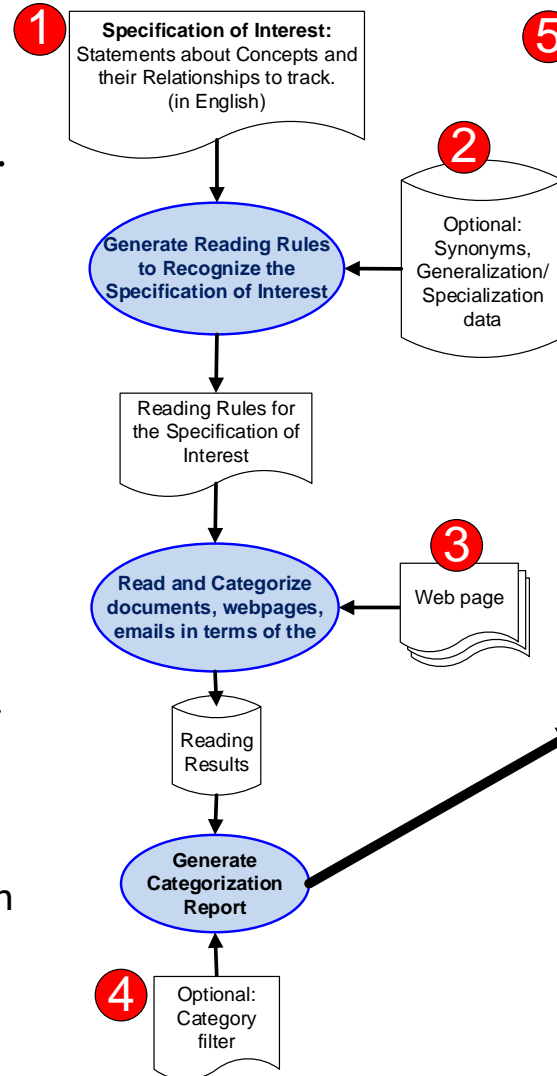
# The SIRA Approach

- SIRA identifies the variety of language patterns and terms in a sentence, and maps them to the concepts and relationships they represent.
- Starting with a statement of investigation (in natural language), SIRA generates a high-speed “reader program” that can recognize all the ways and terms the investigation’s concepts and relationships could be expressed.
- Then, given a corpus of natural language text, SIRA applies this “reader program” to “read” every sentence, looking for the variety of language patterns and terms that can represent the concepts and relationships expressed in the investigation. All matching information is returned for further processing.
- SIRA verifies the returned information (linguistically and semantically) for relevance to the investigation.
- The results can then be rendered in a variety of useful ways.



# Categorization Example

1. “Specification of Interest” is a list of English statements of concepts (can be real or abstract) and how they relate.
2. Providing the synonyms, generalizations and specializations for concepts and relationships can improve completeness of coverage. Common English terms are provided by default.
3. The document source can be identified or individual documents can be supplied.
4. You can filter the report to a subset of all concept data. [Not implemented in current Demo]
5. The generated Categorization Report Shows the % of lines with the given concept/ relationship.



Document	Subcategory	General Category	In % Lines
http://psychcentral.com/lib/an-introduction-to-autism/	autism	Neurodevelopmental Disorder	5.77
http://psychcentral.com/lib/an-introduction-to-autism/	Spectrum	Spectrum	1.92
http://psychcentral.com/lib/an-introduction-to-autism/	diagnosis	diagnosis	7.69
http://psychcentral.com/lib/an-introduction-to-autism/	disorder	disorder	1.92
http://psychcentral.com/lib/an-introduction-to-autism/	spectrum	spectrum	1.92
http://www.webmd.com/brain/autism/autism-spectrum-disorders	diagnosis	diagnosis	3.8
http://www.webmd.com/brain/autism/autism-spectrum-disorders	autism	Neurodevelopmental Disorder	2.53
http://www.webmd.com/brain/autism/autism-spectrum-disorders	Spectrum	Spectrum	1.27
http://www.webmd.com/brain/autism/autism-spectrum-disorders	disorder	disorder	1.27
https://www.psychologytoday.com/basics/autism	diagnosis	diagnosis	1.83
https://www.psychologytoday.com/basics/autism	autism	Neurodevelopmental Disorder	1.83
http://www.doe.virginia.gov/special_ed/disabilities/autism/	diagnosis	diagnosis	7.14
http://www.doe.virginia.gov/special_ed/disabilities/autism/	autism	Neurodevelopmental Disorder	3.57
http://www.doe.virginia.gov/special_ed/disabilities/autism/	spectrum	spectrum	3.57
http://www.autismeurope.org/	learning	learning	1.12
http://www.autismeurope.org/	language	language	1.12
https://en.wikipedia.org/wiki/Autism	Spectrum	Spectrum	0.17
https://en.wikipedia.org/wiki/Autism	individuals	individuals	0.06
https://en.wikipedia.org/wiki/Autism	cause	cause	0.23
https://en.wikipedia.org/wiki/Autism	autism	Neurodevelopmental Disorder	0.64
https://en.wikipedia.org/wiki/Autism	cause	cause	0.12
https://en.wikipedia.org/wiki/Autism	diagnosis	diagnosis	0.52
https://en.wikipedia.org/wiki/Autism	behaviors	behaviors	0.12
https://en.wikipedia.org/wiki/Autism	disorder	disorder	0.06
https://en.wikipedia.org/wiki/Autism	spectrum	spectrum	0.12
https://en.wikipedia.org/wiki/Autism	language	language	0.06
https://www.genome.gov/25522099/learning-about-autism/	individuals	individuals	2.9
https://www.genome.gov/25522099/learning-about-autism/	diagnosis	diagnosis	4.35
https://www.genome.gov/25522099/learning-about-autism/	autism	Neurodevelopmental Disorder	8.7
https://www.genome.gov/25522099/learning-about-autism/	cause	cause	5.8

Example: Recognized Category Output



# What makes SIRA Unique / Different

- Allows for Non-statistical Natural Language Processing (NLP)
- Ultra-high-speed “reader” Each investigation is
- [Domain-specific] Ontology and Dictionary are directly used in Natural Language Understanding (NLU)
- Automated and semi-automated growth of the Ontology and Dictionary
- No pre-indexing the corpus
- SIRA is not a keyword search engine or anything like it.
  - “Search” is at the “Semantic” level not the “term level”
- SIRA can find semantically relevant information even when it doesn't resemble the terms or sentence structure in your investigation.