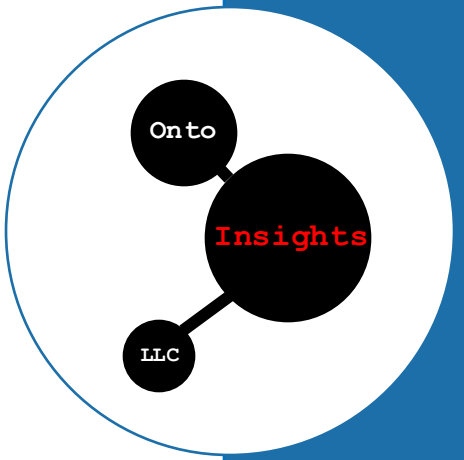
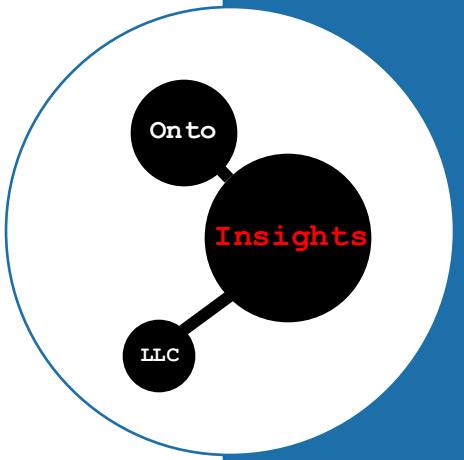
DNA
Deep Narrative Analysis



Andrea Westerinen and Jeff Westerinen
OntoInsights, LLC
arwesterinen@gmail.com, jeff.westerinen@gmail.com

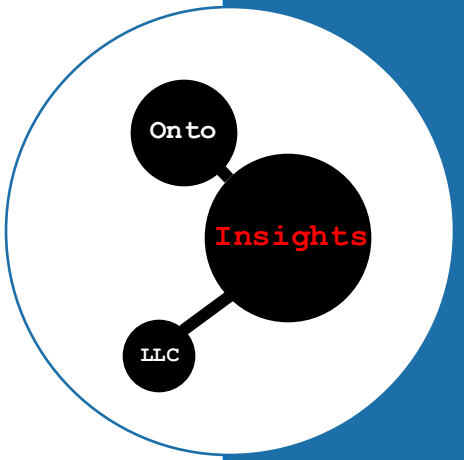
Why Narrative Analysis?

- Storytelling has been the primary mode of human communication for hundreds of thousands of years
  - Humans pre-wired to be influenced by a good story
  - A good story can dramatically increase the virility of a news flash, can catch fire through social media and on-line news sources, and give a voice to the unempowered
- Based on Fisher's narrative theory (1984)
  - Individuals approach the world in a "narrative mode" and make decisions and act within their narrative
  - All communication can be looked at through a narrative lens
    - World is a set of stories from which we choose and re-create our lives



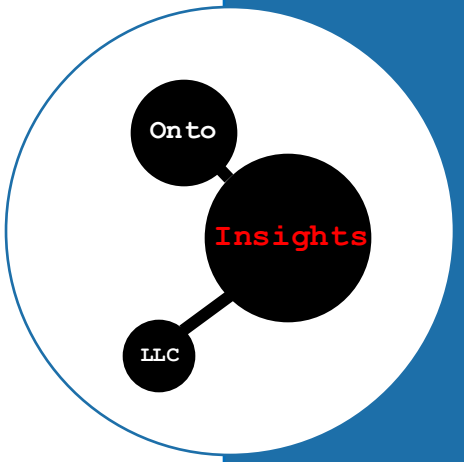
Why DNA?

- Many people no longer trust news
- News' business model has a primary goal of "engaging"/retaining its readers
  - Not reporting a complete and unbiased picture
- Detailed analysis of news and blogs can highlight:
  - Where, how and with what word choices are people and occurrences described and interrelated
  - How news "stories" are constructed from occurrences utilizing storytelling techniques
- This information can inform researchers regarding:
  - What emotions and memories are intended to be triggered
  - Where and how memes, references, quotations, etc. recur and spread
  - How mis-/dis-information is propagated and polarization is reinforced



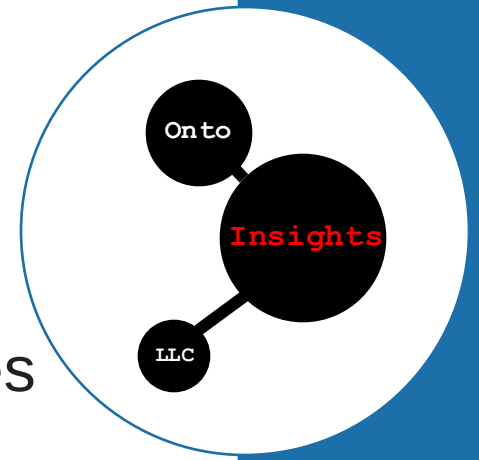
What is DNA?

- Research prototype to analyze news, blogs and audio/video transcripts
- Designed as toolset for researchers
  - Goal to ultimately aid readers in seeing a more complete picture of the world
    - E.g., NATO Information Systems and Technology (IST)-195 Symposium
- Toolset supports:
  - Ingest of news text and rendering as a knowledge graph
    - Using linguistic insights, and AI and semantic technologies
    - Creating consistency and accuracy of parsed results for use in the analyses
  - Aggregation and analysis of information within and across the documents
    - Ultimately executed as an application; currently demo'ed in a Jupyter notebook
  - Inclusion of definitional and contextual background information from Wikipedia/Wikidata, GeoNames and other sources

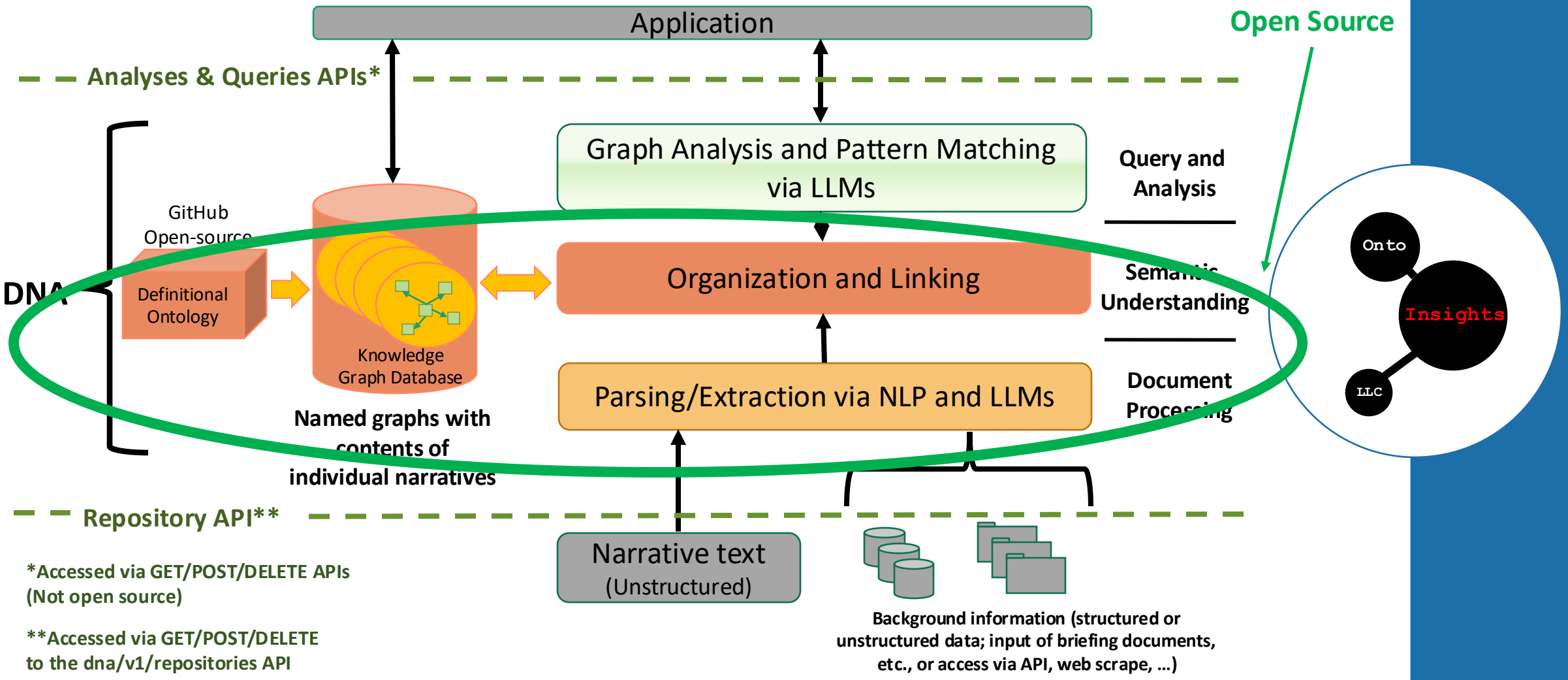


Definition of Terms

- Knowledge graph
  - Graph of interlinked objects, concepts, events/situations, … and the relationships between them
  - Defined according to an ontology which acts as the underlying “schema” for the entities and relationships
- Ontology
  - Formal naming, representation and definition of the concepts, categories, and data and relationship properties in a domain of interest
- Large Language Model
  - Computational model for natural language processing
  - Such as the model behind ChatGPT



DNA Architecture

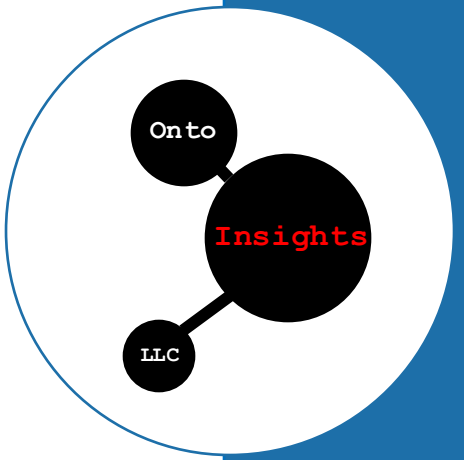


*Accessed via GET/POST/DELETE APIs (Not open source)

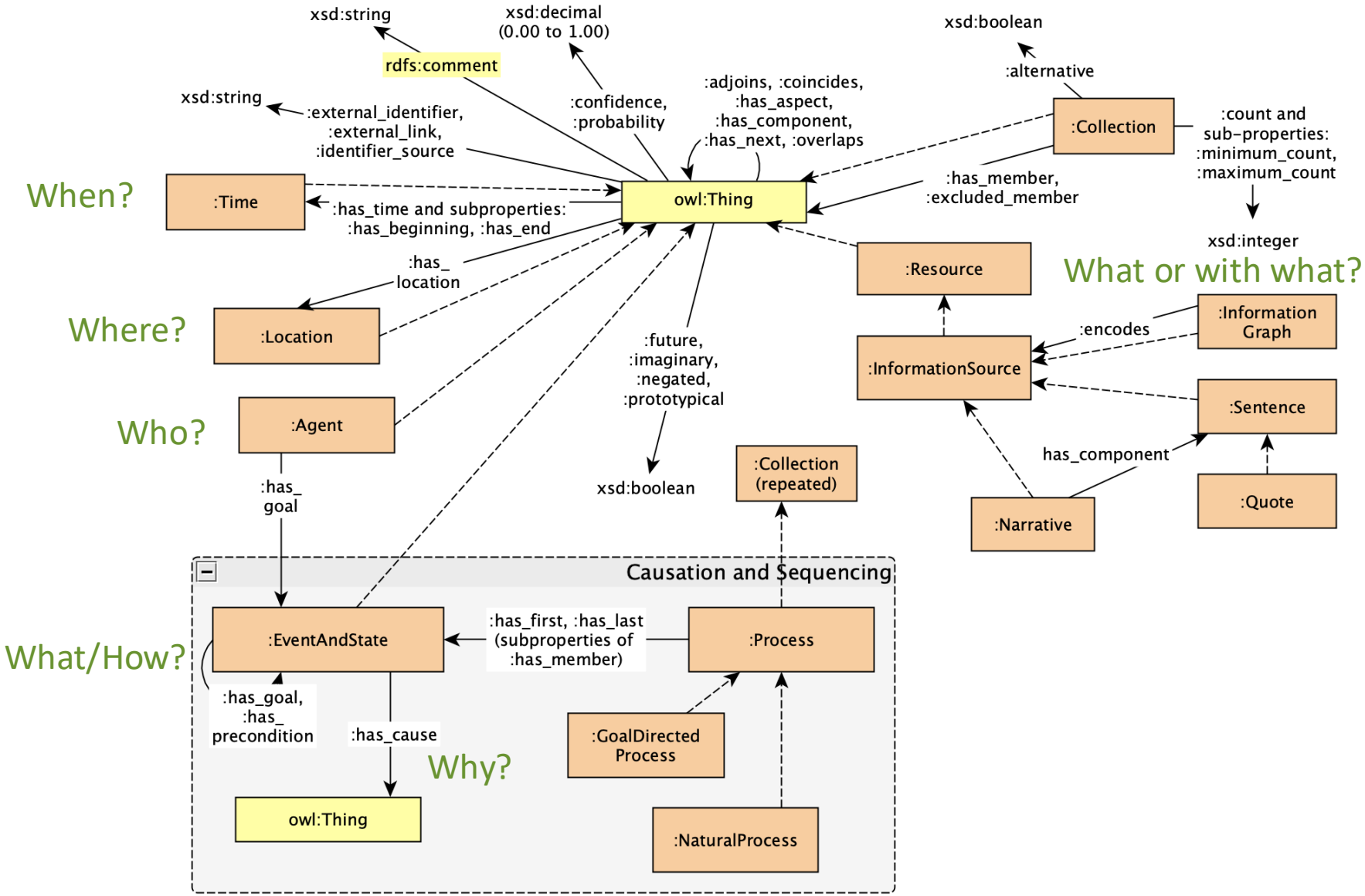
**Accessed via GET/POST/DELETE to the dna/v1/repositories API

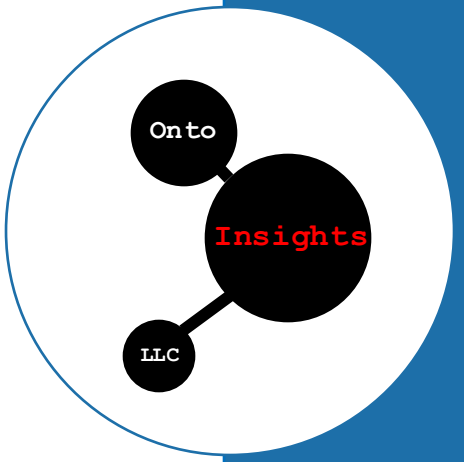
Foundational Technologies

- Linguistic event theory, built into the DNA Ontology
  - Aligned with linguistic and syntactic patterns underlying LLMs
- Natural language processing
  - spaCy for named entity recognition and extraction of sentences
  - LLMs (OpenAI's GPT-4o) for analysis of basic linguistic details, use of rhetorical devices, alignment with the DNA ontology and much more
- Other technologies
  - Ontological reasoning and inference
  - Graph analysis, and machine learning and pattern recognition



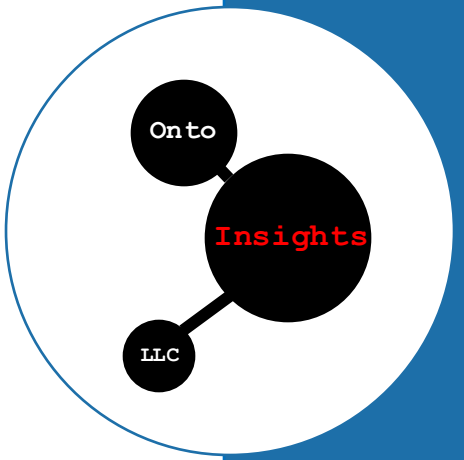
The DNA Ontology





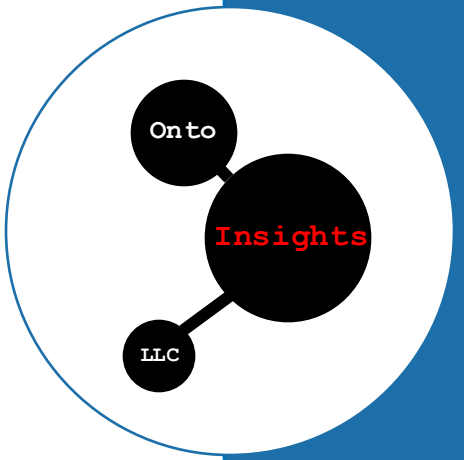
Processing Flow

- Ingest
  - Complete article text and metadata ingested, analyzed and stored in a knowledge graph
- Analysis
  - Focus of the demo
  - Includes overall/narrative and sentence-level analyses



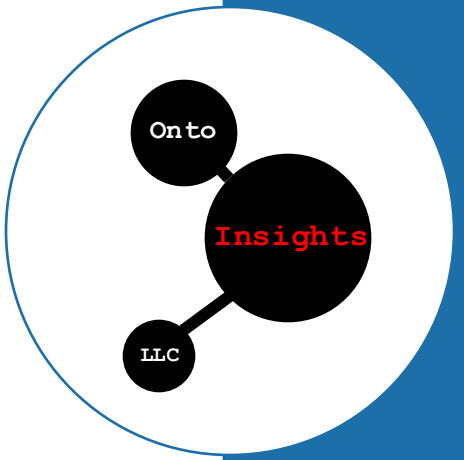
Analysis Target Areas

- Identification of goals to understand purpose
  - E.g., 'advocate' a position, describe a single event in-depth ('describe-single') or 'investigate'
- Detection of subject areas to target/focus semantic analysis
  - E.g., 'crime and law', 'economy and business', or 'politics and international'
- Identification of information flow types to view progression from beginning to end
  - E.g., 'chronological', 'inverted pyramid' or 'question-answer'
- Enumeration of plot lines to create 'narratives' and engage emotionally
  - E.g., 'conflict and resolution', 'rise and fall' or 'justice and revenge'



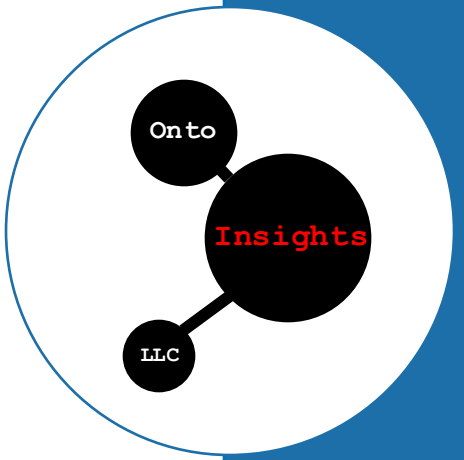
Analysis Target Areas (Continued)

- Extraction of top-level topics and sub-topics to understand what is included/omitted and enable comparison
- Creation of executive summary for quick review
- Identification of sentiment
  - 'positive', 'negative', 'neutral’
- Extraction of named entities (NER)
- Detection of rhetorical devices in sentences
  - E.g., 'ad baculum', 'ad hominem' or 'ad populum'



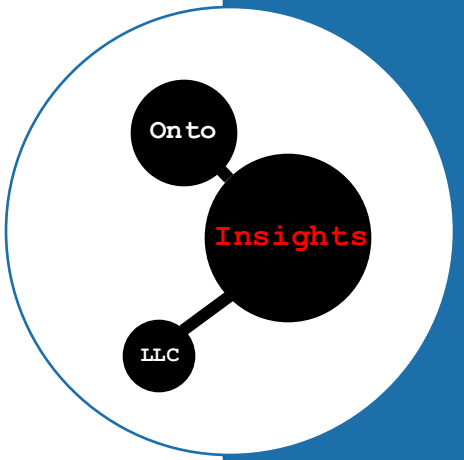
DNA Details

- Open-source
  - https://github.com/ontoinsights/deep_narrative_analysis/tree/master
- Searchable ontology
  - https://ontoinsights.github.io/dna-ontologies/
- API details
  - https://ontoinsights.github.io/dna-swagger/

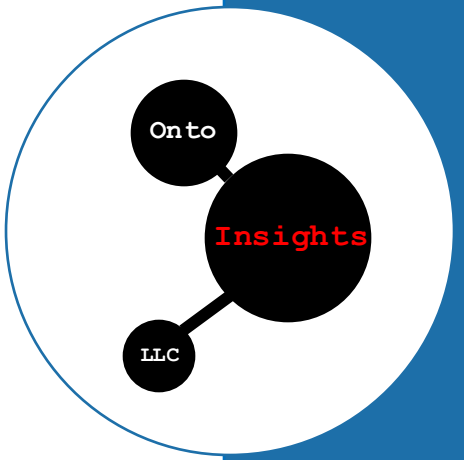


Demo

- Walk-through of the analysis notebook
  - https://github.com/ontoinsights/deep_narrative_analysis/blob/master/notebooks/Demo_Article_Analysis.ipynb
- Can also review ingest, if time permits
  - https://github.com/ontoinsights/deep_narrative_analysis/blob/master/notebooks/Demo_Article_Ingest.ipynb



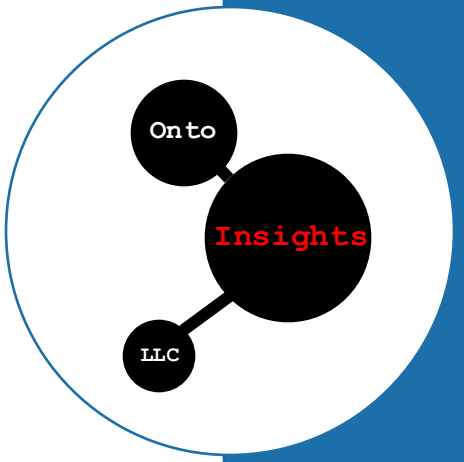
Backup Slides



Ingest Workflow

Example:
https://github.com/ontoinsights/deep_narrative_analysis/blob/master/notebooks/Demo_Article_Ingest.ipynb

- DNA API to create (or reuse) a named ”repository” to hold news article data
  - ”Repositories” reside in a Stardog database on a free cloud endpoint
  - Accessible by name
- DNA API to add background data (text identifying named entities)
  - To assure consistency of entity names/identifying references
- Article text and its metadata obtained
  - Currently manually extract text and metadata (see notebooks/articles in the GitHub repo)
  - Automated news ingest prototyped
- DNA API to ingest article (based on provided text and metadata)



Prompt Format

https://github.com/ontoinsights/deep_narrative_analysis/blob/master/dna/query_openai.py

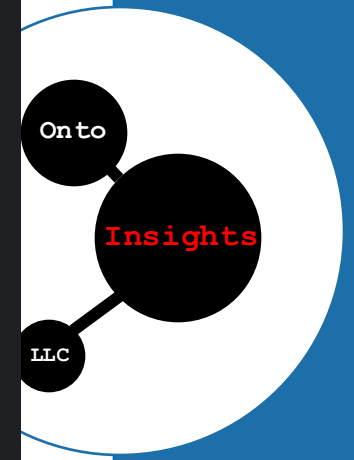
- Includes:
  - System prompt
  - Instructions with potential ”considerations”
  - Inputs
  - Expected JSON output

```
sentence_prompt = \
    f'<Task: You are ChatGPT, a large language model trained by OpenAI using the GPT-4 architecture, with expertise ' \
    'in linguistics and natural language processing (NLP). Your objective is to analyze a sentence from a narrative ' \
    'or news article.> ' + \
    '<Instructions: 1. Input Formats: a) You are given the text of a sentence from an article, where ' \
    'the sentence ends with the string "**" which is ignored. b) A numbered list of rhetorical devices that ' \
    'may be used in the sentence, is also provided. ' \
    '2. Sentence Analysis: a) Indicate the grade level that is expected of a reader to understand the ' \
    'sentence semantics. b) Provide the numbers of the various rhetorical devices used in the sentence, and ' \
    'explain why they are identified. If there are no rhetorical devices used, return an empty array for ' \
    'the "rhetorical_devices" JSON key, specified in the Output. > ' \
    '<Inputs: 1. Sentence text: {sent_text} ** ' + \
    f'2. Rhetorical devices: {rhetorical_device_texts}> ' \
    f'<Output: Return the results as a JSON object using the following structure: {sentence_result}>'
```

Prompt Details 1

```
rhetorical_devices = ['ad baculum', 'ad hominem', 'ad populum', 'allusion', 'exceptionalism', 'expletive',
                      'imagery', 'invective', 'loaded language', 'logos', 'paralipsis', 'pathos',
                      'rhetorical question or accusation']
rhetorical_devices_text = 'The rhetorical device categories are: ' \
    '1. An appeal to force or a threat of force in order to compel a conclusion (ad baculum)' \
    '2. Use of wording that verbally demeans or attacks a person (ad hominem) ' \
    '3. Reference to general or popular knowledge such as "the most popular xyz" or "everyone says xyz" (ad populum) ' \
    '4. Reference to an historical/literary person, place or thing that has symbolic meaning (allusion) ' \
    '5. Use of language that indicates that a particular entity is somehow unique, extraordinary or ' \
    'exemplary (exceptionalism)' \
    '6. Use of emphasis words, such as "in fact", "of course", "clearly" or "certainly" (expletive) ' \
    '7. Use of imagery and descriptive phrases that paint a vivid picture that emotionally engages a reader (imagery)' \
    '8. Use of ridicule, or angry or insulting language (invective) ' \
    '9. Use of "loaded language" such as words like "double-dealing", with strong connotations which invoke ' \
    'emotions and judgments' \
    '10. Use of statistics and numbers (logos) ' \
    '11. Indicating that little or nothing is said about a subject in order to bring attention to it, ' \
    'such as saying "I will not mention their many crimes" (paralipsis)' \
    '12. Wording that appeals to emotion such as fear or empathy (pathos)' \
    '13. Asking rhetorical questions or making an explicit or implicit accusation'
```

```
sentence_result = '{"grade_level": int, ' \
                  '"rhetorical_devices": [{"device_number": int, "explanation": "string"}]}'
```



DNA Processing