TEXAS A&M UNIVERSITY
Engineering

Extracting Science Traceability Graphs from Mission Concept Abstracts using Natural Language Processing

Benjamin Simpson (TAMU), Dr. Daniel Selva (TAMU), Dr. David Richardson (NASA GSFC)

Presentation Outline

1. Background/Motivation
2. AI – Natural Language Processing
3. Methodology – The Processing Pipeline
4. Current Results
5. Summary/Ongoing Work

BACKGROUND/MOTIVATION

Mission Evaluation – The Reviewer's Perspective

- Assessing mission concepts on several factors, including:
  - Objective metrics
    - Cost
    - Schedule
    - Technical Feasibility
  - Subjective metrics
    - **Traceability of decisions**
      - Clarity and completeness
- Mission/portfolio relevance vs. intrinsic mission details
  - The former looks at the correlation between proposed goals/objectives vs. established needs
  - The latter seeks to understand the traceability of engineering decisions *within* the mission concept

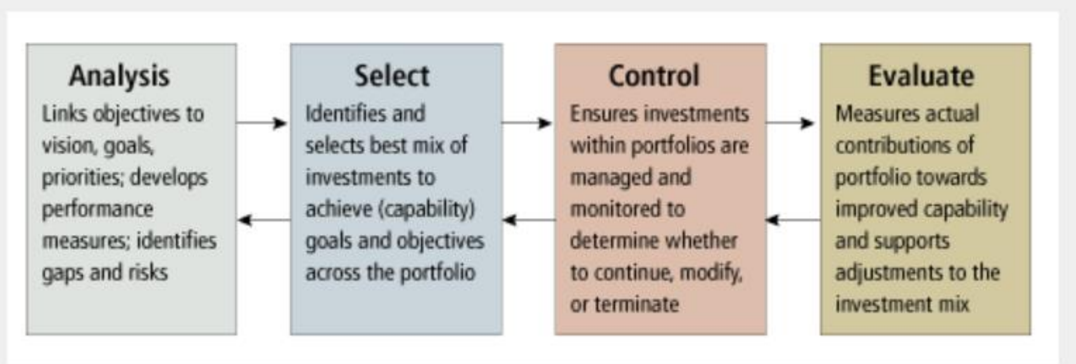


Figure 1. Simplified View of Portfolio Management Activities

Image credit [MITRE]

Traceability of Decisions

TEXAS A&M UNIVERSITY Engineering

- Mapping the flow of programmatic goals -> measurement requirements-> engineering decisions
  - Requirements need to provision bidirectional traceability [1]
  - Reviewers need to question the logical traceability flow of concept documentation
- Some notable methods to capture traceability
  - **Requirement Traceability Matrices (RTMs)**
  - **Science Traceability Matrices (STMs)** [2]

Science Traceability Matrix

TEXAS A&M UNIVERSITY Engineering

- Standard requirement for NASA proposals
  - Establishes a taxonomy in tabular format
    - The bridge between the science and engineering worlds
- Follow-on renditions
  - P-STAF – mapping measurement campaigns to instruments [3]
- Still, STMs/RTMs can vary heavily in complexity and scope:
  - No robust STM standard
  - Jargon and content representation varies significantly in practice
  - Notion of goals and objectives not clearly established
  - Also consider that STMs/RTMs may not yet be available in early documentation!

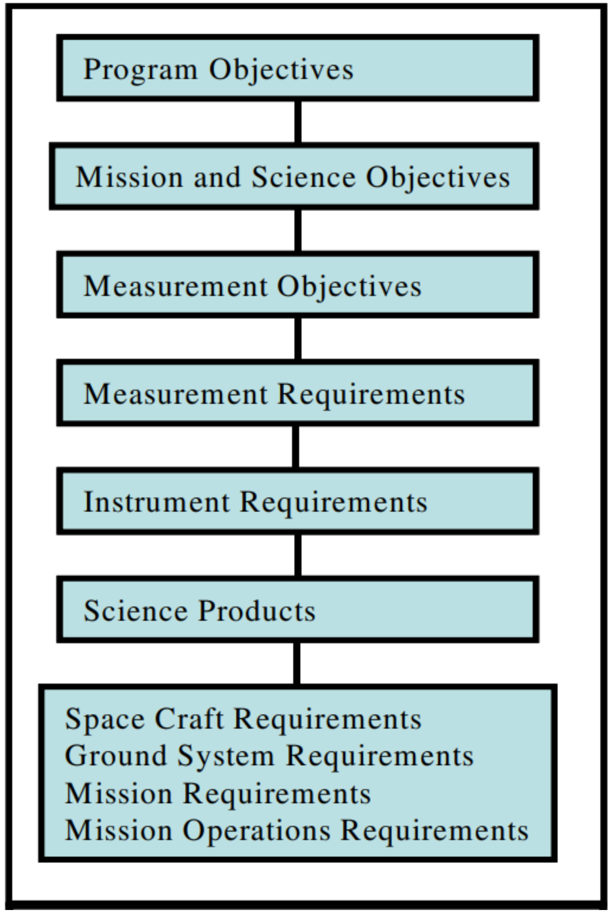


Figure: Contents included in a STM as per [2]

STM Example

TEXAS A&M UNIVERSITY
Engineering

NASA Science Goals	CDIM Science Goals	CDIM Science Objectives	Science Requirements: Physical Parameters	Science Requirements: Observables	Science Requirements: Measurement Requirement	Instrument Requirements: Instrument Parameter	Instrument Requirements: Science Requirement	Instrument Requirements: Capability	Mission Requirements: Driver	Mission Requirements: Parameter
Explore the origin and evolution of the galaxies, stars and planets that make up our universe [NASA Science Plan] How does the Universe work? How did we get here? [NASA 2014 Science Mission Directorate Strategy Document]	Trace the stellar mass buildup, dust production history, and metal enrichment history during cosmic reionization.	Determine if the rate of growth of metals and dust corresponds to the growth of stellar mass at 5 < z < 8.	Metallicity of galaxies via the oxygen abundance, stellar mass, and dust attenuation (extinction rate, dust density)	[OIII], [OII], [NII]/Hα, Hα/Hβ @5< z <8	(i) Wavelength coverage to detect Hα out to z of 10. (ii) Spectral resolving power to resolve [NII] and Hα. (iii) Sensitivity to detect galaxies < 10^9 M_{sun} in a deep survey.	Wavelength range	$2.2 \le \lambda \le 6.0$ μm	$0.75 \le \lambda \le 7.5$ μm	Data Reliability and Systematic Error Control	Deep, medium and wide surveys each with ≥ 90% voxel completeness for internal reliability. Spatial resolution: Effective PSF FWHM ≤ 2" at 1 μm (from science requirements). Stable cooling to < 35 K to control > 5 μm array dark current.
						Spatial resolution (pixel scale)	Θ_{pix} = 1"–2"	Θ_{pix} = 1"		
						Spectral resolving power	$\lambda/\Delta\lambda \ge 300$	$\lambda/\Delta\lambda = 300$		
						Point source broadband photometric sensitivity (R=5, 5σ; deep survey)	24.5 AB mag at J band	25.2 AB mag at J band		
						Spectral line flux sensitivity (3.5σ; deep survey)	5.0×10^{-18} erg s^{-1} cm^{-2} at 4.6 μm	2.7×10^{-18} erg s^{-1} cm^{-2} at 4.6 μm		
	Establish the role of active galactic nuclei (AGN) in cosmic reionization.	Determine the fractional contribution of super-massive black hole/ AGNs to reionization photon budget.	Unbiased UV photon spectral density; black-hole masses via line widths of optical lines.	Rest-frame UV continuum @z = 5–8. [MgII] and other metal lines.	(i) Sensitivity to detect faint quasars in a wide survey. (ii) Spectral resolving power to detect equivalent width of broad metal lines.	Wavelength range	$2.9 \le \lambda \le 6.0$ μm	Same as above	Survey Strategy	Deep survey: 15 deg^2, imbedded in the Wide survey.
						Spatial resolution (PSF; FWHM)	Θ_{FWHM} = 2" at K band	Θ_{FWHM} < 2" at K band		
						Spectral resolving power	$\lambda/\Delta\lambda \ge 300$	Same as above		
						Point source broadband photometric sensitivity (R=5, 5σ; wide survey)	23.5 AB mag at K band	24.0 AB mag at K band		Medium survey: 30 deg^2, to overlap with 21-cm fields from HERA and SKA1-LOW.
	Establish the progression and topology of reionization from cosmic dawn at z = 10 to the end of reionization at z < 6.	Determine the progress of reionization by measuring the ionization fraction in at least 10 redshift bins at 5 < z < 10, with accuracy better than 10%.	Lyα luminosity function, escape fraction, and the spatial distribution.	Lyα	(i) Wavelength coverage to detect Lyα out to z of 10. (ii) Sensitivity to detect faint galaxies.	Wavelength range	$0.75 \le \lambda_{Ly\alpha} \le 0.98$ μm	$0.75 \le \lambda \le 7.5$ μm		
						Spectral resolving power	$\lambda/\Delta\lambda \ge 100$	Same as above		
						Spectral line flux sensitivity (3.5σ; deep survey)	2.9×10^{-17} erg s^{-1} cm^{-2} at 0.85 μm	2.0×10^{-17} erg s^{-1} cm^{-2} at 0.85 μm		Wide survey: 300 deg^2, driven by number of AGN detections.
			Reionization history of the universe.	Lyα and Hα	(i) Ability to perform cross-correlations, including Lyα and Hα, and 21-cm radio measurements.	Wavelength range	$0.75 \le \lambda_{Ly\alpha} \le 1.4$ μm $3.9 \le \lambda_{H\alpha} \le 7.2$ μm	$0.75 \le \lambda \le 7.5$ μm		
						Spectral resolving power	$\lambda/\Delta\lambda \ge 100$	Same as above		
						Surface brightness sensitivity (1σ; medium survey)	1.3×10^{-18} erg s^{-1} cm^{-2} Hz^{-1} sr^{-1} at 1.1 μm	1.5×10^{-19} erg s^{-1} cm^{-2} Hz^{-1} sr^{-1} at 1.1 μm		Read, reduce, and telemeter spectral imaging data.

STM for the Cosmic Dawn Intensity Mapper (CDIM) Probe Concept by NASA

STM Example

TEXAS A&M UNIVERSITY Engineering

Science Goals	Science Objectives	Scientific Measurements	Driving Requirements
1.Measure the spin distribution of accreting black holes	1.1 Measure the spin distribution of accreting black holes	Thermal continuum	Energy range: 0.2–30 keV
		Reflection & X-ray reverberation	Energy resolution: 200 eV
		High frequency QPO	Time resolution: 100 microsec
			Effective Area: 20,000 cm^2
		Transient outbursts	Observe bright sources with full energy and time resolution
			Wide-field monitoring: 75% of sky, 5 mcrab (1 day) sensitivity, 1 keV energy resolution, 2 arcmin position accuracy
			ToO response (< 24 hours)
	1.2 Measure BH spin for 20 AGN to <10%	Reflection & X-ray reverberation	Energy range: 1–30 keV
		Jetted TDE detection	Energy resolution: 200 eV
			Effective Area: 20,000 cm^2
2. Understand the equation of state of dense matter	2.1 Measure the mass and radius to within 5-10% for ~20 pulsars to map the EOS and probe potential phase transitions	Pulse profile modeling for rotation powered pulsars, accretion powered pulsars, and thermonuclear burst oscillation sources	Effective area: 16,300 cm2 @1 keV/38,200 cm^2 @ 6 keV;
			Time resolution: 80 microsec
			Energy resolution: 85-175 eV FWHM (0.2-10 keV)
			TOO response time: hours
	2.2 Search for the fastest spinning pulsars	Search for spin frequencies up to 2 kHz	Time resolution: 50 microsec
3. Explore the properties of the precursors and electromagnetic counterparts of gravitational wave sources	3.1 Enable detection of 5–10 short gamma-ray bursts per year	Detect and localize w/ immediate trigger or ground searches	Wide-field monitor as above with 1ms time resolution
	3.2 Search for signatures of merging supermassive BH		All wide-field monitor data downlinked to ground

STM for the STROBE-X Probe Concept by NASA

Case Study – Astrophysical Decadal Survey

- Introduced early in this project as the target case study
- Survey strives to develop a comprehensive plan of action for the next decade
- 2020 Survey Report slated for release in Fall 2021

Research Question

TEXAS A&M UNIVERSITY Engineering

- How can we leverage AI to help build/visualize traceability in the context of mission proposals?
- What we know about our domain:
  - Contains mission concepts
  - Highly variant verbiage
  - Very mixed representations of 'similar' information

Astro2020 Statement of Task

1. Provide an overview of the current state of astronomy
2. Identify most compelling science challenges

Targeted Task Area!

3. Develop a comprehensive research strategy
4. Utilize and recommend decision rules for the research strategy
5. Assess the state of the profession

AI – NATURAL LANGUAGE PROCESSING

Natural Language Processing (NLP)

- An umbrella term encompassing a wide range of strategies used to analyze text (from structured to unstructured)
- Can be as simple as looking at term frequency
- Some useful NLP tasks
  - Parts of Speech (POS) tagging
  - Named-entity recognition
  - Relation extraction

These are very crucial for this work

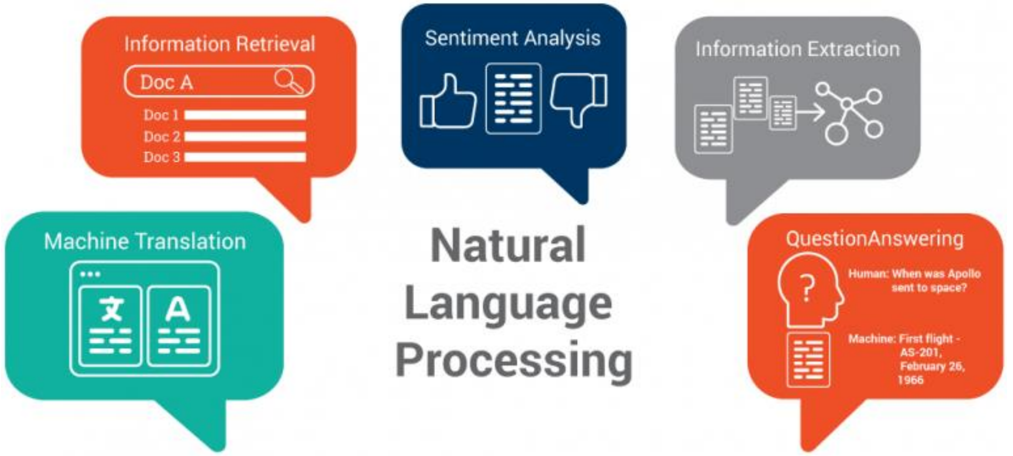


Image credit*

*https://medium.com/@mallrishabh52/natural-language-processing-2913817282c1

Named-Entity Recognition (NER)

TEXAS A&M UNIVERSITY Engineering

- Identify words or phrases from text and assign them a label
  - Example
    - "*Texas A&M* {entity_type: ORG} *defeated No.1 ranked Alabama in a stunning upset.*"
  - State of the art methods utilize machine learning (neural networks, transformers) for NER
    - Models are pre-trained and fined tuned to perform NER
  - Also consider entity disambiguation
    - Entities that may seem similar but are actually different
      - Example: "*Washington*" -> PERSON or LOCATION?

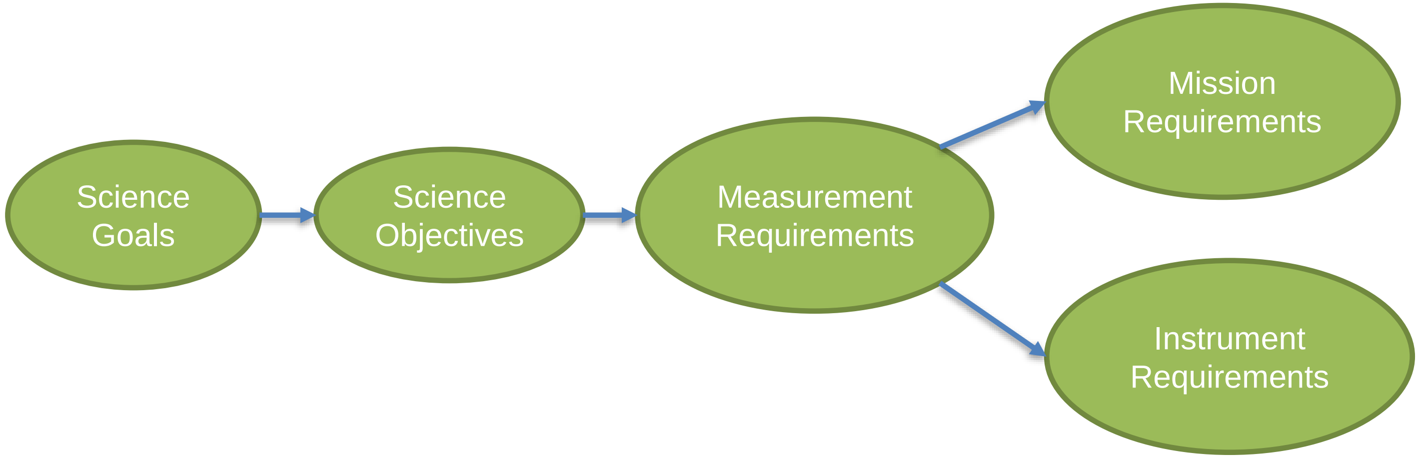
Relation Extraction

TEXAS A&M UNIVERSITY Engineering

- Given entities and their types … are they related (and how)?
- Several methods exists; we have explored:
  - Supervised-methods
    - Train a model to predict relations based on labeled data
    - Labels could be based on features (entity types, parse trees, POS tags)
  - Embeddings (related entities appear spatially close)
    - Transformer models encapsulate these embeddings
- Paired with NER, we attain two useful techniques for **Information Extraction**

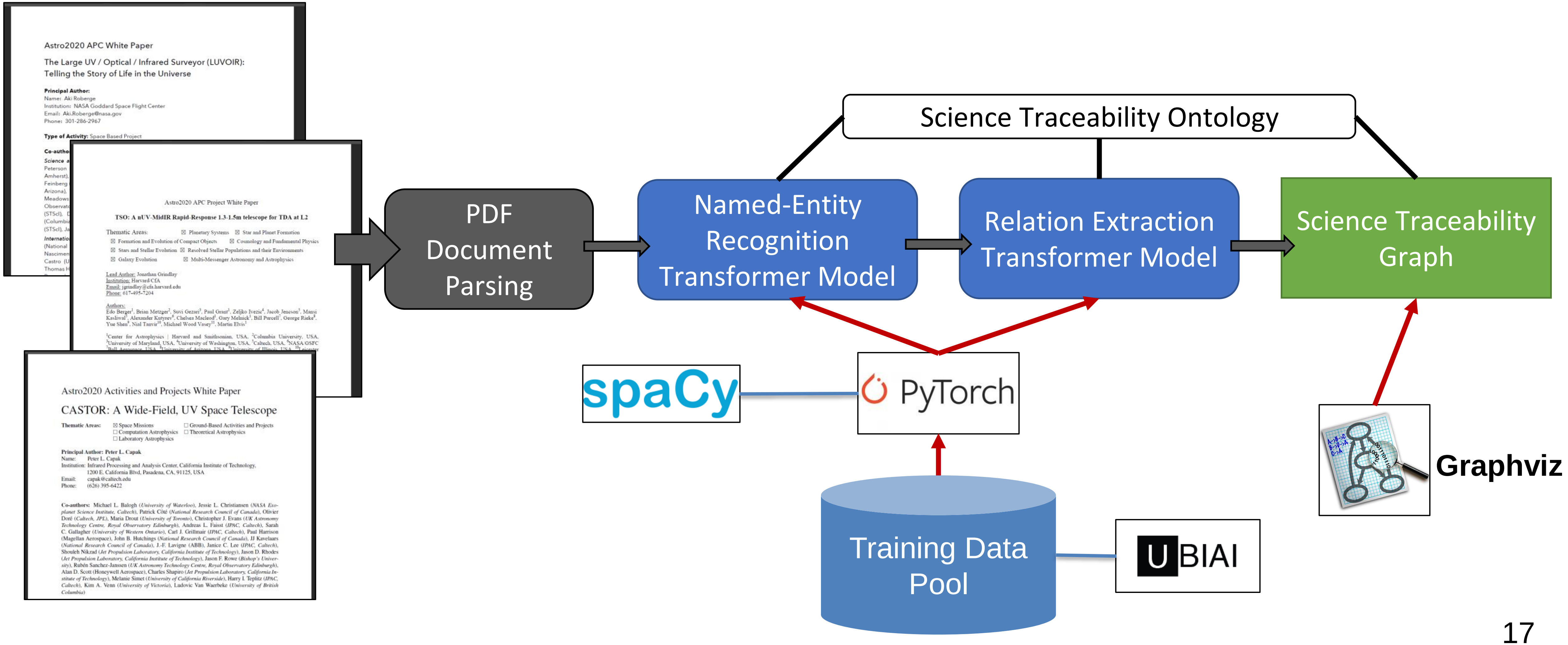
Application to the Astrophysics Decadal Survey

- Science traceability matrices are important when assessing NASA space missions
  - **Postulate**: STMs can be represented as graphs with entities (nodes) and relations (edges)
- We can reduce the STM to a graph-style taxonomy and apply information extraction methods on mission concept documentation
  - I.e. submitted Astrophysics Decadal Proposals



METHODOLOGY – THE PROCESSING PIPELINE

Pipeline Structure



Science Traceability Ontology

- We formulated a custom ontology to serve as the backbone behind the science traceability 'graph'
  - Motivations from STM, P-STAF, and Astro Decadal Survey Process [2-4]
- For other domains (DoD) you can adapt from this taxonomy
  - Brainstorming examples:
    - "Science Theme" -> "Capability Need"
    - "Instrument" -> "Sensor"
    - "Observable Requirement" -> "Capability Requirement"

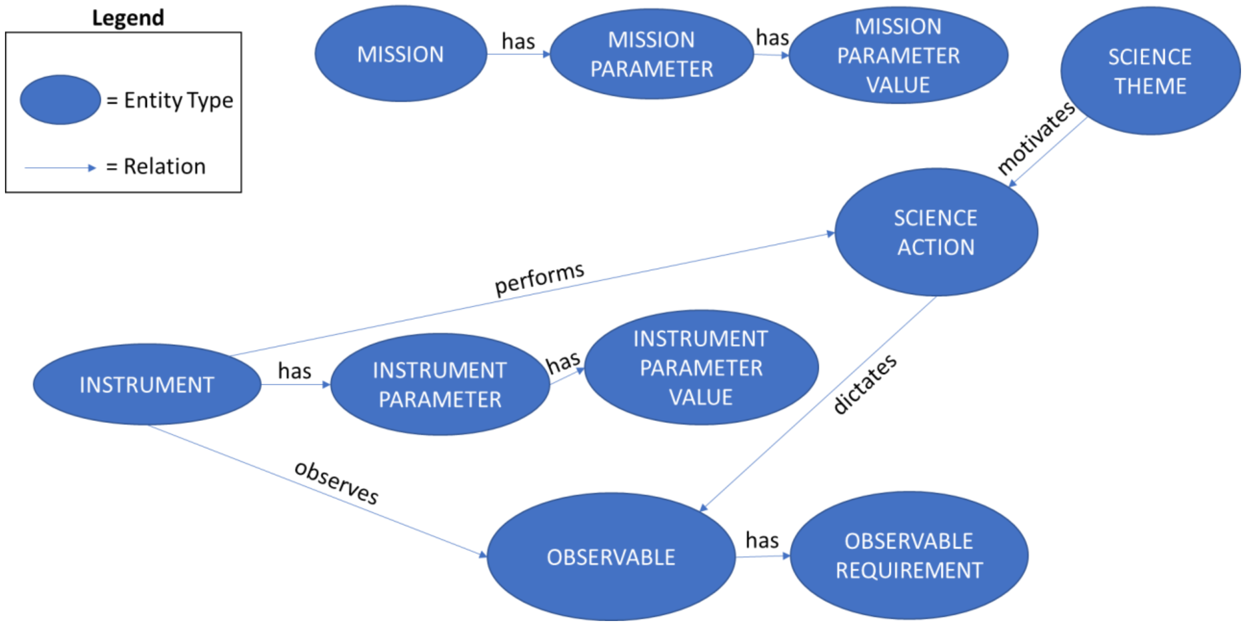


Figure: Guiding ontology for the STM graph taxonomy (Note: We are actively working on an updated version for this graph).

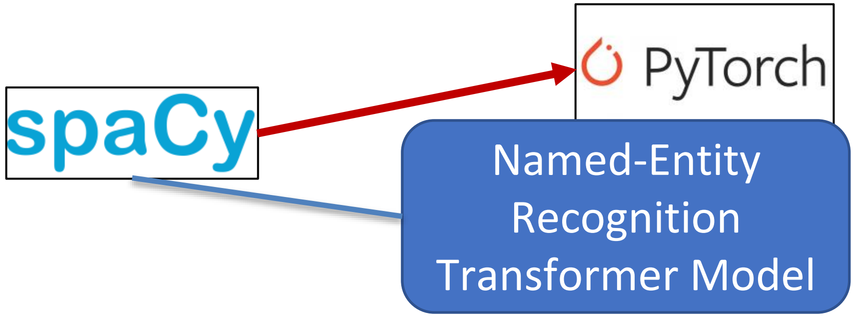
PDF Document Parsing

TEXAS A&M UNIVERSITY Engineering

- Open Access Python library to parse documentation
  - https://pypi.org/project/py-pdf-parser/
- Extract “science” and “engineering” sections from proposal documentation
  - Created custom python script to filter out unwanted sections of the PDF document (namely the author and references lists)

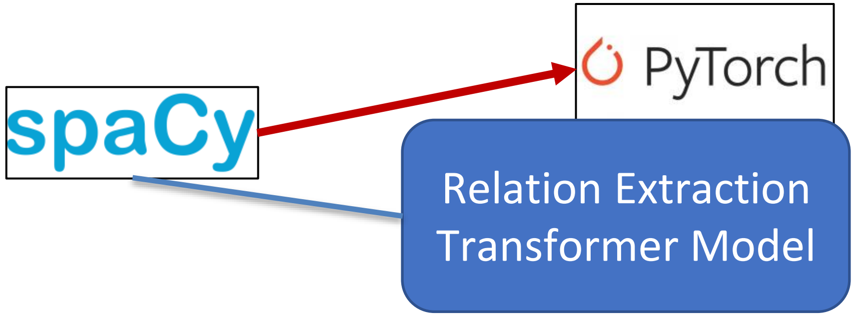
Pre-Trained NER Transformer

- Wrapped through spaCy 3.0 (python)
- Used PyTorch transformer models from huggingface:
  - https://huggingface.co/
- SciBERT [5] pretrained model
  - Transformer model based off BERT architecture
  - Trained on 1.14 million scientific papers (3.17 billion tokens)



Relation Extraction Transformer

- Adds another pre-trained SciBERT model to the pipeline
- spaCy provides example that trains entity-entity pairings (relations) from scratch
  - https://github.com/explosion/projects/tree/v3/tutorials/rel_component



Training the Transformer Models

- Training data development conducted through a web-based annotation tool:
  - UBIAI: https://ubiai.tools/
- Training files are then downloaded and pushed through the training process

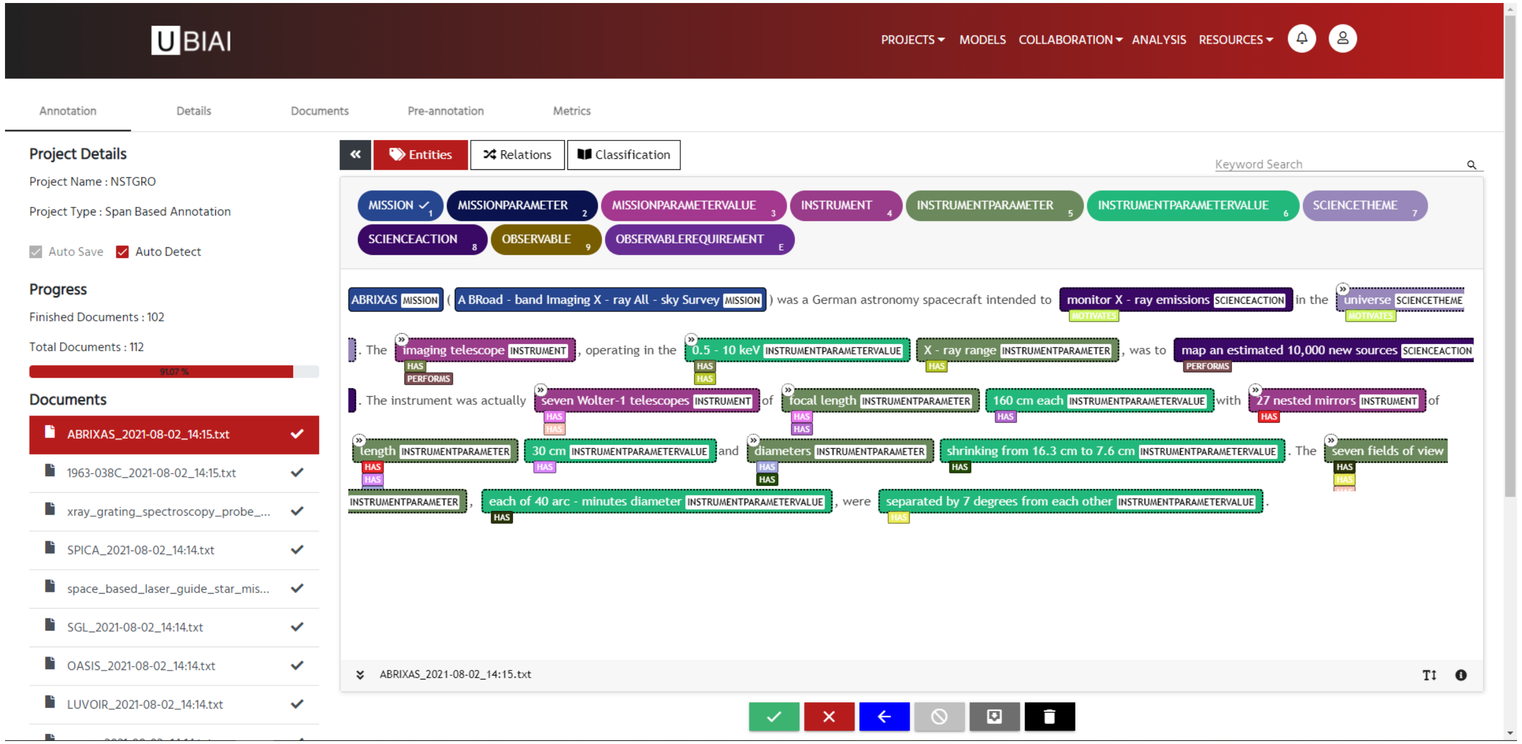
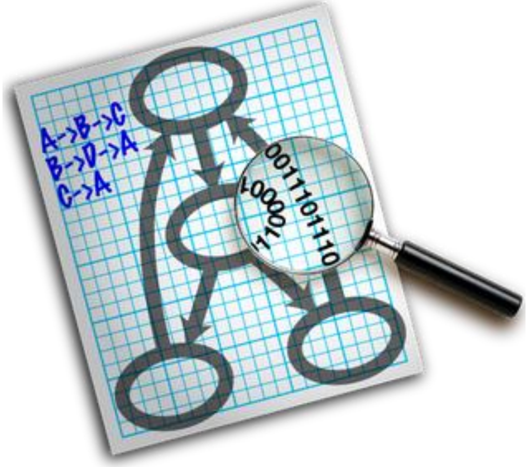


Figure: UBIAI annotation user interface

Visualization through GraphViz

TEXAS A&M UNIVERSITY Engineering

- Graph visualization done through GraphViz
  - Open-source graph visualization software
  - https://graphviz.org/
  - Python wrapper for graphviz: https://pypi.org/project/graphviz/
- Customized graph structure to organize entities and relations to resemble an STM (in graph form)



DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

Tool GUI - AstroNLP

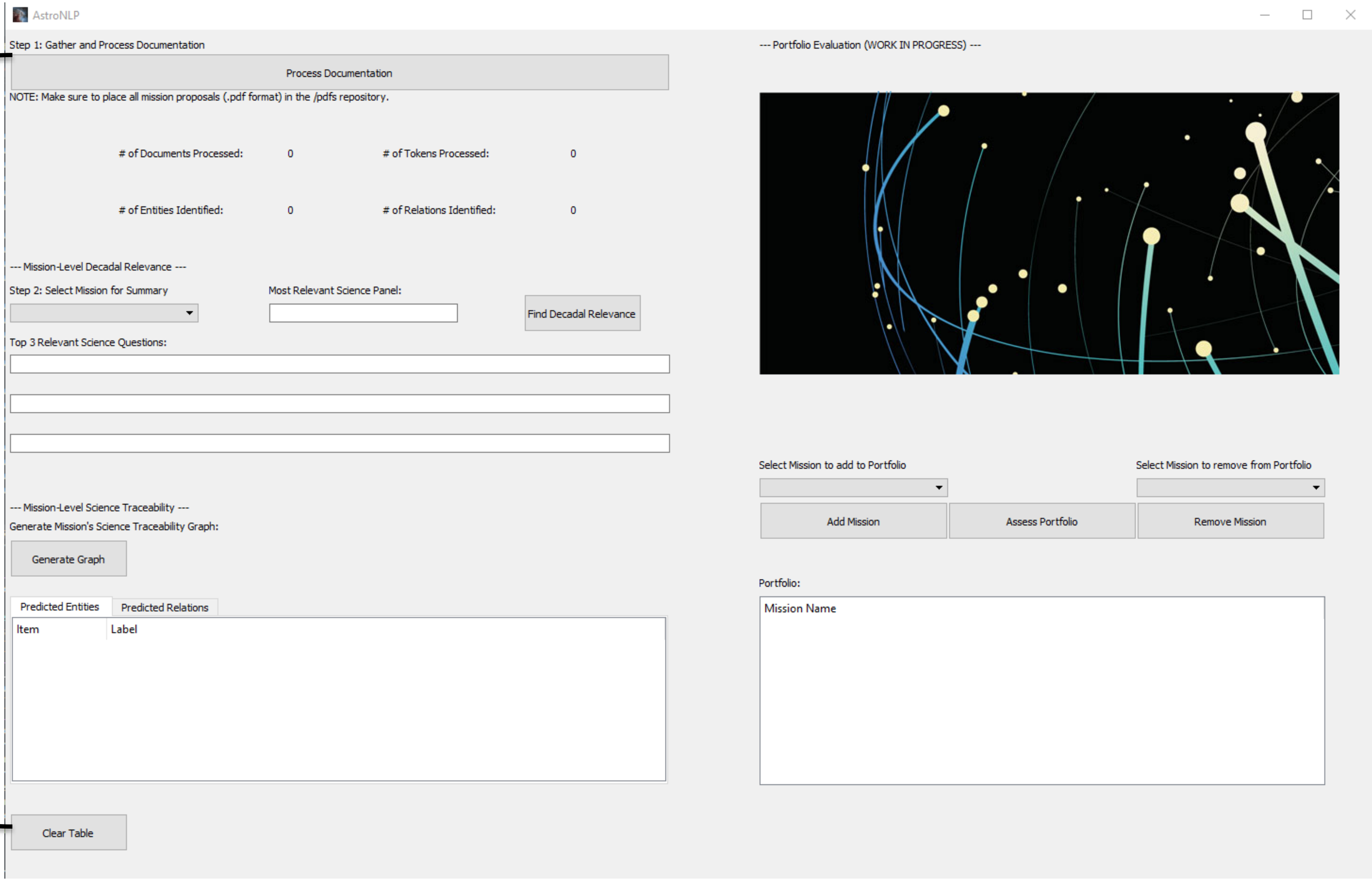
TEXAS A&M UNIVERSITY Engineering

- Processing pipeline implemented in python-based tool
  - GUI Library: PyQt5 (https://pypi.org/project/PyQt5/)
- Capabilities:
  - Process multiple PDFs (documentation contained in local repo is automatically read and parsed)
  - Extract entities and relations
  - Visualize STM in graph form for each proposal

Tool GUI - AstroNLP

TEXAS A&M UNIVERSITY Engineering

Mission-level analysis (**traceability extraction** and mission-level relevance)

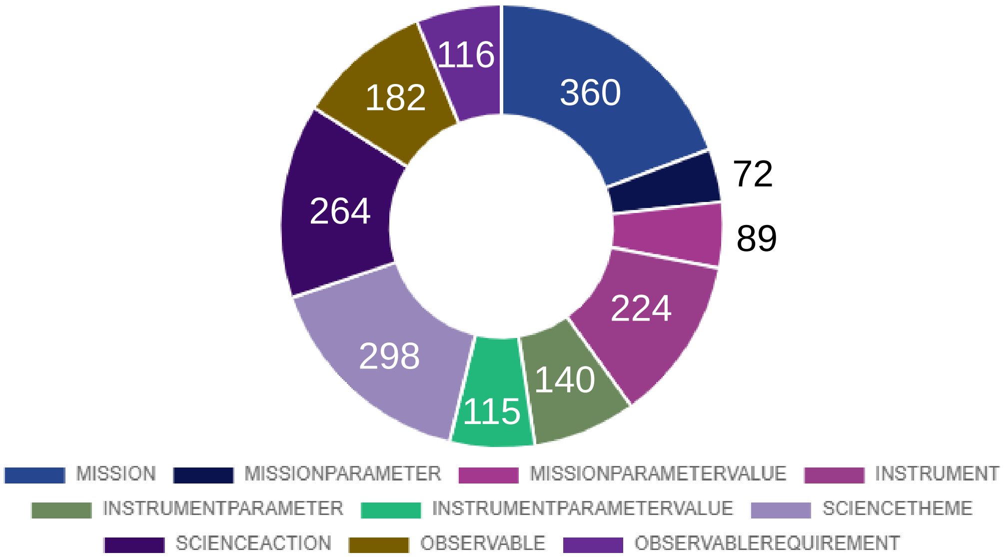


Portfolio-level analysis (relevance only)

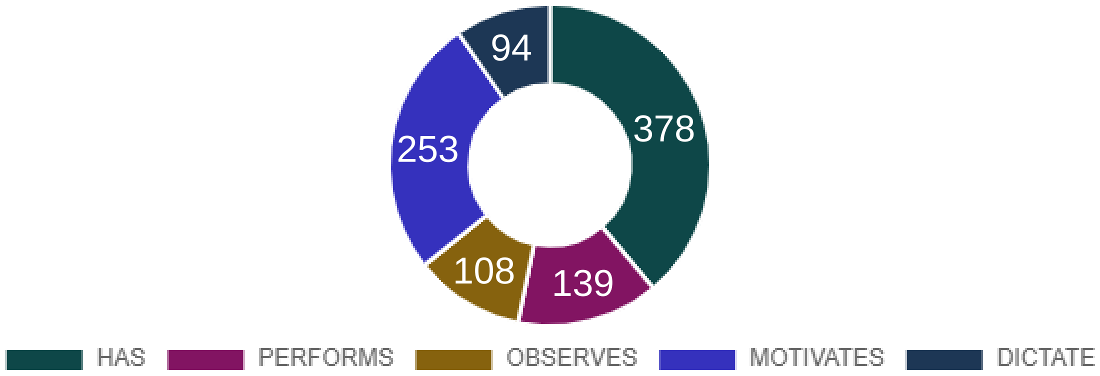
CURRENT RESULTS

Training Data Size

Named-Entity Recognition - # of Examples



Relation Extraction - # of Examples



*These numbers grow weekly as more examples are added!

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

Evaluating Performance

- Precision, Recall, and F1 scores are a common method to evaluate NER and Relation Extraction Tasks
  - Precision (P) – Fraction of relevant instances amongst predicted instances
  - Recall (R) – Fraction of relevant instances that were predicted
  - F1 – Harmonic mean of precision and recall:
- We're, however, using the scoring metrics detailed in the Message Understanding Conference (MUC) [6] with a slight deviation
  - They capture different categories of error beyond the strict correct/incorrect metrics
  - We use these to evaluate the NER transformer only

MUC-5 Definitions

COR	❑ Correct	response = key
PAR	❑ Partial	response $\cong$ key
INC	❑ Incorrect	response $\neq$ key
SPU	❑ Spurious	key is blank and response is not
MIS	❑ Missing	response is blank and key is not
	❑ Noncommittal	key and response are both blank

Figure: Scoring categories as per [6]

***We take a slight deviation from the "Partial" (PAR) metric, preferring to calculate it as the degree of 'coverage' the predicted entity has with a true entity. -> modPAR**

Transformer Performance Scores

- Message Understanding Conference (MUC) metrics [6]:

$$POS = COR + INC + modPAR + MIS$$

$$ACT = COR + INC + modPAR + SPU$$

$$P = \frac{COR + modPAR}{ACT}$$

$$R = \frac{COR + modPAR}{POS}$$

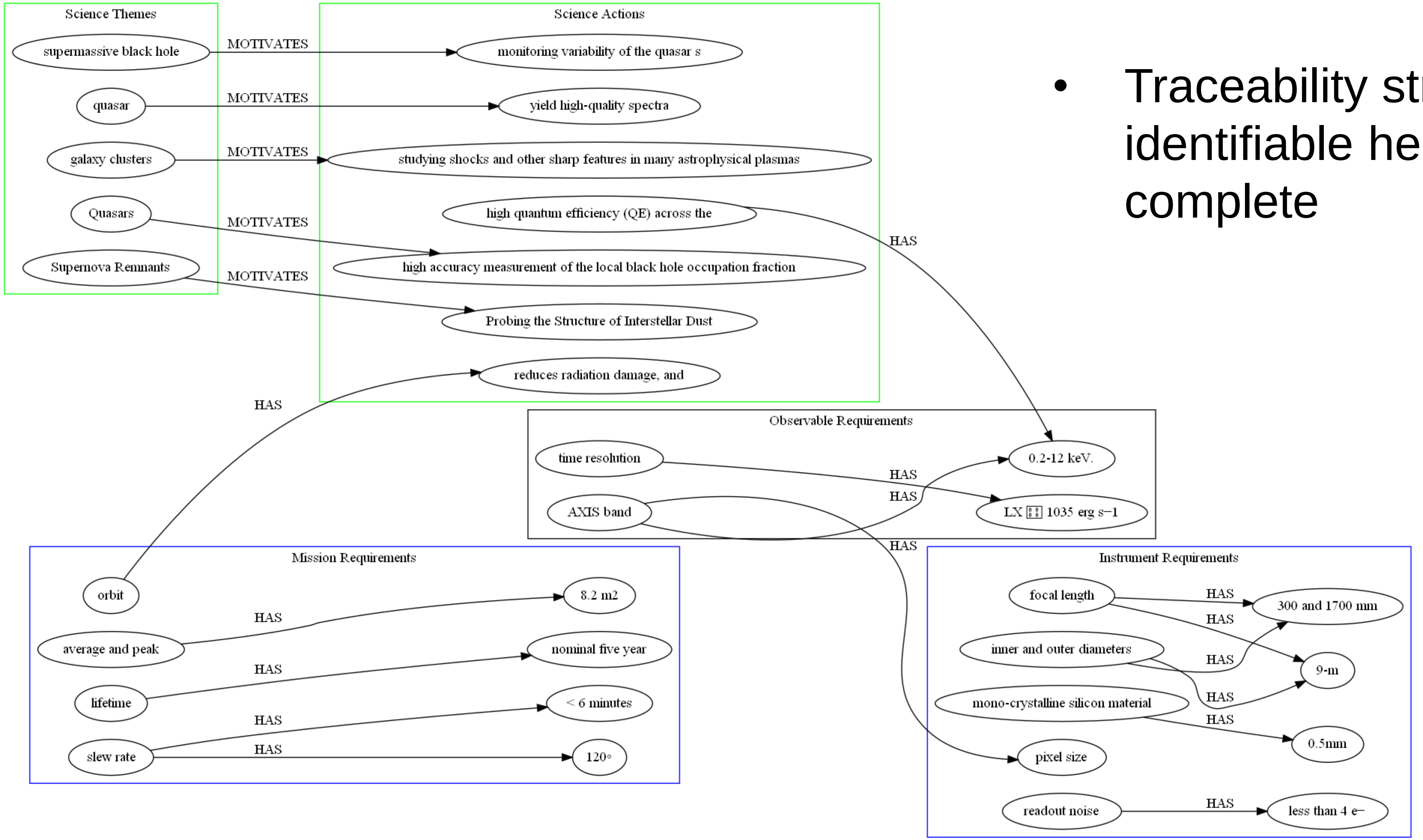
$$F1 = \frac{2PR}{P + R}$$

Scoring Metric	NER Transformer	Relation Extraction Transformer
Precision (P)	0.17	0.49
Recall (R)	0.14	0.28
F1	0.15	0.35

Applied Results: AXIS

- AXIS: The Advanced X-Ray Imaging Satellite
  - Probe Mission Concept
  - 18 page proposal document
- Post-Processed Document Metrics:
  - 7097 tokens analyzed
  - 473 total entities identified
  - 20* relations extracted
    - *Many relations are filtered out if they do not meet a minimum confidence score (currently 50%)

AXIS Extracted STG



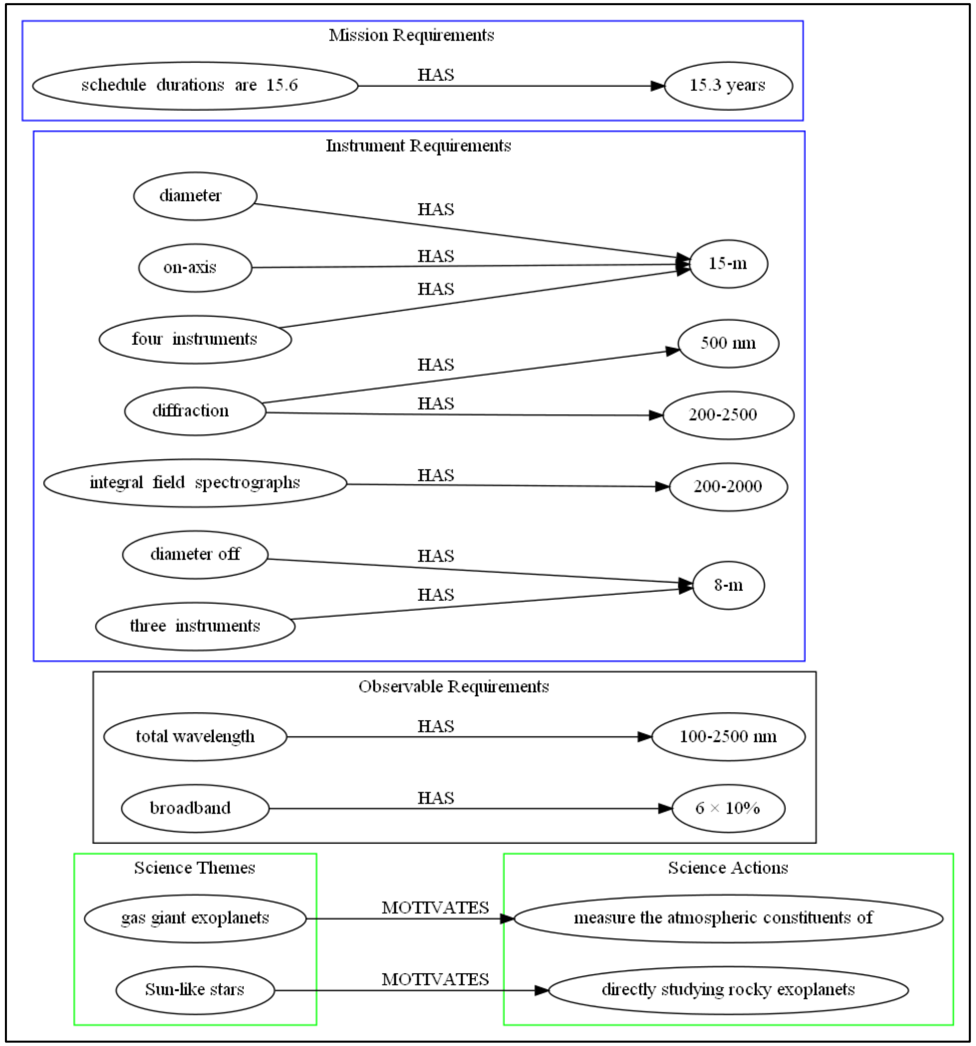
- Traceability structure is slightly identifiable here but by no means complete

Applied Results: LUVOIR

TEXAS A&M UNIVERSITY Engineering

- LUVOIR: Large UV/Optical/IR Surveyor
  - Large Mission Concept
  - 12 page proposal document
- Post-Processed Document Metrics:
  - 5788 total tokens analyzed
  - 220 total entities identified
  - 13 total relations extracted

LUVOIR Extracted STG



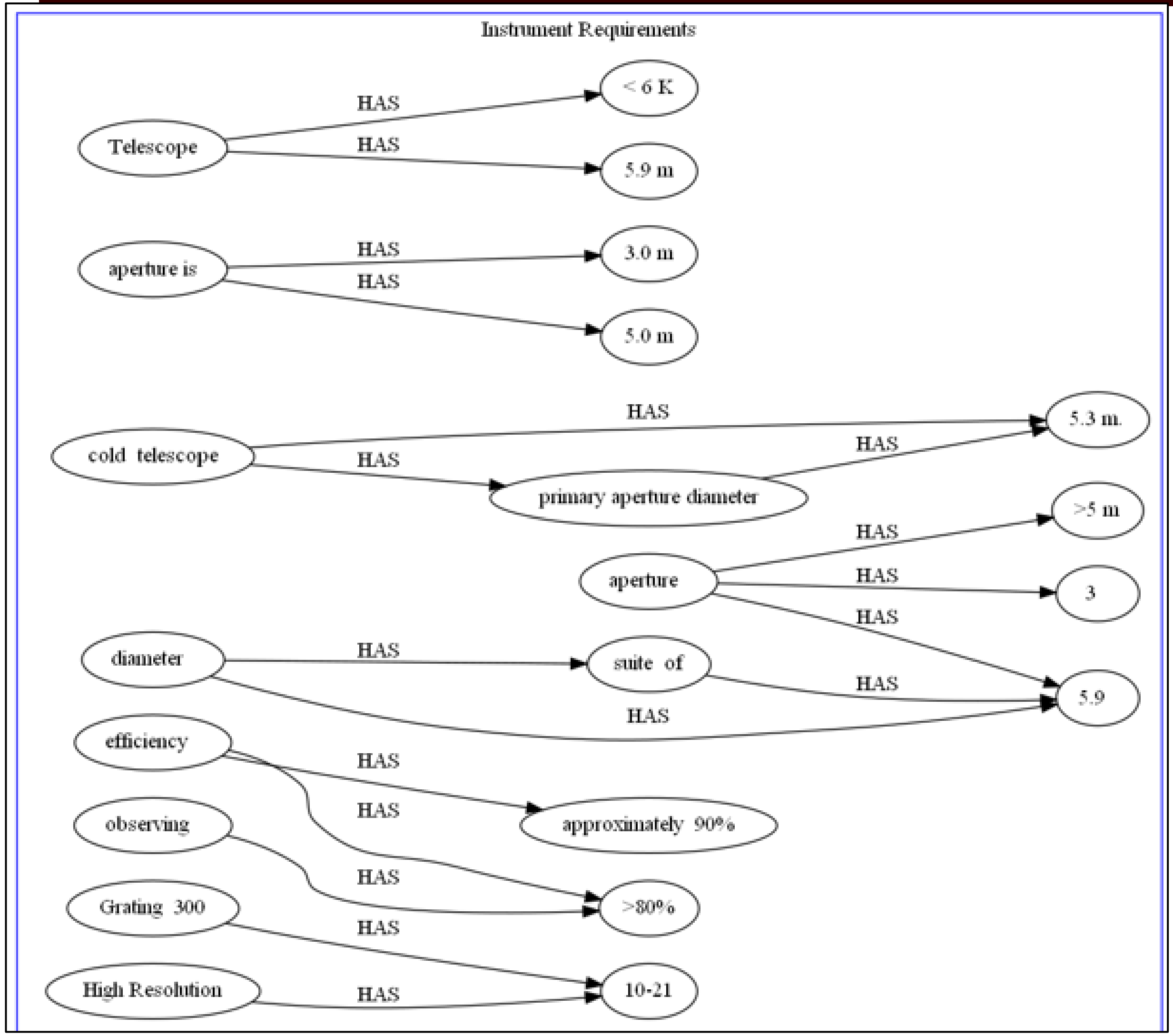
- Some comments:
  - Certain relations were not extracted (e.g. no clarity on traceability from science actions, to observable requirements, to instrumentation)

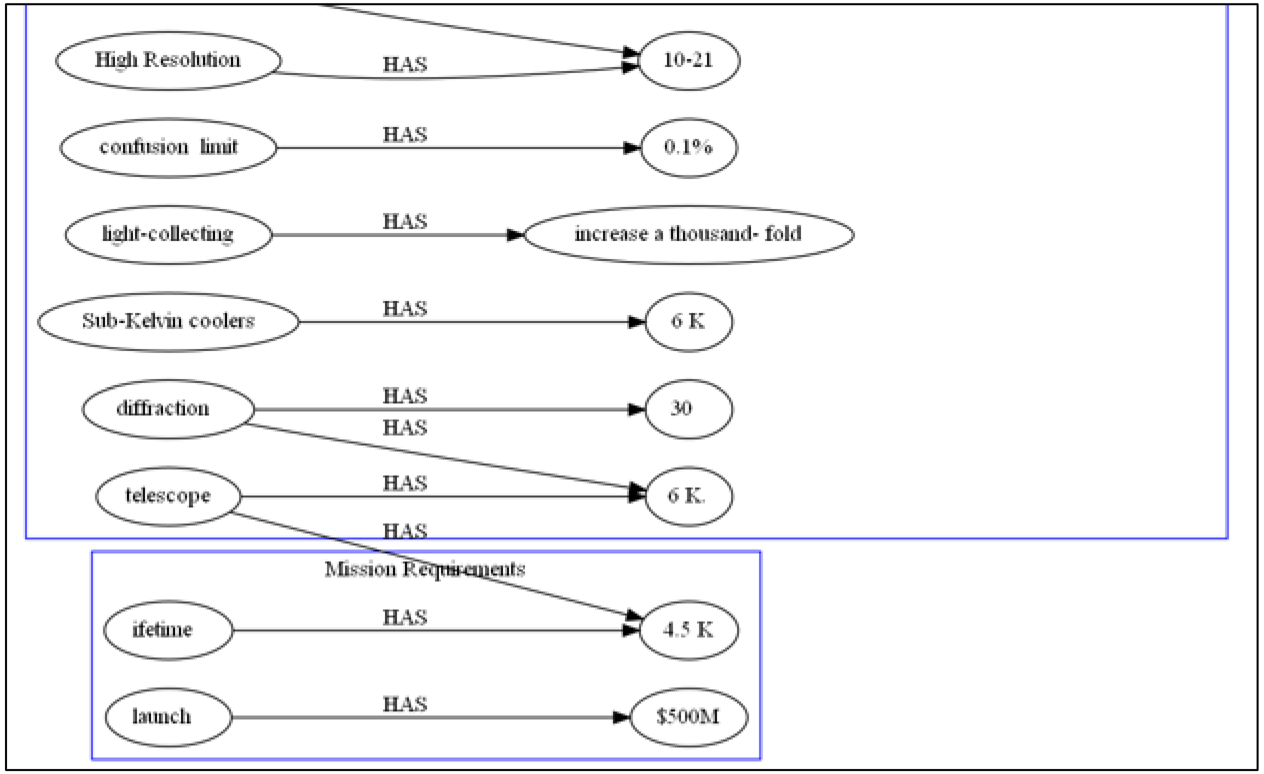
Applied Results: OST

- OST: Origins Space Telescope
  - Large Mission Concept
  - 15 page proposal document
- Post-Processed Document Metrics:
  - 9429 tokens analyzed
  - 531 total entities identified
  - 35 total relations extracted

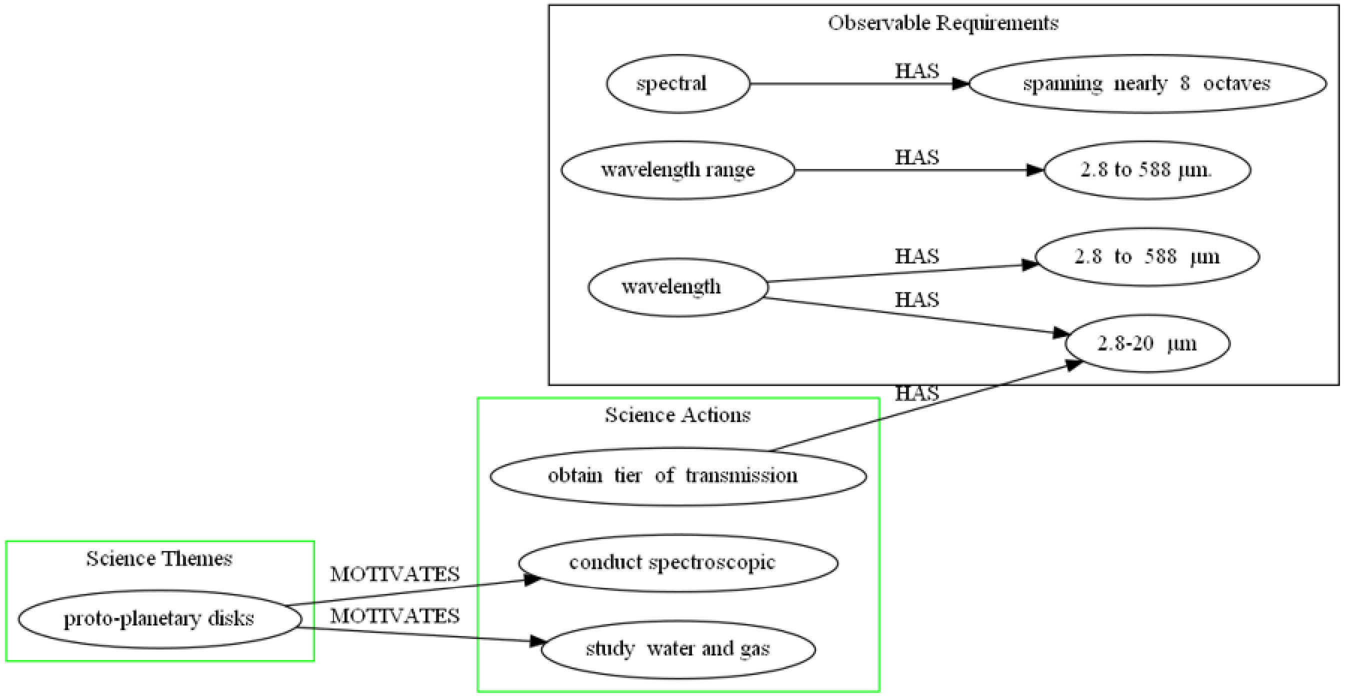
OST Extracted STG

TEXAS A&M UNIVERSITY Engineering





OST Extracted STG cont.



- More spurious information
- Very little 'scientific' information extracted
- A few instances of 'incorrect' relations
  - i.e. not compliant with predefined ontology

Discussion

- Valid instances of extracted relations and valid entities show promise however, more work is needed
- Limitations on training data size affect quality of outputs
  - NER is more complicated a task than RE
    - More label types, more variations of entities within an entity type
- Graph visualization limited to items with sufficient relation confidence scores
  - I.e. relations with low confidence are filtered out (50% confidence threshold)
  - Certain ‘actual’ relations could be missed by the model

SUMMARY/ONGOING WORK

Summary

TEXAS A&M UNIVERSITY Engineering

- Area of proposal evaluation could very well leverage support from natural language processing
  - Semantic technology provides a very viable use case for this area
- We created a pipeline for the generation of science traceability graphs within the astrophysics space mission domain
- We have demonstrated, to a low degree of fidelity, the ability of extracting a traceability structure from raw documentation

Application to Other Domains

TEXAS A&M UNIVERSITY Engineering

- Whilst our tool is not immediately applicable to other domains (e.g. Earth Science, DoD, Human Spaceflight), the methods represented can be applied elsewhere
- What you will need:
  - An ontology of sorts to guide your annotation and training processes
    - Talk with experts … get a good framework established early and follow it strictly
    - Be mindful of ambiguities that could arise
  - Plenty of training examples:
    - Most of the training data development has been conducted by myself and an undergraduate student
    - If you have access to a rich data pool, and have some extra hands on deck, this need could be met more swiftly

Future Work

- Improvements of performance within the pipeline:
  - Increasing the number of training examples
    - Also updating ontology to better map mission parameters into the graph STM
  - Filtering rules to exclude invalid relations and inappropriate entities
- Assessments on a portfolio level:
  - Multi-mission summaries
  - Determine science relevance to the Astrophysics Decadal Survey Decadal for >1 missions
- This work was selected for publication at the 2022 AIAA SciTech Conference:
  - Title: "*Extracting Science Traceability Graphs from Mission Concept Abstracts using Natural Language Processing*"

Acknowledgements

This work was supported by a NASA Space Technology Graduate Research Opportunities Award (Grant No. 80NSSC20K1226).

References

[1] – NASA, 2016, *Systems Engineering Handbook,* Revision 1, Washington, DC, USA: National Aeronautics and Space Administration (NASA), NASA/SP-2016-6105 Rev2.

[2] – J. R. Weiss, W. D. Smythe and Wenwen Lu, "Science traceability," *2005 IEEE Aerospace Conference*, 2005, pp. 292-299

[3] – L. Jones-Wilson, S. Susca and K. Reinholtz, "Project-domain Science Traceability and Alignment Framework (P-STAF): Analysis of a payload architecture," *2018 IEEE Aerospace Conference*, 2018, pp. 1-16

[4] – National Research Council. 2010. New Worlds, New Horizons in Astronomy and Astrophysics. Washington, DC: The National Academies

References

[5] – Beltagy, Iz et al. “SciBERT: Pretrained Contextualized Embeddings for Scientific Text.” ArXiv abs/1903.10676 (2019): n. pag.

[6] – Chinchor, N. Beth Sundheim, B. 1993. “MUC-5 evaluation metrics”. *In Proceedings of the 5th conference on Message understanding (MUC5 '93). Association for Computational Linguistics*, USA, 69–78.

BACKUP CHARTS

Updated STM Ontology