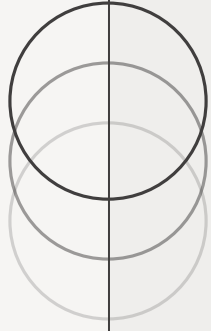


INTRODUCTION TO
KNOWLEDGE ENGINEERING

CS RESEARCH KNOWLEDGE GRAPH

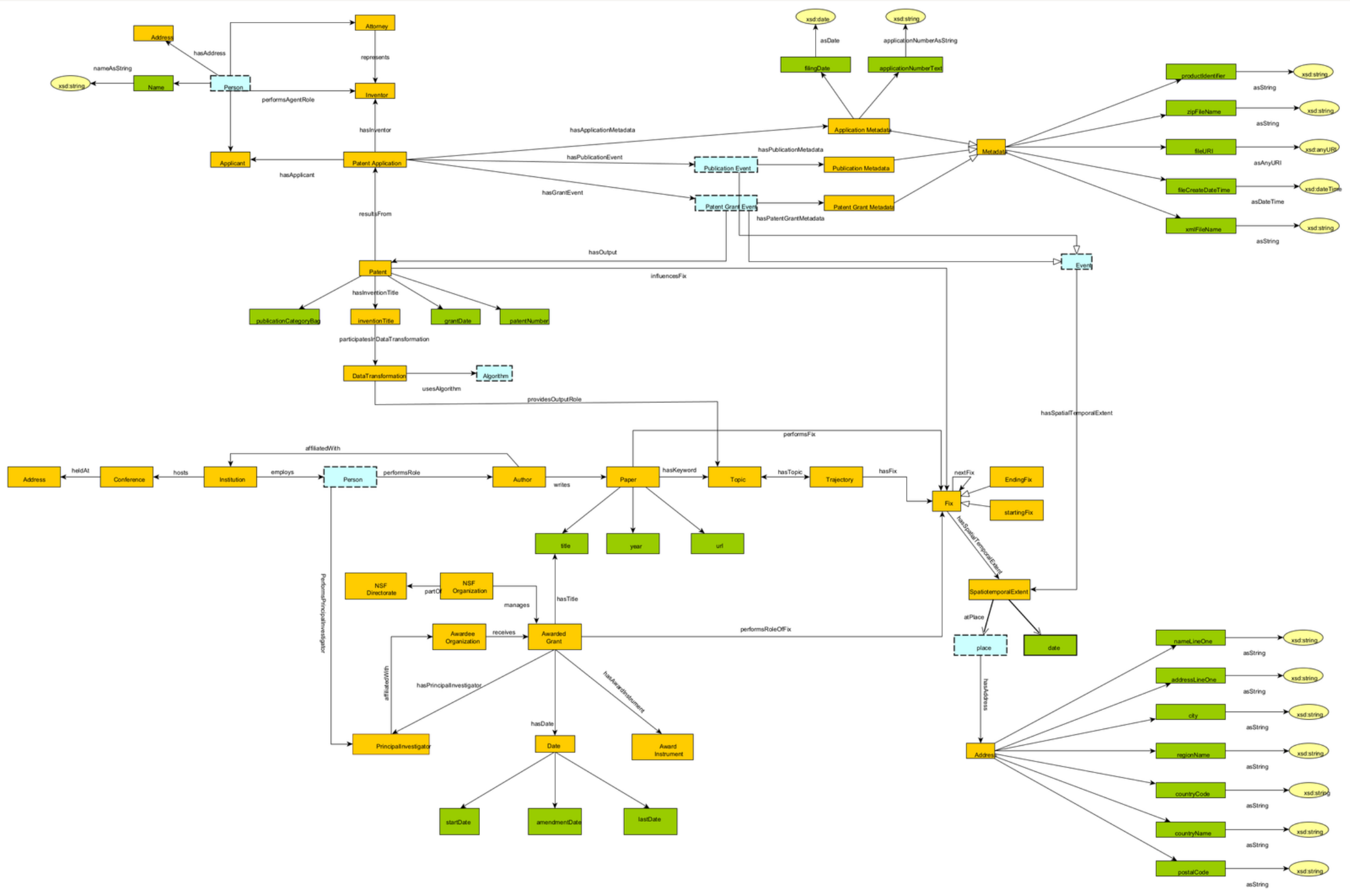
JULIA GRACE M
MOSES RAJ M
SUBRITI
SUMANTH

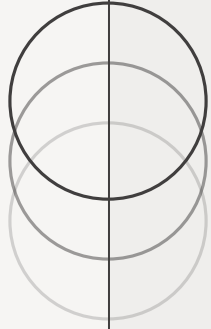


Introduction

- Research in CS generates thousands of papers, patents, and grants every year.
- All of these are stored in different datasets: DBLP, NSF, USPTO.
- **Challenge:** lack of unified structure for analyzing CS research evolution.
- **Goal:** Create a semantic knowledge graph connecting funding, publications, innovation in the field of CS

Schema Diagram





Classes and Relations

- **Core classes:**

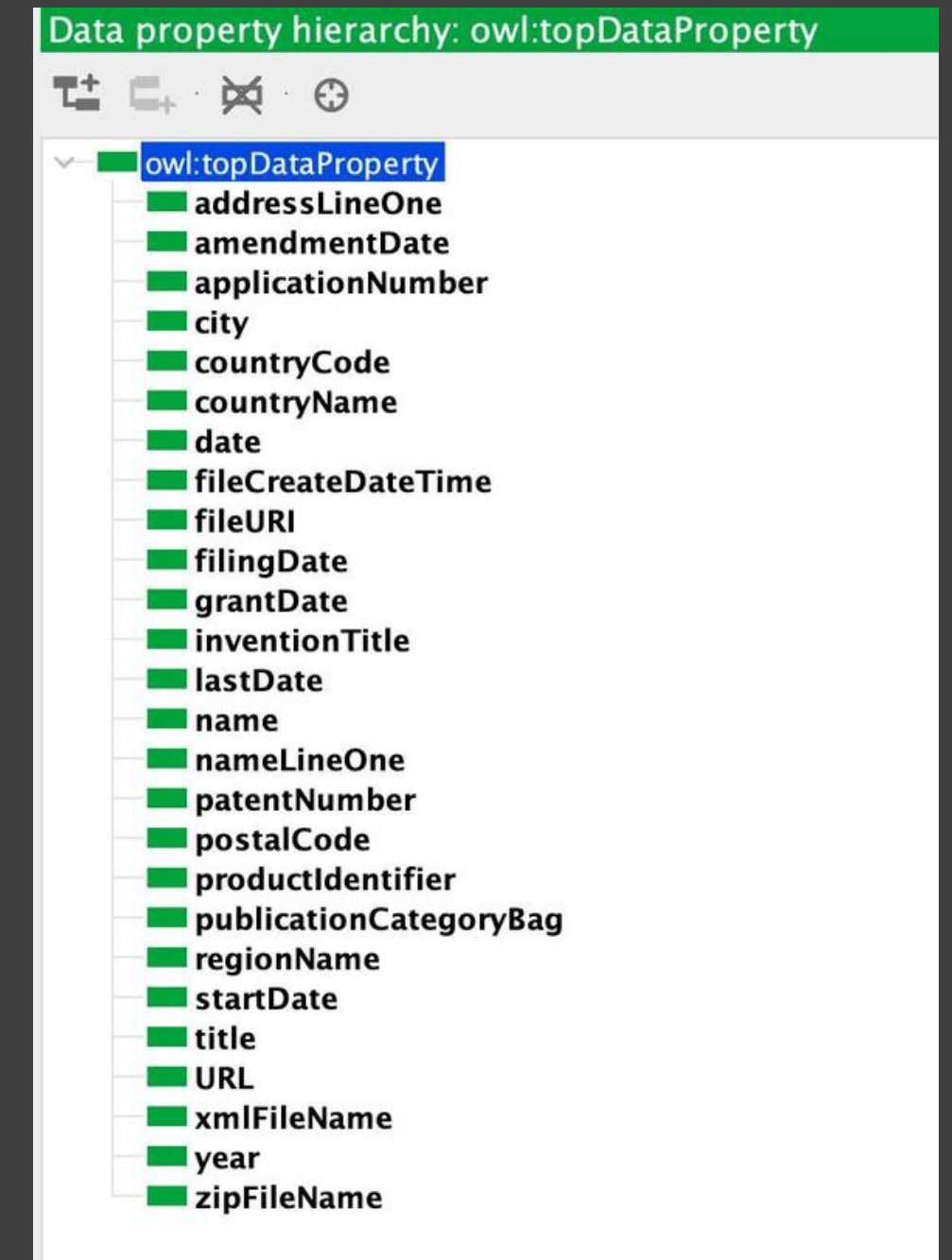
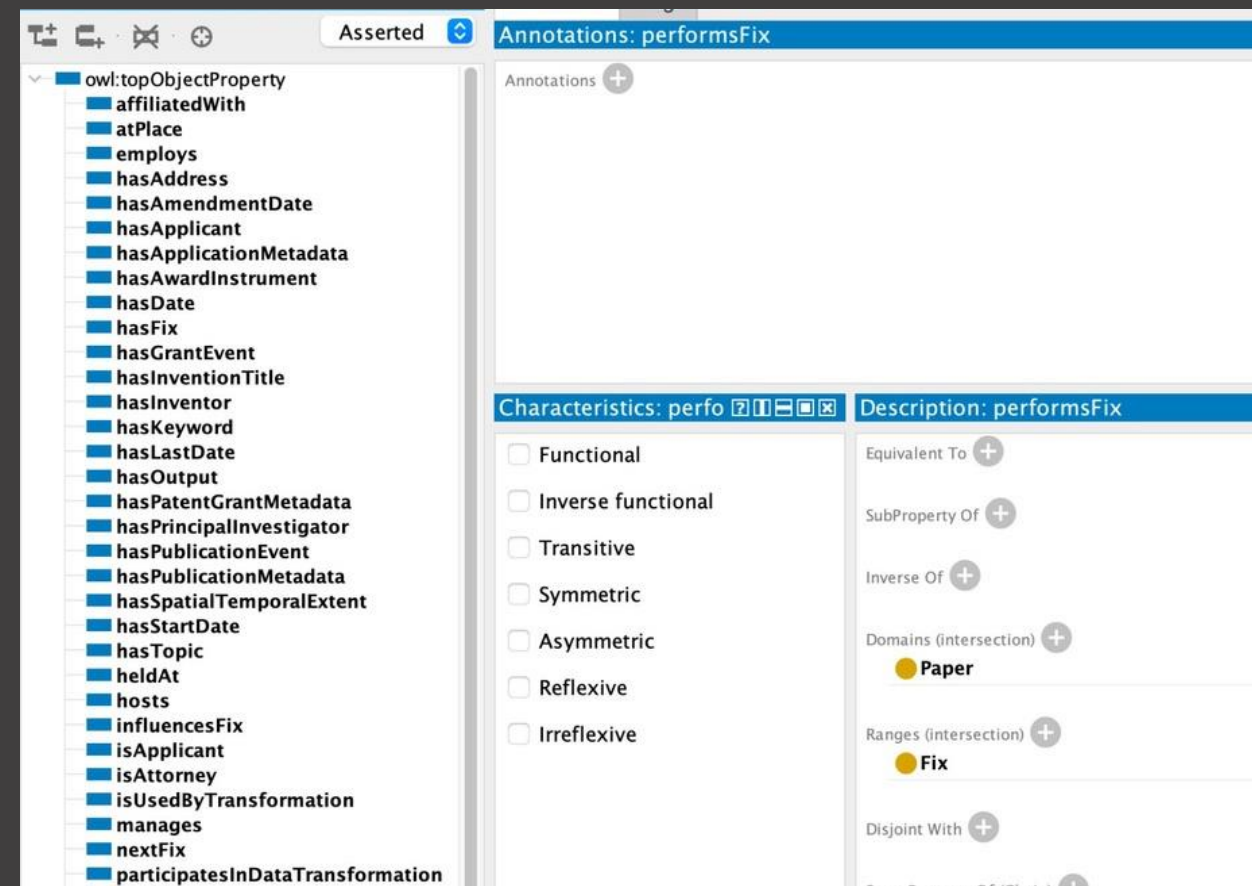
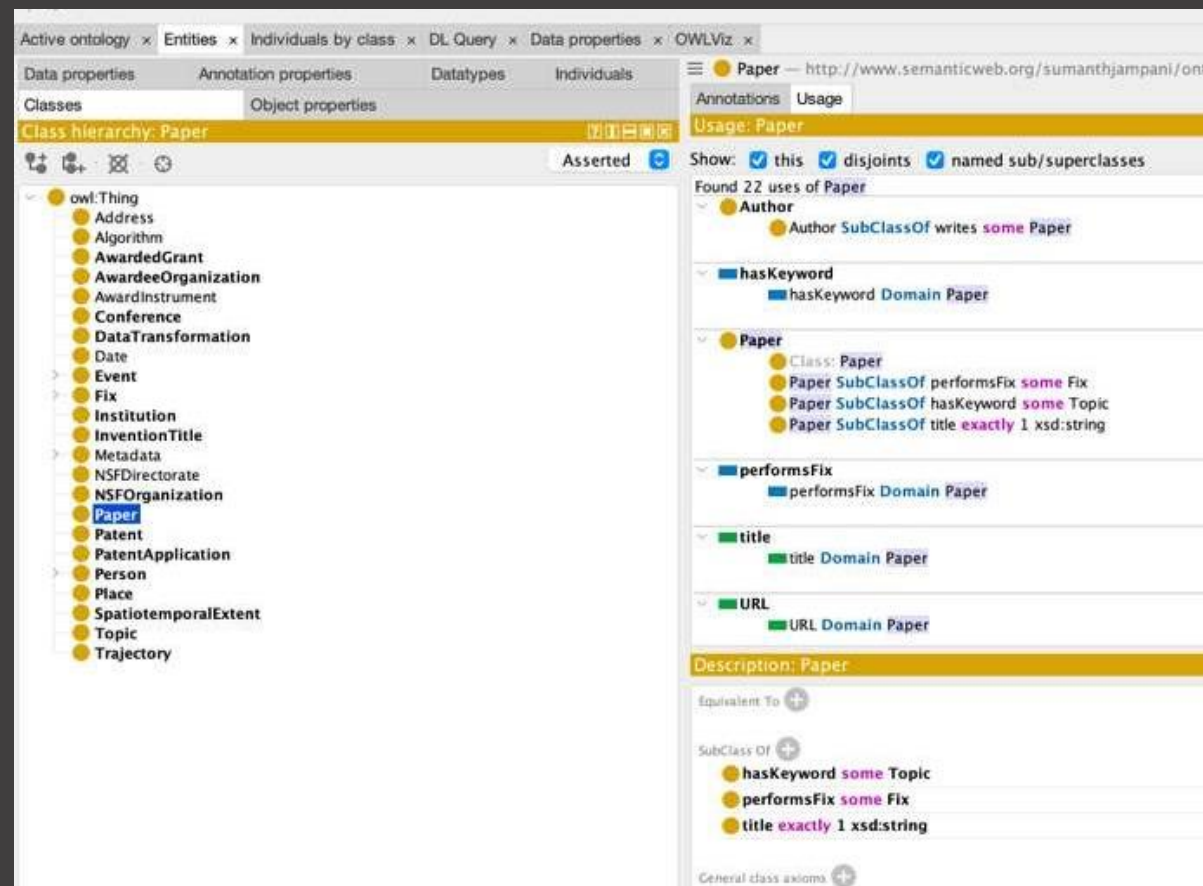
Paper, Patent, Grant, Author, Institution, Topic, Trajectory, Fix, SpatiotemporalExtent

- **Key relations:**

writes, performs, manages, hasTopic, hasFix, influences, hasGrantEvent, partOfTrajectory, etc.

Axiomatization

Axiomatization is the process of defining the core rules and constraints of a knowledge domain so that all other facts can be logically derived.



Why is Axiomatization important?

Because once these axioms are in place, the reasoner can start doing the heavy lifting.

It can infer new relationships, detect errors, and verify whether our model is logically sound—almost like having a built-in quality check for the ontology.

```
### http://www.semanticweb.org/sumanthjampani/ontologies/2025/10/Project/manages
:manages rdf:type owl:ObjectProperty ;
        rdfs:domain :NSFOrganization ;
        rdfs:range :AwardedGrant .

### http://www.semanticweb.org/sumanthjampani/ontologies/2025/10/Project/nextFix
:nextFix rdf:type owl:ObjectProperty ;
        rdfs:domain :Fix ;
        rdfs:range :Fix .

### http://www.semanticweb.org/sumanthjampani/ontologies/2025/10/Project/partOf
:partOf rdf:type owl:ObjectProperty ;
        rdfs:domain :NSFOrganization ;
        rdfs:range :NSFDirectorate .

### http://www.semanticweb.org/sumanthjampani/ontologies/2025/10/Project/participatesInDataTransformation
:participatesInDataTransformation rdf:type owl:ObjectProperty ;
        rdfs:domain :InventionTitle ;
        rdfs:range :DataTransformation .

### http://www.semanticweb.org/sumanthjampani/ontologies/2025/10/Project/performsAgentRole
:performsAgentRole rdf:type owl:ObjectProperty ;
        rdfs:domain :Person ;
        rdfs:range :Inventor .

### http://www.semanticweb.org/sumanthjampani/ontologies/2025/10/Project/performsFix
:performsFix rdf:type owl:ObjectProperty ;
        rdfs:domain :Paper ;
        rdfs:range :Fix .

### http://www.semanticweb.org/sumanthjampani/ontologies/2025/10/Project/performsRole
:performsRole rdf:type owl:ObjectProperty ;
        rdfs:domain :Person ;
        rdfs:range :Author .

### http://www.semanticweb.org/sumanthjampani/ontologies/2025/10/Project/performsRoleOfFix
:performsRoleOfFix rdf:type owl:ObjectProperty ;
        rdfs:domain :AwardedGrant ;
        rdfs:range :Fix .

### http://www.semanticweb.org/sumanthjampani/ontologies/2025/10/Project/providesOutputRole
:providesOutputRole rdf:type owl:ObjectProperty ;
        rdfs:domain :DataTransformation ;
        rdfs:range :Topic .
```

The screenshot displays a web-based ontology editor interface. The main window shows a class hierarchy for 'Patent' under 'owl:Thing'. The hierarchy includes classes like Address, Algorithm, AwardedGrant, AwardeeOrganization, AwardInstrument, Conference, DataTransformation, Date, Event, Fix, Institution, InventionTitle, Metadata, NSFDirectorate, NSFOrganization, Paper, Patent (highlighted), PatentApplication, Person, Place, SpatiotemporalExtent, Topic, and Trajectory. The 'Annotations' tab is active, showing a list of annotations for 'Patent'. The 'Log' window is open, displaying a series of log messages including 'Saving Workspace and Ontologies', 'Running Reasoner', and 'Pre-computing inferences:'. The log messages indicate that the reasoner is active and has processed the ontology in 41 ms by Hermit.

Project (http://www.semanticweb.org/sumanthjampani/ontologies/2025/10/Project) : [/Users/subriti/Downloads/axiomatizationFinal 11.29.25.ttl]

Patent

Active ontology x Entities x Classes x Object properties x Data properties x Individuals by class x

Class hierarchy: Patent [Annotations] [Usage]

Annotations: Patent

Annotations +

Description: Patent

Equivalent To +

SubClass Of +

- hasInventionTitle exactly 1 InventionTitle
- influencesFix some Fix
- patentNumber exactly 1 xsd:string
- resultsFrom only PatentApplication

General class axioms +

SubClass Of (Anonymous Ancestor)

Instances +

Target for Key +

- patentNumber

Disjoint With +

Disjoint Union Of +

Log

INFO 13:10:50 Saving Workspace and Ontologies

INFO 13:10:50 Saved tab state for 'Data properties' tab

INFO 13:10:50 Saved tab state for 'Active ontology' tab

INFO 13:10:50 Saved tab state for 'Object properties' tab

INFO 13:10:50 Saved tab state for 'Individuals by class' tab

INFO 13:10:50 Saved tab state for 'Entities' tab

INFO 13:10:50 Saved tab state for 'Classes' tab

INFO 13:10:50 Saved workspace

INFO 13:10:56 Saving Workspace and Ontologies

INFO 13:10:56 Saved tab state for 'Data properties' tab

INFO 13:10:56 Saved tab state for 'Active ontology' tab

INFO 13:10:56 Saved tab state for 'Object properties' tab

INFO 13:10:56 Saved tab state for 'Individuals by class' tab

INFO 13:10:56 Saved tab state for 'Entities' tab

INFO 13:10:56 Saved tab state for 'Classes' tab

INFO 13:10:56 Saved workspace

INFO 13:10:58 Running Reasoner

INFO 13:10:58 Pre-computing inferences:

- class hierarchy
- object property hierarchy
- data property hierarchy
- class assertions
- object property assertions
- same individuals

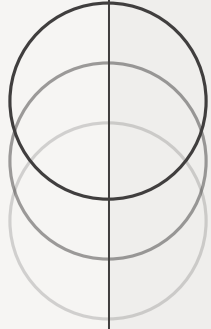
INFO 13:10:58 Ontologies processed in 41 ms by Hermit

INFO 13:10:58

Show log file Preferences Time stamp Clear log

OK

Reasoner active Show Inferences



Data Transformation

Common Entities

- **Authors**

DBLP → authors of papers

Patents → inventors

NSF → Principal Investigators (PI/Co-PI)

- **Research Topics / Keywords**

DBLP → extracted from paper titles/keywords

Patents → extracted from abstracts/claims

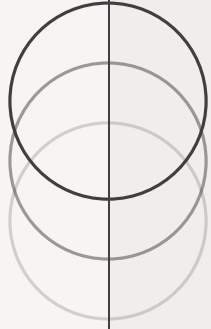
NSF → topics from award abstracts/keywords

- **Temporal Information**

DBLP → publication year

NSF → award year & duration

Patents → filing date, grant date



Data Transformation

Data Cleaning

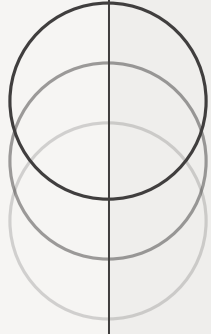
- Convert all dates to a single format
- Standardize author/institution names
 - Normalize topics/keywords

Topic Extraction

- TF-IDF keyword extraction
- Contextual embeddings (BERT)
 - Topic modeling

Clustering to Find Topic Groups

- Group similar papers, patents, and NSF awards together
 - Creates unified topic clusters



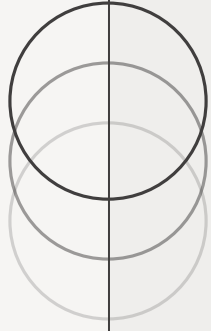
Competency Questions

1Q. Which institutions have received the most NSF grants related to AI research?

Ans) Grants are identified through the `hasKeyword` relation connecting a Paper to a Topic.

Each Grant is then linked to an Institution via the `awardedTo` relation.

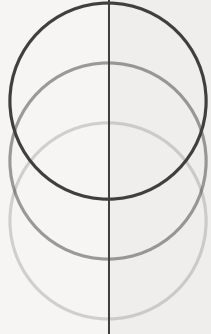
Counting the number of grants for a particular Topic gives the answer of how much funding is going into one Research Topic.



2Q. How has CS topics evolved over time?

Ans) Each Topic has a Trajectory, made up of time-ordered Fixes, and each Fix corresponds to a Paper with a publication date.

By aggregating these fixes along the trajectory, the ontology provides a timeline showing how a CS topic has evolved over time.

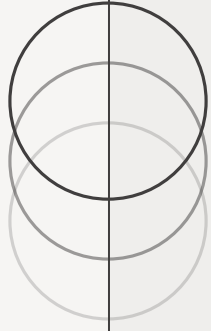


3Q. Which NSF projects produced the highest innovation velocity?

Ans) The number of papers and patents that are being released in a particular Topic after a grant is issued is the innovation velocity

Using the awardDate of the grant and the hasDate values of resulting papers/patents, the ontology measures how quickly new outputs appear.

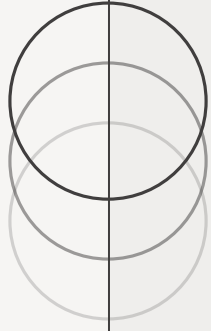
This allows computation of innovation velocity.



4Q. Which authors frequently publish in both academic venues (conferences/journals) and also file patents?

Ans) The schema links an Author to Paper and Inventors to their Patents.

Person who appear in both sets—those who have writes → Paper (Venue) and files → Patent (hasInventor) —are identified as publishing academically while also generating patents.

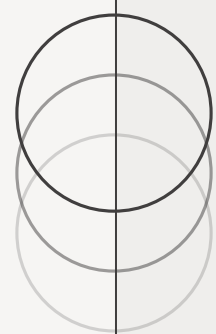


5Q. Which NSF-funded areas have produced patents most rapidly (shortest time lag)?

Ans) Each Grant is linked to one or more Topics and to resulting Patents.

By comparing the awardDate of a grant with the hasDate of patents connected to it, the ontology can compute the shortest time lag.

This shows which NSF-funded areas generate patents the fastest.



THANKYOU