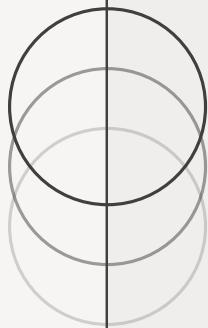


INTRODUCTION TO
KNOWLEDGE ENGINEERING

CS RESEARCH KNOWLEDGE GRAPH

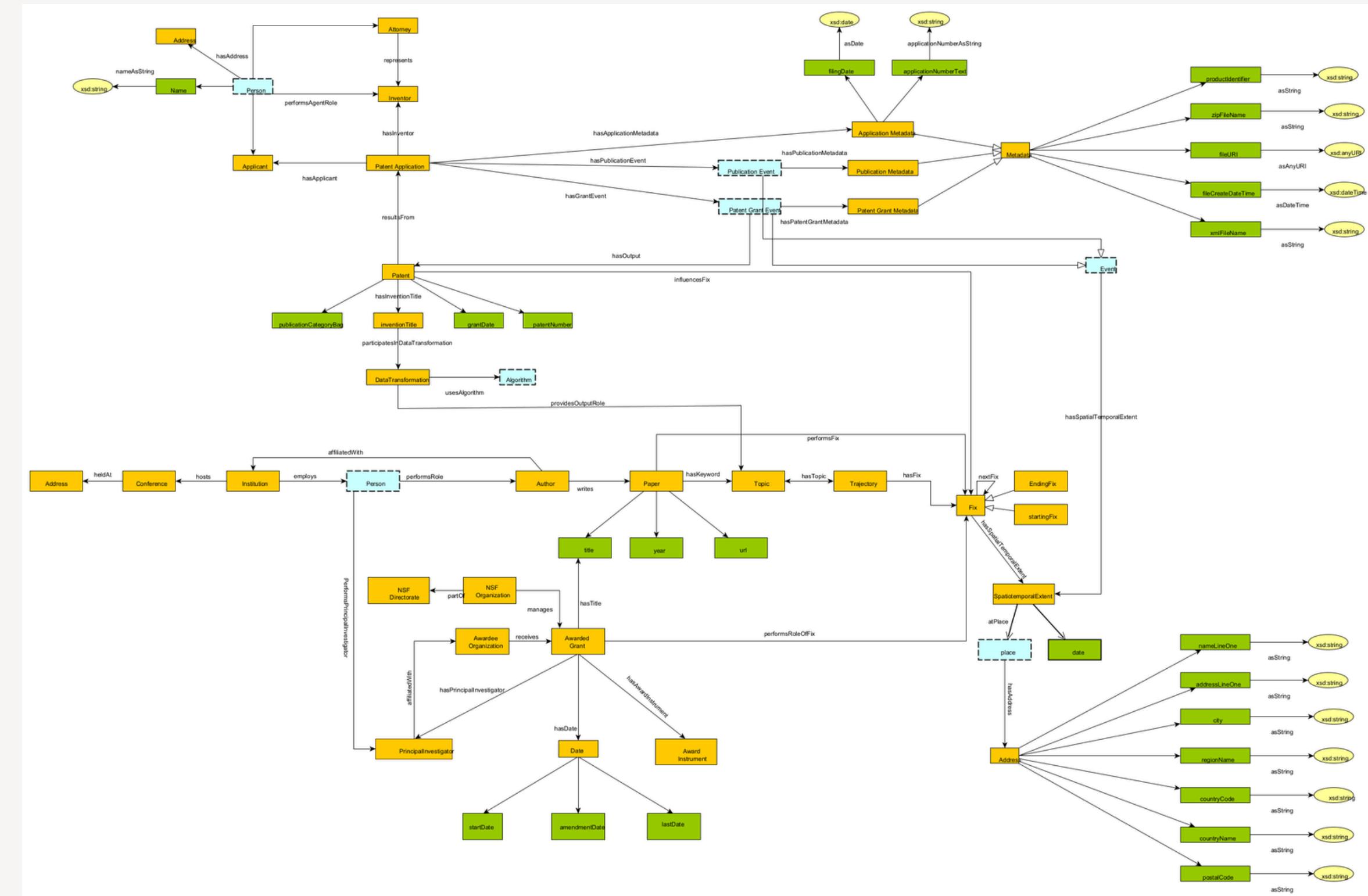
JULIA GRACE M
MOSES RAJ M
SUBRITI
SUMANTH

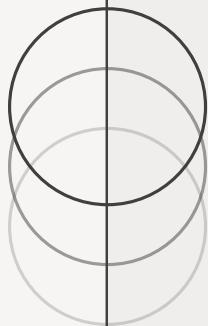


Introduction

- 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.

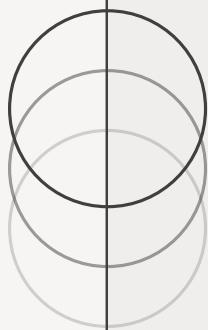
Axiomatization

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

This screenshot shows the OWLViz interface with the 'Class hierarchy: Paper' tab selected. The left sidebar lists various ontology classes. The main pane displays the inheritance chain for the 'Paper' class, showing it inherits from 'owl:Thing' and numerous other classes like 'Address', 'Algorithm', 'AwardedGrant', etc. A search bar at the top right allows for filtering results.

This screenshot shows the OWLViz interface with the 'Annotations: performsFix' tab selected. It lists various annotations for the 'performsFix' property, such as 'owl:topObjectProperty', 'owl:topDataProperty', and several specific annotations involving 'Paper' and 'Fix' classes. On the right side, there are tabs for 'Characteristics', 'Description', and 'General axioms'.

This screenshot shows the OWLViz interface with the 'Data property hierarchy: owl:topDataProperty' tab selected. It displays a hierarchical tree of data properties under 'owl:topDataProperty'. The tree includes properties like 'addressLineOne', 'amendmentDate', 'applicationNumber', 'city', 'countryCode', 'countryName', 'date', 'fileCreateDateTime', 'fileURI', 'filingDate', 'grantDate', 'inventionTitle', 'lastDate', 'name', 'nameLineOne', 'patentNumber', 'postalCode', 'productIdentifier', 'publicationCategoryBag', 'regionName', 'startDate', 'title', 'URL', 'xmlFileName', 'year', and 'zipFileName'. Each property is represented by a green square icon.



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

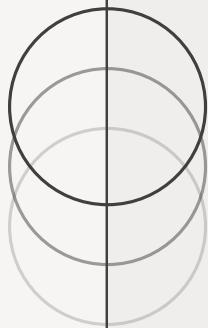
NSF → topics from award abstracts/keywords

- **Temporal Information**

DBLP → publication year

NSF → award year & duration

Patents → filing date, grant date



Data Transformation

Data Cleaning

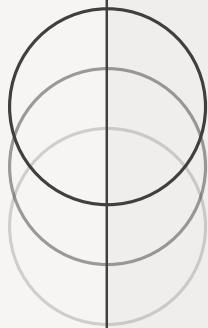
- Convert all dates to a single format
- Standardize author/institution names
- Normalize topics/keywords

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



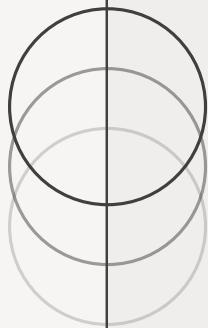
Competency Questions

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.

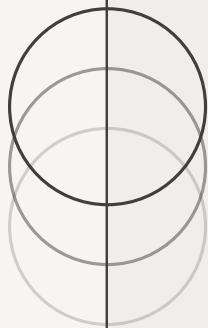
Counting the number of grants for a particular Topic gives the answer of how much funding is going into one Research Topic.



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.

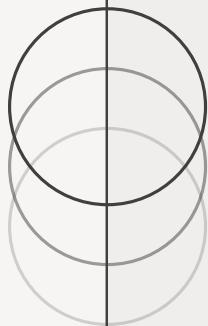


3Q. Which NSF projects produced the highest innovation velocity?

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.

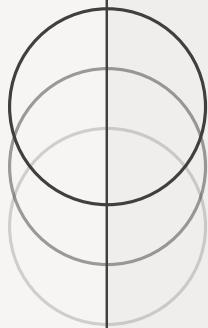
This allows computation of innovation velocity.



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