Mathematical Theories as Abstractions for Ontology Learning

Michael Grüninger

Semantic Technologies Lab
University of Toronto

July 11, 2025

Answering natural language questions about datasets

```
Alice lectures MATH456
MATH456 is a ScienceCourse
Bob teaches MIE123
MIE123 is an EngineeringCourse
```

- *Is Alice a professor in the Math Department?*
- *Is Bob enrolled in any courses?*

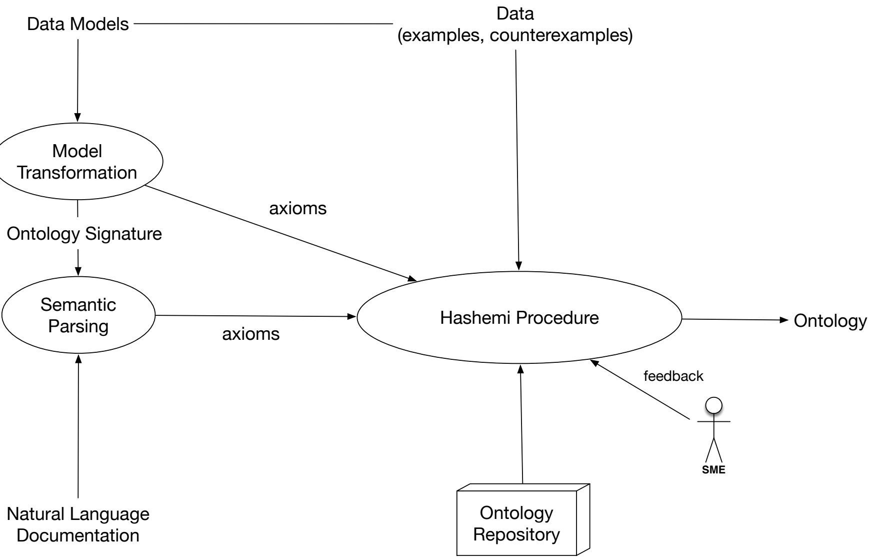
Ontologies

- What do “teaches”, “lectures”, and “professor” mean?
- How do we represent this meaning and use it to answer questions?

Challenge

- Rather than (manually) design the ontology first and then associate it with data, can we (semiautomatically) design the ontology from the data itself?

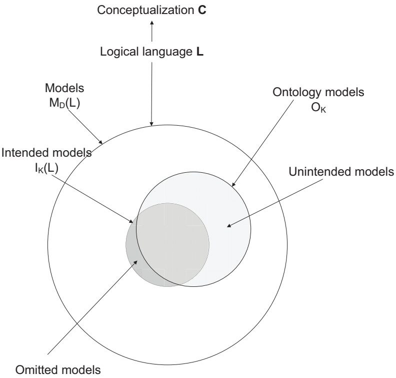
SEADOO Architecture



Intended Models

- Ontologies are designed with respect to a set of semantic requirements.
- There are various ways to specify requirements, such as competency questions and use cases.
- An ontology can be considered to be a set of logical theories whose purpose is to capture the intended interpretations corresponding to a certain conceptualization and to exclude the unintended interpretations.

Ontology Design



Datasets as Intended Semantics

- Examples
  - Clean datasets without any errors should be considered to be consistent with the ontology, and hence they can be considered to be a specification of intended models for the ontology.
- Counterexamples
  - Datasets with quality problems should be inconsistent with the ontology. They can be considered to be a specification of interpretations that falsify one or more axioms of the ontology.

Hashemi Procedure: Discovery Phase

- Given a set of examples (intended models) and counterexamples (falsifying interpretations), search through the ontology repository to find the best match – an ontology whose models match all of the examples and none of the counterexamples.

Hashemi Procedure: Dialogue Phase

- Once a potential match has been found, generate a set of models and falsifying interpretations for the user.
- If the user agrees that the models are correct, then we conjecture that we have found the right ontology.
- If the user disagrees, then we refine the search.
- If we cannot find an exact match, then an ontology designer needs manually modify the best match to construct the ontology.

COLORE

- The COLORE (Common Logic Ontology Repository) project is building an open repository of ontologies specified using Common Logic (ISO 24707).
  - Testbed for ontology evaluation and integration techniques, and that can support the design, evaluation, and application of ontologies in first-order logic.

Seek and You Will Find?

- Since we are designing new ontologies, it will not be the case that the ontology we seek already exists, so how can the Hashemi Procedure possibly work?
- The answer can be found in ontology verification ...

Ontology Verification

- With verification, we want to characterize the models of an ontology up to isomorphism and determine whether or not these models are equivalent to the intended models.
- Relationships between first-order ontologies within a repository can be used to support ontology verification.
- The fundamental insight is that we can use the relationships between ontologies to assist us in the characterization of the models of the ontologies.
- The objective is the construction of the models of one ontology from the models of another ontology by exploiting the relationships between these ontologies and their modules in the repository.

Structures Méres

Bourbaki proposed the following idea:

> *At the centre of our universe are found the great types of structures – orderings, algebra, topology; they might be called the mother-structures.*
>
> *Beyond this first nucleus, appear the structures which might be called multiple structures. They involve two or more of the great mother-structures ... combined organically by one or more axioms which set up a connection between them.*

Structures for Binary Relations

- Orderings
- Graphs
- Successor
- Incidence

Structures for Ternary Relations

- Betweenness
- Cyclic Betweenness
- Magmas
- Multigraphs
- Quivers
- Order Bundles
- Graph Bundles
- Incidence Bundles

Amalgamations of Structures

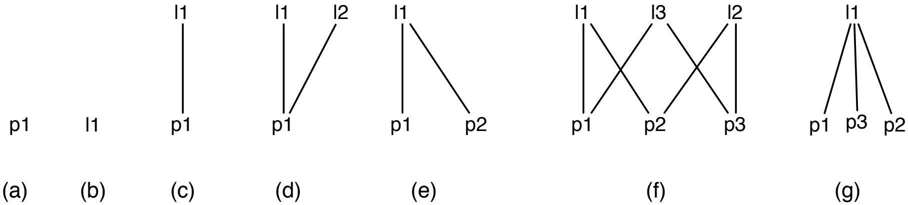
- Mereographs
- Subposets
- Subgraphs
- Ordered Geometry
- Mereological Geometry
- Multigeometries

COLORE Hypothesis

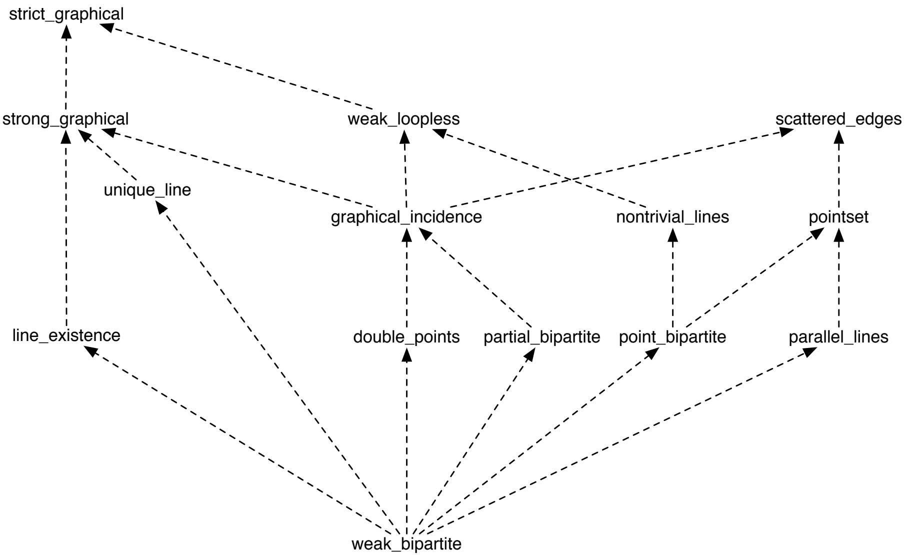
- *Every consistent first-order theory used in practice is logically synonymous to the combination of mathematical theories that serve as ontology patterns in COLORE.*

Bipartite Incidence Structures

- Bipartite incidence structures generalize geometry
- Points and lines are disjoint sets of elements
- There is an incidence relation between points and lines



Bipartite Incidence Theories



Translation Definitions

- In practice, users present their examples and counterexamples using the signature of their domain
  e.g. Course, Professor, teaches
- We use translation definitions to map the domain signature to the signatures of the mathematical theories.

$$(\forall x)\ Course(x) \equiv point(x) \tag{1}$$

$$(\forall x)\ Professor(x) \equiv line(x) \tag{2}$$

$$(\forall x, y)\ teaches(x, y) \equiv in(x, y) \wedge line(x) \wedge point(y) \tag{3}$$

Use Case 2: Datasets

- Examples:

```
Professor(Alice).
Course(MATH456)
teaches(Alice,MATH456).
```

Use Case 2: Demo

- One of the models that the system generates in the Dialogue Phase is:

```
line(0), line(1), point(2),
 in(0,0),  in(0,2),  in(1,1), in(1,2),
 in(2,0), in(2,1), in(2,2)
```

which the user classifies as an example.

Use Case 2: Solution

The best matching theory is the combination of

`colore.oor.net/bipartite_incidence/partial_bipartite.clif`
`colore.oor.net/bipartite_incidence/point_bipartite.clif`
`colore.oor.net/bipartite_incidence/strong_graphical.clif`
`colore.oor.net/bipartite_incidence/parallel_lines.clif`

Discussion

- This is not a reductionist approach (in which ontologies simply are mathematical theories), nor is it a foundationalist approach (in which mathematical theories play the role of foundational ontologies); we are simply acknowledging the role that mathematical structures play in the semantics of ontologies.
- Any consistent first-order theory has a model, and in the classical Tarskian semantics of first-order logic, a model is specified by a set (the universe of discourse) together with a set of relations that are sets of n-tuples of elements of the universe of discourse.
- Insofar as we impose conditions on membership in these relations, models of first-order axiomatizations are isomorphic to mathematical structures.

Summary

- We use a repository of existing mathematical theories as ontology patterns.
- We can semi-automatically design new ontologies through an interactive procedure by which a user provides datasets as a way of specifying intended and unintended models.
- If the ontology that we are designing does not correspond to any theory in the repository, then we have effectively identified a new theory which has never been identified within the mathematical literature i.e. we have uncovered new mathematics.