DATA+AI SUMMIT 2022

The Semantics of Biology

Vaccine and Drug Research with Knowledge Graphs and Logical Inferencing on Apache Spark

John Hunter

Senior Product Director, Ai Operations

ORGANIZED BY databricks

Last year’s talk

Querying Bio Knowledge Graph with Spark + Sparql

This year’s talk

Querying Bio Knowledge Graph with Spark + Sparql + **Logical inferencing**

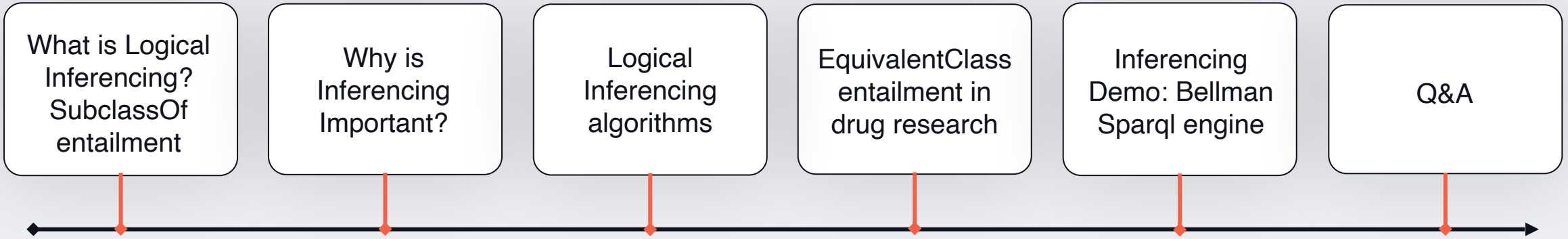
Takeaway

Takeaway

There is no such thing as a….

FISH

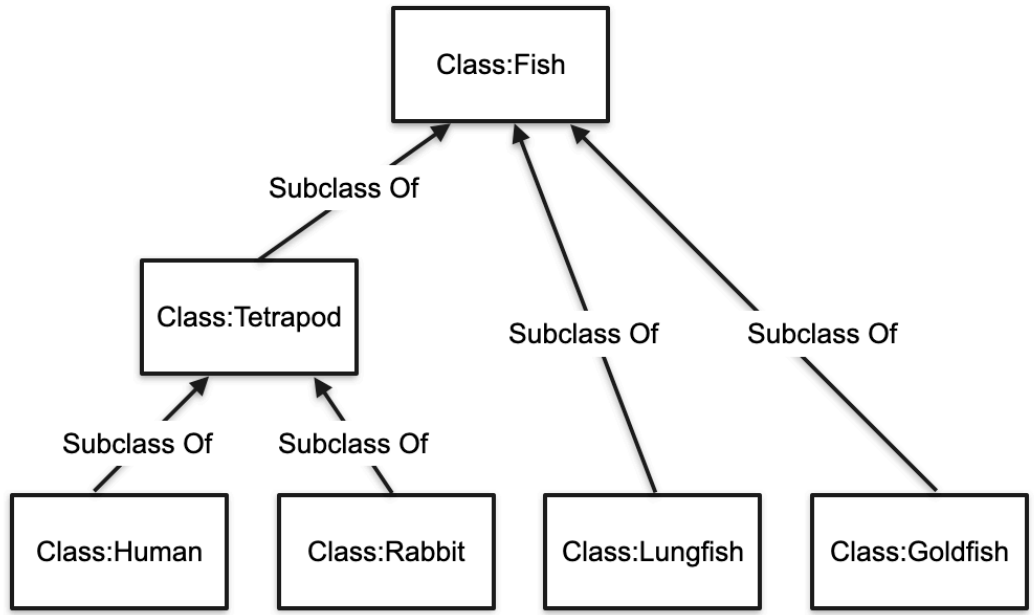
Agenda


What is Logical Inferencing? SubclassOf entailment
Why is Inferencing Important?
Logical Inferencing algorithms
EquivalentClass entailment in drug research
Inferencing Demo: Bellman Sparql engine
Q&A

What is Logical Inferencing?

Schema + Instances = Knowledge Graph

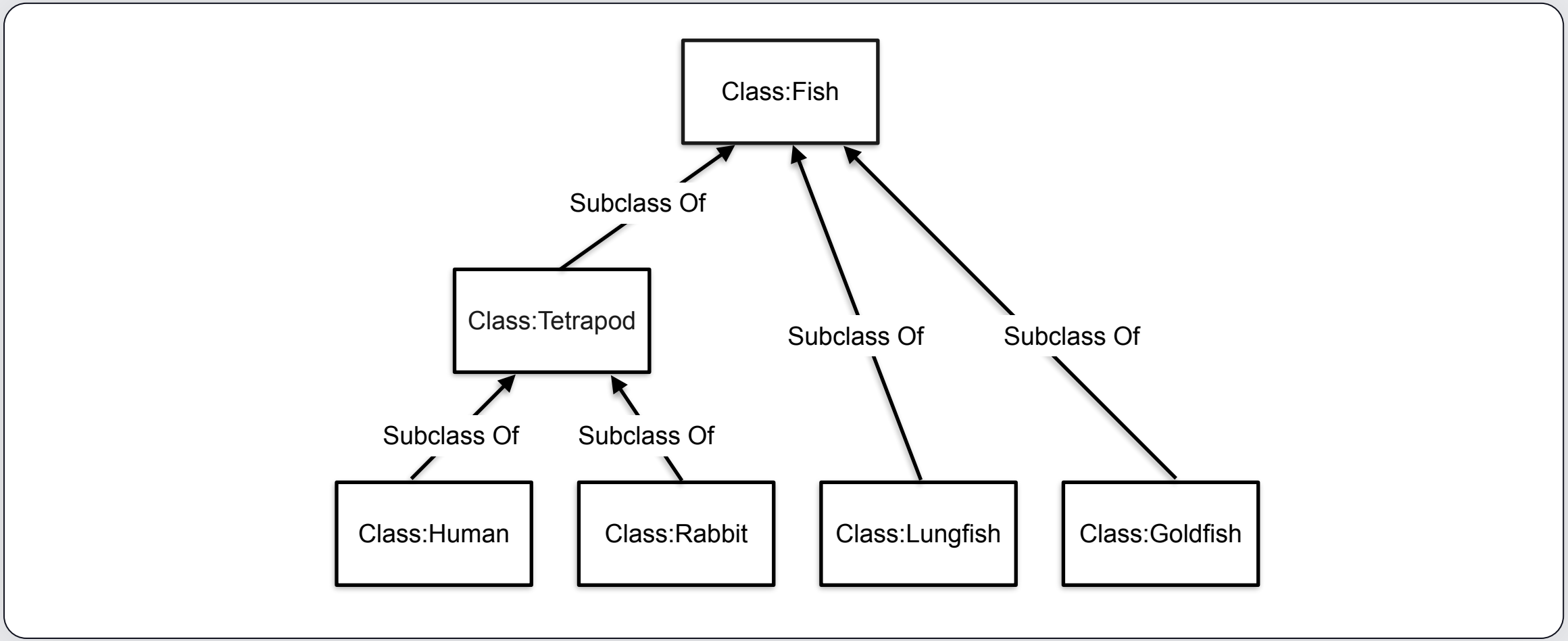
The Schema



The Instances

	Subject	Predicate	Object
1	Bubbles	Type	Goldfish
2	Slippy	Type	Lungfish
3	Alice	Type	Human
4	Bob	Type	Rabbit

The Schema: Terminological knowledge of a domain



The Instances: Assertional knowledge of a domain.

	Subject	Predicate	Object
1	Bubbles	Type	Goldfish
2	Slippy	Type	Lungfish
3	Alice	Type	Human
4	Bob	Type	Rabbit

Querying **WITHOUT** logical inferencing

Please give me all instances of Tetrapods in the Knowledge Graph

	Subject	Predicate	Object
1	Bubbles	Type	Goldfish
2	Slippy	Type	Lungfish
3	Alice	Type	Human
4	Bob	Type	Rabbit

Querying **WITHOUT** logical inferencing

Results:
{ }

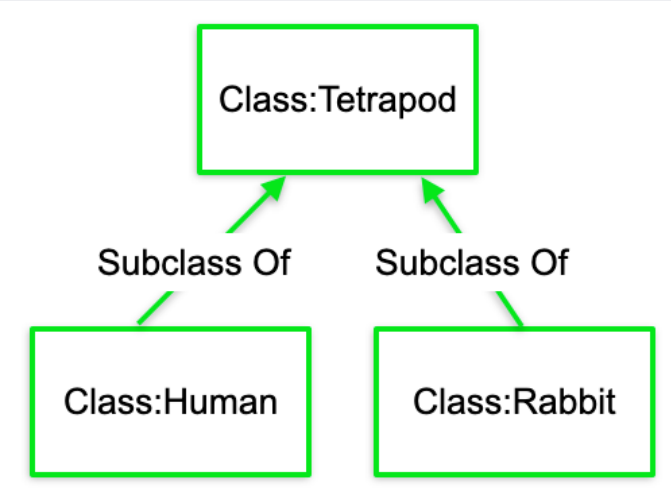
	Subject	Predicate	Object
1	Bubbles	Type	Goldfish
2	Slippy	Type	Lungfish
3	Alice	Type	Human
4	Bob	Type	Rabbit

Querying **WITH** logical inferencing

Please give me all instances of Tetrapods in the Knowledge Graph

	Subject	Predicate	Object
1	Bubbles	Type	Goldfish
2	Slippy	Type	Lungfish
3	Alice	Type	Human
4	Bob	Type	Rabbit

+

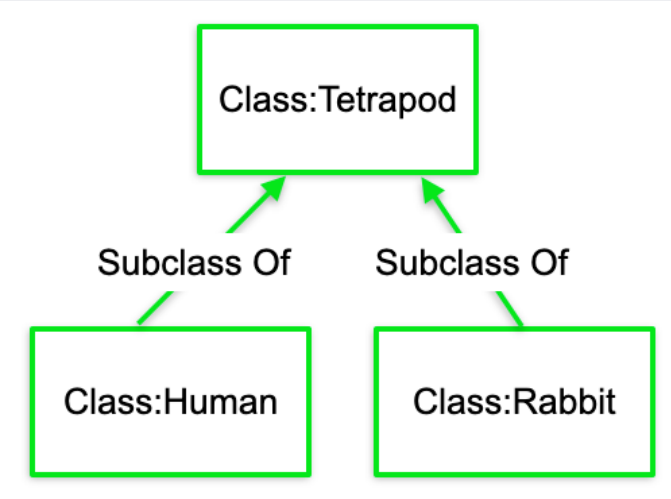


Querying **WITH** logical inferencing

Results:
- Alice
- Bob

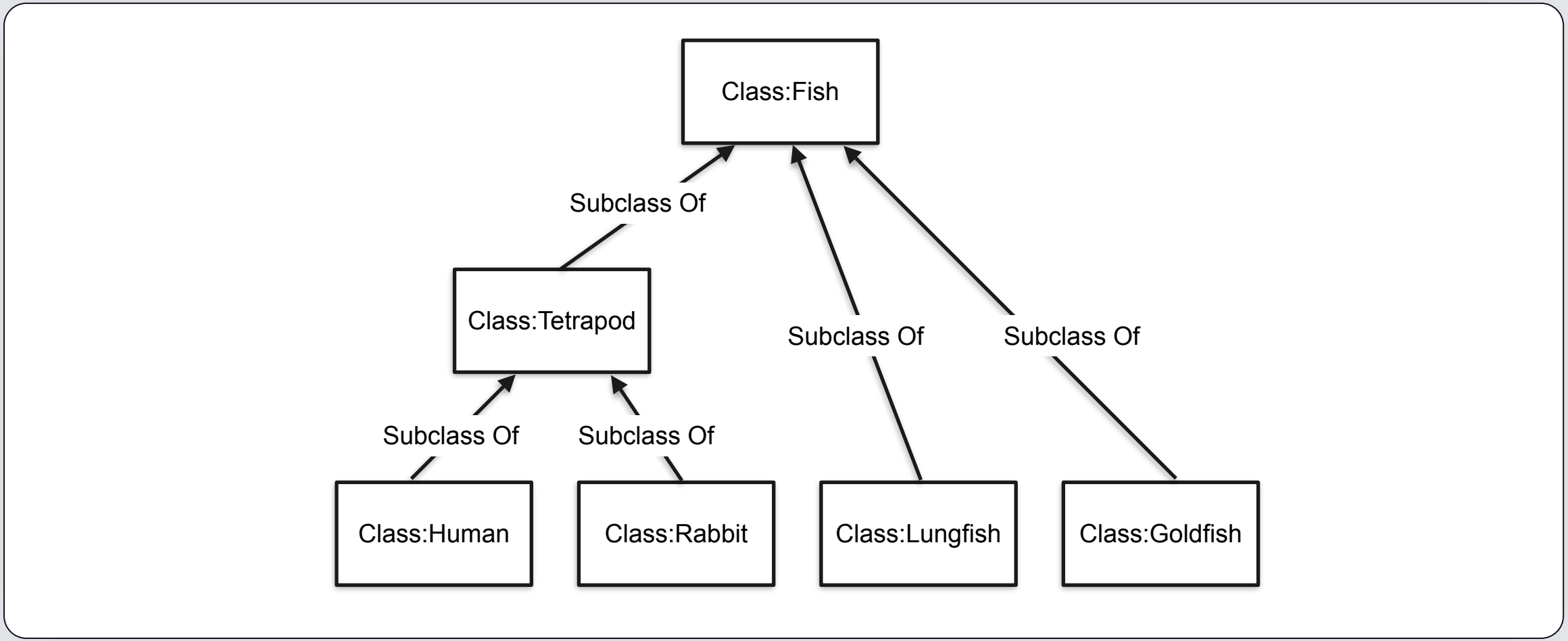
	Subject	Predicate	Object
1	Bubbles	Type	Goldfish
2	Slippy	Type	Lungfish
3	Alice	Type	Human
4	Bob	Type	Rabbit

+



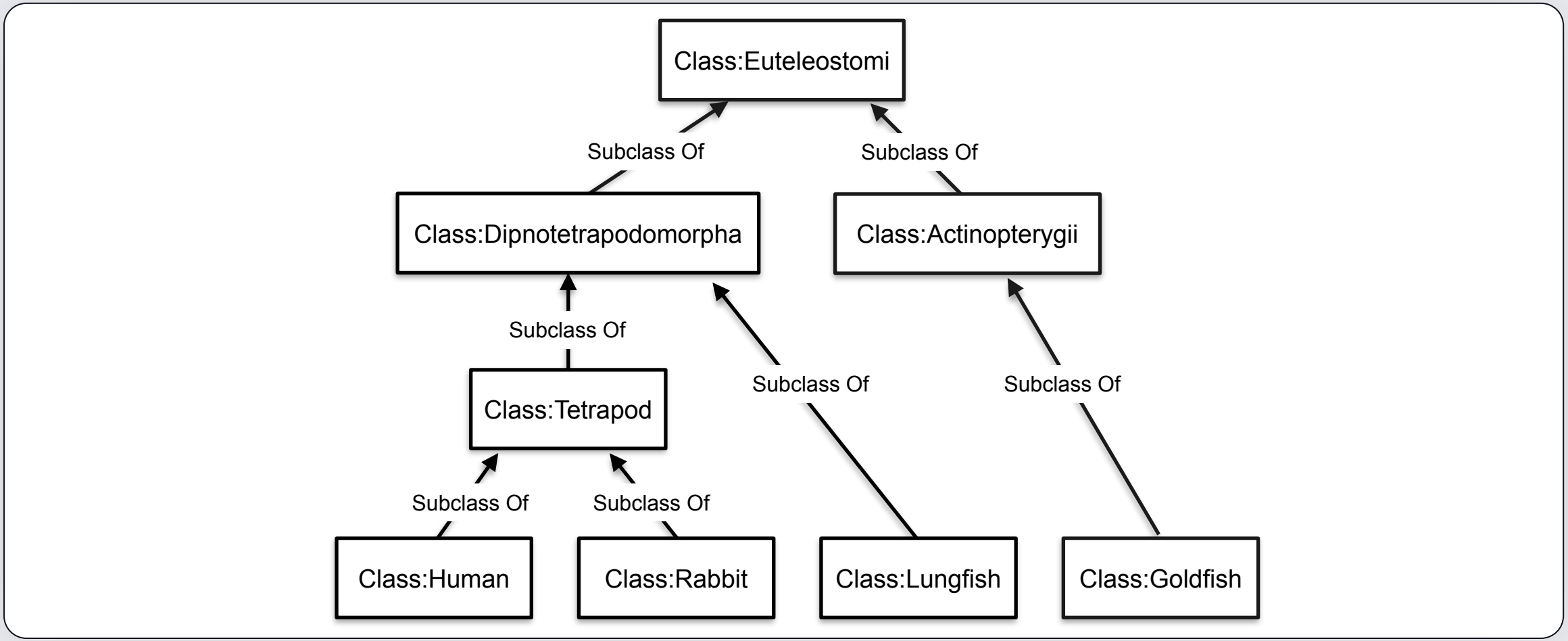
Why is Logical Inferencing important?

The Schema: Terminological knowledge of a domain



No such thing as a fish
- Science

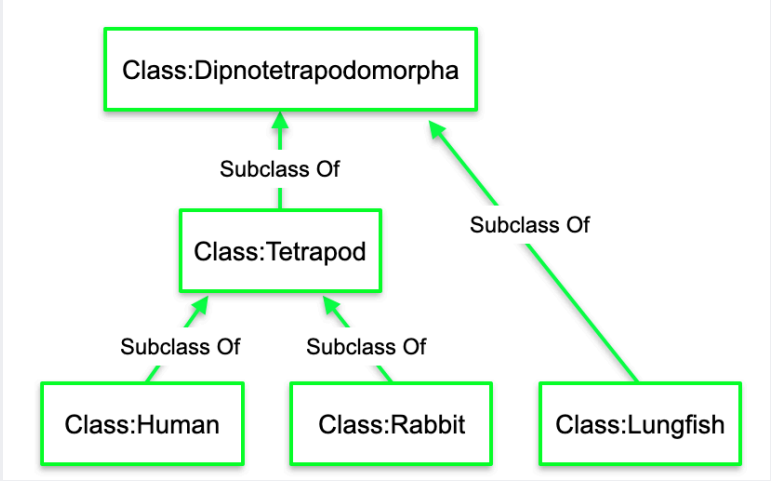
The Schema: Terminological knowledge of a domain



Please give me all instances of ***Dipnotetrapodomorpha*** in the Knowledge Graph

	Subject	Predicate	Object
1	Bubbles	Type	Goldfish
2	Slippy	Type	Lungfish
3	Alice	Type	Human
4	Bob	Type	Rabbit

+

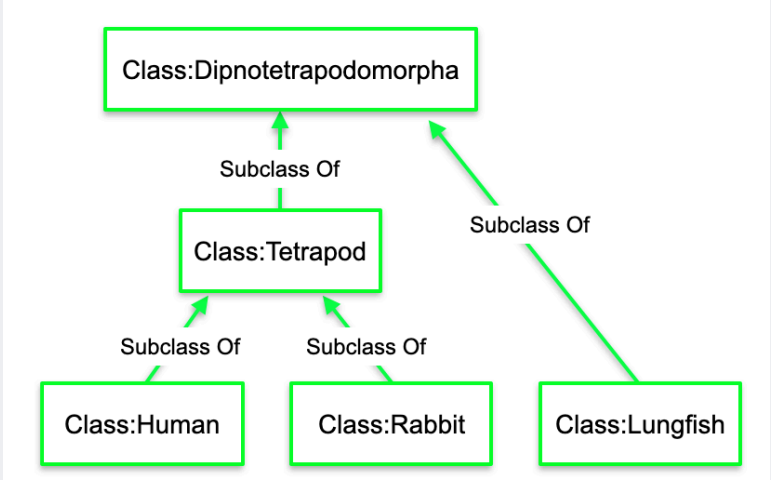


Results:

- Alice
- Bob
- Slippy

	Subject	Predicate	Object
1	Bubbles	Type	Goldfish
2	Slippy	Type	Lungfish
3	Alice	Type	Human
4	Bob	Type	Rabbit

+



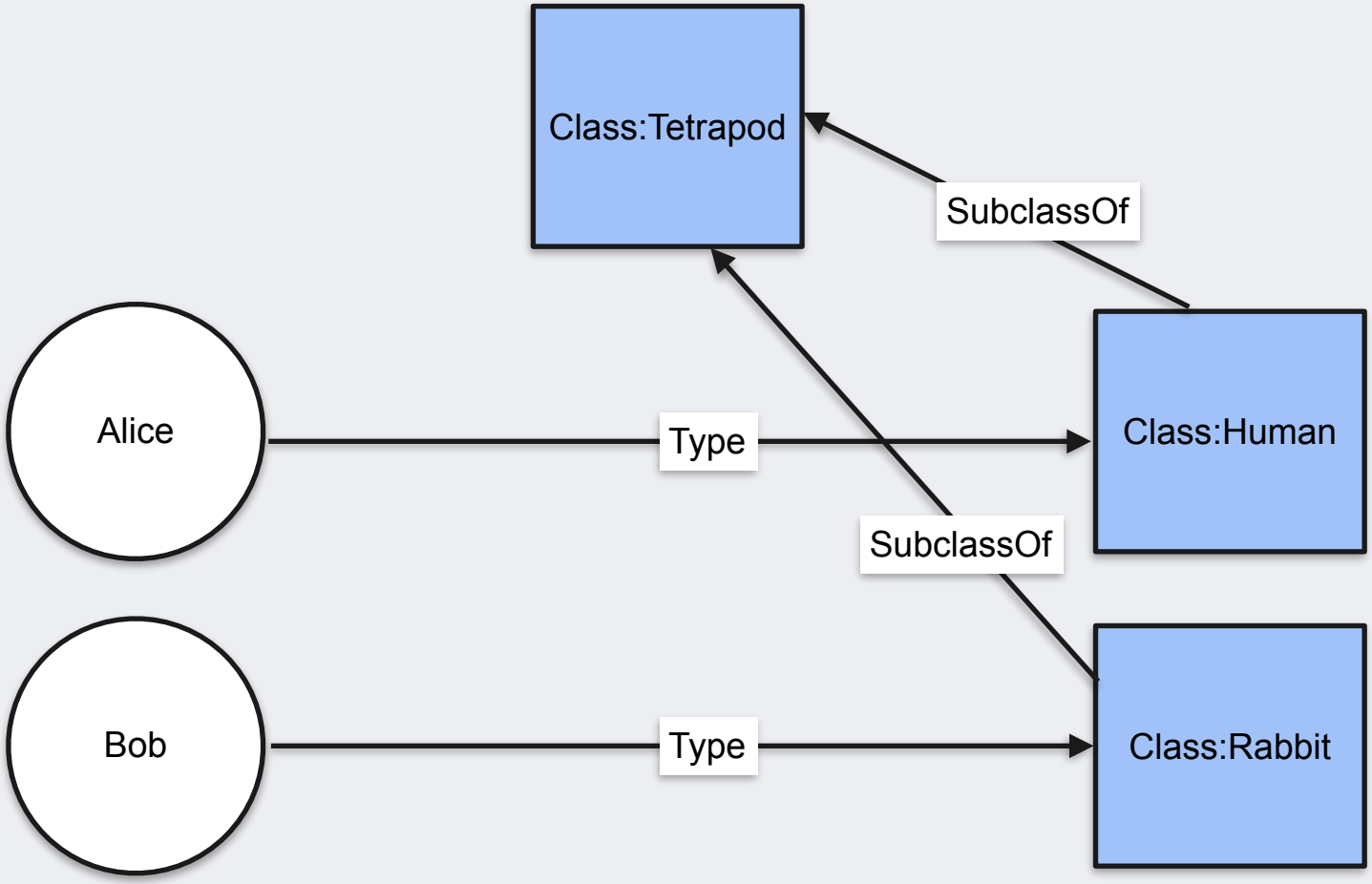
Logical inferencing in code

Forward Chaining Algorithm

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX class: <http://example.org/class/>

SELECT ?tetrapod
WHERE {
 ?tetrapod rdf:type class:Tetrapod
}
```

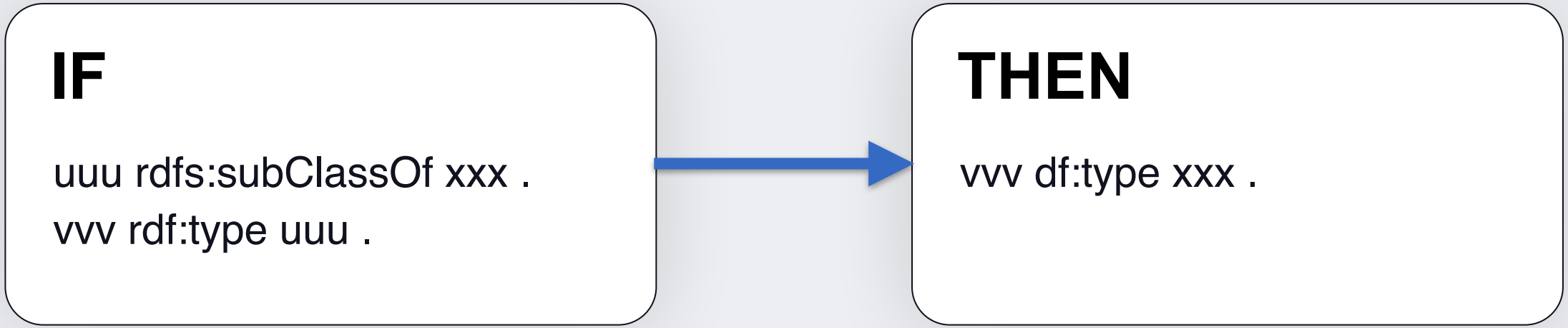
Please give me all instances of Tetrapods in the Knowledge Graph

Forward Chaining Algorithm



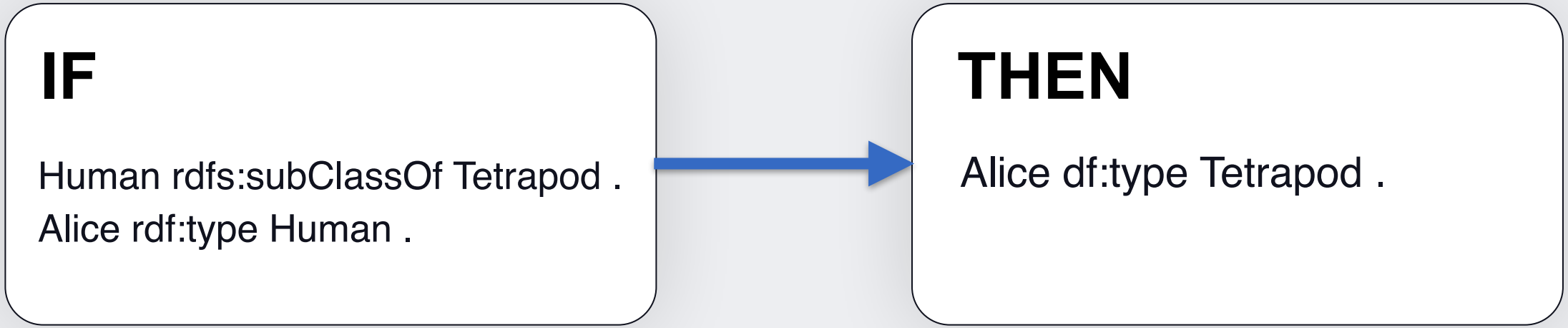
Forward Chaining Algorithm

RDFS9 entailment: SubclassOf



Forward Chaining Algorithm

Modus Ponens Applied



Forward Chaining Algorithm

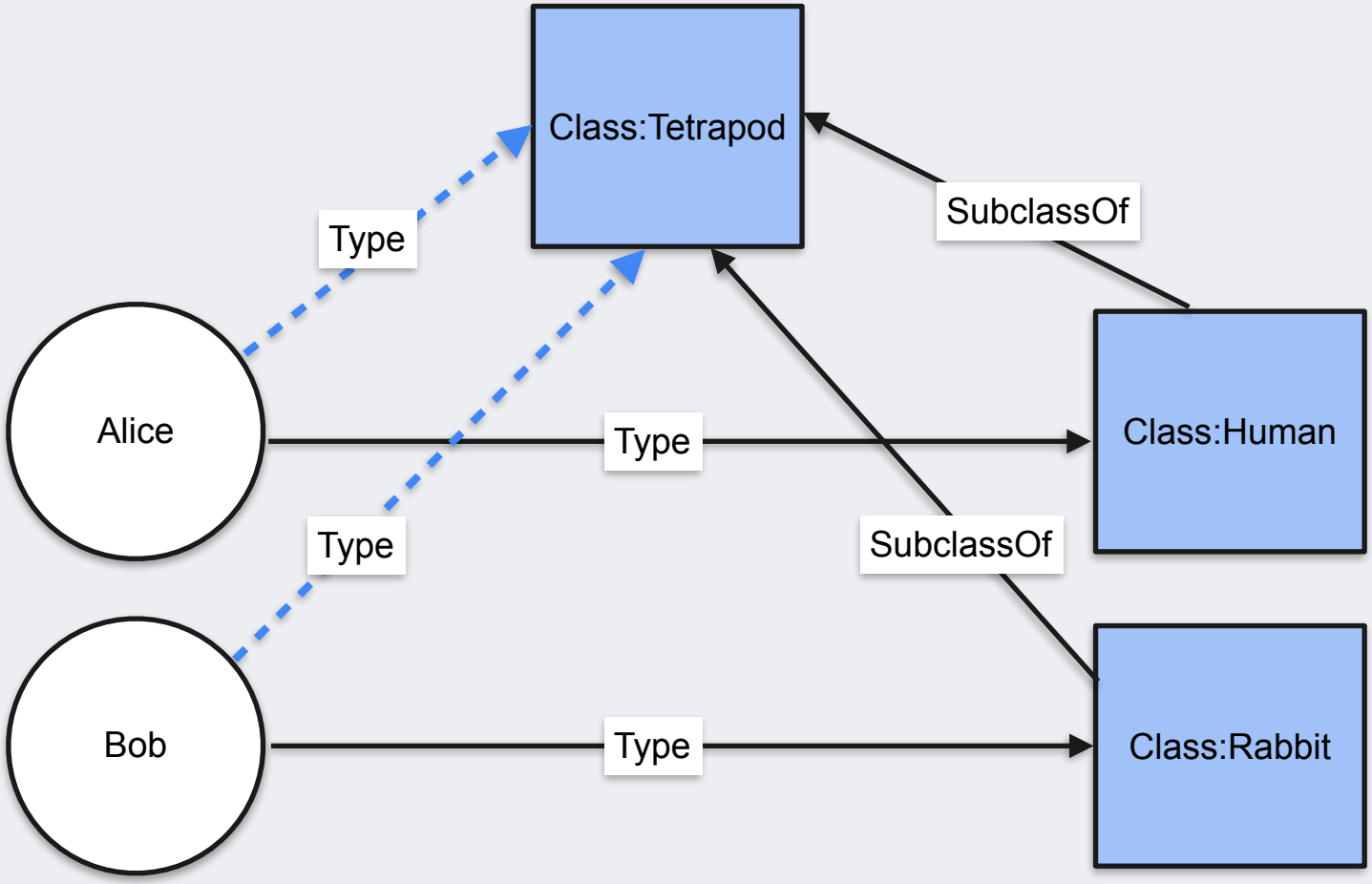
```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX class: <http://example.org/class/>

CONSTRUCT { ?z rdf:type ?y }
WHERE {
 uuu rdfs:subClassOf class:Tetrapod .
 ?tetrapod ref:type ?uuu .
}
```

Materialize inferred triples

Left hand side of the entailment

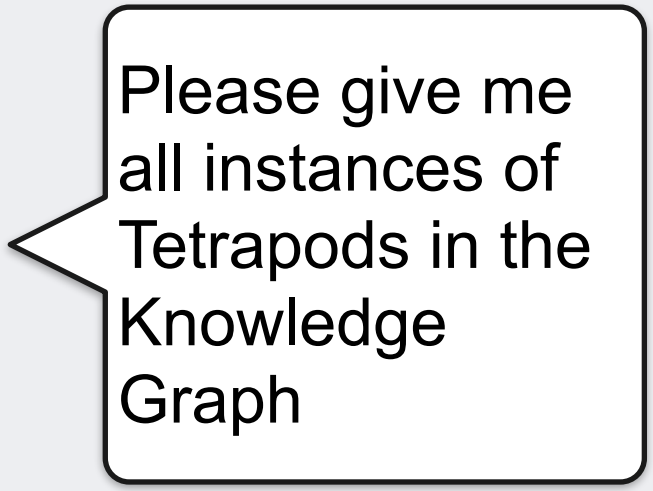
Forward Chaining Algorithm

Runtime connection materialization applied

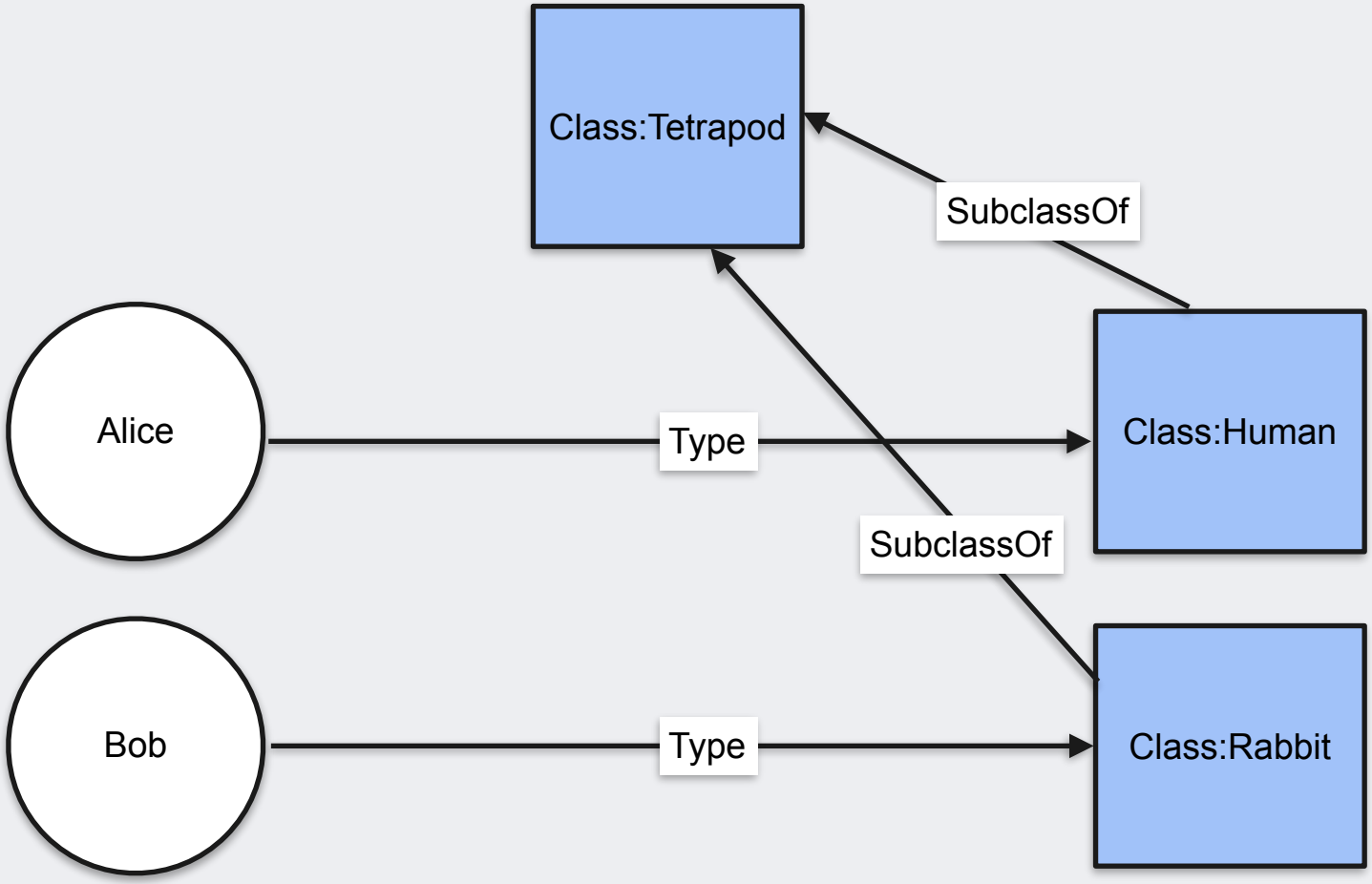


Backward Chaining Algorithm

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX class: <http://example.org/class/>

SELECT ?tetrapod
WHERE {
 ?tetrapod rdf:type class:Tetrapod
}
```

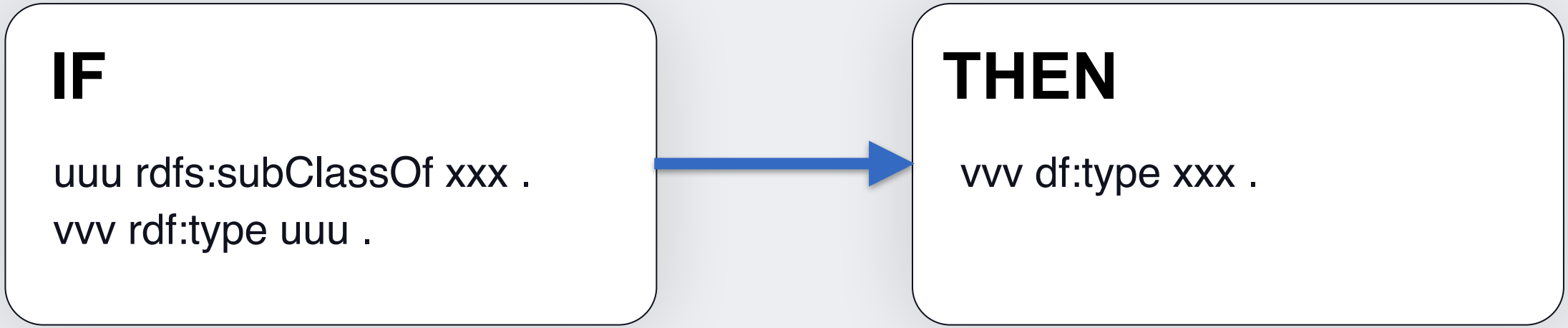


Backward Chaining Algorithm



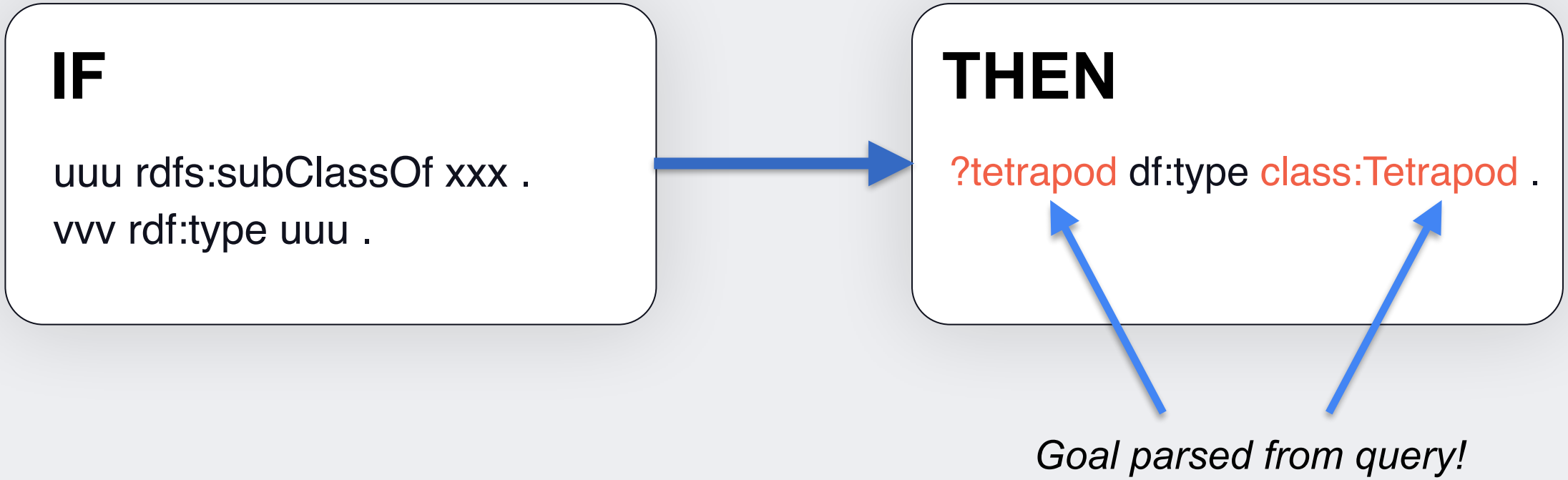
Backward Chaining Algorithm

RDFS9 entailment: subclassOf



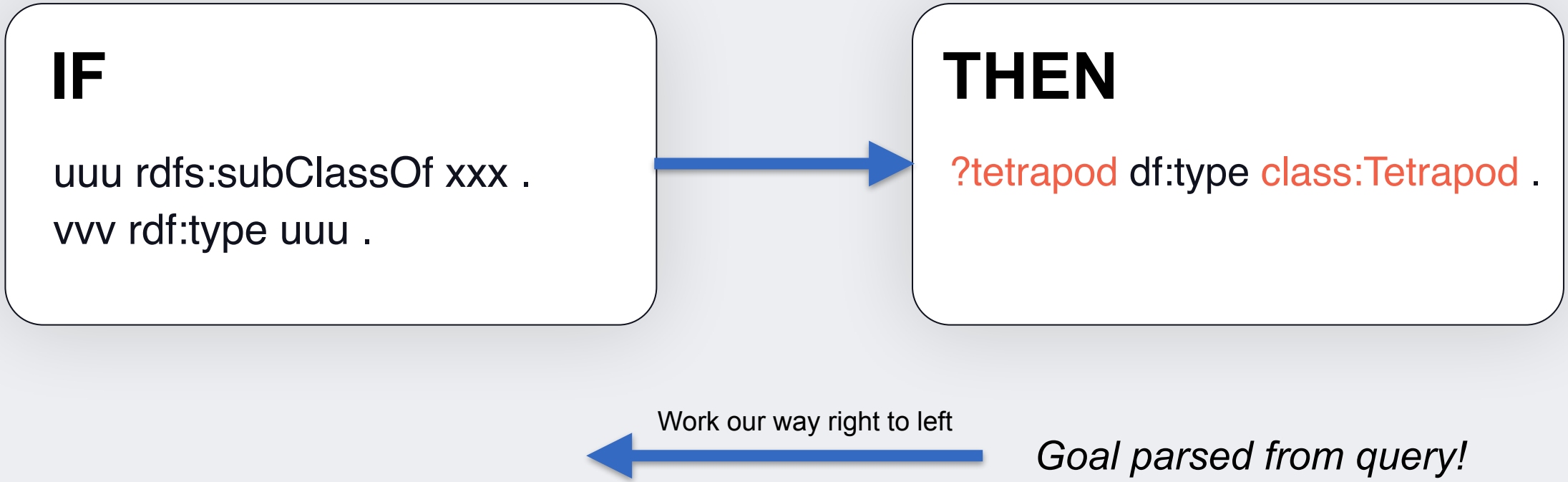
Backward Chaining Algorithm

RDFS9 entailment (non transitive subclassOf)



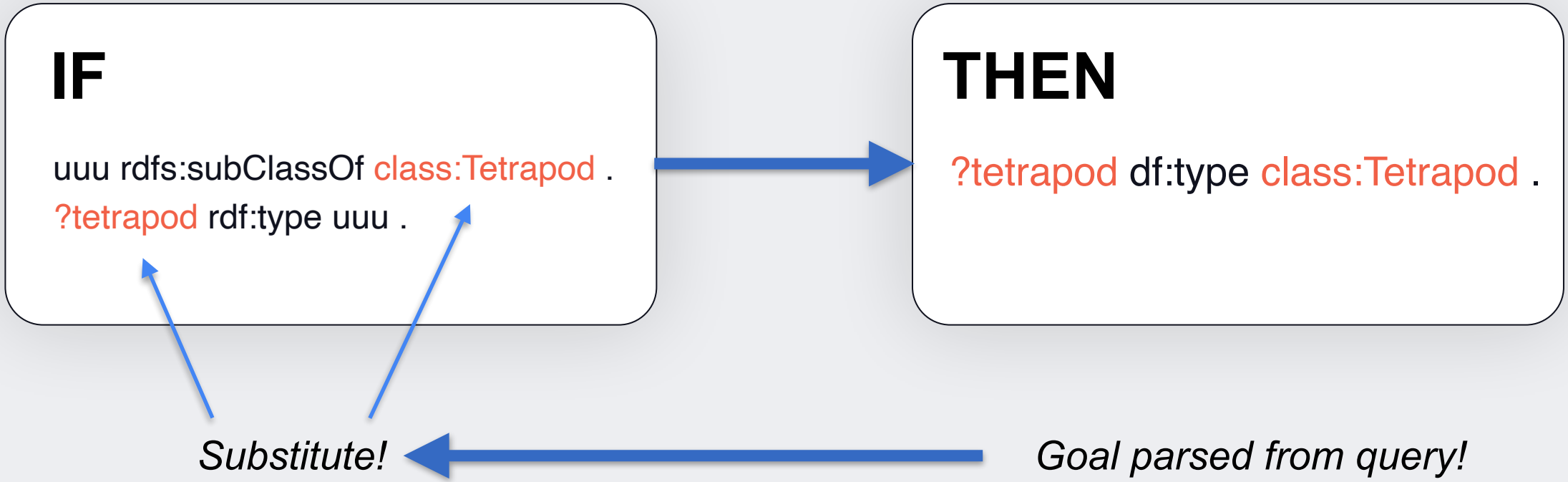
Backward Chaining Algorithm

RDFS9 entailment (non transitive subclassOf)



Backward Chaining Algorithm

RDFS9 entailment (non transitive subclassOf)



Backward Chaining Algorithm

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX class: <http://example.org/class/>

SELECT ?tetrapod
WHERE {
   {
     uuu rdfs:subClassOf class:Tetrapod .
     ?tetrapod ref:type ?uuu .
   } UNION {
     ?tetrapod rdf:type class:Tetrapod
   }
}
```

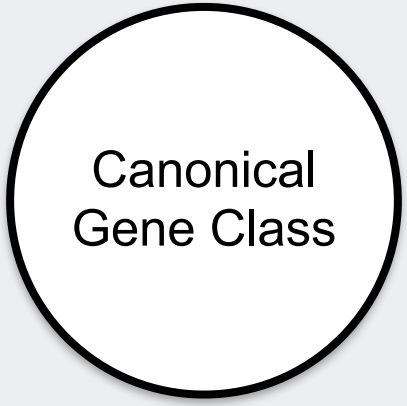
Rewrite the query with the left hand side of the entailment

Union with the right hand side of the entailment

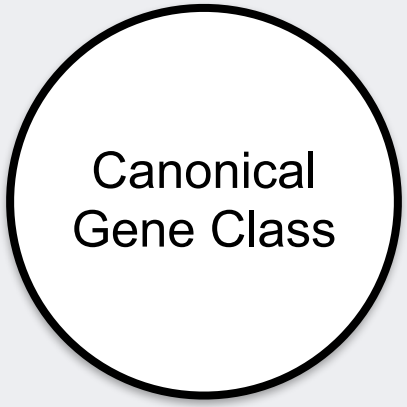
“The great thing about standards is that you have so many to choose from”

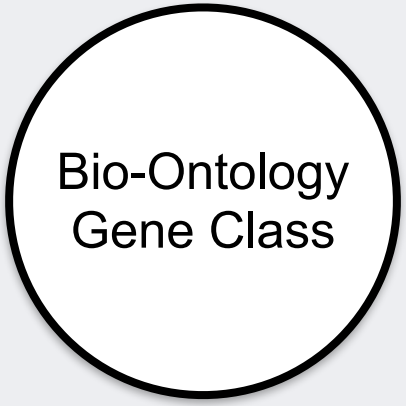
- Either Grace Hopper
- Or Andrew Tanenbaum

Equivalent Class Entailment

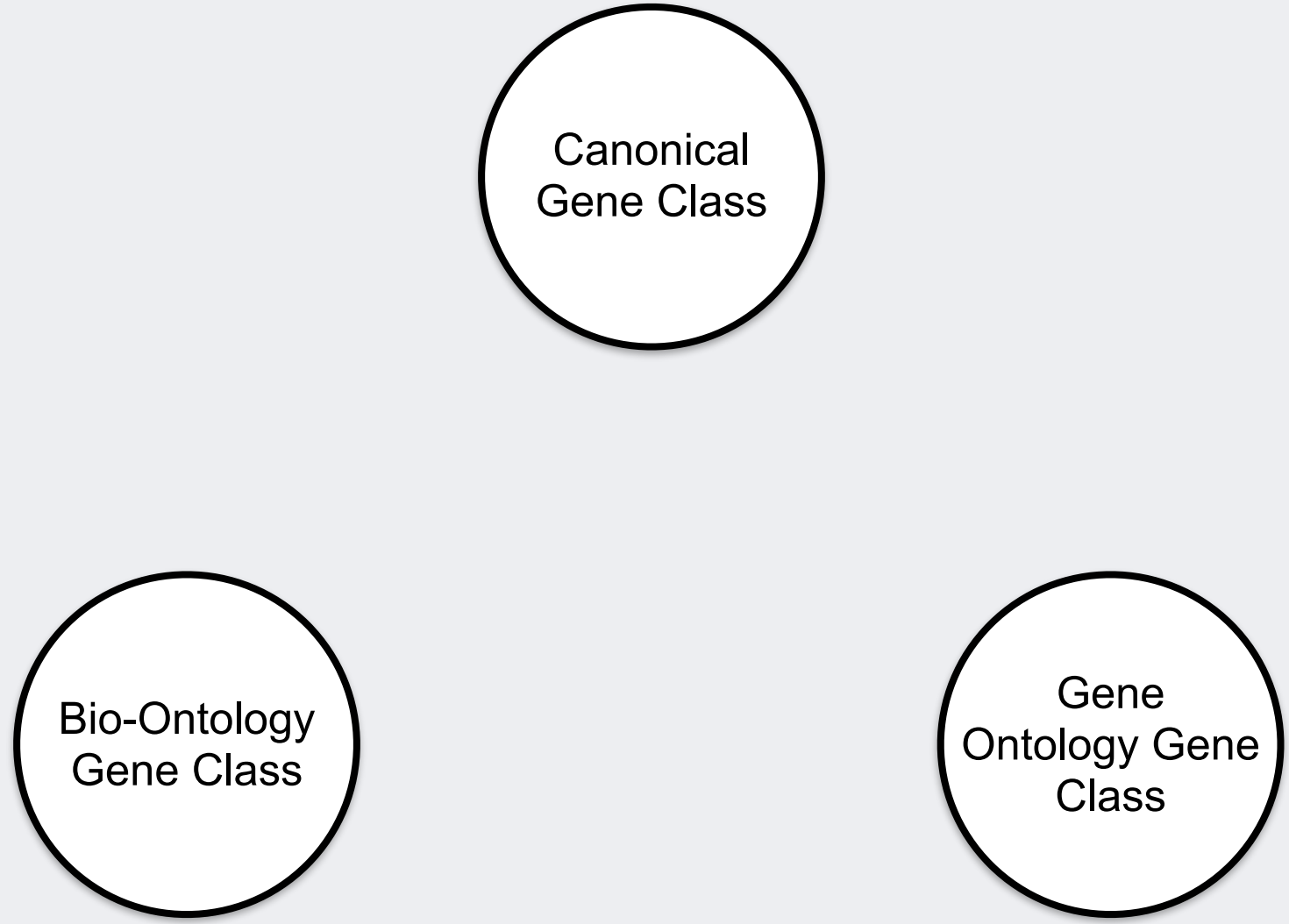


Equivalent Class Entailment



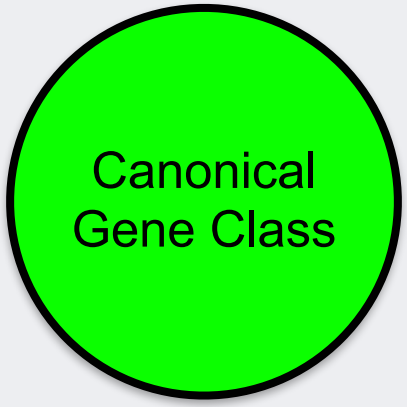


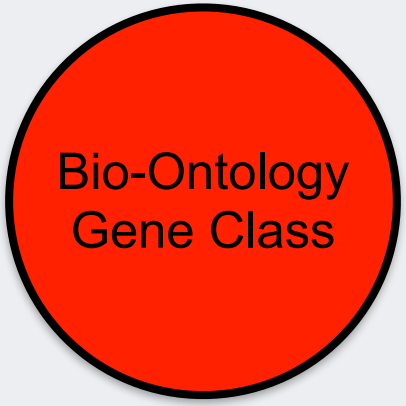
Equivalent Class Entailment

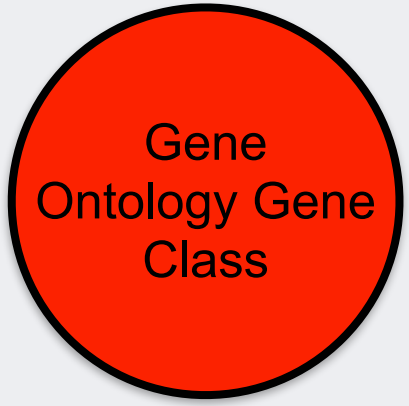


Equivalent Class Entailment

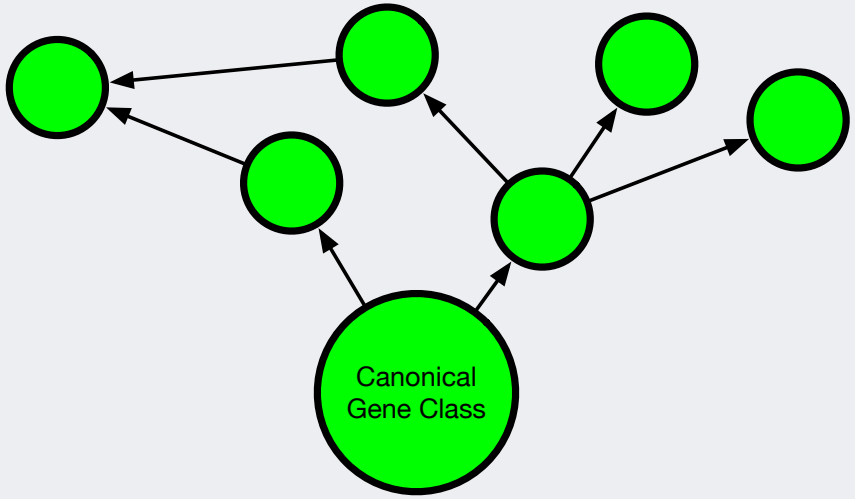
Please give me all instances of genes and their variants in the Knowledge Graph

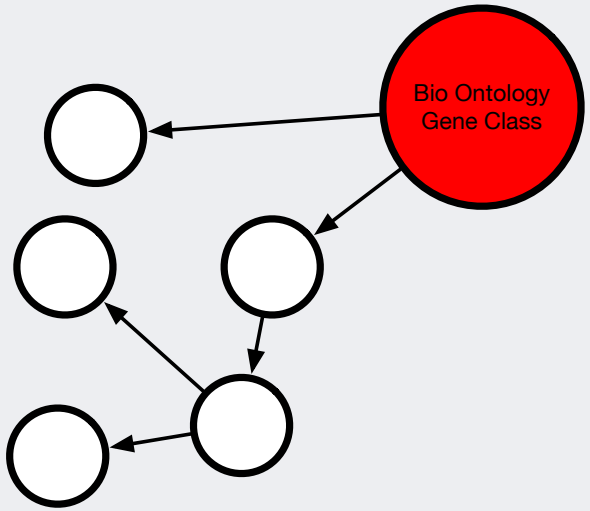


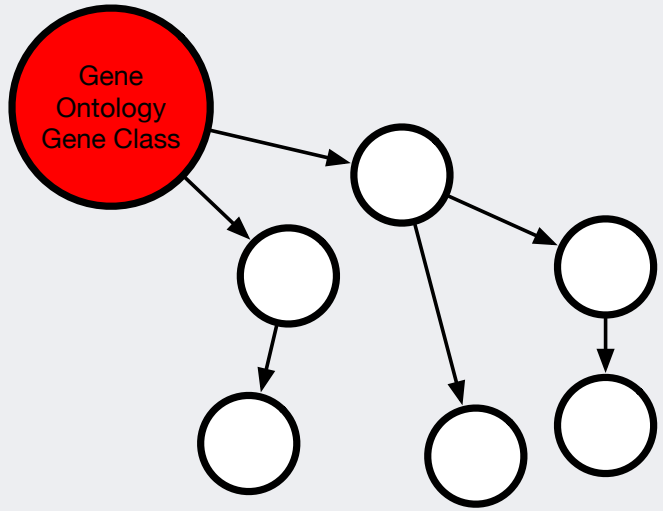




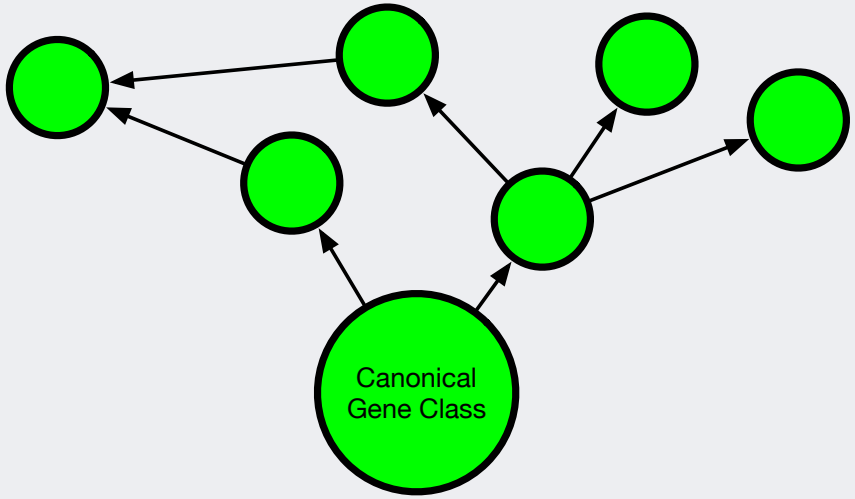
Equivalent Class Entailment

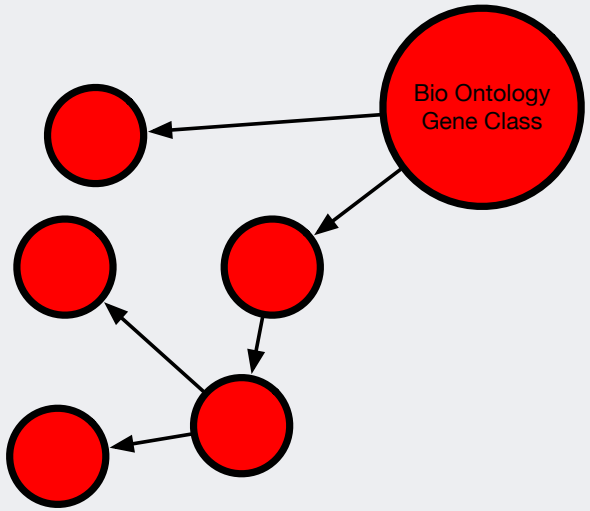


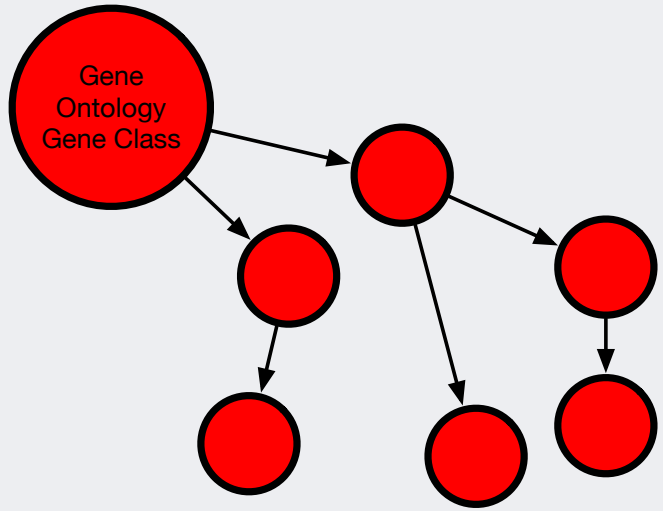




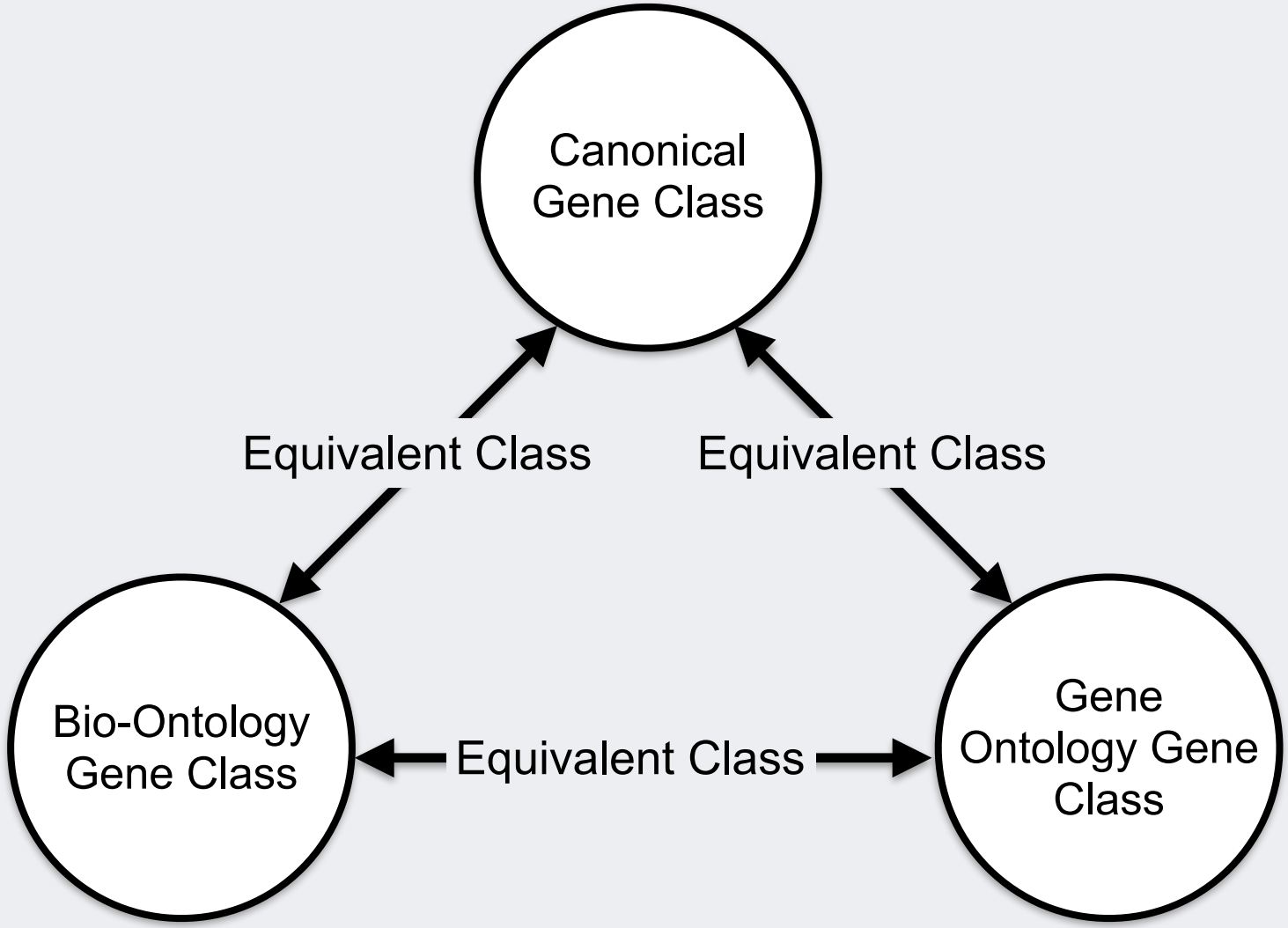
Equivalent Class Entailment





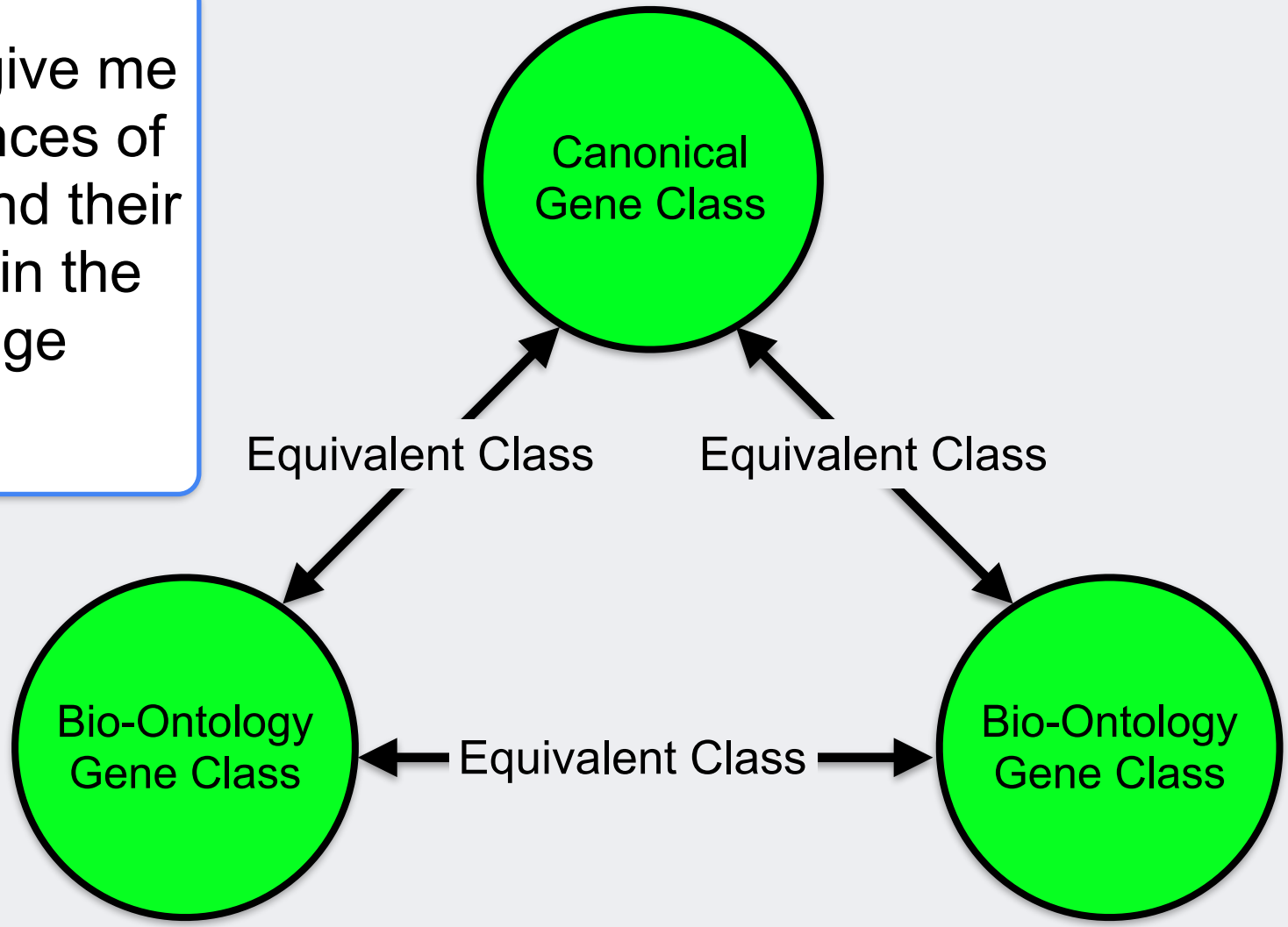


Equivalent Class Entailment

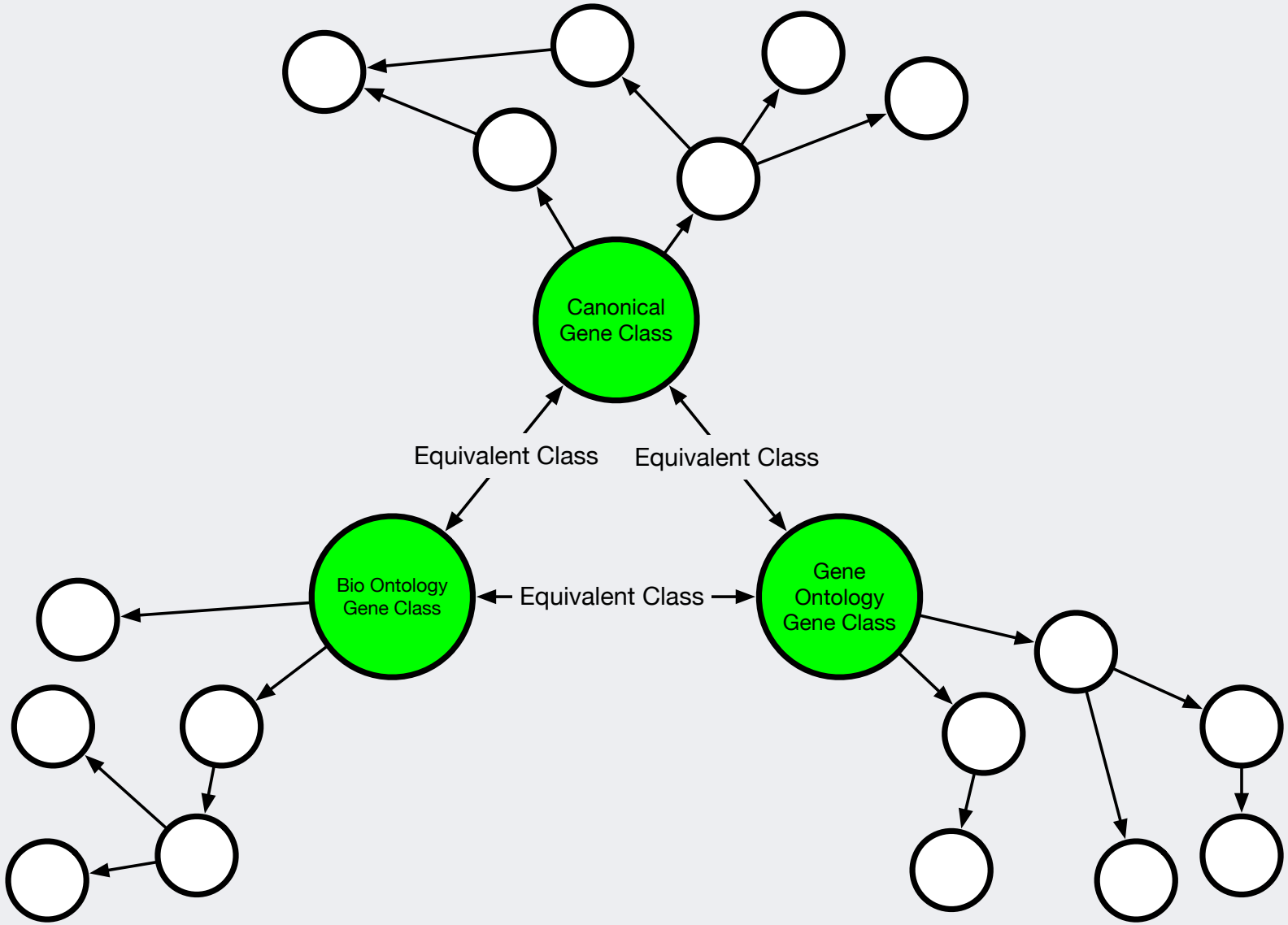


Equivalent Class Entailment

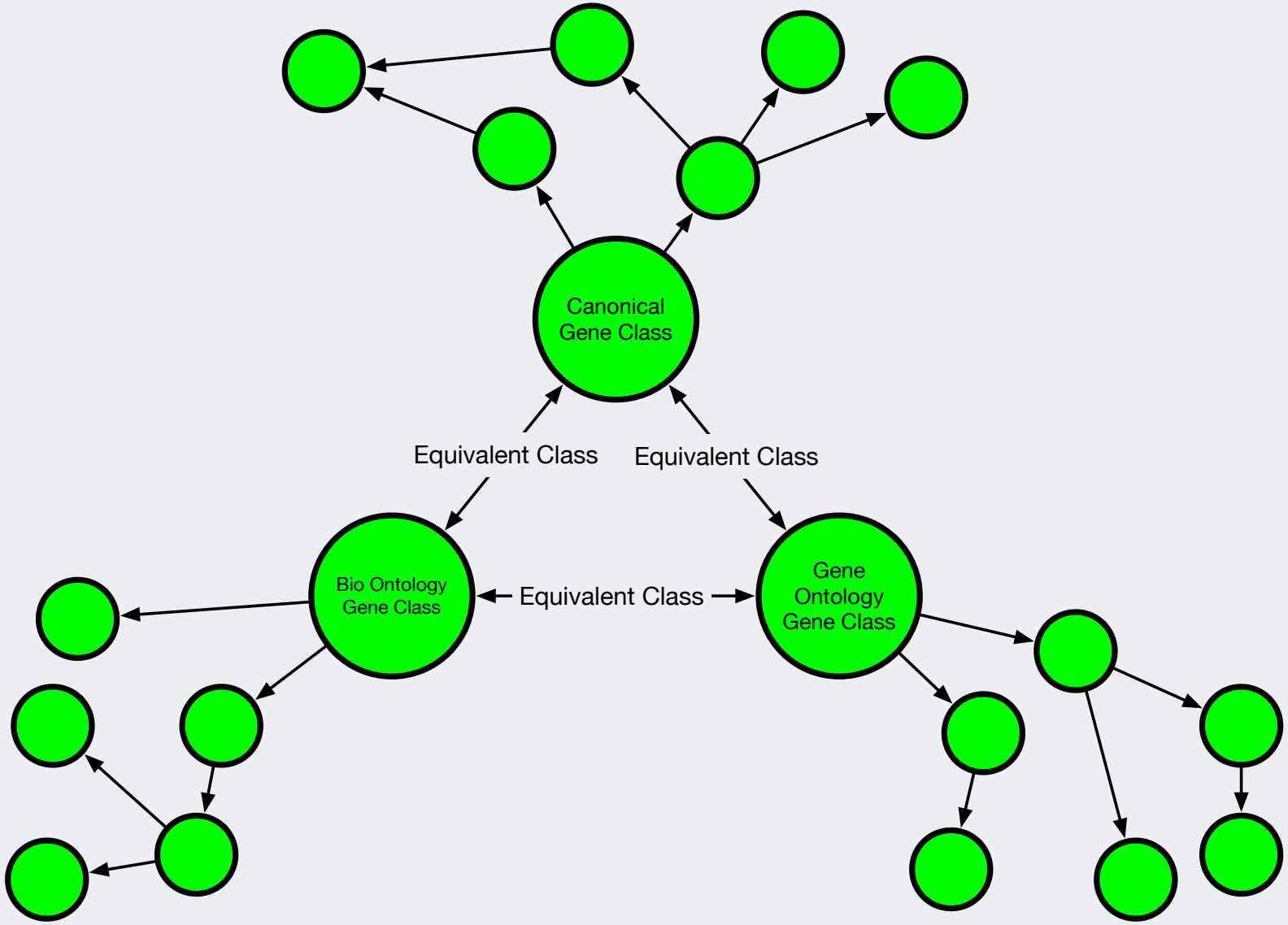
Please give me all instances of genes and their variants in the Knowledge Graph



Equivalent Class Entailment



Equivalent Class Entailment



Logical inferencing demo

Bellman engine on Github

https://github.com/gsk-aiops/bellman

Last Year's Talk

https://www.youtube.com/watch?v=Kj5bZ2afWSU

DATA+AI
SUMMIT 2022

Thank you

John Kakoulides Hunter

Senior Product Director - GSK