(Sub)symbolic AI Representations

Bloomberg
Engineering

ECIR 2023
6 April, 2023

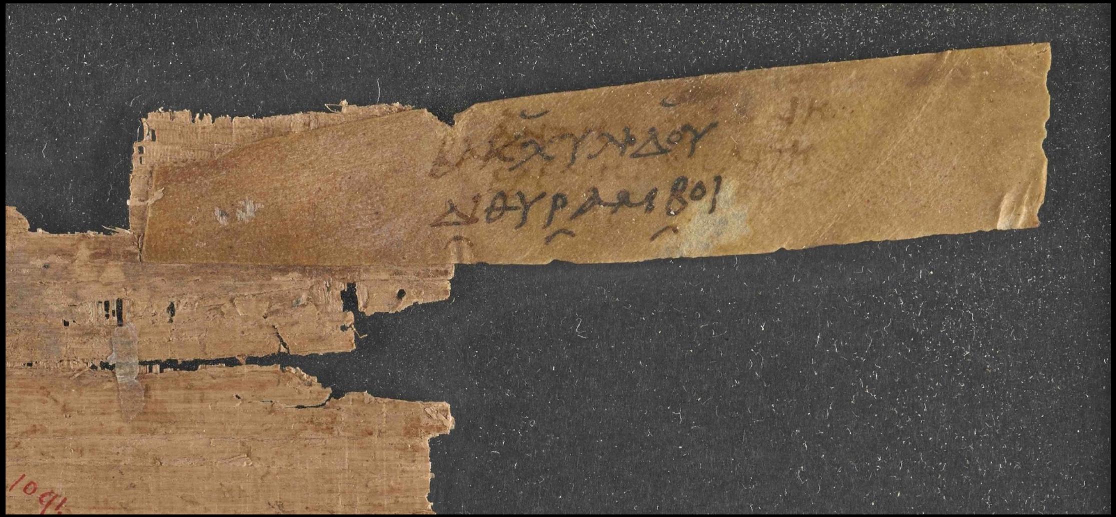
Edgar Meij, Ph.D.
Head of AI Search
@edgarmeij

TechAtBloomberg.com

A Little History of AI

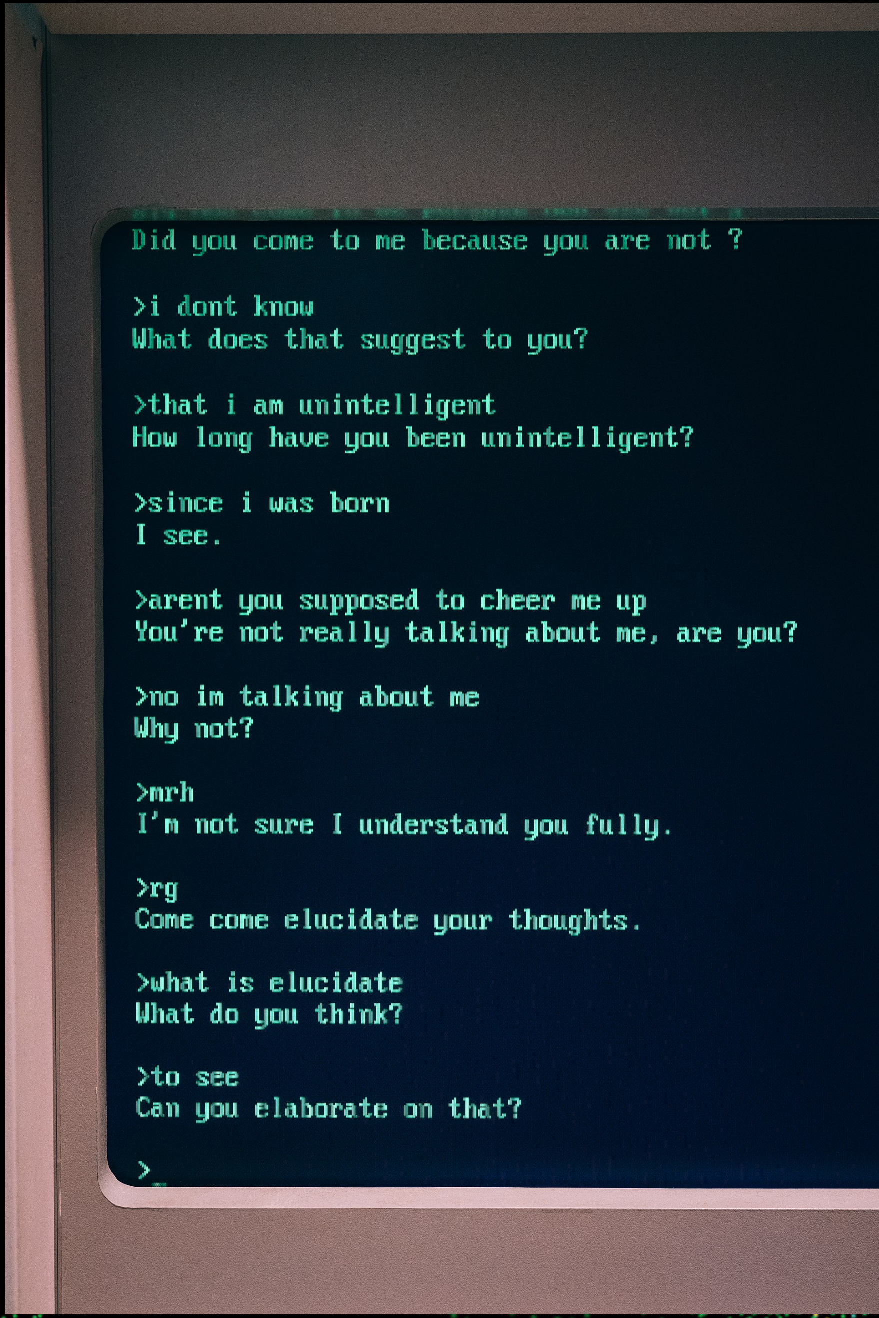
TechAtBloomberg.com

Library Science 101



A Little History of AI – 1950-1960s

- General problem solving and automated reasoning
  - A* algorithm, perceptron
  - Arthur Samuel: Checkers program, “machine learning”
  - ELIZA

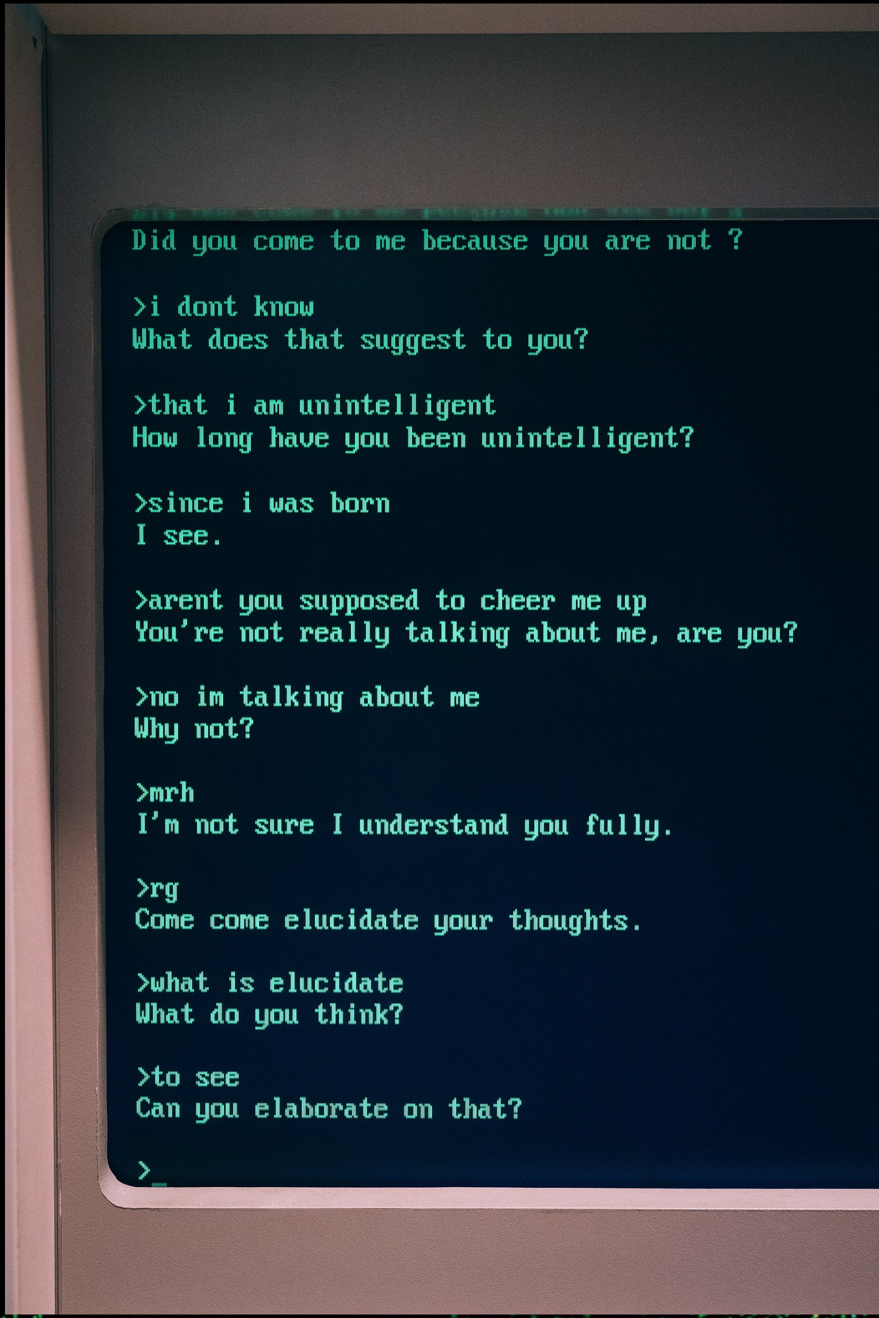


By Marcin Wichary from San Francisco, Calif. - Flickr, CC BY 2.0

TechAtBloomberg.com

A Little History of AI – 1950-1960s

- General problem solving and automated reasoning
  - A* algorithm, perceptron
  - Arthur Samuel: Checkers program, “machine learning”
  - ELIZA
- Some issues though
  - Did not live up to promises/expectations
  - Limited storage & computing power
- Winter (1960s)



By Marcin Wichary from San Francisco, Calif. - Flickr, CC BY 2.0

TechAtBloomberg.com

A Little History of AI – 1970-1980s

- Symbolic AI, knowledge engineering, and rule-based expert systems
  - “Semantic networks”
  - WordNet

KNOWLEDGE GRAPHS:
representation and structuring of scientific knowledge

proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
Prof.dr.ir. H.H. van den Kroonenberg,
volgens besluit van het College van Dekanen
in het openbaar te verdedigen
op vrijdag 18 december 1987 te 16.00 uur

door

René Ronald Bakker

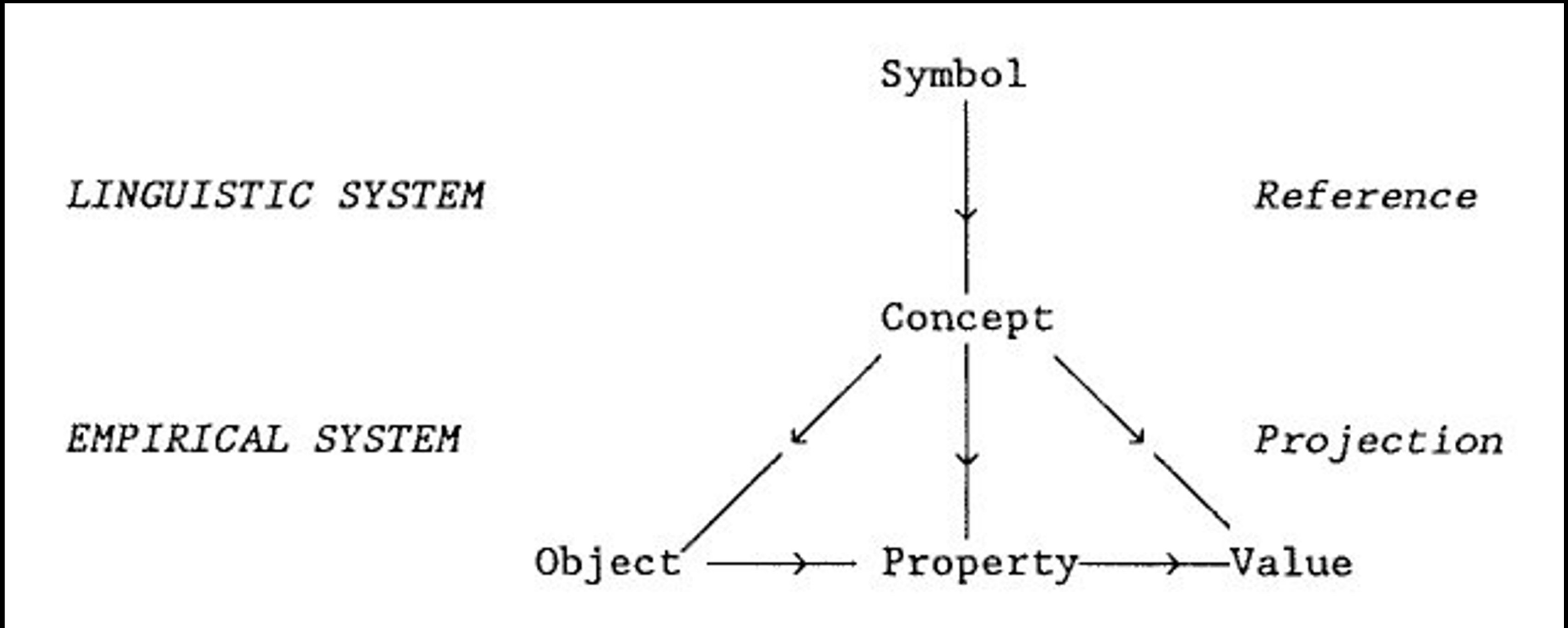
geboren op 8 april 1956 te Utrecht

"A knowledge graph is a labeled directed graph D(P,A) with
(a) P is a set of points that represent concepts, relations or frameworks;
(b) $A \subseteq P \chi P$ is a set of arcs that form the connections between the entities. An arc exists for $p_1 \in P$ to $p_2 \in P$ if:
(b1) p_1 represents a concept that is the tail of a relation represented by p_2;
(b2) p_2 represents a concept that is the head of a relation represented by p_1;
(b3) p_1 represents an element in the contents of a framework and p_2 represents the FPAR-relation;
(b4) p_2 is a framework and p_1 represents the accompanying FPAR-relation."

From: RR Bakker, 1987. "*Knowledge Graphs: representation and structuring of scientific knowledge*".

Bloomberg
Engineering

From that same dissertation:



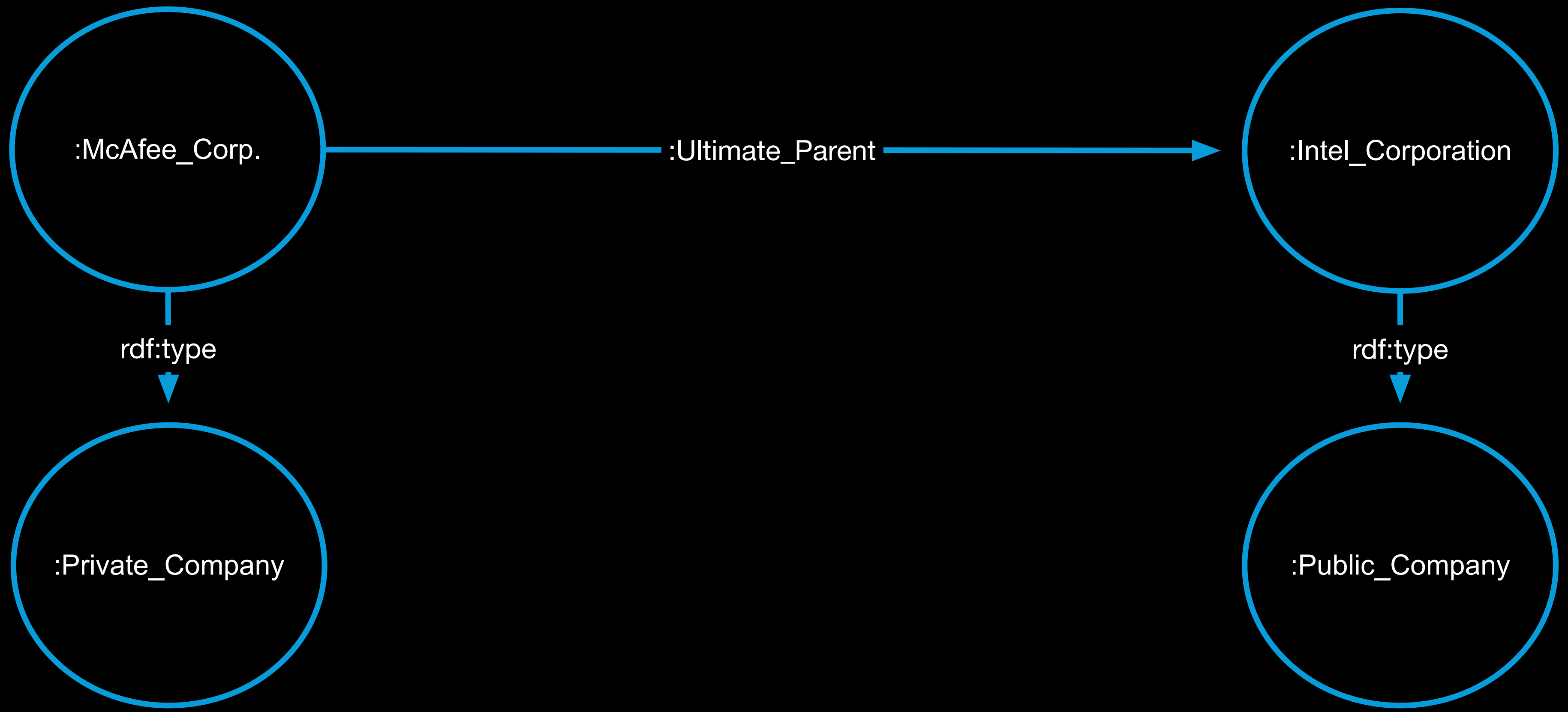
From: RR Bakker, 1987. "*Knowledge Graphs: representation and structuring of scientific knowledge*".

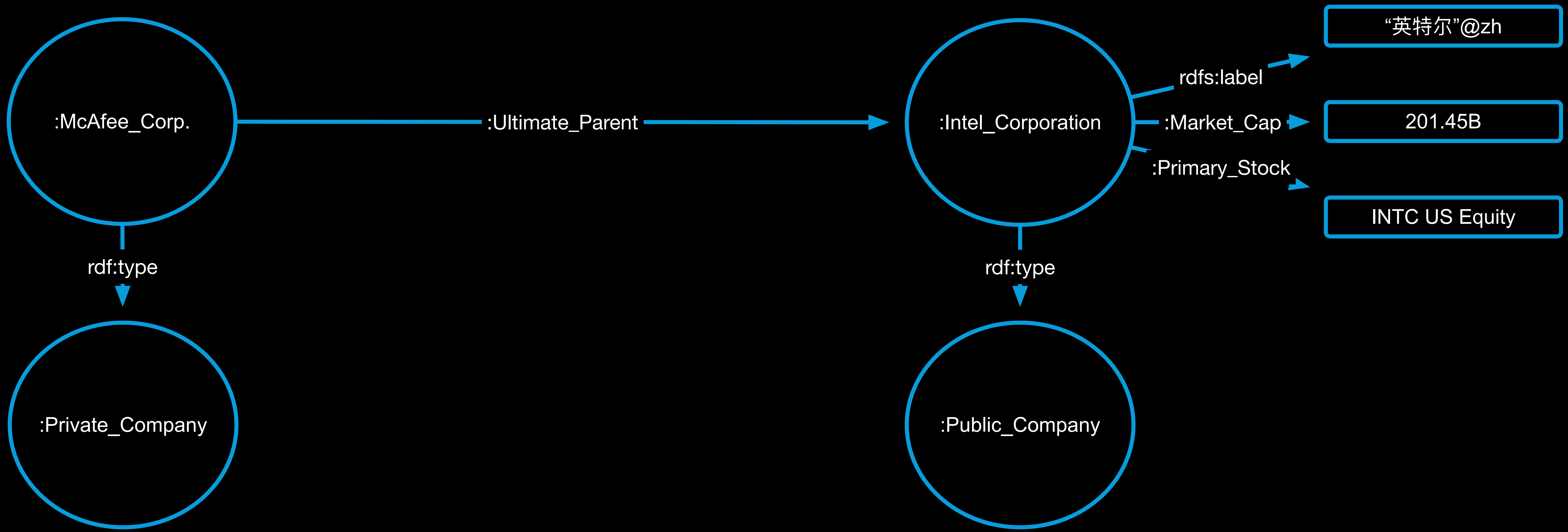
Symbolic AI

- Logical reasoning and rule-based inference

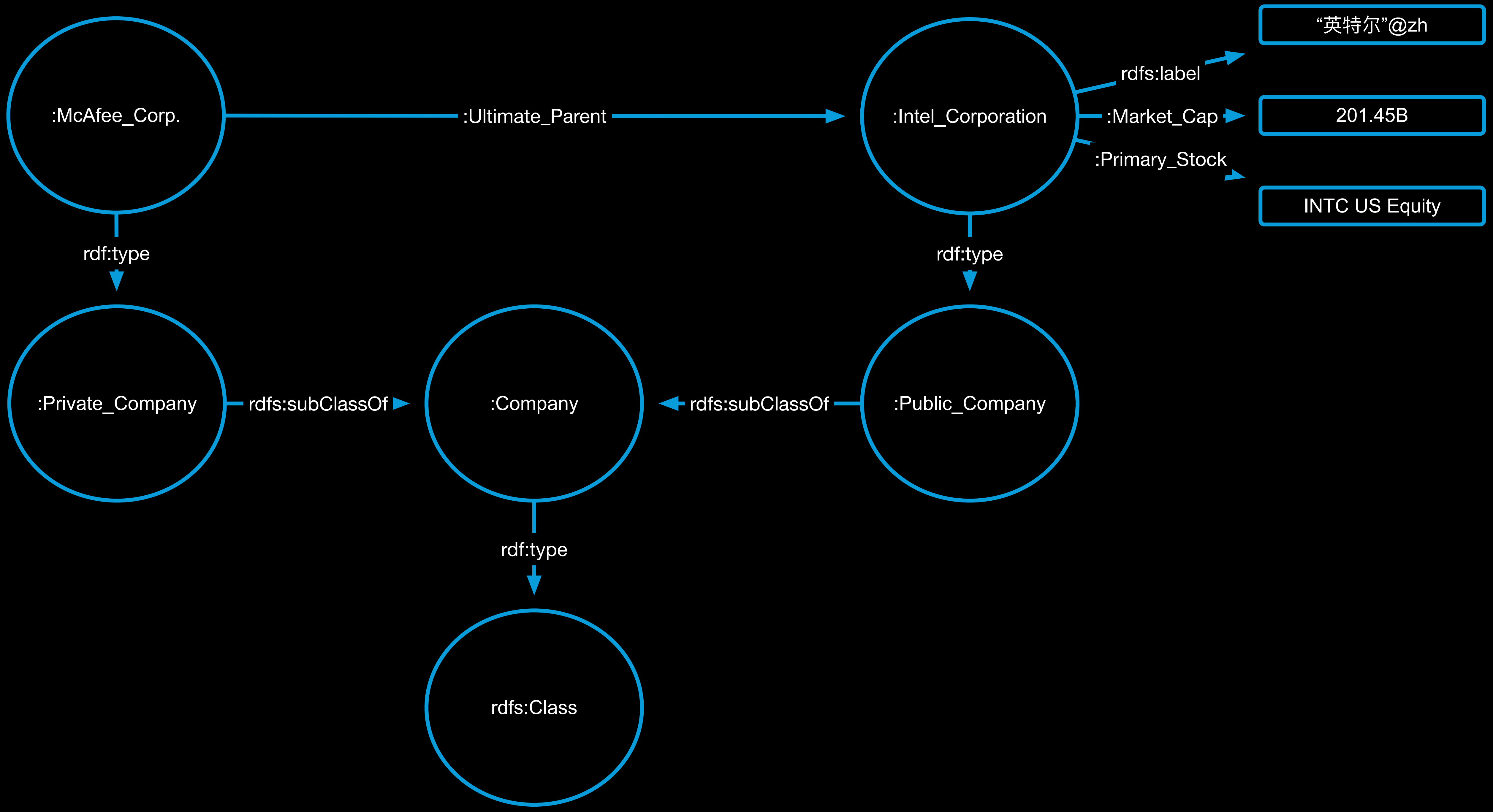
Symbolic AI

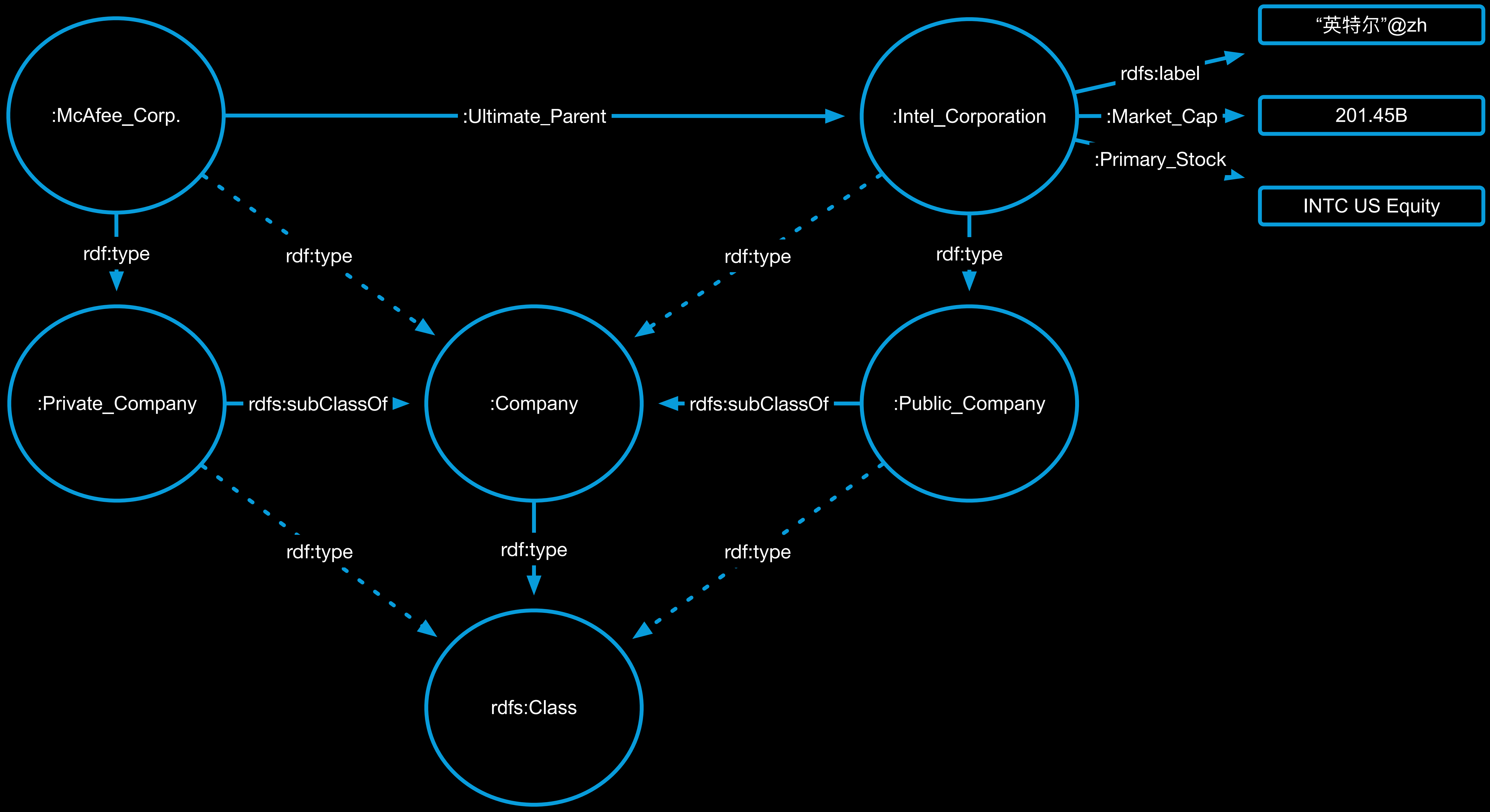
- Logical reasoning and rule-based inference
  - Structural/Hierarchical knowledge representation
  - Easily interpretable models and predictions

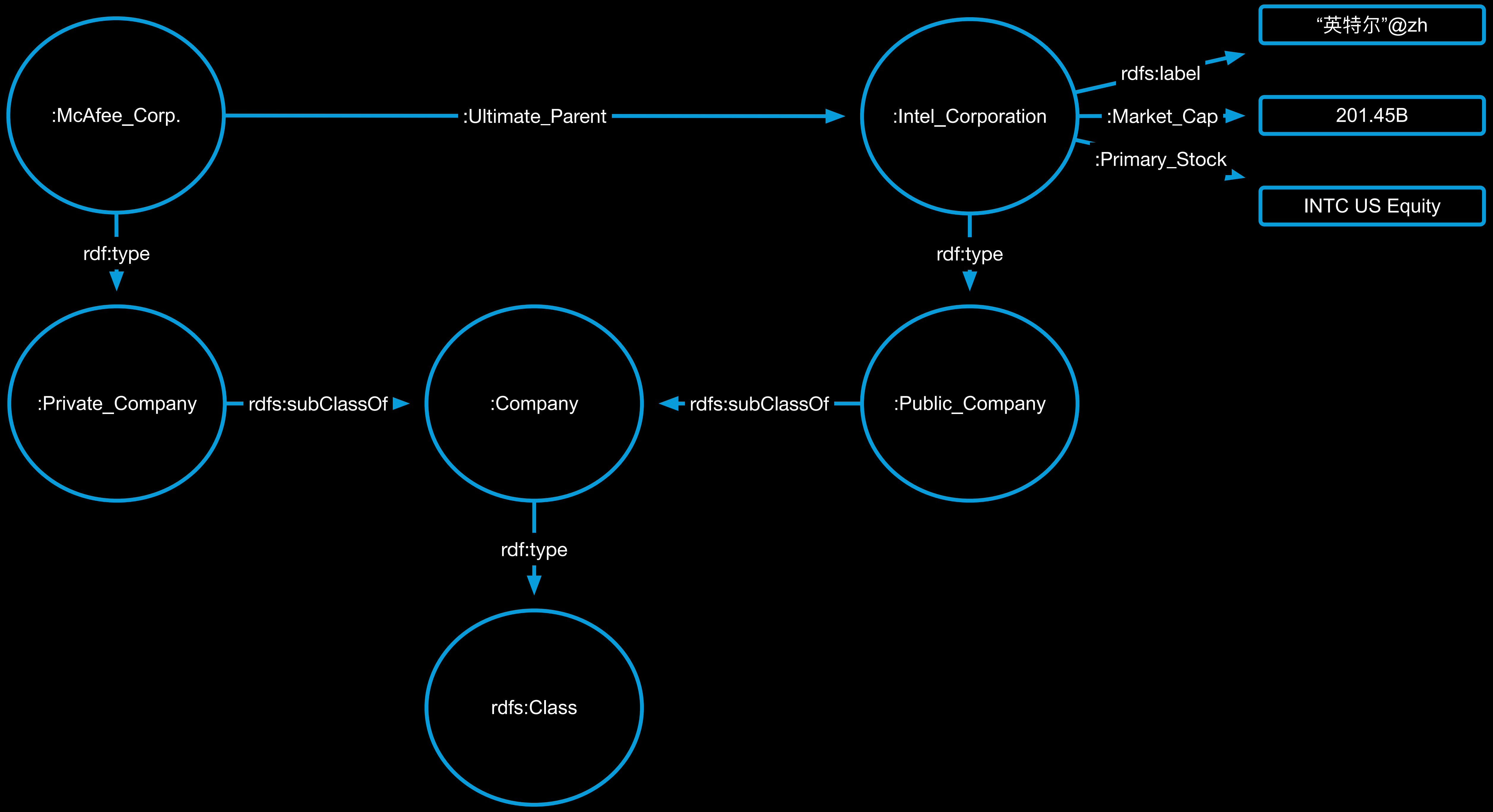

:McAfee_Corp.
:Ultimate_Parent
:Intel_Corporation
rdf:type
rdf:type
:Private_Company
:Public_Company



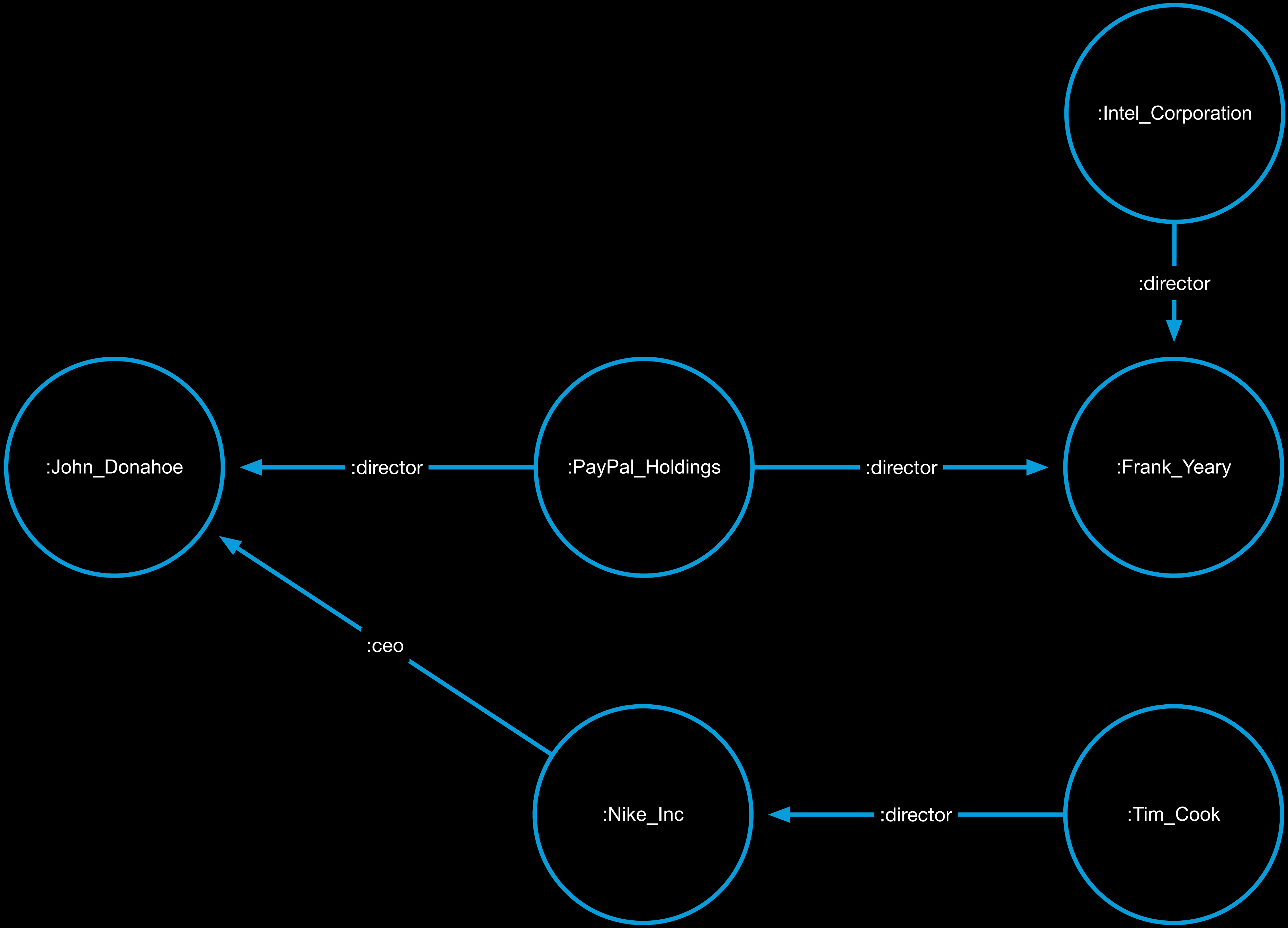
:McAfee_Corp.
:Ultimate_Parent
:Intel_Corporation
rdfs:label
"英特尔"@zh
:Market_Cap
201.45B
:Primary_Stock
INTC US Equity
rdf:type
rdf:type
:Private_Company
:Public_Company


“英特尔”@zh
rdfs:label
:McAfee_Corp.
:Ultimate_Parent
:Intel_Corporation
:Market_Cap
201.45B
:Primary_Stock
INTC US Equity
rdf:type
rdf:type
:Private_Company
rdfs:subClassOf
:Company
rdfs:subClassOf
:Public_Company
rdf:type
rdfs:Class

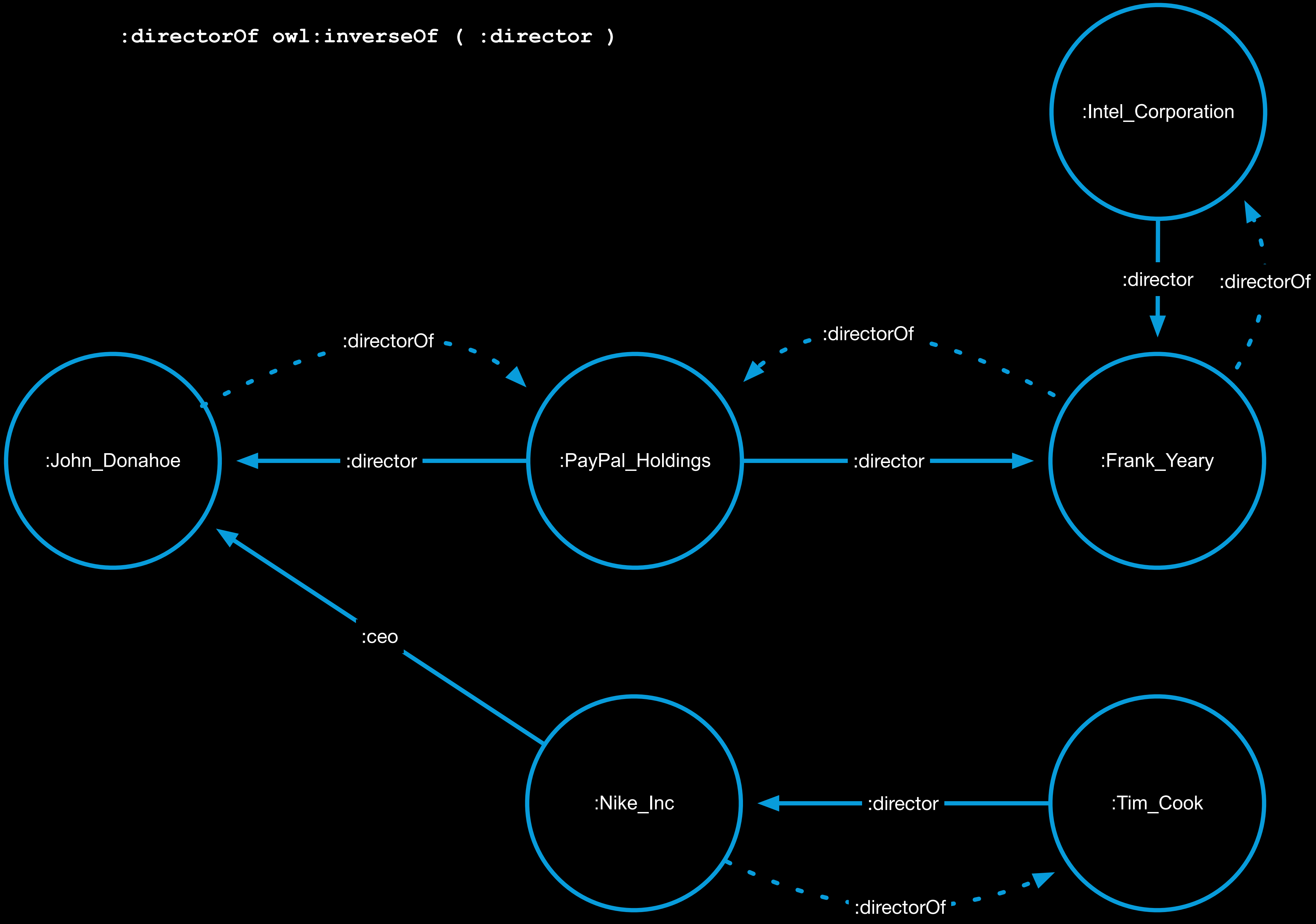

:McAfee_Corp.
:Ultimate_Parent
:Intel_Corporation
rdfs:label
“英特尔”@zh
:Market_Cap
201.45B
:Primary_Stock
INTC US Equity
rdf:type
rdf:type
rdf:type
rdf:type
:Private_Company
rdfs:subClassOf
:Company
rdfs:subClassOf
:Public_Company
rdf:type
rdf:type
rdf:type
rdfs:Class



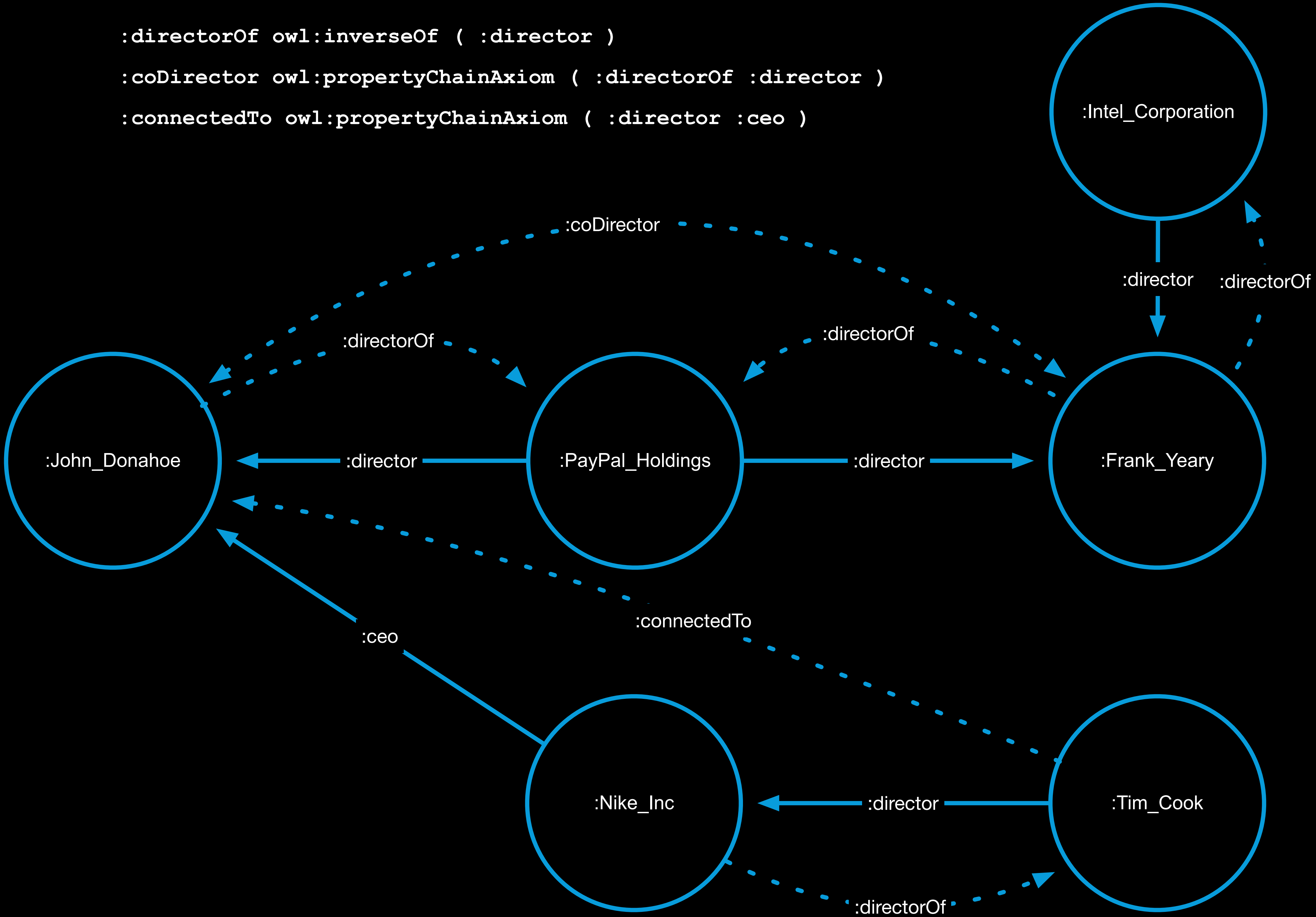
:McAfee_Corp.
:Ultimate_Parent
:Intel_Corporation
rdfs:label
"英特尔"@zh
:Market_Cap
201.45B
:Primary_Stock
INTC US Equity
rdf:type
rdf:type
:Private_Company
rdfs:subClassOf
:Company
rdfs:subClassOf
:Public_Company
rdf:type
rdfs:Class


:Intel_Corporation
:director
:John_Donahoe
:director
:PayPal_Holdings
:director
:Frank_Yeary
:ceo
:Nike_Inc
:director
:Tim_Cook

```
:directorOf owl:inverseOf ( :director )
```



```
:directorOf owl:inverseOf ( :director )
:coDirector owl:propertyChainAxiom ( :directorOf :director )
:connectedTo owl:propertyChainAxiom ( :director :ceo )
```



A Little History of AI – 1970-1980s

- Symbolic AI, knowledge engineering, and rule-based expert systems

- Logical reasoning and rule-based inference
  - Structural/Hierarchical knowledge representation
  - Easily interpretable models and predictions

- Some issues though
  - Hard to (re)model to new domains
  - Knowledge acquisition
  - Generalizing to unseen cases

- Winter (1990s)

A Little History of AI – 2000s

- Handling uncertainty in knowledge acquisition and representation (2000s)
  - LMs, statistical learning

- Deep learning and end-to-end representation learning (2010s-now)
  - Word2vec
  - LSTMs
  - BERT
  - “Fine tuning”

BUSINESS | JOURNAL REPORTS: LEADERSHIP

The Optimistic Promise of Artificial Intelligence

Andrew Ng and Tong Zhang on how AI is going to be like electricity, transforming every industry

THE AI REVOLUTION IS ON

Microsoft researchers achieve speech recognition milestone

China's AI Agenda Advances

Google Translate Has Reached Human-Like Accuracy Thanks To Neural Machine Translation Engine

Alibaba's AI Outguns Humans in Reading Test

The Sublime and Scary Future of Cameras With A.I. Brains

Stanford team creates computer vision algorithm that can describe photos

A Little History of AI

- ChatGPT, GPT-4, BARD (now)
  - Multi-modal
  - Dialogs
  - Search integration
  - Seemingly disruptive!

Representations

TechAtBloomberg.com

Representations

- Prime goal of metadata ("data about data") is to assist users (or machines!) in finding and discovering data, to help find, organize, and display, and thus to make better use of the data

TechAtBloomberg.com

Bloomberg
Engineering

Representations

- Prime goal of metadata ("data about data") is to assist users (or machines!) in finding and discovering data, to help find, organize, and display, and thus to make better use of the data

- Different representations of a *unit of retrieval* (such as an entity, document, video, etc) enable effective and efficient ways to help with the above

- For instance,
  - Bag of words, TF-IDF, embeddings
  - Graphs with binary relationships, weighted edges, or neural/embeddings-based
  - Feature vector(s)
  - Etc

Representations

- **Symbolic AI** – Logical reasoning and rule-based inference
  - Structural/Hierarchical knowledge representation
  - Easily interpretable models and predictions
  - Hard to (re)model to new domains

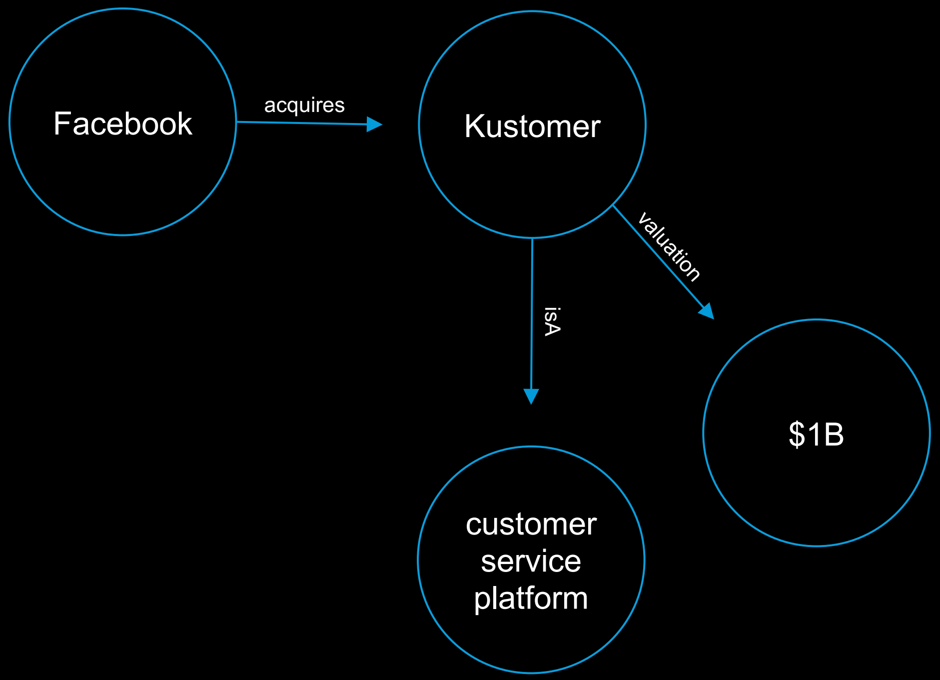
- **Statistical (“Subsymbolic”) AI** – Learning from examples and pattern-based inference
  - Easier to scale since no explicit representations need to be created
  - Only two types of representation need to be *learned*: <u>feature</u> and <u>class</u>
  - Poor at reasoning, hard to interpret
  - Needs a lot of training material

TechAtBloomberg.com

Bridging symbolic + subsymbolic AI

- “Information extraction”, "Knowledge base population”, “Entity linking”, “Named entity disambiguation”, “Semantic parsing”, etc.

Facebook to acquire customer service platform Kustomer for $1B.

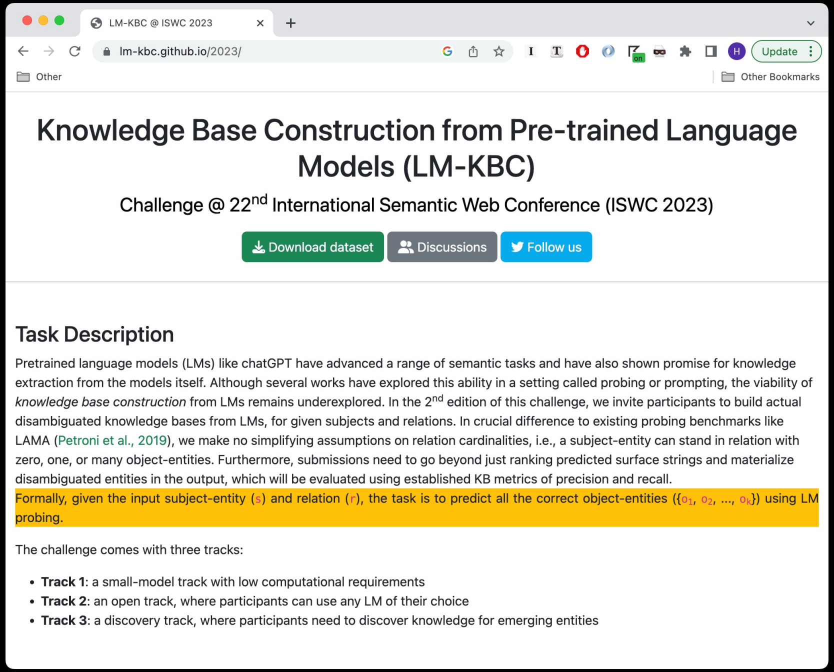


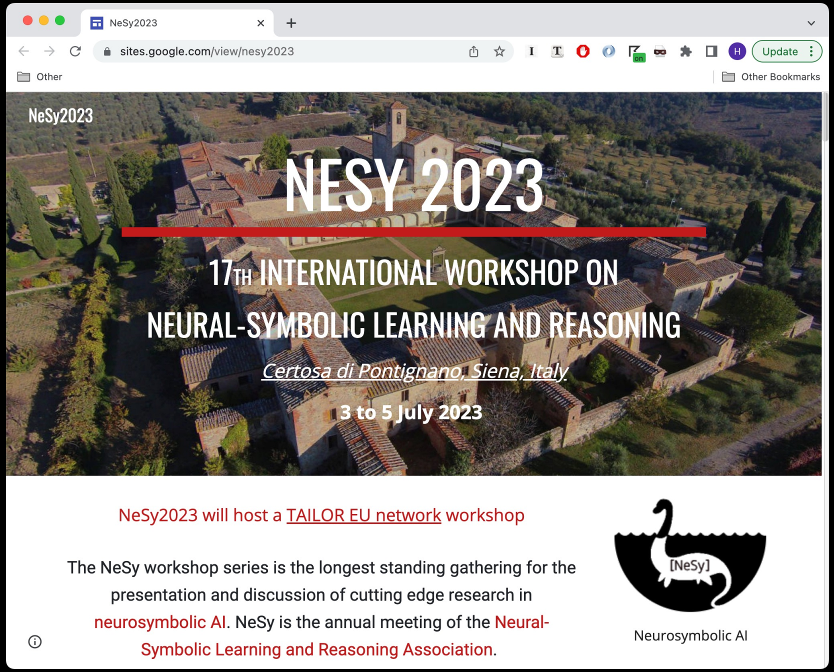
Neurosymbolic AI

- Deep learning's prime innovation: *learn* structural representations, end-to-end, (possibly) task-specific, from data

Neurosymbolic AI

- Deep learning's prime innovation: *learn* structural representations, end-to-end, (possibly) task-specific, from data





Thank You!

Part of https://github.com/laura-dietz/neurosymbolic-representations-for-IR/

Contact me: emeij@bloomberg.net

https://TechAtBloomberg.com/ai
https://TechAtBloomberg.com/data-science-research-grant-program

https://www.bloomberg.com/careers

Bloomberg Engineering

TechAtBloomberg.com

Agenda

Part 1: Symbolic AI representations and tasks

- ~~(Sub)symbolic AI, and representations~~
- Foundations for this tutorial
- Question Answering on Knowledge Graphs

Part 2: Text-to-symbols and Ranking

- Neural Text and Graph Representations
- Text-Symbol Alignment and Semantic Annotations
- Entity Representations and Entity Ranking

Part 3: Neuro-symbolic representations for Reasoning

- Reasoning about Relevance
- Neuro Pseudo-Relevance Feedback with Explainability

Part 4: Applications for Neuro-symbolic approaches

- Industry Use Case: Knowledge Discovery
- Panel & Discussion

TechAtBloomberg.com