

(Sub)symbolic AI Representations

Engineering

Bloomberg

ECIR 2023
6 April, 2023

Edgar Meij, Ph.D.
Head of AI Search
[@edgarmeij](https://twitter.com/edgarmeij)

TechAtBloomberg.com

A Little History of AI

TechAtBloomberg.com

© 2023 Bloomberg Finance L.P. All rights reserved.

Library Science 101



TechAtBloomberg.com

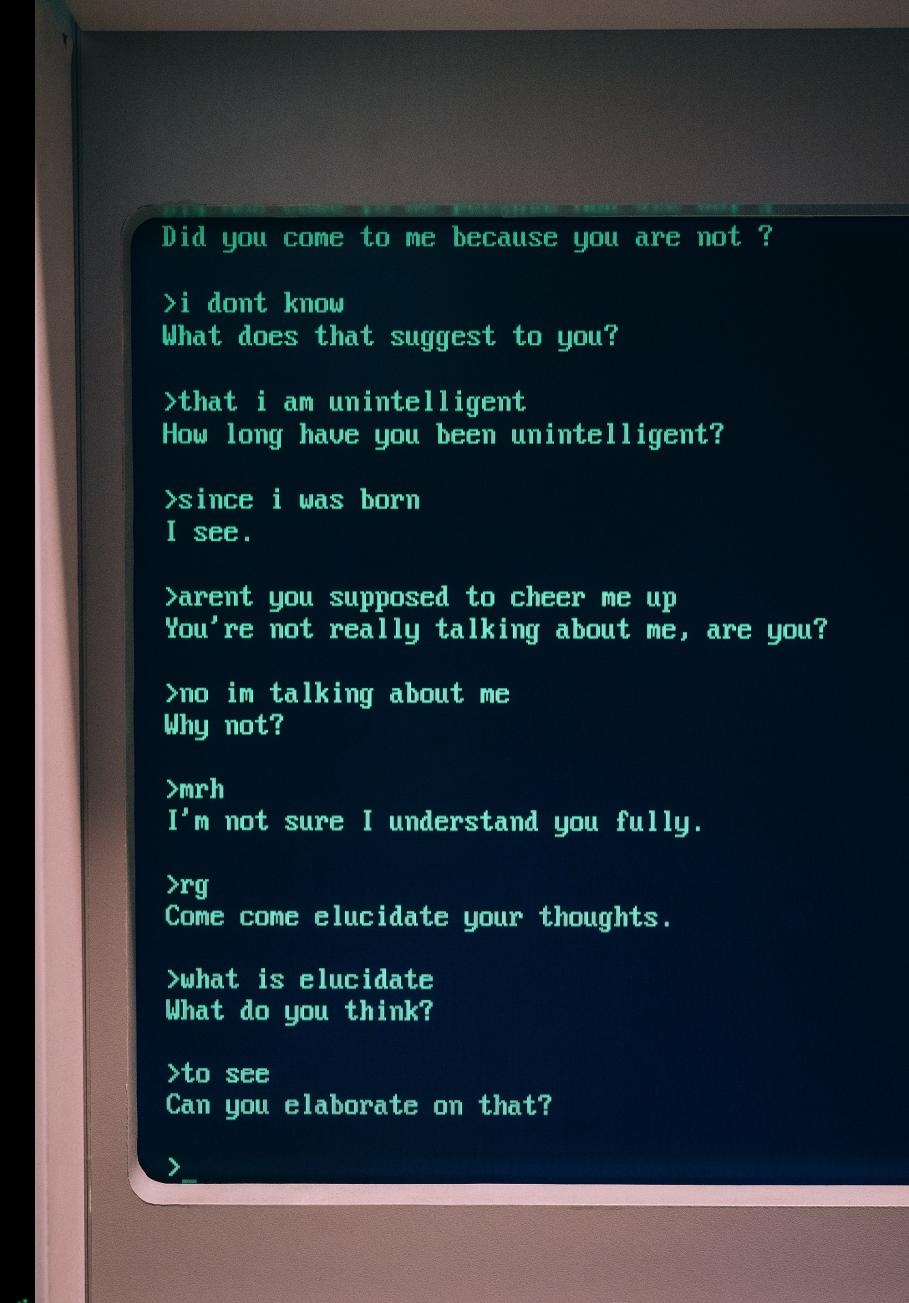
© 2023 Bloomberg Finance L.P. All rights reserved.

Edgar Meij – ECIR 23 Tutorial on Neuro-Symbolic Representations for IR

Bloomberg
Engineering

A Little History of AI – 1950-1960s

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



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)

TechAtBloomberg.com

```
Did you come to me because you are not ?  
>i dont know  
What does that suggest to you?  
>that i am unintelligent  
How long have you been unintelligent?  
>since i was born  
I see.  
>arent you supposed to cheer me up  
You're not really talking about me, are you?  
>no im talking about me  
Why not?  
>mrh  
I'm not sure I understand you fully.  
>rg  
Come come elucidate your thoughts.  
>what is elucidate  
What do you think?  
>to see  
Can you elaborate on that?  
>_
```

A Little History of AI – 1970-1980s

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

TechAtBloomberg.com

© 2023 Bloomberg Finance L.P. All rights reserved.
Edgar Meij – ECIR 23 Tutorial on Neuro-Symbolic Representations for IR

Bloomberg
Engineering

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

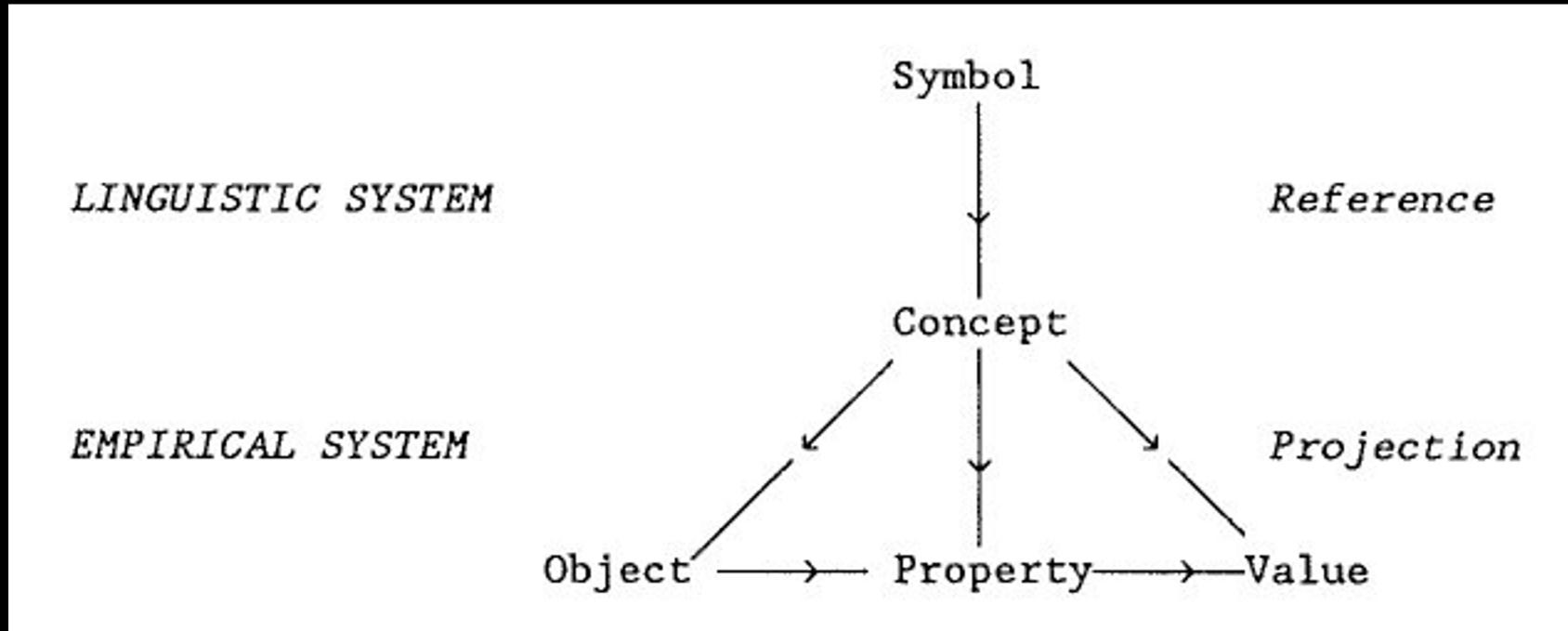
"A knowledge graph is a labeled directed graph $D(P,A)$ with

- (a) P is a set of points that represent concepts, relations or frameworks;
- (b) $A \subseteq P \times P$ 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

TechAtBloomberg.com

© 2023 Bloomberg Finance L.P. All rights reserved.
Edgar Meij – ECIR 23 Tutorial on Neuro-Symbolic Representations for IR

Bloomberg
Engineering

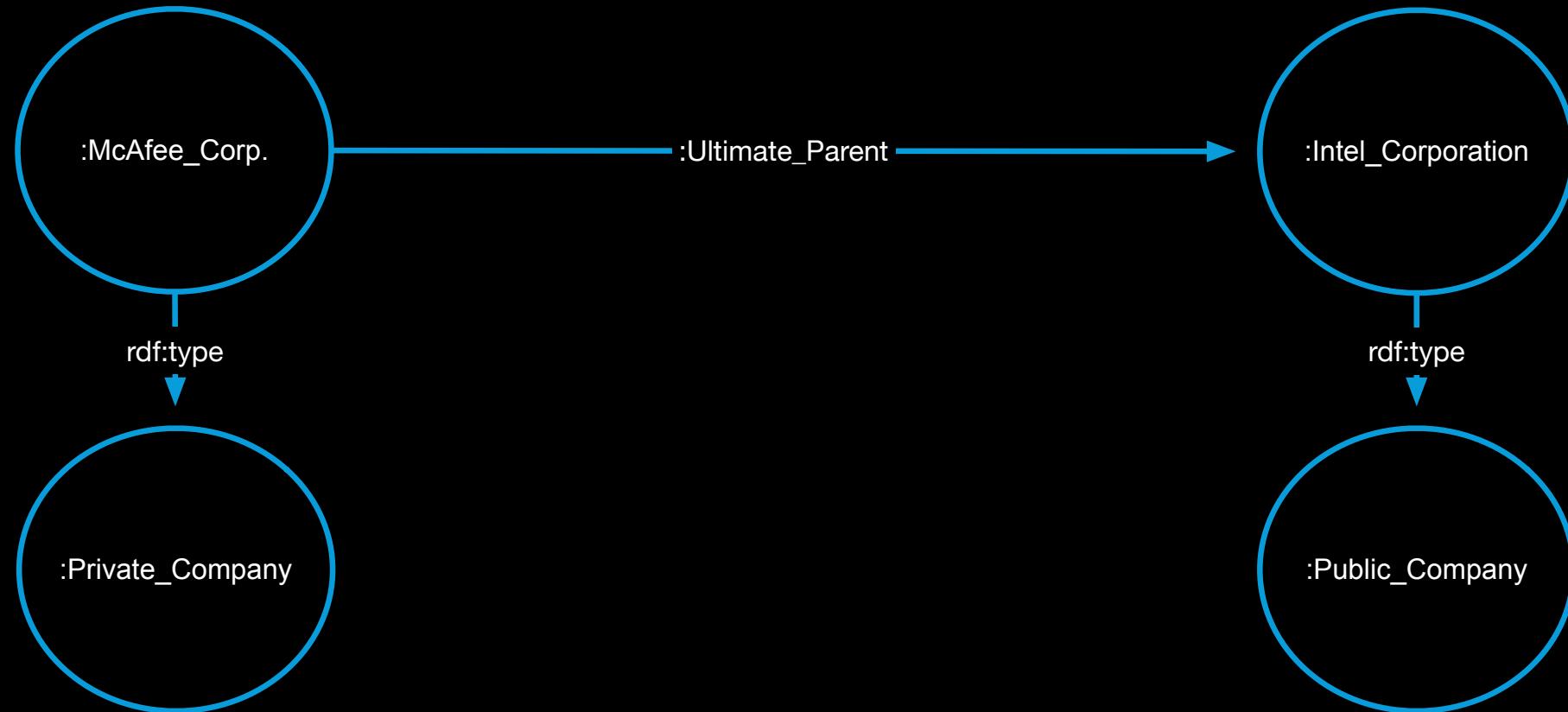
Symbolic AI

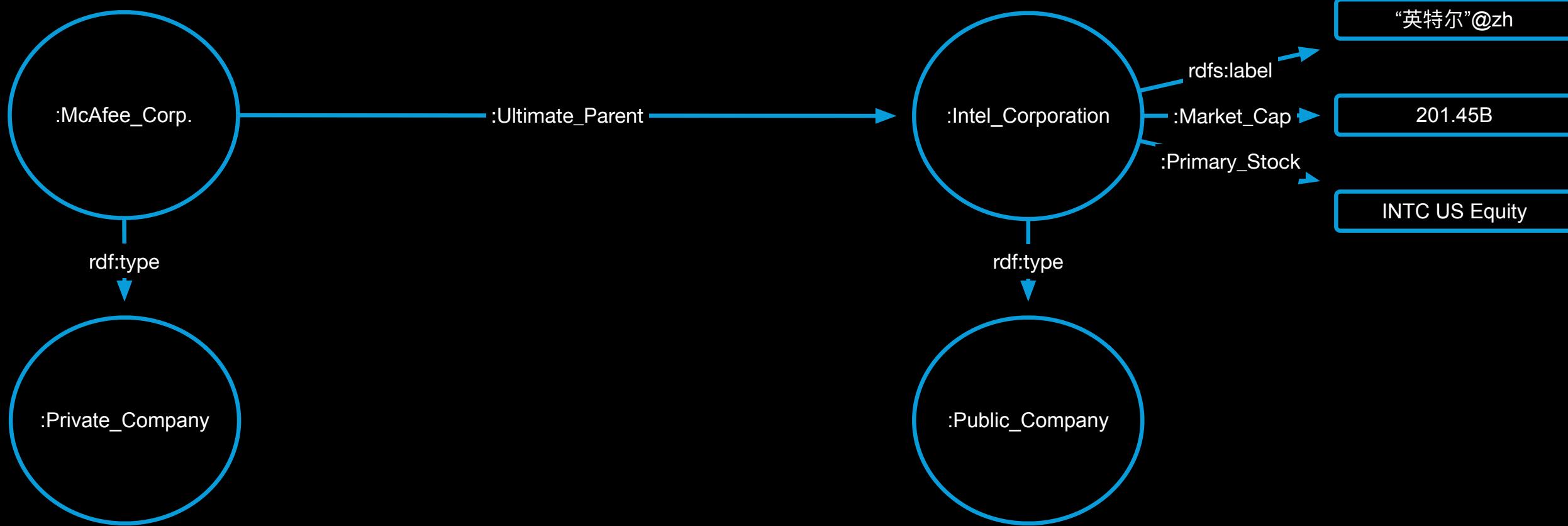
- Logical reasoning and rule-based inference
 - Structural/Hierarchical knowledge representation
 - Easily interpretable models and predictions

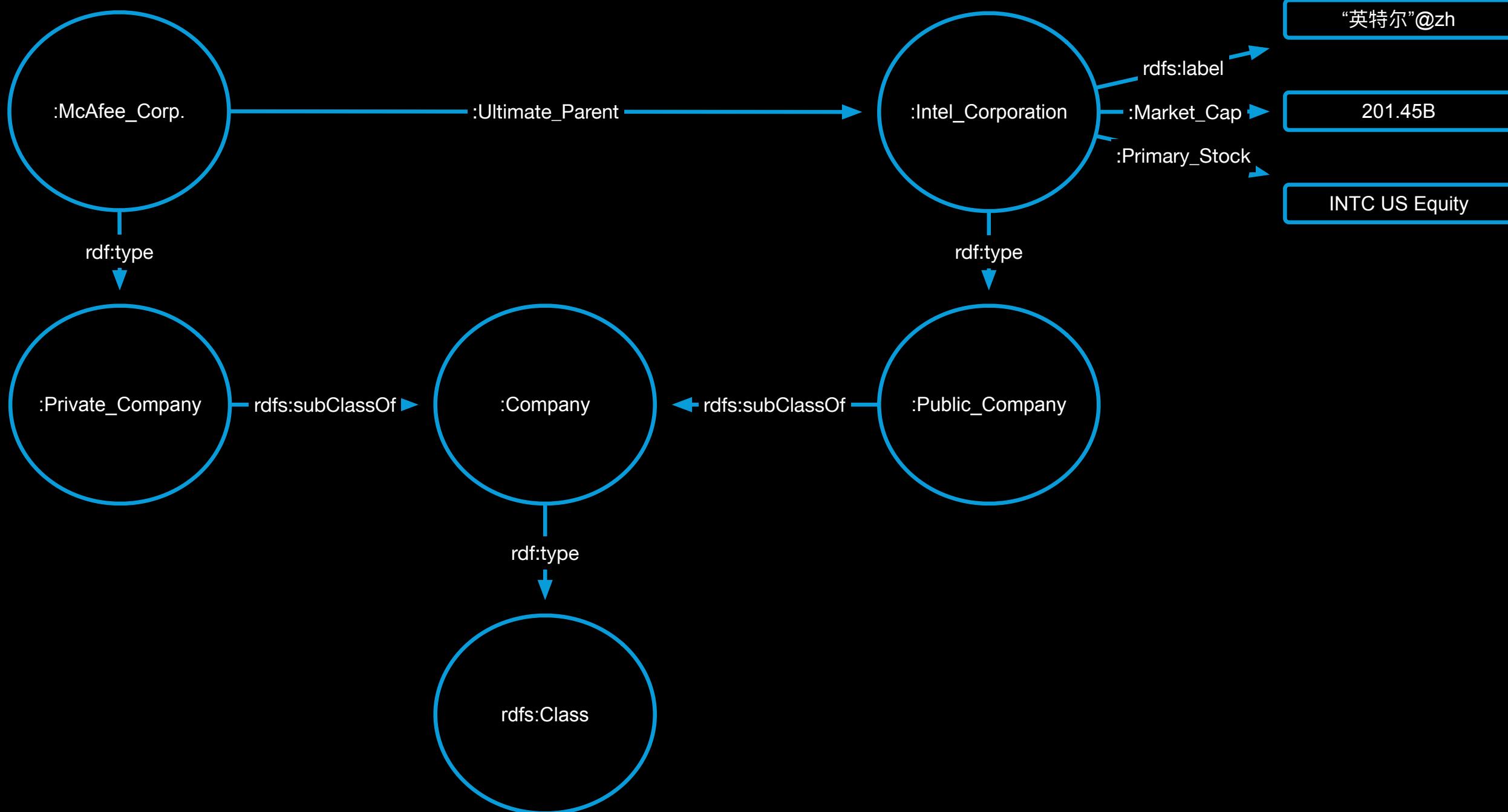
TechAtBloomberg.com

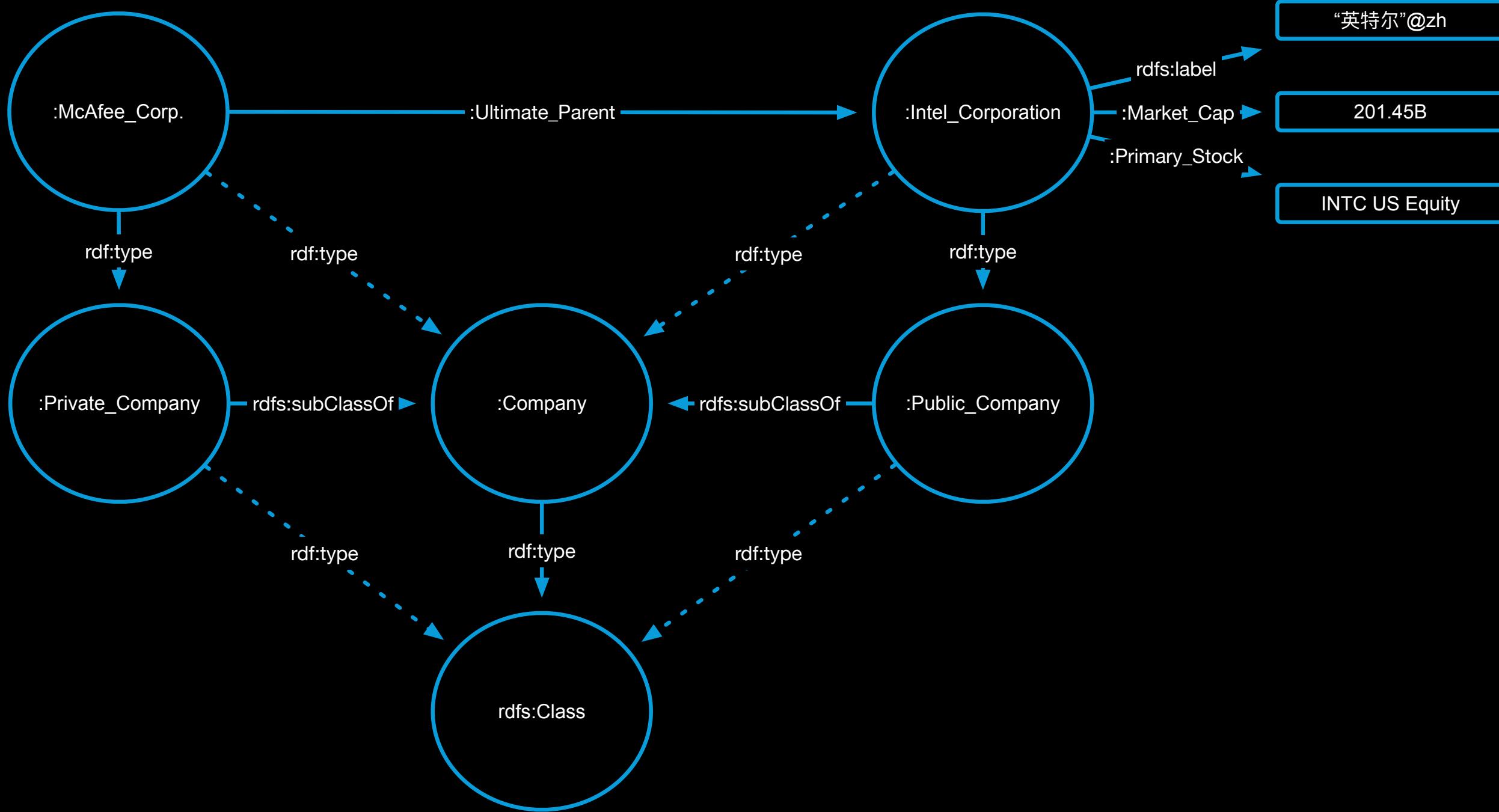
© 2023 Bloomberg Finance L.P. All rights reserved.
Edgar Meij – ECIR 23 Tutorial on Neuro-Symbolic Representations for IR

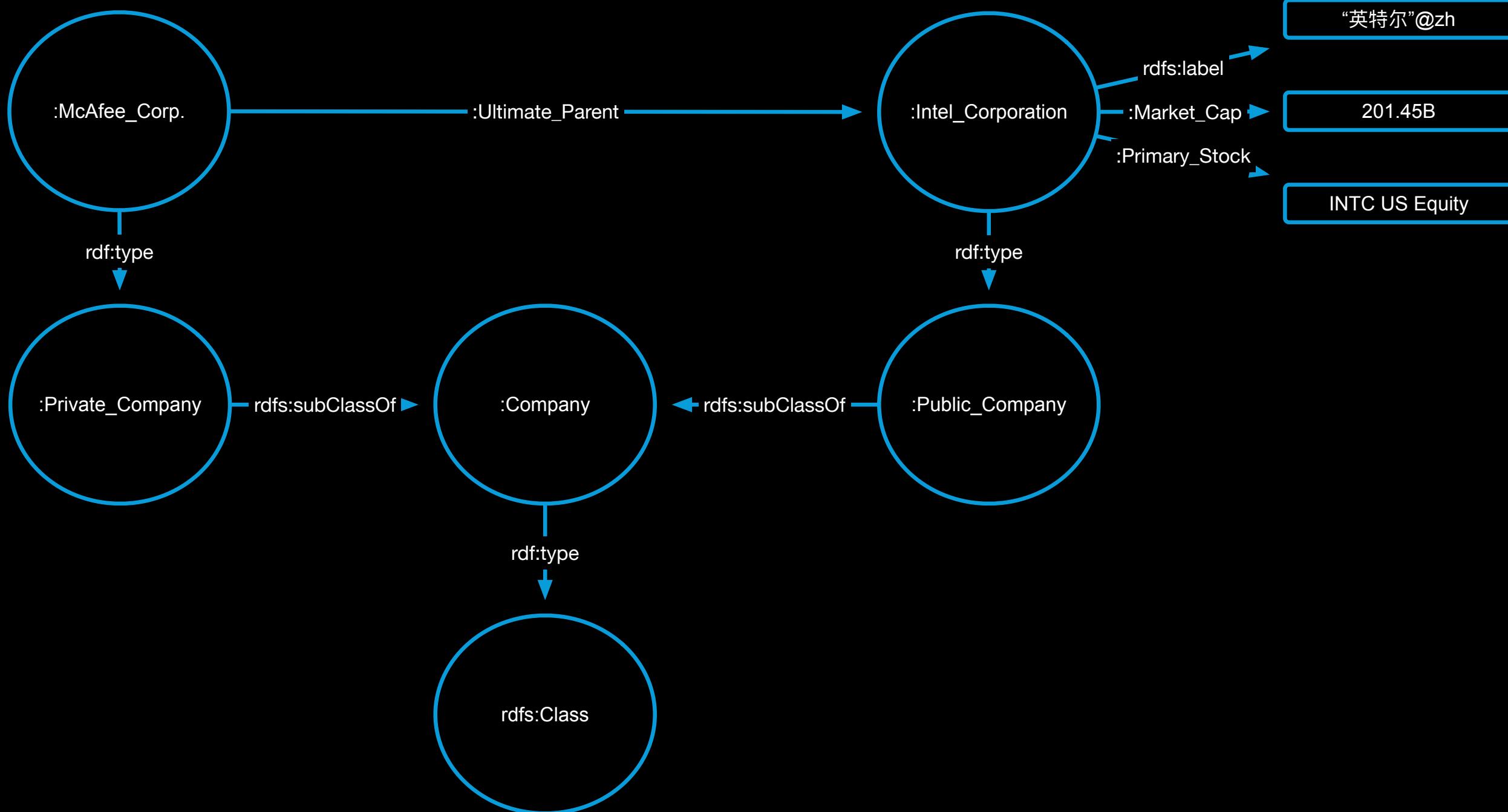
Bloomberg
Engineering

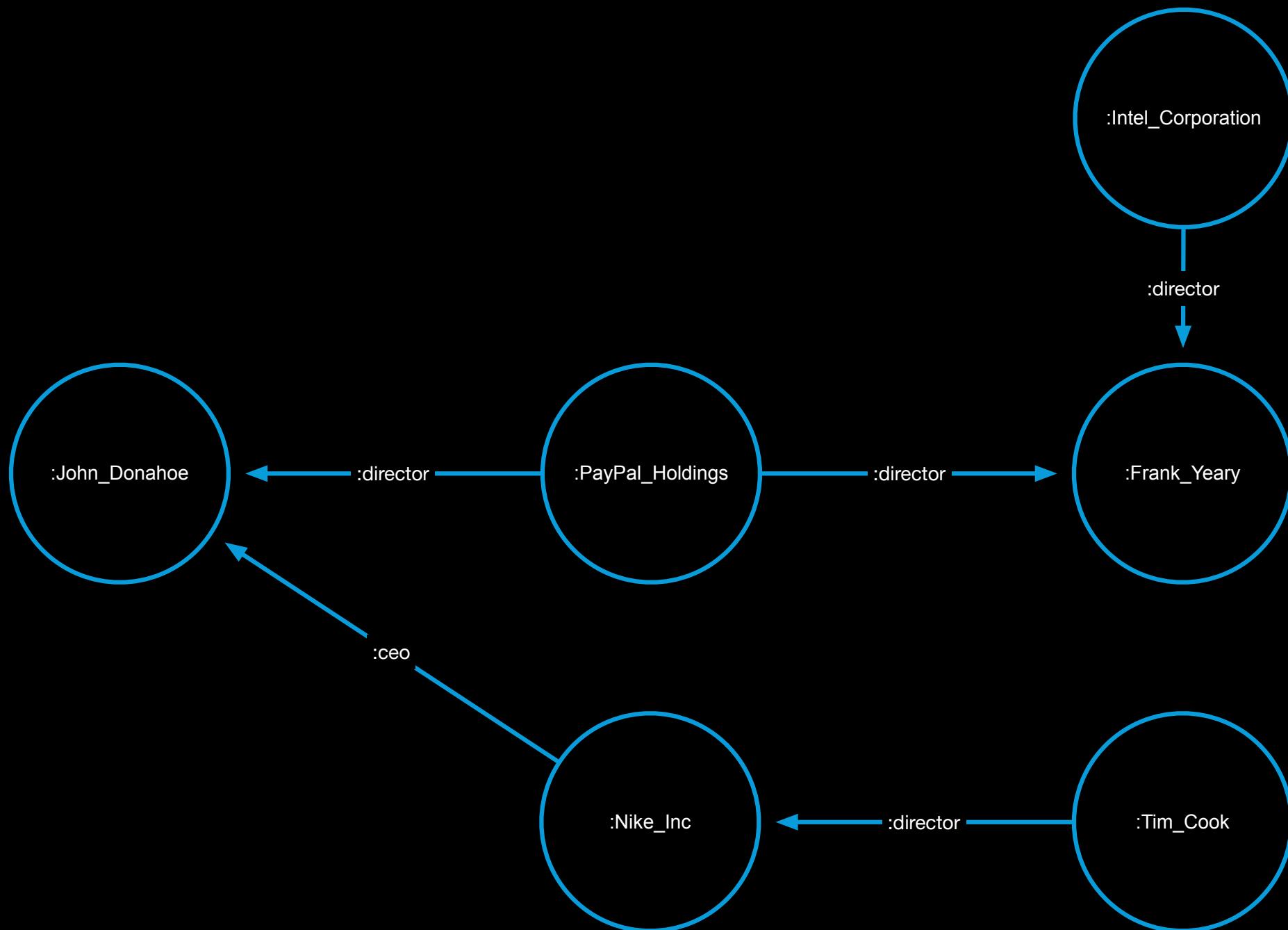




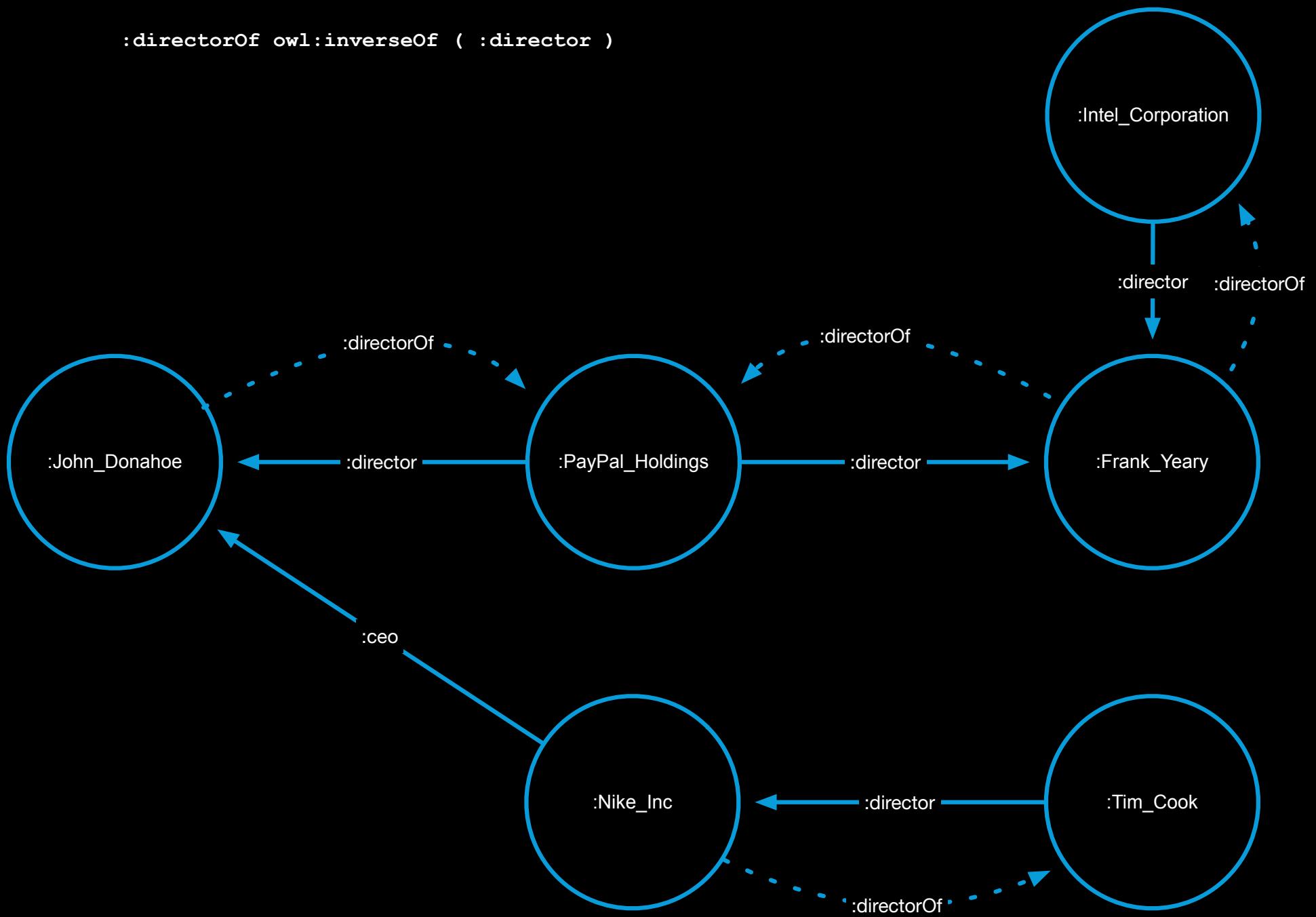




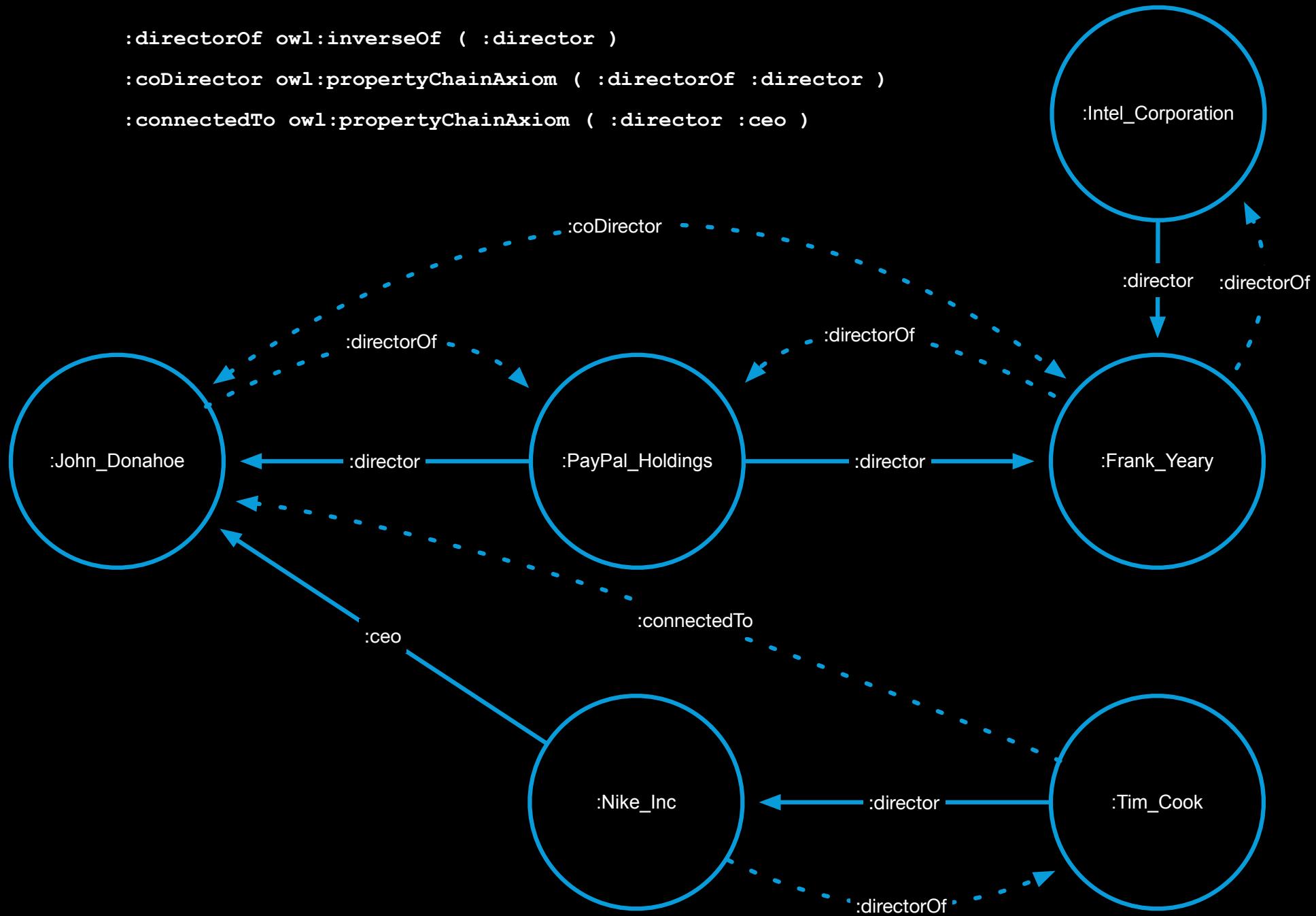




:directorOf owl:inverseOf (:director)



```
: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

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

THE AI REVOLUTION IS ON

A Little History of AI

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

TechAtBloomberg.com

© 2023 Bloomberg Finance L.P. All rights reserved.
Edgar Meij – ECIR 23 Tutorial on Neuro-Symbolic Representations for IR

Bloomberg
Engineering

Representations

TechAtBloomberg.com

© 2023 Bloomberg Finance L.P. All rights reserved.

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

© 2023 Bloomberg Finance L.P. All rights reserved.
Edgar Meij – ECIR 23 Tutorial on Neuro-Symbolic Representations for IR

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

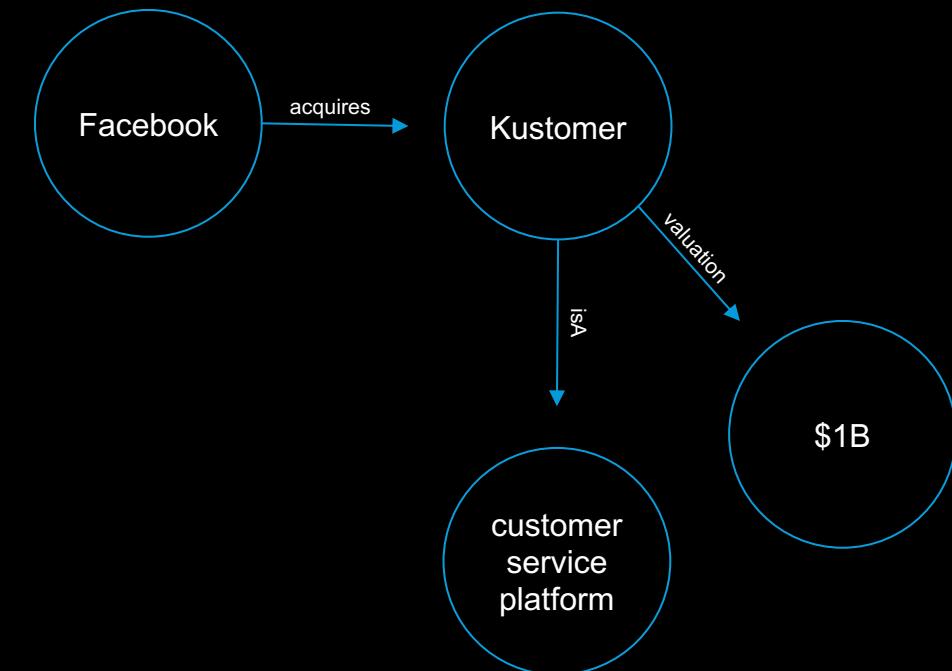
Representations

- **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*: feature and class
 - Poor at reasoning, hard to interpret
 - Needs a lot of training material

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.



TechAtBloomberg.com

Neurosymbolic AI

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

TechAtBloomberg.com

© 2023 Bloomberg Finance L.P. All rights reserved.
Edgar Meij – ECIR 23 Tutorial on Neuro-Symbolic Representations for IR

Bloomberg
Engineering

Neurosymbolic AI

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

The screenshot shows the homepage of the LM-KBC challenge at ISWC 2023. The title is "Knowledge Base Construction from Pre-trained Language Models (LM-KBC)". Below it, it says "Challenge @ 22nd International Semantic Web Conference (ISWC 2023)". There are three buttons: "Download dataset", "Discussions", and "Follow us". A section titled "Task Description" explains the challenge: "Pretrained language models (LMs) like chatGPT have advanced a range of semantic tasks and have also shown promise for knowledge extraction from the models itself. Although several works have explored this ability in a setting called probing or prompting, the viability of knowledge base construction from LMs remains underexplored. In the 2nd edition of this challenge, we invite participants to build actual disambiguated knowledge bases from LMs, for given subjects and relations. In crucial difference to existing probing benchmarks like LAMA (Petroni et al., 2019), we make no simplifying assumptions on relation cardinalities, i.e., a subject-entity can stand in relation with zero, one, or many object-entities. Furthermore, submissions need to go beyond just ranking predicted surface strings and materialize disambiguated entities in the output, which will be evaluated using established KB metrics of precision and recall." It then describes the task: "Formally, given the input subject-entity (s) and relation (r), the task is to predict all the correct object-entities ($\{o_1, o_2, \dots, o_k\}$) using LM probing." Below this, it lists three tracks: "Track 1: a small-model track with low computational requirements", "Track 2: an open track, where participants can use any LM of their choice", and "Track 3: a discovery track, where participants need to discover knowledge for emerging entities".

The screenshot shows the homepage of the NeSy 2023 workshop. The title is "NESY 2023" with a subtitle "17th INTERNATIONAL WORKSHOP ON NEURAL-SYMBOLIC LEARNING AND REASONING". It says "Certosa di Pontignano, Siena, Italy" and "3 to 5 July 2023". Below this, it states "NeSy2023 will host a TAILOR EU network workshop". A paragraph explains the workshop's purpose: "The NeSy workshop series is the longest standing gathering for the presentation and discussion of cutting edge research in neurosymbolic AI. NeSy is the annual meeting of the Neural-Symbolic Learning and Reasoning Association." To the right is the NeSy logo, which is a black umbrella with a white "NeSy" monogram.

TechAtBloomberg.com

© 2023 Bloomberg Finance L.P. All rights reserved.

Edgar Meij – ECIR 23 Tutorial on Neuro-Symbolic Representations for IR

Bloomberg
Engineering

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>

TechAtBloomberg.com

© 2023 Bloomberg Finance L.P. All rights reserved.

Engineering

Bloomberg

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