# Introduction to Knowledge Graph

-- The basics, storing schema, others

Luyue Zhao 11/27/2022

*"The one thing better than perfection, is standardization."*

# Abstract

- These slides will introduce the basic concept of knowledge graphs, applications, storing mechanisms, and others, of knowledge graphs.

# Index

- 1. Why knowledge graph?
- 2. What is knowledge graph?
- 3. Usage of knowledge graphs
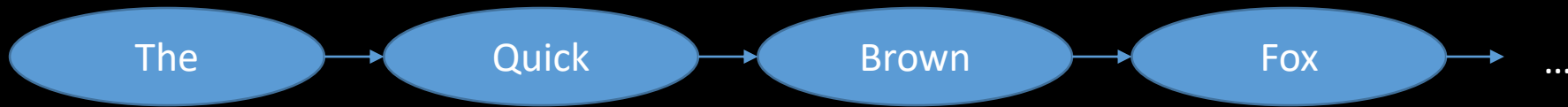- 4. Describe a knowledge graph

# Why knowledge graph?

- Graph is one of the most intuitive and omnipresent ways to represent information and data.
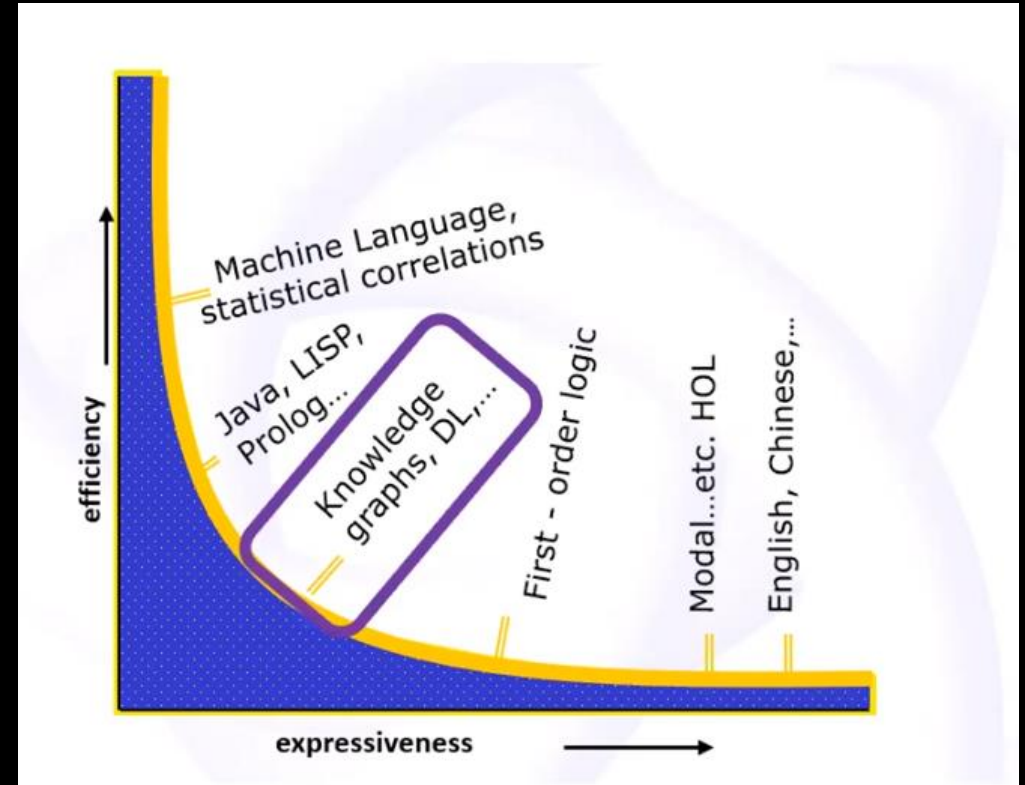
- Image is a graph.
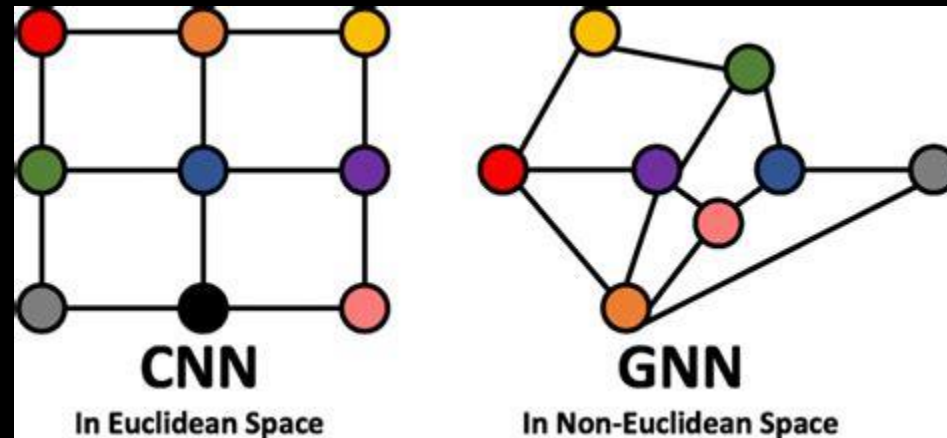
# Why knowledge graph?

- Text is a graph:

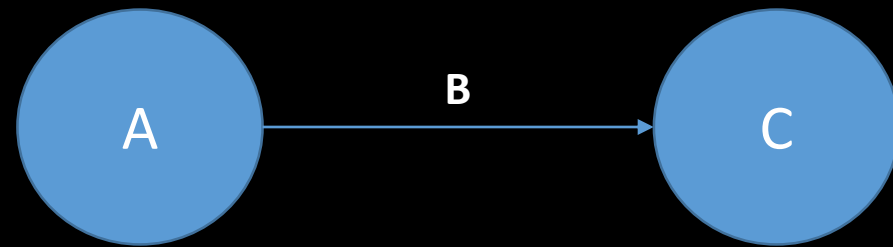```
The → Quick → Brown → Fox → …
```

# Why knowledge graph?

- Graphs stand in the middle between human and machine.

- Graph is understandable by human, and interpretable by machine.

- With the implementation in NLP, graphs are better suited than sentence embedding or Word2Vec in interpretability.

# What is knowledge graph?



Directed Labeled Graph
Nodes and Edges have well-defined meanings

# What is knowledge graph?



Directed Labeled Graph
Nodes and Edges have well-defined meanings
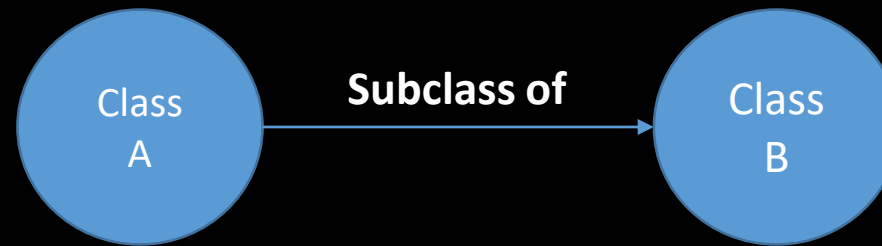
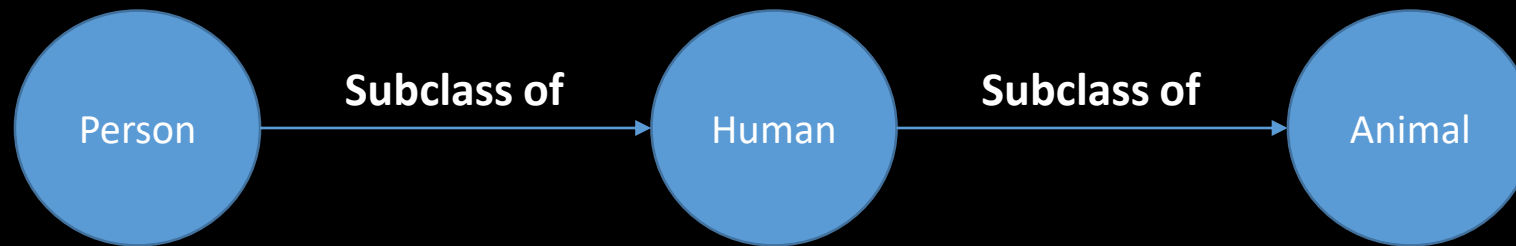# What is knowledge graph?



Directed Labeled Graph
Nodes and Edges have well-defined meanings

# What is knowledge graph?



Directed Labeled Graph
Nodes and Edges have well-defined meanings

# What is knowledge graph?



Directed Labeled Graph
Nodes and Edges have well-defined meanings

# How to define meanings?

- User's actions
  - "friend (request)" action
- Human Understandable languages
  - Wordnet
- Logical Specifications
  - Set of Axioms
- Associative definition
  - Cat – Images of cats
- Embeddings
  - Statistics on a corpus of texts; BERT

# Usage – Natural Language Processing

- Entity Extraction, Relation Extraction and reasoning

- **Albert Einstein** is a **German-born theoretical physicist** who developed the **theory of relativity**.
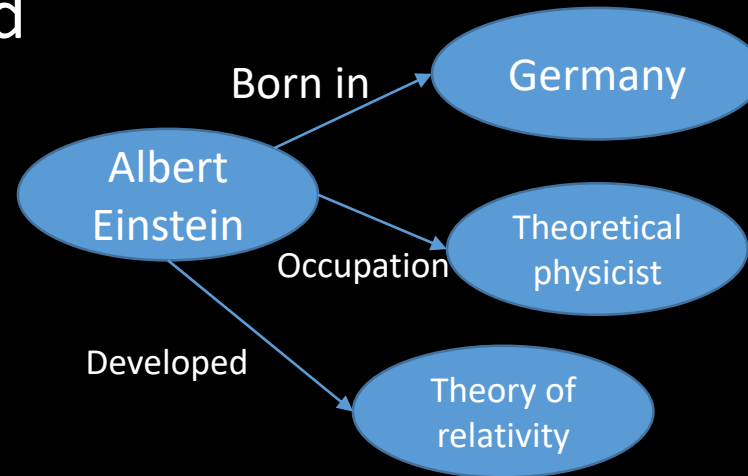
# Usage – Natural Language Processing

- Entity Extraction, Relation Extraction and reasoning

- Albert Einstein is a German-born theoretical physicist who developed the theory of relativity.
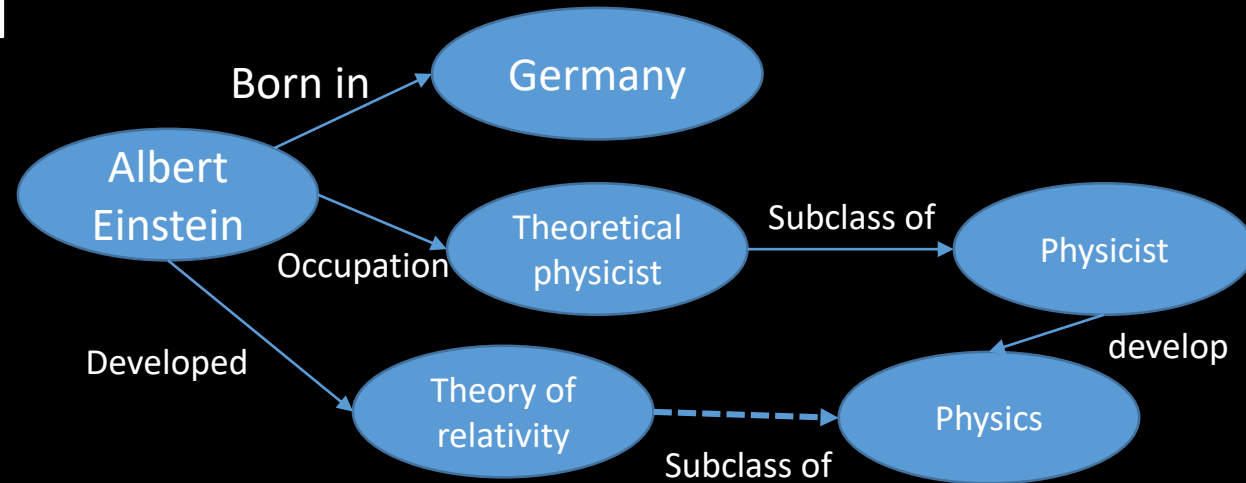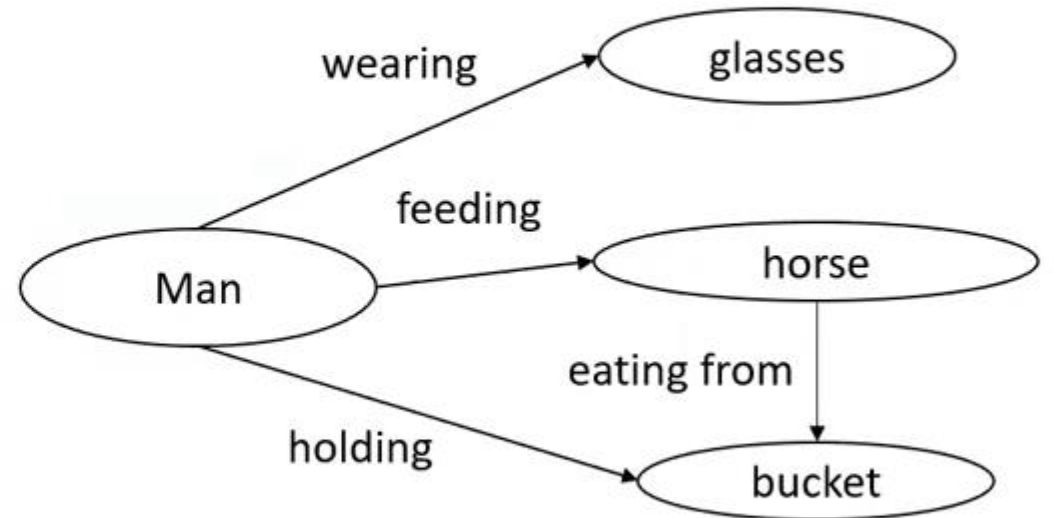
# Computer Vision
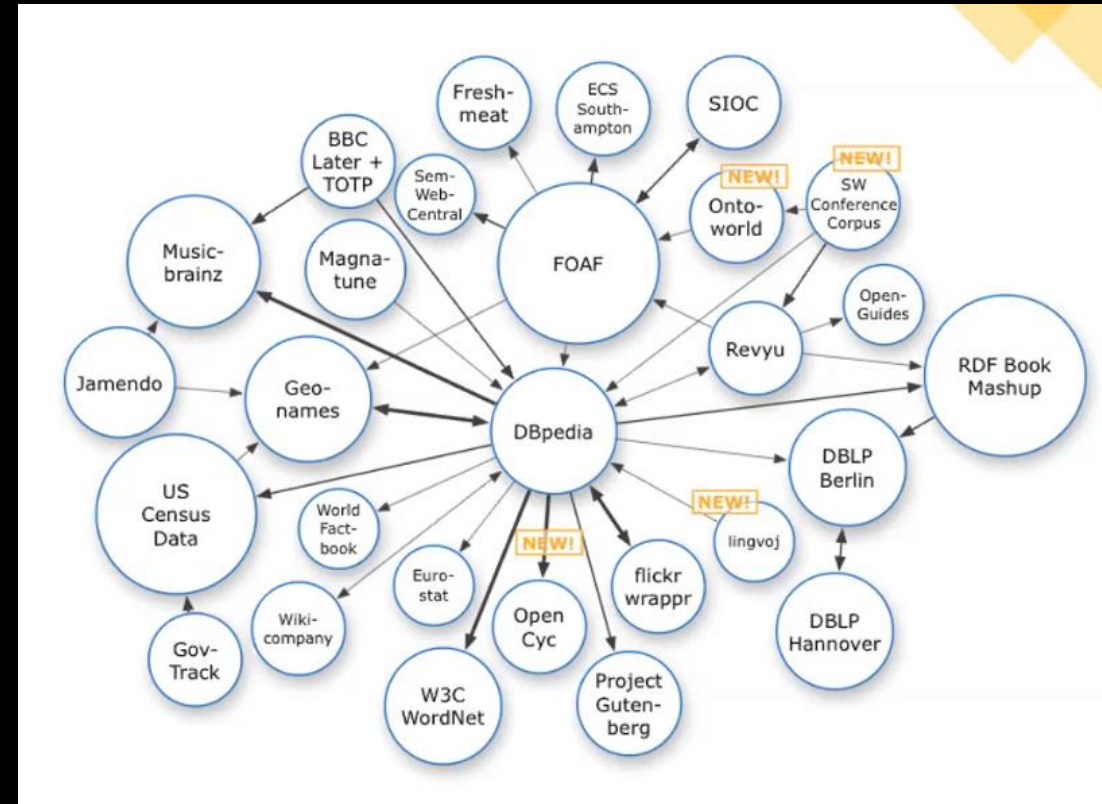
# Current Development

- From online heterogeneous, unconstructed data

- To structured information which can be reasoned with.

# Other Usage of Knowledge graph

- Web search engine (Google Knowledge Graph, Microsoft KG,…)

- Production Information (Amazon, Ebay,…)

- Question & Answer

- Scientific analysis:
  - Search, discovery, data exploration & dissection
  - Heterogeneous data connections

# What is "knowledge graph?"

- Closed-world assumption: "what we know."

- Open-world assumption: "what we have".

- The "atom" of knowledge graphs are always the entities and relations between them (i.e. triplets).

# What is "knowledge graph?"

| Definition | Source |
|---|---|
| "A knowledge graph (i) mainly describes real world entities and their interrelations, organized in a graph, (ii) defines possible classes and relations of entities in a schema, (iii) allows for potentially interrelating arbitrary entities with each other and (iv) covers various topical domains." | Paulheim [16] |
| "Knowledge graphs are large networks of entities, their semantic types, properties, and relationships between entities." | Journal of Web Semantics [12] |
| "Knowledge graphs could be envisaged as a network of all kind things which are relevant to a specific domain or to an organization. They are not limited to abstract concepts and relations but can also contain instances of things like documents and datasets." | Semantic Web Company [3] |
| "We define a Knowledge Graph as an RDF graph. An RDF graph consists of a set of RDF triples where each RDF triple $(s, p, o)$ is an ordered set of the following RDF terms: a subject $s \in U \cup B$, a predicate $p \in U$, and an object $U \cup B \cup L$. An RDF term is either a URI $u \in U$, a blank node $b \in B$, or a literal $l \in L$." | Färber et al. [7] |
| "[...] systems exist, [...], which use a variety of techniques to extract new knowledge, in the form of facts, from the web. These facts are interrelated, and hence, recently this extracted knowledge has been referred to as a knowledge graph." | Pujara et al. [17] |

Table 1: Selected definitions of knowledge graph

# Resource Description Framework

- RDF stands for **R**esource **D**escription **F**ramework

- designed to be read and understood by computers

- written in XML

- Constructed on triplets

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:cd="http://www.recshop.fake/cd#">

<rdf:Description
rdf:about="http://www.recshop.fake/cd/Empire Burlesque">
  <cd:artist>Bob Dylan</cd:artist>
  <cd:country>USA</cd:country>
  <cd:company>Columbia</cd:company>
  <cd:price>10.90</cd:price>
  <cd:year>1985</cd:year>
</rdf:Description>

<rdf:Description
rdf:about="http://www.recshop.fake/cd/Hide your heart">
  <cd:artist>Bonnie Tyler</cd:artist>
  <cd:country>UK</cd:country>
  <cd:company>CBS Records</cd:company>
  <cd:price>9.90</cd:price>
  <cd:year>1988</cd:year>
</rdf:Description>
.
.
.
</rdf:RDF>
```
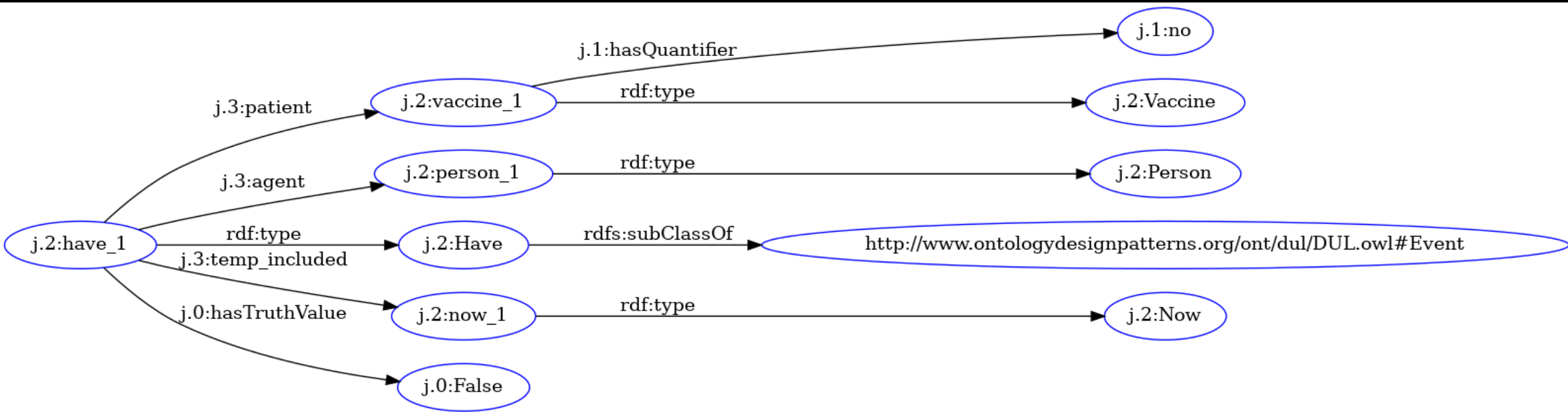
# Resource Description Framework

URI(Uniform Resource Identifier)

j.0: Abbreviations

The word "bad", begin from index
of 15(nonnegativeinteger), end at 18, and
writes like a string "bad", which is
interpreted as the fred domain object:
"ont/fred/domain.owl#bad", which are
referred elsewhere.

```
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:j.0="http://www.ontologydesignpatterns.org/ont/boxer/boxing.owl#"
    xmlns:j.1="http://ontologydesignpatterns.org/cp/owl/semiotics.owl#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:j.2="http://www.essepuntato.it/2008/12/earmark#"
    xmlns:j.3="http://www.ontologydesignpatterns.org/ont/fred/domain.owl#"
    xmlns:j.4="http://www.ontologydesignpatterns.org/ont/boxer/boxer.owl#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:j.5="http://www.ontologydesignpatterns.org/ont/fred/pos.owl#"
    xmlns:j.6="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#" >
<rdf:Description rdf:about="http://www.essepuntato.it/2008/12/earmark#hasContent">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.essepuntato.it/2008/12/earmark#refersTo">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#hasQuality">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.ontologydesignpatterns.org/ont/fred/domain.owl#docuverse">
    <rdf:type rdf:resource="http://www.essepuntato.it/2008/12/earmark#StringDocuverse"/>
    <j.2:hasContent rdf:datatype="http://www.w3.org/2001/XMLSchema#string">coronavirus is bad</j.2:hasContent>
</rdf:Description>
<rdf:Description rdf:about="http://www.ontologydesignpatterns.org/ont/fred/domain.owl#offset_15_18_bad">
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">bad</rdfs:label>
    <j.5:pennpos rdf:resource="http://www.ontologydesignpatterns.org/ont/fred/pos.owl#JJ"/>
    <j.2:begins rdf:datatype="http://www.w3.org/2001/XMLSchema#nonNegativeInteger">15</j.2:begins>
    <j.2:ends rdf:datatype="http://www.w3.org/2001/XMLSchema#nonNegativeInteger">18</j.2:ends>
    <rdf:type rdf:resource="http://www.essepuntato.it/2008/12/earmark#PointerRange"/>
    <j.1:hasInterpretant rdf:resource="http://www.ontologydesignpatterns.org/ont/fred/domain.owl#Bad"/>
    <j.2:refersTo rdf:resource="http://www.ontologydesignpatterns.org/ont/fred/domain.owl#docuverse"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.essepuntato.it/2008/12/earmark#begins">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</rdf:Description>
<rdf:Description rdf:about="http://ontologydesignpatterns.org/cp/owl/semiotics.owl#hasInterpretant">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.ontologydesignpatterns.org/ont/fred/domain.owl#now_1">
    <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/fred/domain.owl#Now"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.ontologydesignpatterns.org/ont/fred/domain.owl#Bad">
    <j.5:boxerpos rdf:resource="http://www.ontologydesignpatterns.org/ont/fred/pos.owl#a"/>
```

Namespaces:
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
j.0: http://www.ontologydesignpatterns.org/ont/boxer/boxing.owl#
j.1: http://www.ontologydesignpatterns.org/ont/fred/quantifiers.owl#
j.2: http://www.ontologydesignpatterns.org/ont/fred/domain.owl#
j.3: http://www.ontologydesignpatterns.org/ont/boxer/boxer.owl#
rdfs: http://www.w3.org/2000/01/rdf-schema#

We have no vaccine

# But this is not enough…

- Having a unique URI doesn't mean computer can reason with the data.

- For example:
  - What is "begins", labeled by "http://www.essepuntato.it/2008/12/earmark#begins"?
  - URI is not URL, you cannot open it.

- To truly achieve a globally recognizable knowledge graph, we need a standard.

# Schema.org

- Schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond.

- Schema.org types defines the restraints that could be utilized for knowledge reasonings.

# OWL (Web Ontology Language)

- OWL is a ontology programming and reasoning language enabling sophisticated reasoning in knowledge graphs.

- Unlike schema.org, OWL is designed with knowledge reasoning as a primary objective. Each axiom defined by OWL can be directly translated into precise mathematical logic axioms. Such that for example, $rdfs:domain$ $and$ $rdfs:range$ are precisely mathematically defined, while $schema:domaininclude, schema:rangeinclude$ are not.

# Conclusion

- Knowledge graph: using in translate human knowledge to machine interpretable graphs.

- Directed Edge Graph, Triplets, RDF format.

- Broad usage and bottom-up development.

- Standards: schema.org, OWL,…