

Foundations of Knowledge Graphs

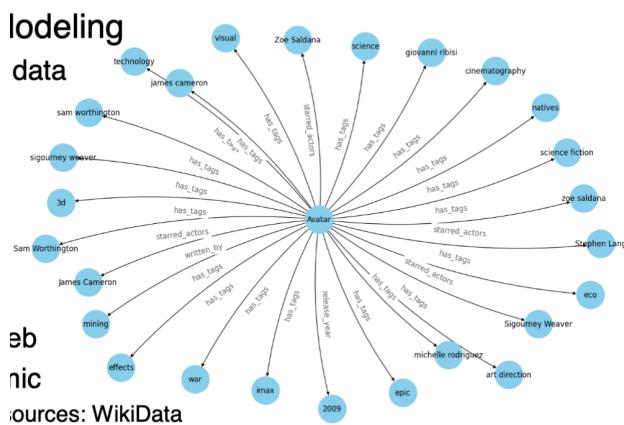
MIE223
Winter 2025

1 Knowledge Graphs

1.1 Knowledge Graphs

Knowledge Graphs

- “Triple” data on web
- Distributed, dynamic
 - Some controlled sources: WikiData
- Nodes are entities
- Edges are labeled by relations (linked data)
- Foundation in RDF



1.2 The Resource Description Framework Is All About Making Statements

- A statement can be thought of as an ordered triple composed of three items: (resource, predicate, value)
- A Resource is anything that can be identified.
- A Predicate is a property name that has a URI. The Predicate may or may not actually be resolvable.
- A Value is another Resource or a literal.
- Statements may be represented in RDF XML, abbreviated RDF XML, or (locally) graphs
- We want to state world knowledge in a structured format

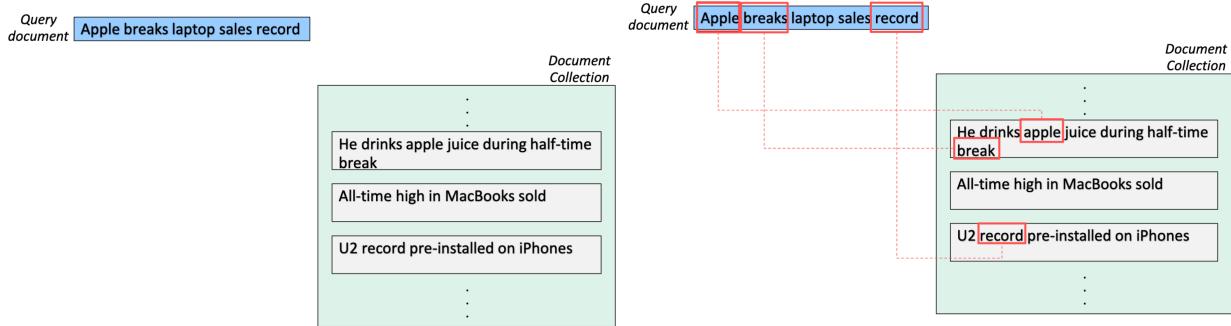
Why?

- A Dynamic Source of Structured Data
 - Query it dynamically on the web from different sources (SPARQL)
 - Inherently linked (entities and relations are shared) across sources
 - * See schema.org

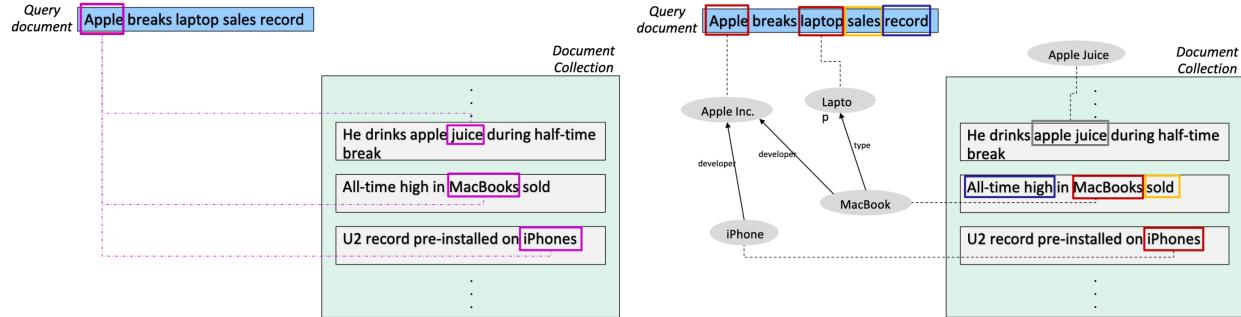
- Question Answering
 - When was William G. Davis born?
 - Who was the first woman in space? What year?
- Entity Resolution
 - You search for Michael Jordan
 - The basketball player, the professor, the general?
 - Important for search

2 Entity Resolution

2.1 Common task: Information Retrieval (i.e., Google)



Matching words do not always indicate similarity. the query does not match the ideal article
Word co-occurrence can be misleading, too. apple still matches apple juice.



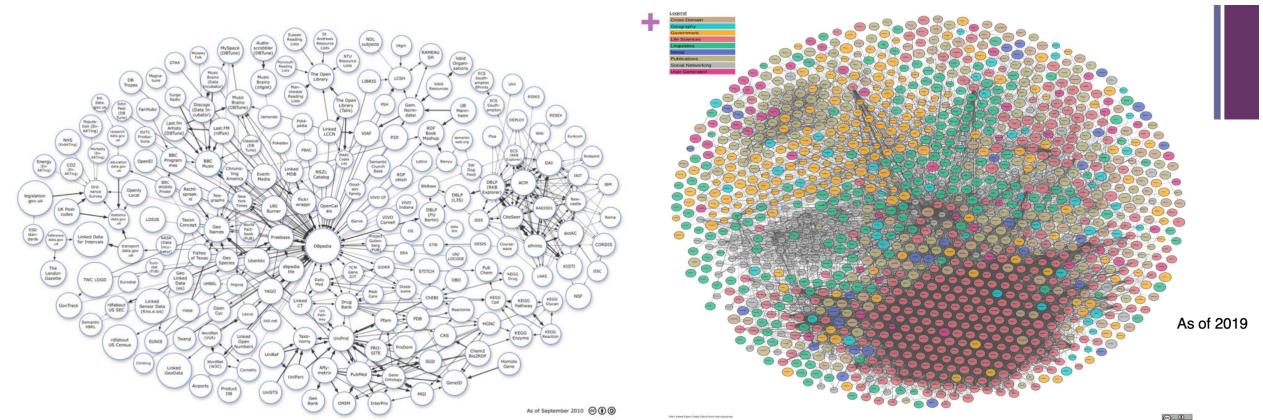
Knowledge Graph Technologies: resolve ambiguity & exploit relational knowledge We want apple to refer to the Apple company. Extract noun phrases and resolve conflicts

3 Resource Description Framework (RDF)

3.1 First: Let's get some RDF

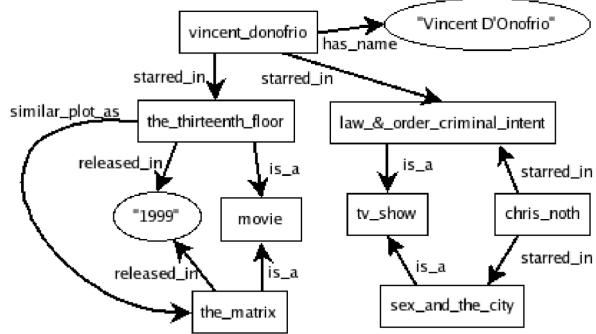
```
curl --include --location --header "Accept:application/rdf+xml"
http://dbpedia.org/resource/Yukihiro_Matsumoto
How about Satoshi_Nakamoto?
For human readable information, see the corresponding
https://en.wikipedia.org/wiki/Satoshi_Nakamoto
curl --include --location --header "Accept:application/rdf+xml"
http://dbpedia.org/resource/Sirius
Who is this for?
The web for programmers.
Can we describe things like we do people and stars?
```

3.2 Each of these stores many many RDF triples.



3.3 A Knowledge Graph from Triples

most graphs are directional, right is triple format



Start Node	Edge Label	End Node
(vincent_donofrio, (law_&_order_ci, the_thirteenth_floor, similar_plot_as, the_matrix)	starred_in, is_a, starred_in, starred_in	law_&_order_ci), tv_show), the_matrix)

In Dbpedia, most widely used linking predicates are owl:sameAs, rdfs:seeAlso, foaf:knows

3.4 Different Formats (RDF/XML)

```

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ex="http://www.example.org/">
  <rdf:Description rdf:about="http://www.example.org/vincent_donofrio">
    <ex:starred_in>
      <ex:tv_show rdf:about="http://www.example.org/law_and_order_ci" />
    </ex:starred_in>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.example.org/the_thirteenth_floor">
    <ex:similar_plot_as rdf:resource="http://www.example.org/the_matrix" />
  </rdf:Description>
</rdf:RDF>
  
```

In XML

```

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
  xmlns:edu="http://www.example.org/">
  <rdf:Description rdf:about="http://www.princeton.edu">
    <geo:lat>40.35</geo:lat>
    <geo:long>-74.66</geo:long>
    <edu:hasDept rdf:resource="http://www.cs.princeton.edu"/>
    <dc:title>Department of Computer Science</dc:title>
  </rdf:Description>
</rdf:RDF>
  
```

Another example

3.5 As A Table

Subject	Predicate	Object
<http://www.princeton.edu>	edu:hasDept	<http://www.cs.princeton.edu>
<http://www.princeton.edu>	geo:lat	"40.35"
<http://www.princeton.edu>	geo:long	"-74.66"
<http://www.cs.princeton.edu>	dc:title	"Department of Computer Science"

3.6 A triple may be 3 URI's

How do we say "Billy Holiday was a songwriter" ?

With three URI's:

 http://dbpedia.org/resource/Billie_Holiday
 http://dbpedia-owl:occupation
 http://dbpedia.org/page/Songwriter

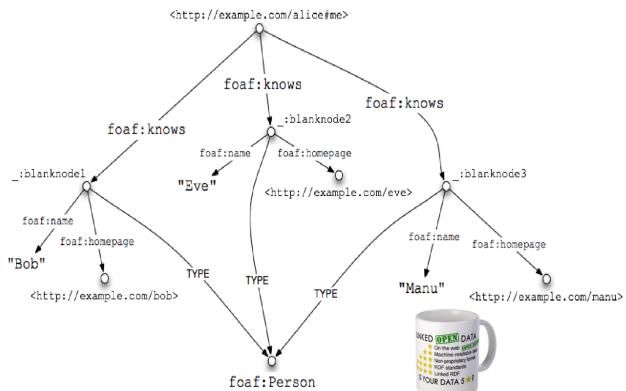
In an ontology, we might learn that that all
Songwriters are People.

What deduction could we make?

3.7 RDFA and RDF

- RDF stands on its own. Has many representations.
- RDFA is a lightweight version of RDF for web pages.
- RDFA is being used today by search engines like Google and sites like Best Buy.
- JSON-LD is being supported by W3C for WoT.

With RDFA, we can make these statements in a web page.



3.8 Adding RDFA

Describing persons

```
<div xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <ul>
    <li typeof="foaf:Person">
      <a href="http://example.com/bob/">Bob</a>
    </li>
    <li typeof="foaf:Person">
      <a href="http://example.com/eve/">Eve</a>
```

```

        </li>
    <li typeof="foaf:Person">
        <a href="http://example.com/manu/">Manu</a>
    </li>
</ul>
</div>

```

3.9 Add Homepages

Describing relationships with rel attribute.

```

<div xmlns:foaf="http://xmlns.com/foaf/0.1/">
    <ul>
        <li typeof="foaf:Person">
            <a rel="foaf:homepage" href="http://example.com/bob/">Bob</a>
        </li>
        <li typeof="foaf:Person">
            <a rel="foaf:homepage" href="http://example.com/eve/">Eve</a>
        </li>
        <li typeof="foaf:Person">
            <a rel="foaf:homepage" href="http://example.com/manu/">Manu</a>
        </li>
    </ul>
</div>

```

3.10 Summary

- In MIE223, we focus on Knowledge Graphs (KGs) as a structured data source
 - I.e., we query SPARQL and load it into Pandas table
 - * We're often querying one or two sources
 - * This is probably the most common use of Knowledge Graphs
 - You'll see a lot of SQL next year in Data Modelling
 - * For data stored internally within an organization
- But KGs are part of the Semantic Web
 - The idea that structured data can be distributed around the web
 - Updated dynamically
 - Retrieved like you would with a Google Search
 - * This use of KGs has not quite taken off the way single well-curated sources have taken off... consider why