Medical Informatics

Lecture 12: Linked Data

Dr Areti Manataki

Nanjing Medical University

In the previous lecture

- Triple & graph patterns

- SPARQL query structure

- Using OPTIONAL & solution modifiers

- Querying multiple sources

SPARQL query
```
PREFIX : <http://usher.ed.ac.uk/medinf/resource>
PREFIX uv: <http://usher.ed.ac.uk/medinf/vocab/>
SELECT ?film
WHERE {:JamesDean    uv:playedIn    ?film . }
```

In this lecture

- Translating data in a relational database into the RDF model
- Linked Data principles

From RDB to RDF

From RDB to RDF

There is a standard mapping from records in a DB to RDF triples:

- each field (column) label is mapped to an RDF Predicate

- the data in each corresponding field is mapped into the Object

- the Subject of the triple corresponds to the primary key of the record (or it can be a blank node, but we're not covering these in this course)

DB Records as Triples

	Predicate	
Subject	Object	

DB Records as Triples

	Name	Artist	Place	
ID0039	"The Red Vineyard"	V. Van Gogh	Arles	

```
:Paintings_0039   uv:name    "The Red Vineyard" ;
                  uv:artist  db:Vincent_Van_Gogh ;
                  uv:place   db:Arles .
```

Linked Data principles

Linked Data Principles

1. Use URIs as names for things.
2. Use HTTP URIs, so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things.

1ˢᵗ principle

Use URIs as names for things.

- Use URI references to identify not just digital content, but also real world objects, abstract concepts, as well as relationships between objects or concepts.

- URIs allow us to uniquely identify resources, thus avoiding any ambiguity.

2nd principle

Use HTTP URIs, so that people can look up those names.

- Globally unique names can be created in a decentralised fashion by domain name owners; no central naming authority is required.

- Not just a name, but a means of accessing information describing the identified entity.

  - Any HTTP URI should be dereferenceable: HTTP clients can look up the URI using the HTTP protocol and retrieve a description of the resource that is identified by the URI

3ʳᵈ principle

When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).

Why RDF?

- Easy to insert RDF links between data from different sources.

- Information from different sources can be combined by graph merging.

- Information using different schemas can be expressed in a single graph, i.e. by mixing different vocabularies.

- Data can be tightly or loosely structured.

4ᵗʰ principle

Include links to other URIs, so that they can discover more things.

Kinds of links:

- Relationship Links point at related things in other data sources.

- Identity Links point at URI aliases used by other data sources to identify the same real-world object or abstract concept.

- Vocabulary Links point from data to the definitions of the vocabulary terms that are used to represent the data.

4ᵗʰ principle

Include links to other URIs, so that they can discover more things.

Kinds of links:

- Relationship Links point at related things in other data sources.
    - `:areti   foaf:based_near   dbpedia:Edinburgh .`
- Identity Links point at URI aliases used by other data sources to identify the same real-world object or abstract concept.
- Vocabulary Links point from data to the definitions of the vocabulary terms that are used to represent the data.

4ᵗʰ principle

Include links to other URIs, so that they can discover more things.

Kinds of links:

- Relationship Links point at related things in other data sources.

- Identity Links point at URI aliases used by other data sources to identify the same real-world object or abstract concept.
    - `dbpedia:James_Dean owl:sameAs freebase:James_Dean .`

- Vocabulary Links point from data to the definitions of the vocabulary terms that are used to represent the data.

4ᵗʰ principle

Include links to other URIs, so that they can discover more things.

Kinds of links:

- Relationship Links point at related things in other data sources.
- Identity Links point at URI aliases used by other data sources to identify the same real-world object or abstract concept.
- Vocabulary Links point from data to the definitions of the vocabulary terms that are used to represent the data.

```
:James_Dean a terms:Actor .
    terms:Actor a rdfs:Class ;
        rdfs:subClassOf umbel-rc:Artist .
```

Linked Data: Why bother?

- Linked Data provides a more generic, more flexible publishing paradigm which makes it easier for data consumers to discover and integrate data from large numbers of data sources.

- A unifying data model (RDF), which has been designed for global data sharing.

- A standardised data access mechanism (HTTP), which allows data sources to be accessed using generic data browsers and enables the complete data space to be crawled by search engines.

Linked Data: Why bother?

- Hyperlink-based data discovery: links connect all Linked Data into a single global data space and enable Linked Data applications to discover new data sources at run-time.

- Self-descriptive data: vocabulary definitions are recoverable like other data, and vocabulary terms can be linked to one another, hence easing the integration of data from different sources .

Linked Open Data

Linked Open Data

Linked Datasets as of August 2014
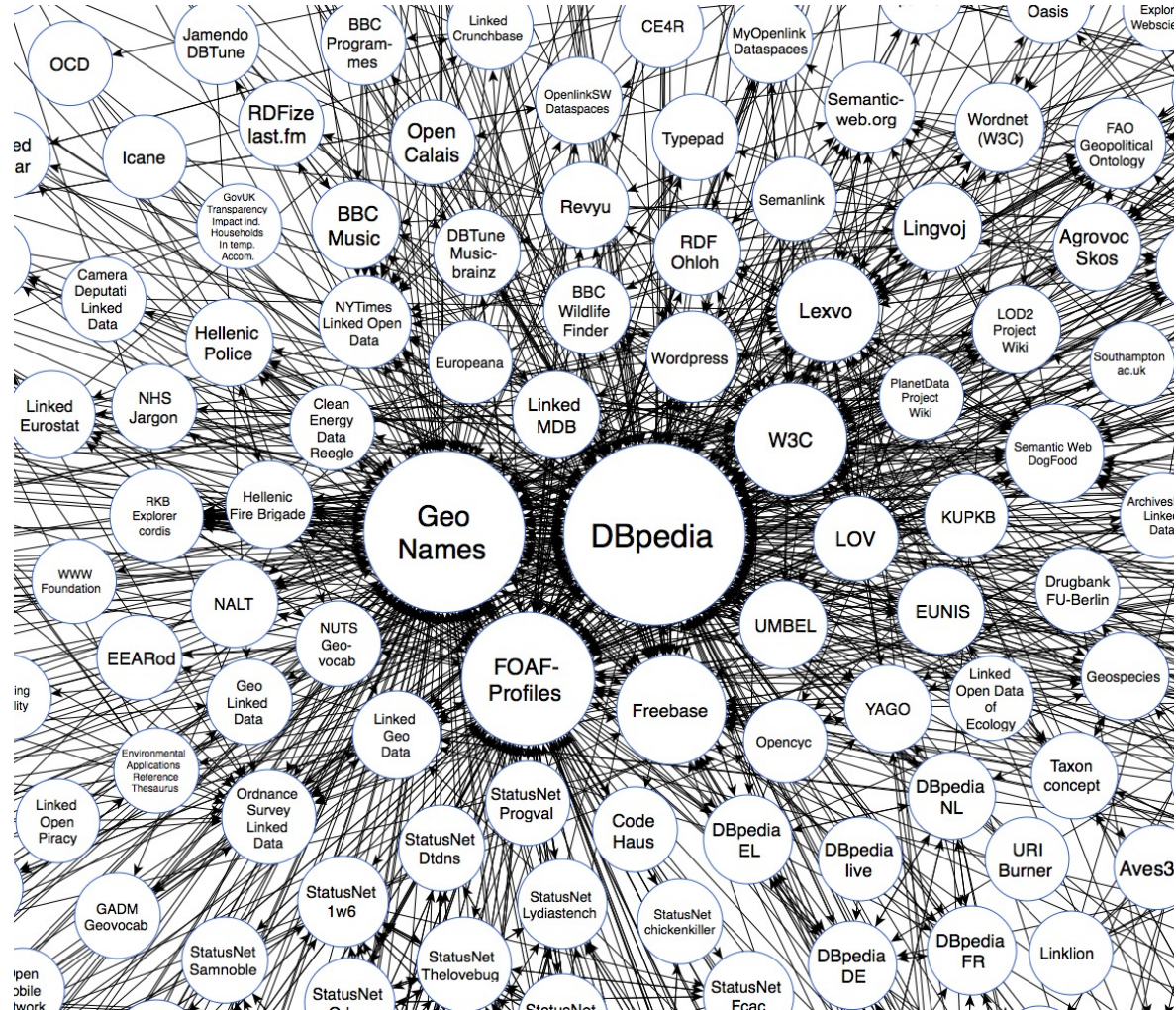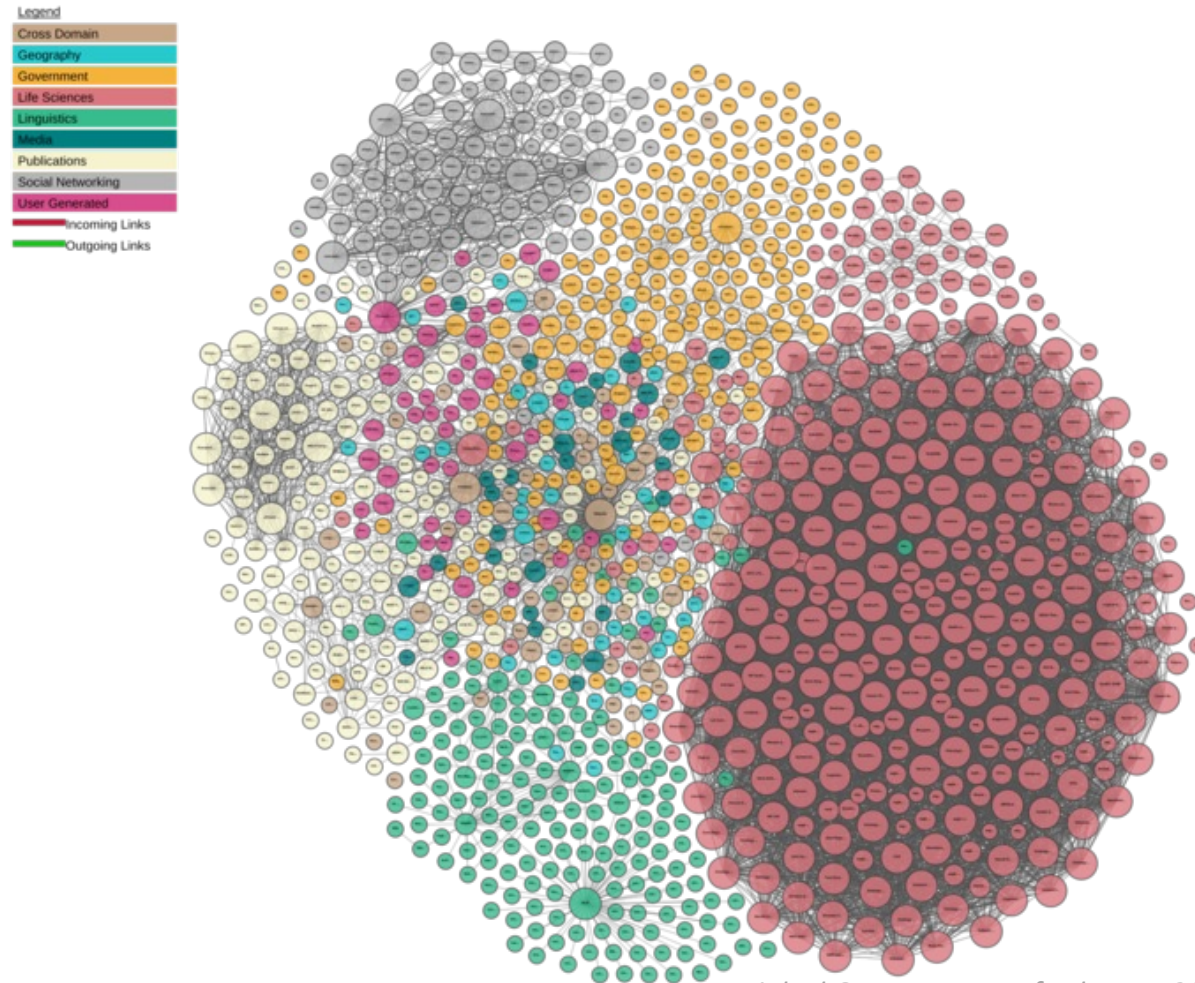
Linked Open Data

Linked Open Data as of February 2017.
[Image by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch
and Richard Cyganiak on Wikimedia, licensed as CC BY-SA 3.0]

Conclusions

- Data represented in the relational model can be translated to RDF triples.

  - Several conversion tools available: https://www.w3.org/wiki/ConverterToRdf#SQL

- Linked Data principles:

  1. Use URIs as names for things.

  2. Use HTTP URIs, so that people can look up those names.

  3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).

  4. Include links to other URIs, so that they can discover more things.

Acknowledgements

The content of these slides was originally created for the Medical Informatics course from The University of Edinburgh, which is licensed under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) license.

These lecture slides are also licensed under a CC BY-SA 4.0 license.