U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Semantic Web and Linked Data

M.EIC FEUP

Liliana Ferreira / Mariana Malta

S1 2025/26

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Class 2: Learning Objectives

- Review some fundamental concepts of Web Semantics and Linked data
- Real-World Applications & Group Exercise
- Practical work by-laws and group selection

The Semantic Web

"The **Semantic Web** is an extension of the current web in which information is given well-defined **meaning,** better enabling computers and people to **work in co-operation**."

[Berners-Lee *et al*, 2001]

Key Concepts

- What is Web Semantics?
  - Turning the web into a database through structured, meaningful data.

- What is Linked Data?
  - Connecting data across different domains for more meaningful, interoperable use.

Web Semantics

- Semantic Web: Giving data meaning (metadata)

- Key Technologies:
  - RDF (Resource Description Framework)
  - OWL (Web Ontology Language)
  - SPARQL (Query Language)

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Linked Open Data Principles

Linked Data

Tim Berners-Lee's 4 Principles:

1. Use Uniform Resource Identifiers (URI) to identify resources.
2. Use HTTP URIs so people can look up data.
3. Provide useful information in standard formats (RDF).
4. Include links to other URIs to connect data.

Concrete, Recent Applications

1. **Wikidata**: linking diverse knowledge, open, heavily used in downstream AIs, tools.
2. **Schema.org**: used in search engines for structured data; markup on the Web
3. **Enterprise Knowledge Graphs**: internal knowledge management in companies (e.g. Google Knowledge Graph / Microsoft / Amazon / financial / biotech).
4. **Biomedical Knowledge Graphs**: for drug discovery, disease mapping, patient data, research literature.
5. **Semantic search / QA systems:** systems that answer questions over linked data (e.g. DBpedia SPARQL endpoint, or hybrid systems combining Knowledge Graphs + text).
6. **Provenance / Trust systems**: act-checking, supply chain traceability, data governance tools

Group exercise

- 6 groups; 1 topic per group.
- Take some time to read and discuss the topic:
  - 1st Individually: read the provided links and search for additional information on the topic; try to give preliminary answers to the questions of the following slide;
  - After, in group: discuss with each other information found and complete. Answer and prepare together the questions of the following slides.
- Finally, prepare a 5 min presentation/demo to the class
  - This presentation will kick off the class on Sept 30.

Group exercise.

- In the following slides you will find some concrete, recent use cases for inspiration.
- While searching and learning, you can try to answer some of the following questions:
  - What the system does?
  - Why is it useful? What problem does it solve in the real world?
  - Where does the information come from? What kinds of data are collected?
  - How is the information organized? Which semantic technologies are used (RDF, ontologies, SPARQL, provenance)?
  - Architecture/workflow (diagram encouraged)?
  - Example query or markup snippet?
  - Strengths + limitations?

Wikidata

- Wikidata is a collaborative **knowledge base** that links structured data across various sources.
- It powers numerous applications, such as Wikipedia, and is used to create a semantic web of knowledge.
- Useful in querying linked data to gather information from different languages and domains.
- You can explore Wikidata directly here: https://www.wikidata.org

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Wikidata Use Case

- **Case:** Scholia – Profiles and Visualizations for Scholars using Wikidata https://scholia.toolforge.org/

- **What to study:** How Wikidata data is queried (SPARQL endpoint), used to build researcher/publication/topic profiles. How open linking (ORCID, DOI, VIAF) makes this possible.
- **Learning Point:** See how an open Knowledge Graph powers concrete, usable services.

Schema.org

- A semantic vocabulary used by major search engines (Google, Bing, Yahoo) to interpret and display information more effectively in search results.
- Research how organizations implement schema.org to improve their website's search visibility and how it enhances the web's semantic infrastructure.
- The main site for exploring the vocabulary and examples of its usage: https://schema.org

Schema.org

- **Case:** *Google's Organization Structured Data & Knowledge Panels*

https://developers.google.com/search/docs/appearance/structured-data/organization

- **What to study:** JSON-LD examples; how adding schema.org markup impacts Google search results (rich snippets, knowledge panels).
- **Learning Point:** Research a "before/after" on any website that uses schema markup.

Enterprise Knowledge Graphs

- **Case:** *JEL – JPMorgan Chase's Neural Entity Linking to Enterprise KGs* (2024)

https://arxiv.org/abs/2411.02695

- **What to study:** How financial news is linked to internal enterprise entities (not in Wikipedia). Importance of disambiguation, embeddings, Knowledge Graph deployment at enterprise scale.
- **Learning Point:** Connect *NLP + KG* in a high-impact business scenario.

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Biomedical Knowledge Graphs

- **Case:** *KG-Hub: A Platform for Building Biomedical Knowledge Graphs* (2024)

https://arxiv.org/abs/2302.10800

https://github.com/Knowledge-Graph-Hub

- **What to study:** Integration of many biomedical knowledge resources (Gene Ontology, MeSH, MONDO, etc.); workflows for building/maintaining graphs; downstream tasks like drug discovery.
- **Learning Point:** Real biomedical impact; data quality + ontology reuse.

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Semantic Search / Question Answering

- **Case:** *Open Research Knowledge Graph (ORKG)*

https://orkg.org/

https://en.wikipedia.org/wiki/Open_Research_Knowledge_Graph

- **What to study:** Semantic search over scientific contributions; ability to compare research results at a structured level (beyond PDFs).
- **Learning Point:** How semantic search improves over keyword-only approaches.

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Provenance / Trust Systems

- **Case:** *Trust Provenance – Farm-to-Fork Traceability Platform*

https://www.trustprovenance.com/

- **What to study:** Provenance metadata in agriculture supply chains; how blockchain and semantic traceability ensure trust, food safety, and regulatory compliance.
- **Learning Point:** Importance of provenance standards (W3C PROV-O) in applied contexts.