Medical Ontologies

In the previous video, we were introduced to the RDF data model. And we had a look at some simple RDF statements, such as Lucy lives in Edinburgh. The examples that we discussed, expressed information about individuals, such as Lucy and Edinburgh. In addition to such ground level objects, we may also want to capture terminological knowledge, for instance, to refer to classes and relations between them. Such information is typically captured in an ontology. An ontology is defined to be a formal explicit specification of a shared conceptualization. Essentially, it is a way of encoding domain knowledge. I sometimes think of ontologies as enhanced dictionaries, where you can look up the meaning of different terms, and find relations between them. Ontologies share many structural similarities, regardless of the domain that they describe or the language in which they are expressed. Most ontologies describe classes, individuals, attributes and relations. We've already explained that classes are kinds of things or concepts such as woman, while individuals are specific instances such as Lucy. Attributes are characteristics that classes and instances can have, such as age. While relations describe how different classes relate to one another, such as MotherOf. It is very common for ontologies to include a class taxonomy, specifying sub-classes of existing classes, such as woman is a sub-class of person. Classes may also be formally defined in an ontology. For example, we can define a woman to be a person whose sex is female. So why are ontologies useful in the context of precision medicine? Ontologies allow us to attach meanings to data, and they enable the standardization of terminology. So, when different datasets used the term "Viral pneumonia", we know what is meant. Ontologies also allow us to infer new knowledge from existing data. So if our dataset includes that James is suffering from viral pneumonia, and our ontology specifies that viral pneumonia is a subclass of infectious pneumonia, which is a subclass of lung disease, then we can infer that James is suffering from a lung disease. Several ontologies exist in the medical field covering a wide range of topics, from diseases and vaccines to anatomy and genes. The Gene Ontology aims to address the need for consistent descriptions of gene products, across databases. It represents information about biological processes, cellular components, and molecular functions. The Disease Ontology provides descriptions of human disease terms, phenotype characteristics, and related medical vocabulary disease concepts. It includes among others, a hierarchy of diseases specifying for instance, that melanoma is a cell type cancer, which is a disease of cellular proliferation. Another widely used ontology is SNOMED-CT, which is a collection of medical terms providing codes, terms, synonyms, and definitions, using clinical documentation and reporting. SNOMED-CT is considered to be the most comprehensive multilingual clinical healthcare terminology in the world. Ontologies allow clinical practitioners and researchers all over the world to understand each other when using medical terms. This is crucial when sharing data, as it allows for a common and unambiguous understanding of different concepts in the datasets. And, as already discussed, by representing biomedical datasets as graph data, it is possible to easily link data and discover previously unknown relationships between different concepts. When it comes to connected data, graph databases outperform traditional relational databases, allowing for simple and fast retrieval of complex structures. Finally, the graph data model is very flexible, allowing us to add new nodes, new kinds of relationships, and new subgraphs to an existing structure without disturbing existing functionality. These benefits make the graph data model an excellent candidate for the Precision Medicine era. where large-scale data sharing is essential.