

# Dataset labelling guidelines

The objective of the dataset is to identify how properties in a given ontology are expressed in natural language.

## Definitions

### Ontology related:

- Ontology is a modelling of a fragment of the world. It is comprised of concepts (classes), relations (properties), rules (logical axioms) and instance data.
- A property in an ontology is often associated with a domain and range by means of axioms.
- The idea of domain and range are based on their set-theoretic definitions.
- The domain of a property is the class or collection of classes to which the subject of the property would belong. For instance, property 'birthDate' has domain 'Animal'.
- The range of a property is the set of classes or datatypes to which the object of the property would belong. For instance, the property 'birthDate' has range 'date' which is a datatype.
- Most properties would have a single class as domain and a single class or a datatype as range.
- Properties connect instances in an ontology and instances are assigned classes or datatype labels.

### Property, domain and range compliance

- First, we try to identify whether the property itself appears in the text. The property often will occur in synonymous terms and not literally. However, the relation conveyed should be along the same lines as the original property implies.
- An entity or its reference, which appears as a subject of a property in a given natural language text should be compared with the domain defined for that property.
- Compliance with ontology would mean:
  - Entity can be said to be of type specified by the given domain.
  - Entity is of type which is a subclass of the given domain.
  - For instance, if the entity, "Tom Cruise", appears as the subject of the property 'birthDate' then it complies with domain 'animal' since 'Person' is a subclass of 'Animal'.
  - As DBpedia is a general-purpose ontology describing common entities in the real world, we will go with common-sense checks for verifying this compliance.
  - For any points of confusion, the DBpedia ontology is available on the web and class and property definitions can be found from links that are their corresponding IRIs. As far as possible, IRIs will be provided while labelling.  
For instance, these IRIs provide a direct link to their detailed description:
    - Property: <https://dbpedia.org/ontology/doctoralStudent>
    - Class: <https://dbpedia.org/ontology/Scientist>
- Similarly, an entity or its reference, which appears as an object of a property in a given natural language text should be compared with the range defined for that property.
- Compliance with ontology would be similar with one additional consideration. There can be data properties where the object is a datatype value so we just have to see if the value as described is of the type specified in range.

## Alignment:

What we mean by alignment is to verify how closely a property is expressed in a sentence (or group of sentences) in terms of its ontological definition of domain and range.

Given a Property(Domain, Range) & a sentence (or group of sentences) to label, we provide 5-labels:

- Full alignment:  $p(D, R)$ 
  - When the property is expressed and the textual entities or their references comply with the domain and range.
- Property and domain are aligned:  $p(D, ?)$
- Property and range are aligned:  $p(?, R)$
- Property expressed, but both domain and range do not align:  $p(?, ?)$
- No alignment

## Illustrative examples:

Correct alignment of both domain and range

<b>Property(Domain, Range)</b>	<i>dbo:birthDate(dbo:Animal, xsd:date)</i>
<b>Evidence</b>	<i>Feynman was born on May 11, 1918, in New York City,[2] to Lucille (née Phillips), a homemaker, and Melville Arthur Feynman, a sales manager.</i>
<b>Subject entity</b>	<i>Richard Feynman</i>
<b>Object entity</b>	<i>1918-05-11</i>
<b>Classification label</b>	<i>Since both, the subject and the object entities are compliant with the specified domain and range, we label it as full alignment, <math>P(D,R)</math>.</i>

Similarly, the following cases for each property(Domain, Range):

- **dbo:birthplace**(dbo:Animal, dbo:Place)
  - Evidence: Feynman was born on May 11, 1918, in New York City,[2] to Lucille (née Phillips), a homemaker, and Melville Arthur Feynman, a sales manager.
- **dbo:almaMater**(dbo:Person, dbo:EducationalInstitution)
  - Evidence: Instead, he attended the Massachusetts Institute of Technology, where he joined the Pi Lambda Phi fraternity.
  - Evidence: Feynman received a PhD from Princeton in 1942; his thesis advisor was John Archibald Wheeler.
- **dbo:doctoralAdvisor**(dbo:Scientist, dbo:Person)
  - Evidence: Feynman received a PhD from Princeton in 1942; his thesis advisor was John Archibald Wheeler.

**Annotation:**

The spans of the subject and object, between which the property (relation) is expressed, are to be marked. Remember that the subject could be a valid pronoun and the object could be a numeric or non-numeric value.

## Agreement

By accepting this document, you agree to using the above guidelines and the ontology descriptions available to the best of your abilities to provide your input. If you do not know the response or the data seems implausible, you are expected to select the best option or skip, but avoid guesswork. You agree not to use any AI model to complete this task.