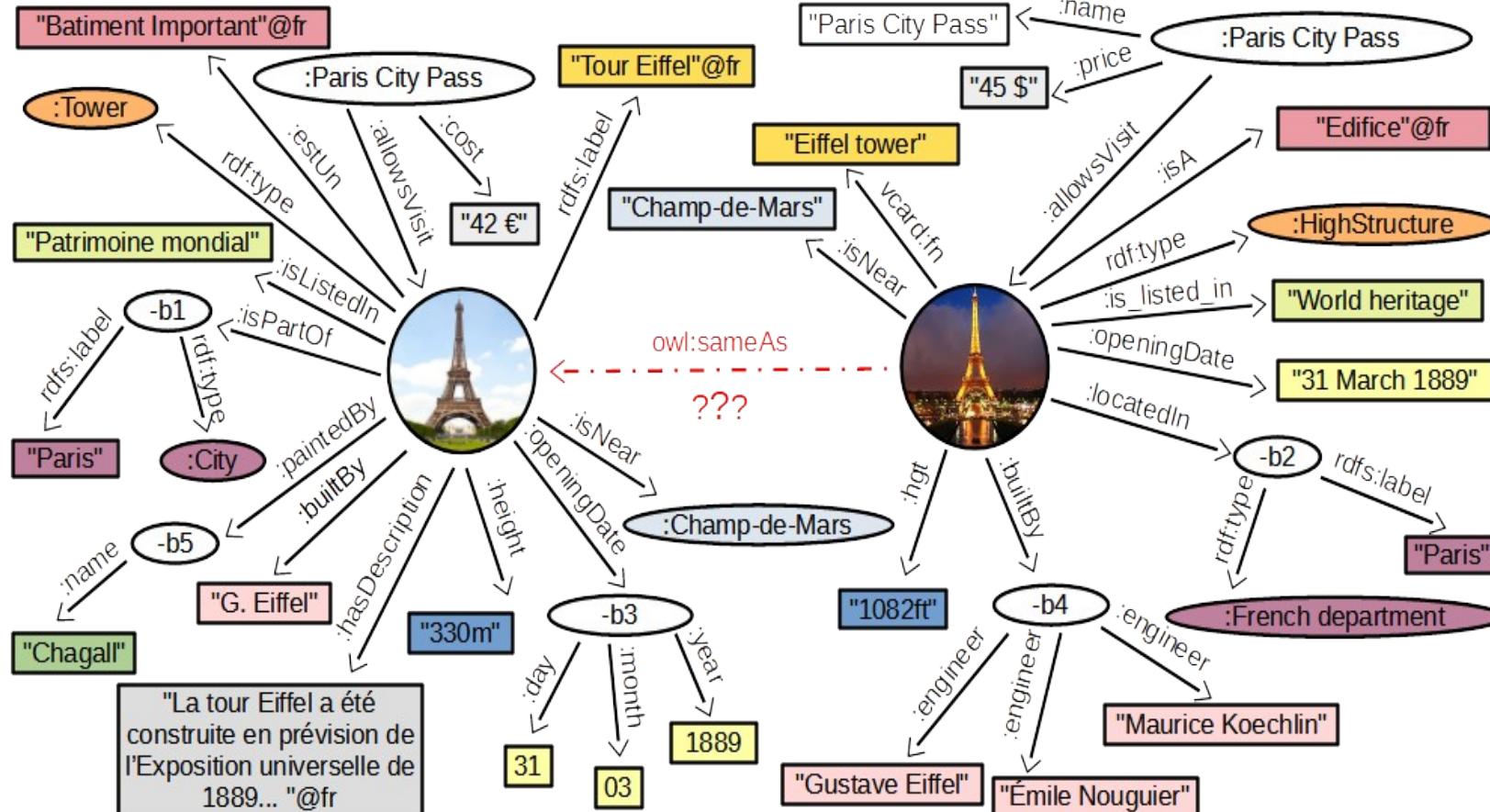


Two different descriptions of the Eiffel Tower

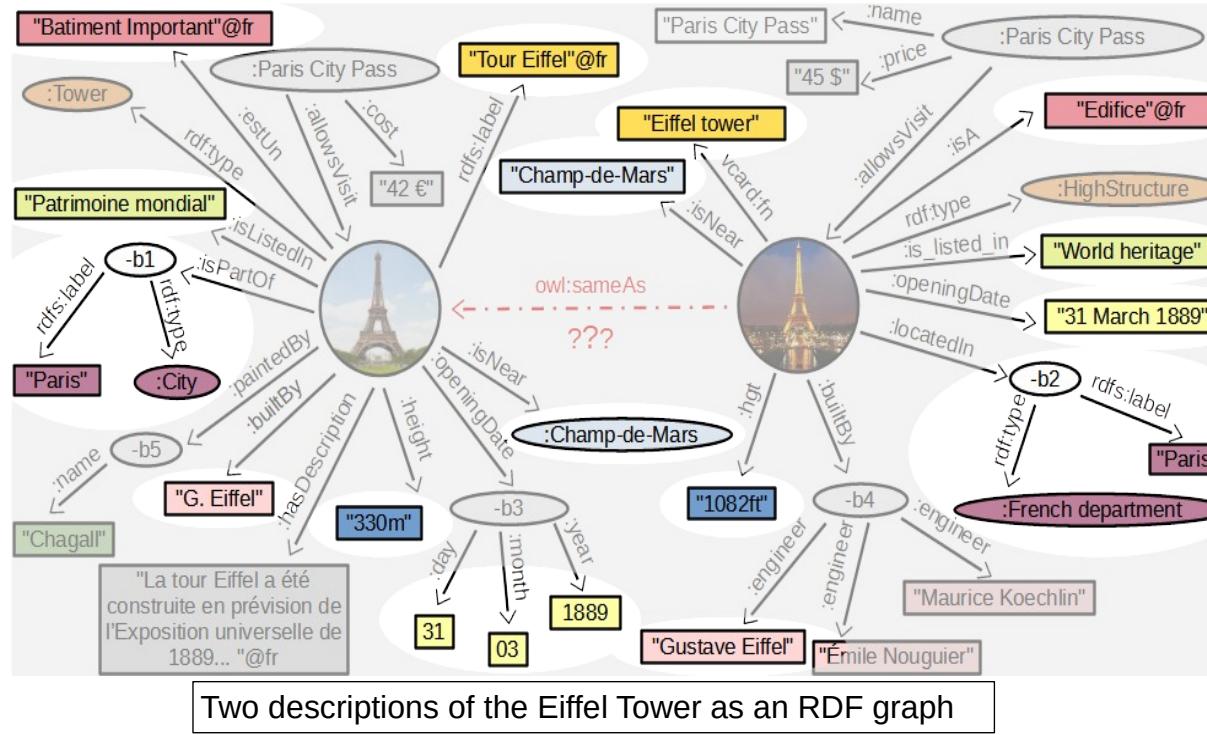


Two descriptions of the Eiffel Tower as an RDF graph

LPT 1

LPT 1: Predicat value problem

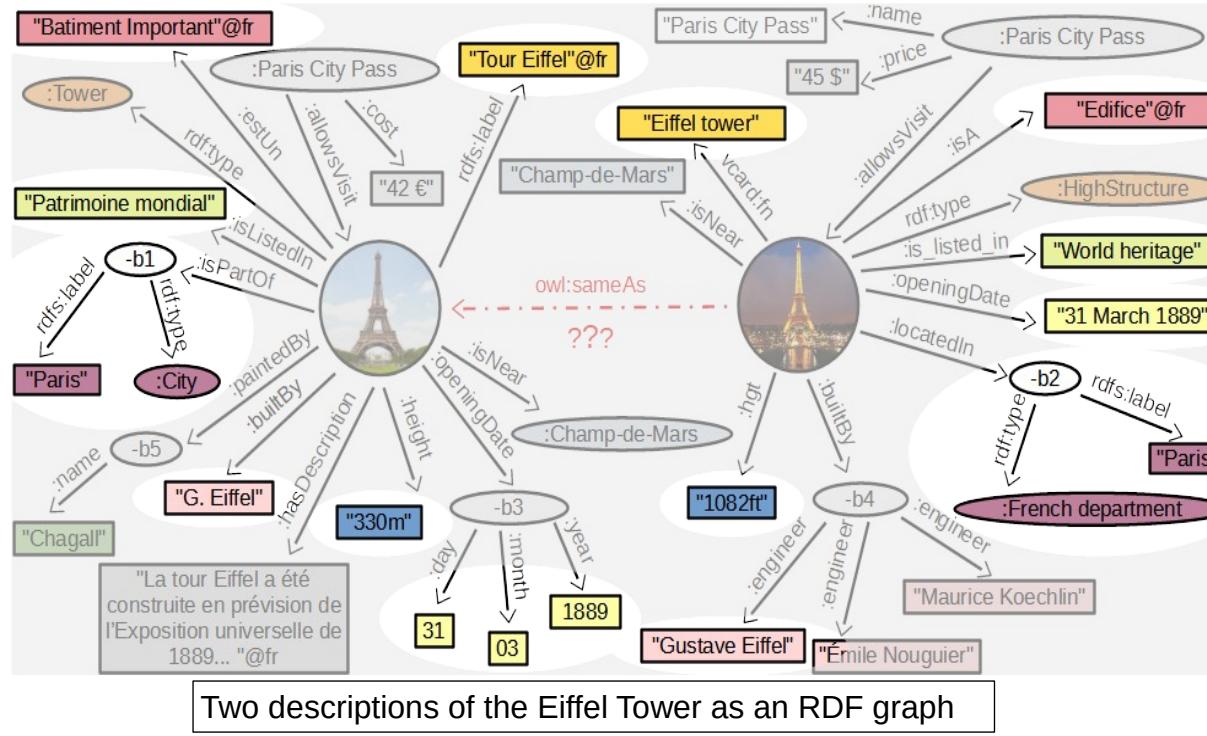
- Any problems encountered when searching for similarity between two values of the characteristic of an entity described in two distinct data sets.



LPT 1.1

LPT 1.1: Predicat value terminological problem

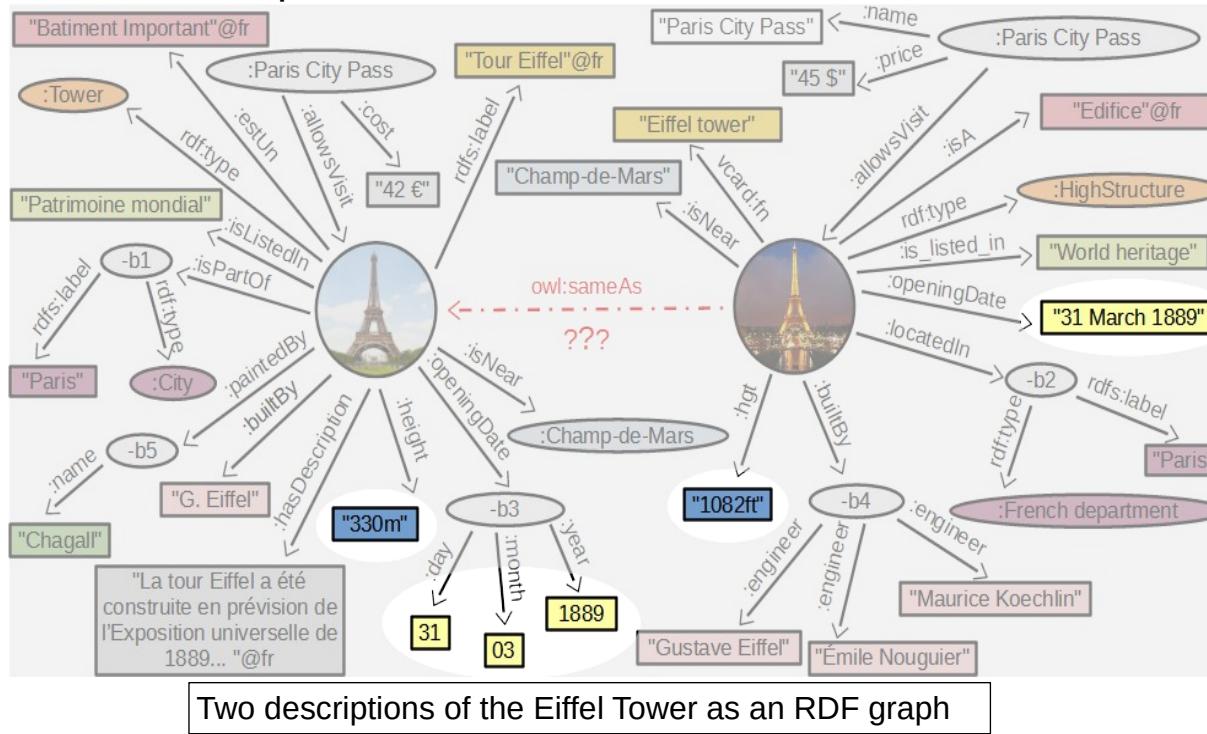
- Any problem encountered when searching for similarity between two values of the characteristic of an entity described in two separate datasets in relation to terminology.



LPT 1.1.1

LPT 1.1.1: Predicat value data format problem

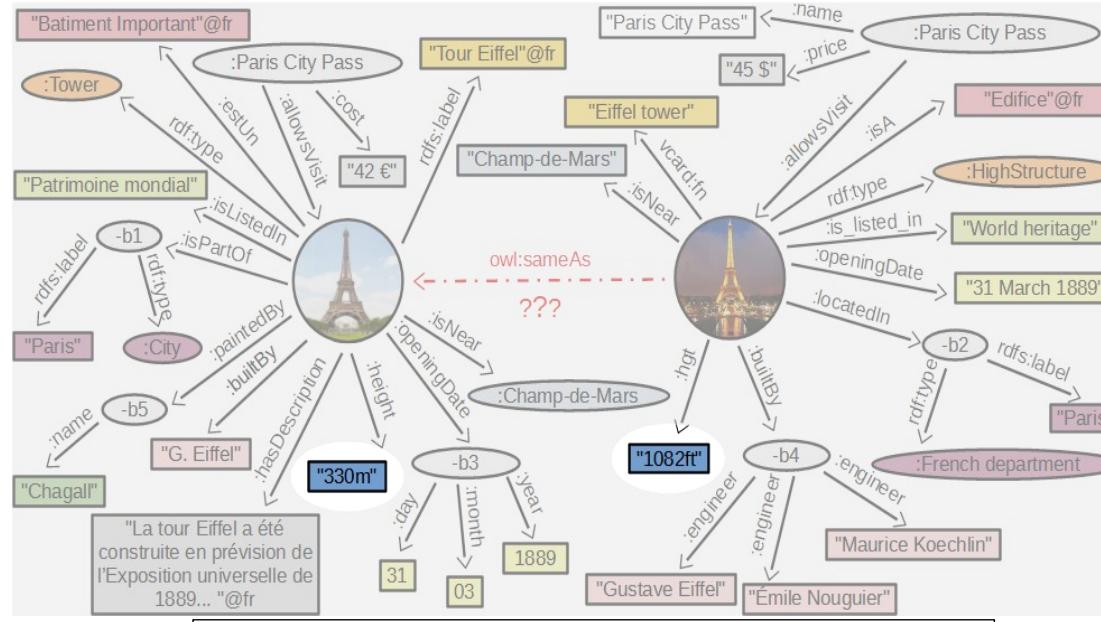
- Any problems encountered when the format of the data is different and thus gives rise to textual variation in the values of the predicates.



LPT 1.1.1.1

LPT 1.1.1.1: Predicat value format encoding problem

- Any problem encountered when searching for similarity between two values of the characteristic of an entity described in two distinct datasets in relation to the encoding of the data.
 - A height can be expressed in meters or feet, a date can be represented by "dd/mm/yyyy" or "mm-dd-yy", etc.

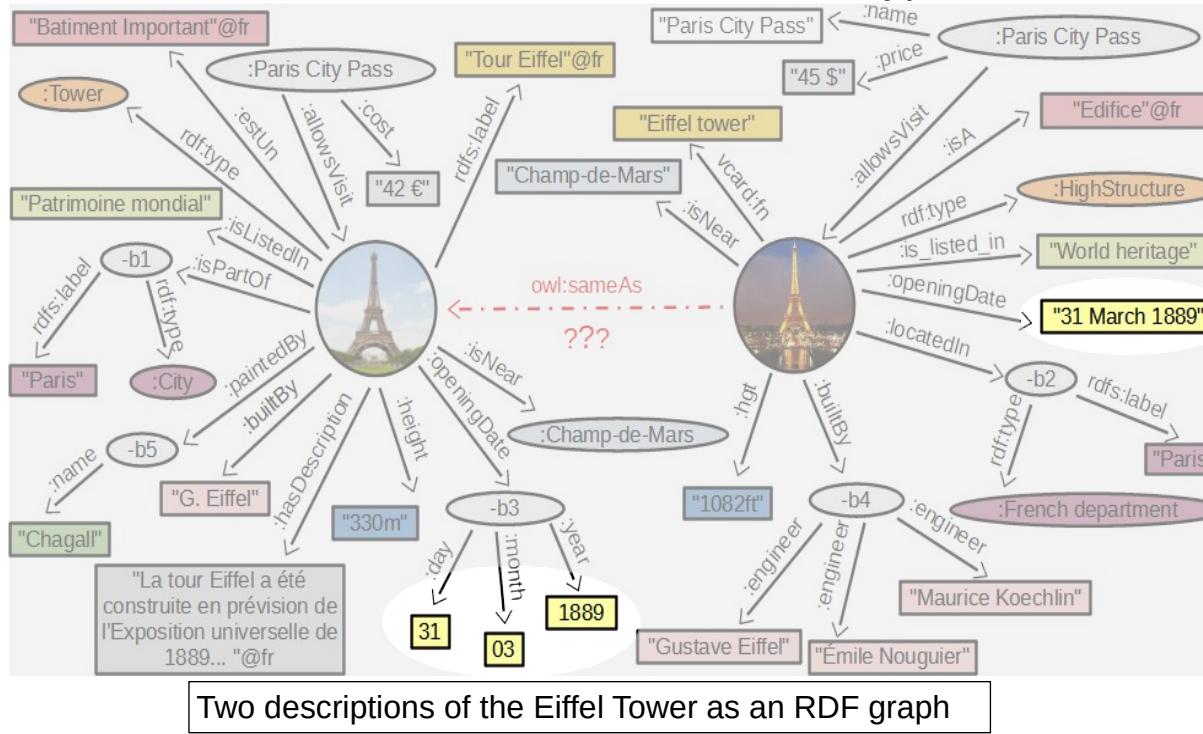


Two descriptions of the Eiffel Tower as an RDF graph

LPT 1.1.1.2

LPT 1.1.1.2: Predicat value format value type problem

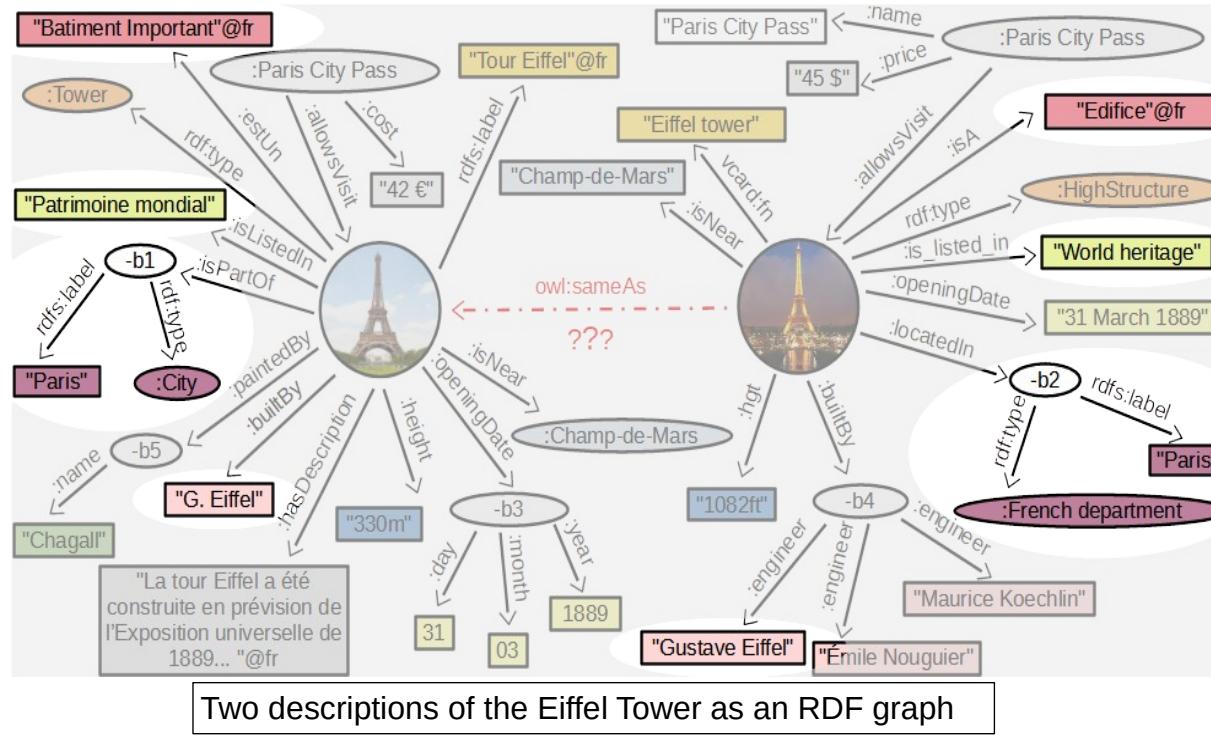
- Any problem encountered when searching for similarity between two values of the characteristic of an entity described in two distinct data sets in relation to the type of data.



LPT 1.1.2

LPT 1.1.2: Predicat value data value problem

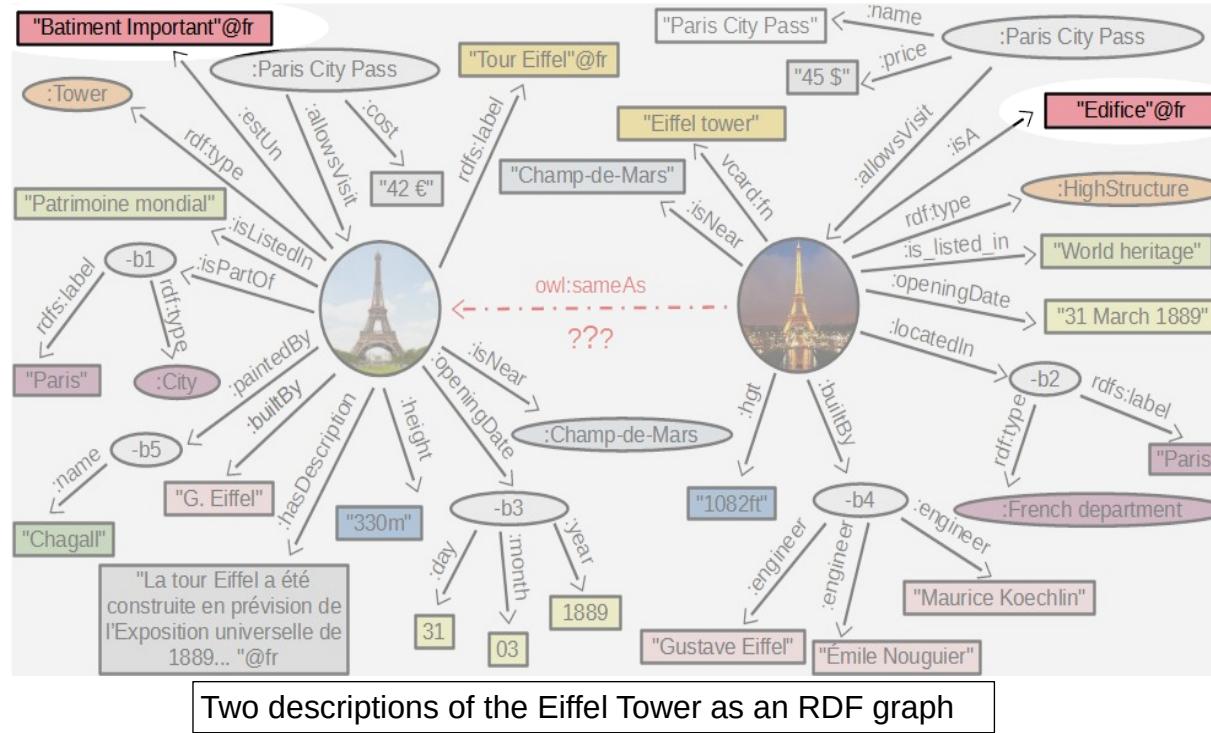
- Any problem encountered when searching for similarity between two values of the characteristic of an entity described in two distinct data sets in relation to the intrinsic value of the data.



LPT 1.1.2.1

LPT 1.1.2.1: Predicat value terminological synonymy problem

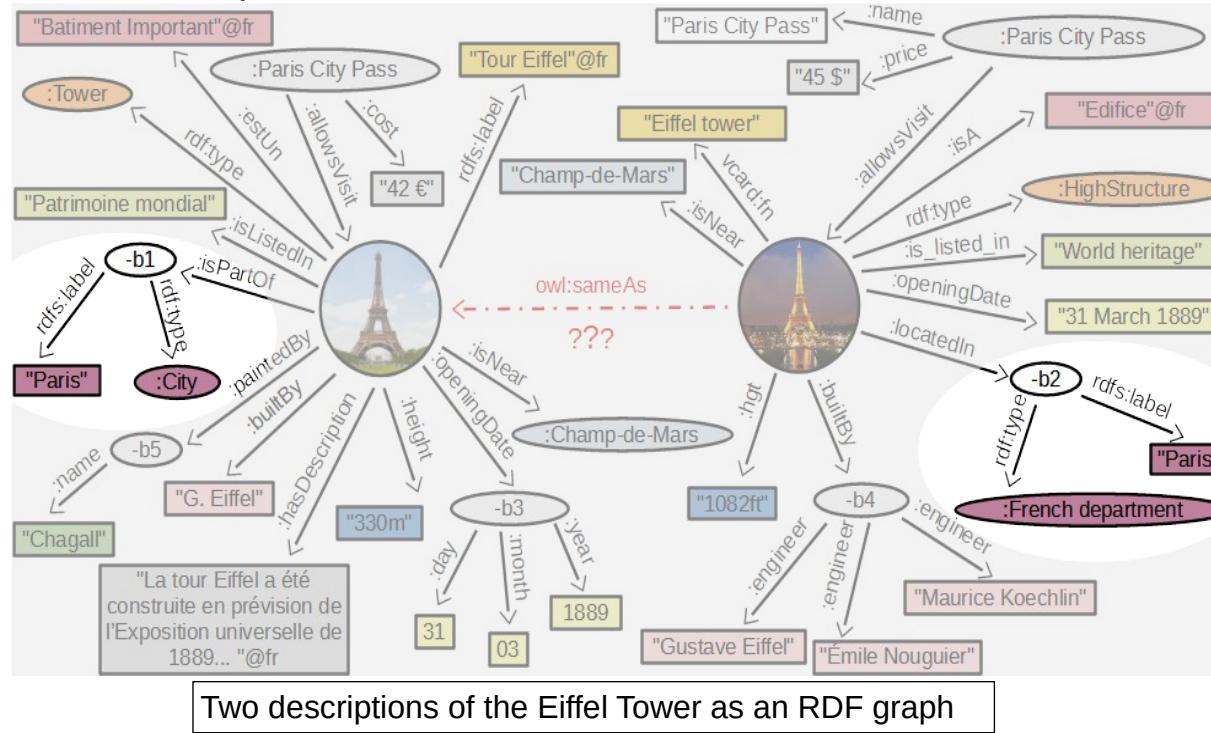
- Any problem encountered when searching for similarity between two terms that are synonymous.



LPT 1.1.2.2

LPT 1.1.2.2: Predicat value terminological homonymy problem

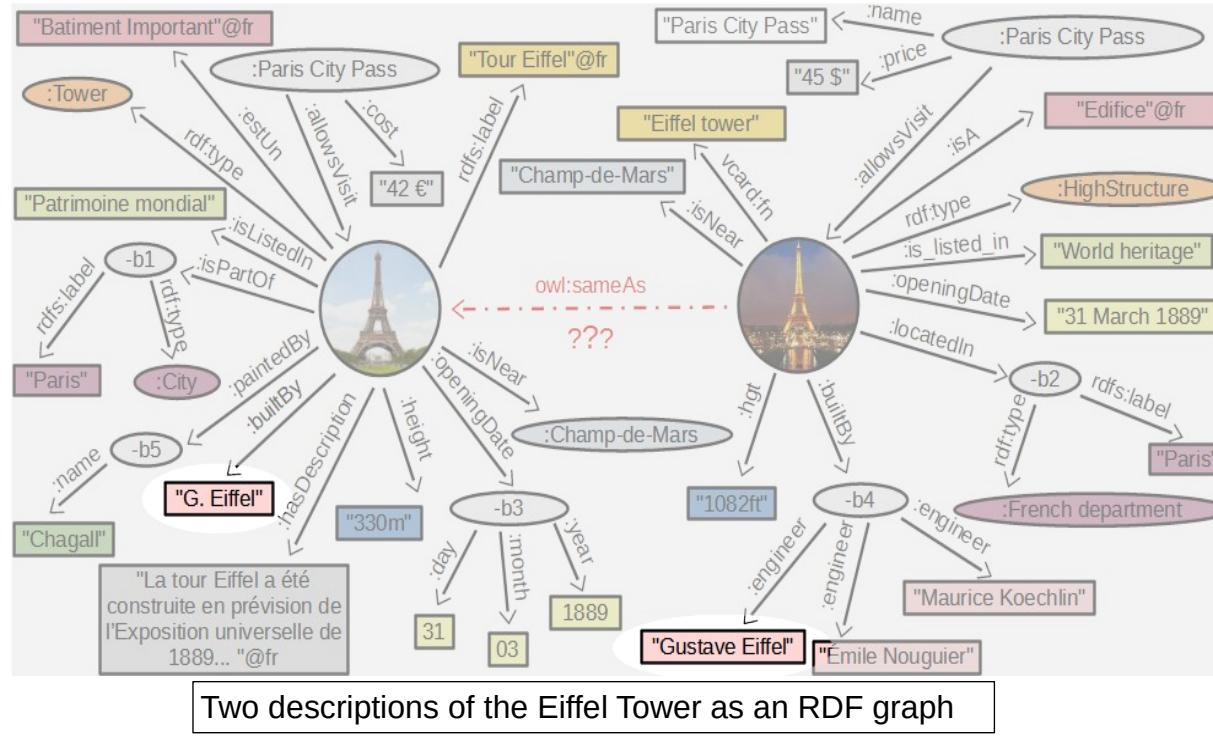
- Any problem encountered when searching for similarity between two identical terms that are used to describe different concepts.



LPT 1.1.2.3

LPT 1.1.2.3: Predicat value terminological acronymy problem

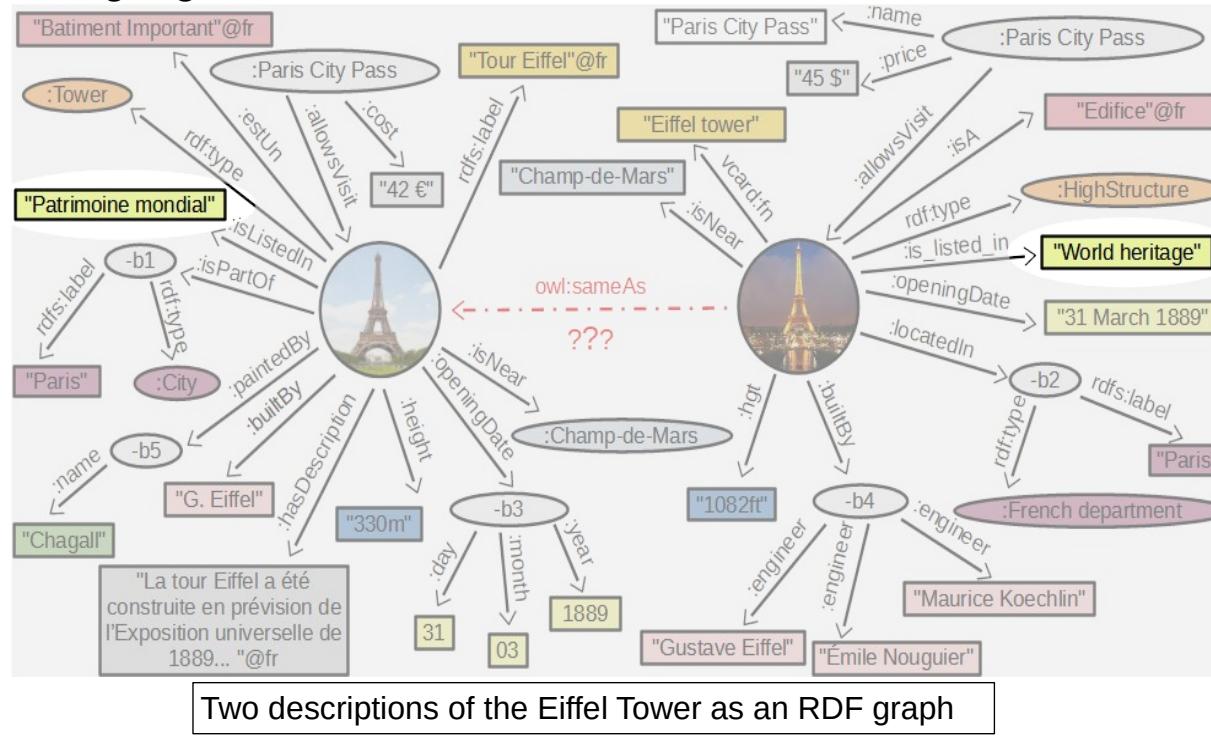
- Any problem encountered when searching for similarity between two terms, one of which is the acronym of the other.



LPT 1.1.2.4

LPT 1.1.2.4: Predicat value terminological multilingual problem

- Any problems encountered when searching for similarity between two values who are expressed in different natural languages.



LPT 1.1.2.5

LPT 1.1.2.5: Predicat value terminological data quality problem

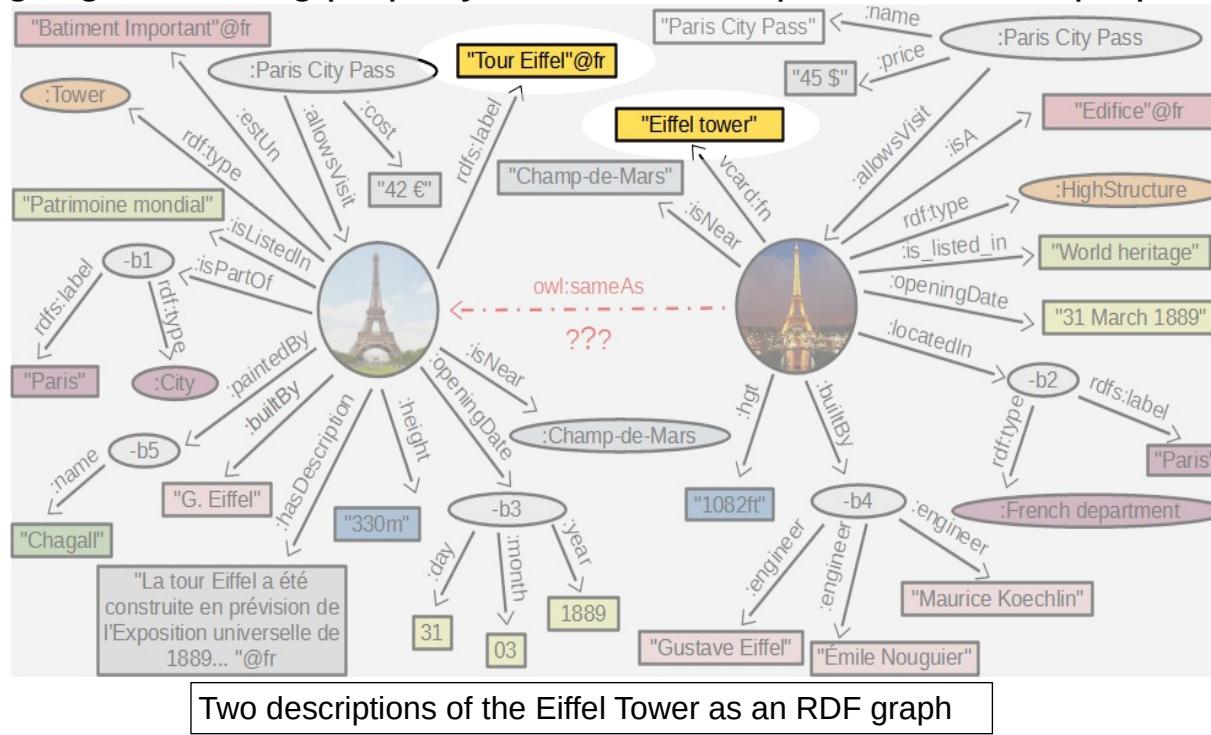
- Inconsistency : « Values of the same properties in two similar descriptions can hold conflicting data. For example, DBpedia identifies the population of Montreal as 1 649 519, while it is 1 600 000 according to Geonames. »¹
- Imcompleteness : « This results from the fact that similar descriptions carry partial data for a similar property »¹
- Incorrectness : « The incorrectness simply refers to the data typographical errors.»¹
- Outdated data : « The compared KBs can include “correct” data taken at different time periods.»¹

¹ ASSI, Ali, MCHEICK, Hamid, et DHIFLI, Wajdi. Data linking over RDF knowledge graphs: A survey. Concurrency and Computation : Practice and Experience, 2020, vol. 32, no 19, p. e5746.

LPT 1.1.3

LPT 1.1.3: Predicat value best practice problem

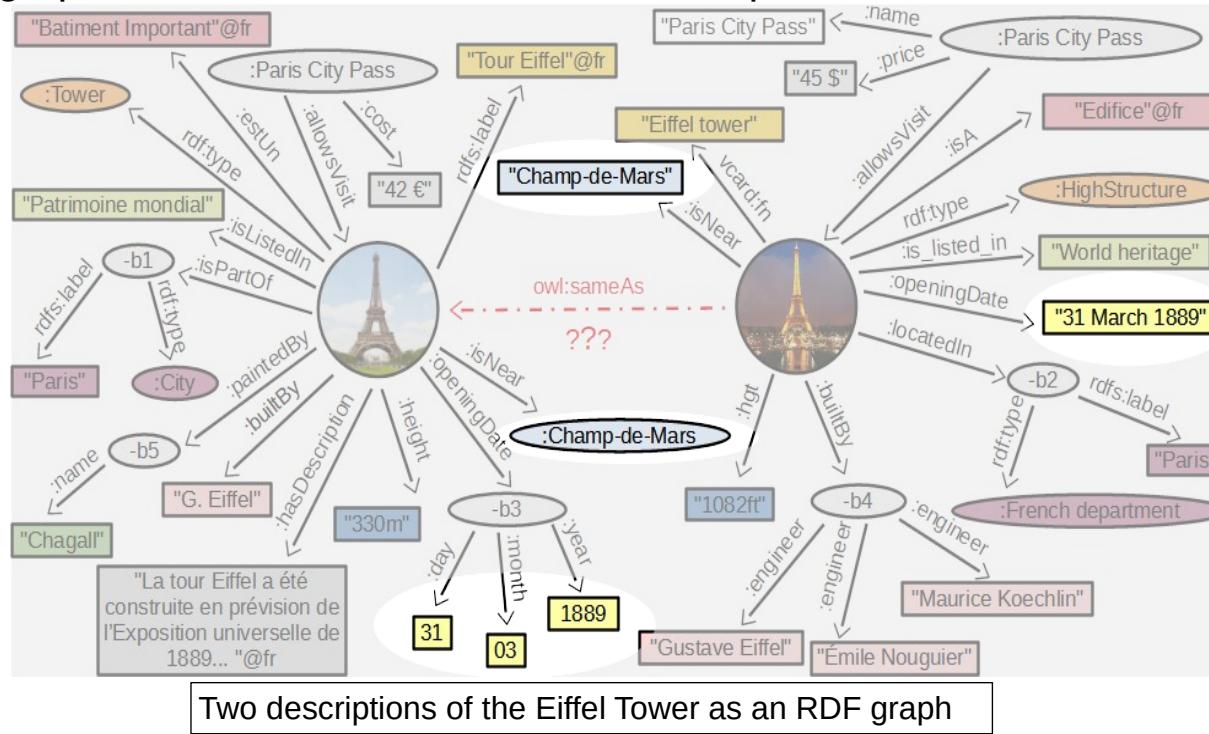
- Any problems encountered in the search for similarity due to non-compliance with good practice (No tag for languages, missing property value, non-unique value for ID properties, ...).



LPT 1.2

LPT 1.2: Predicat value structural problem

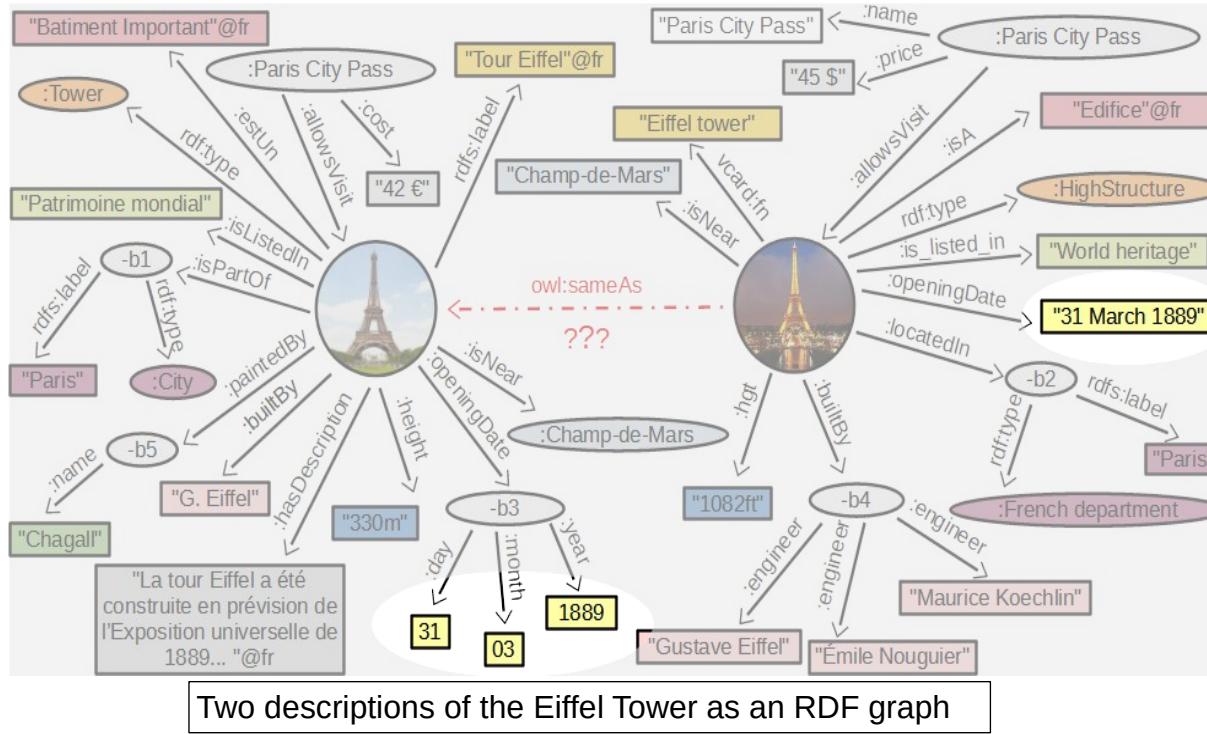
- Any problem encountered when searching for similarity between two entities described in two different data graphs in relation to the structure of the predicate value.



LPT 1.2.1

LPT 1.2.1: Predicat value granularity problem

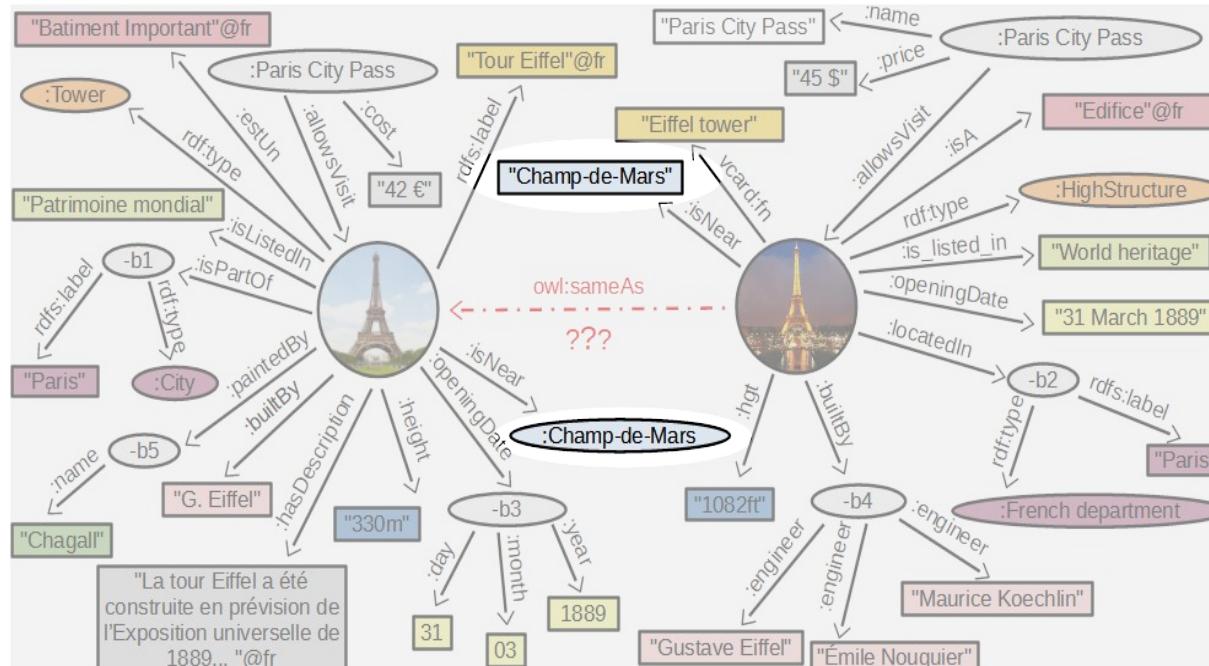
- Any problem encountered when the level of granularity of a property of an entity is different.



LPT 1.2.2

LPT 1.2.2: Predicat value type mismatch problem

- Any problem encountered when the type of a property value of an entity is different.

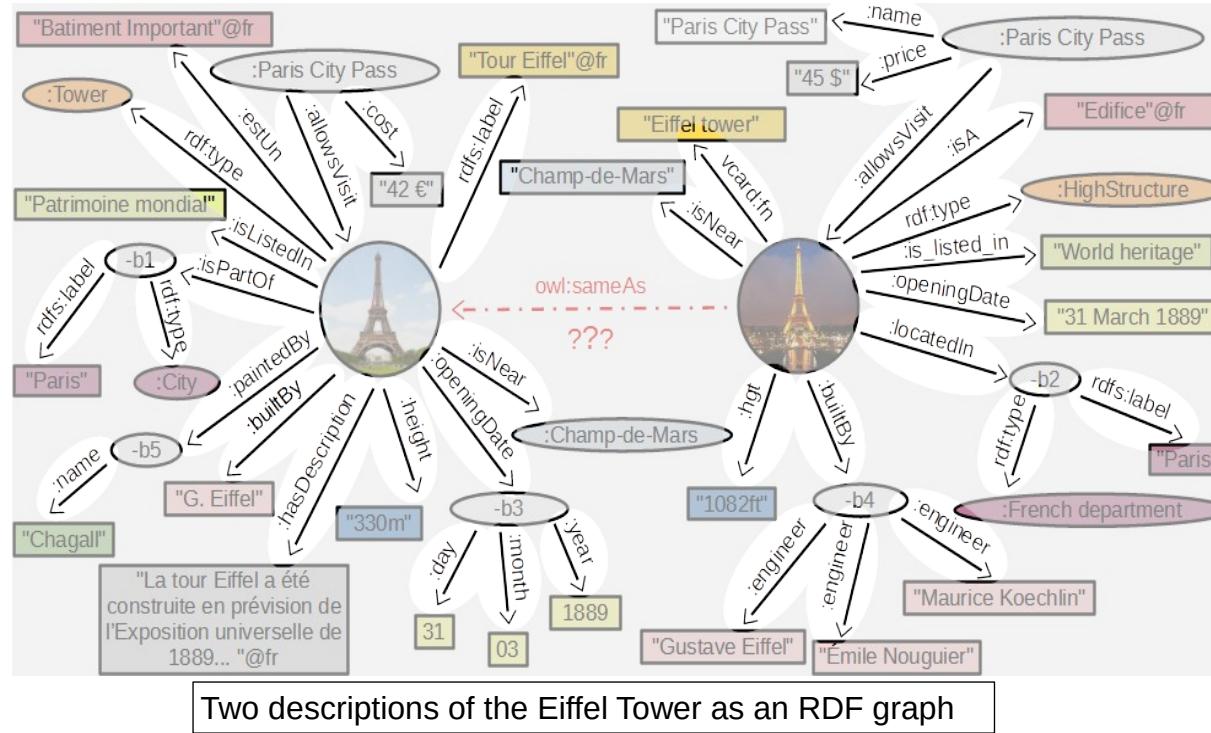


Two descriptions of the Eiffel Tower as an RDF graph

LPT 2

LPT 2: Predicat problem

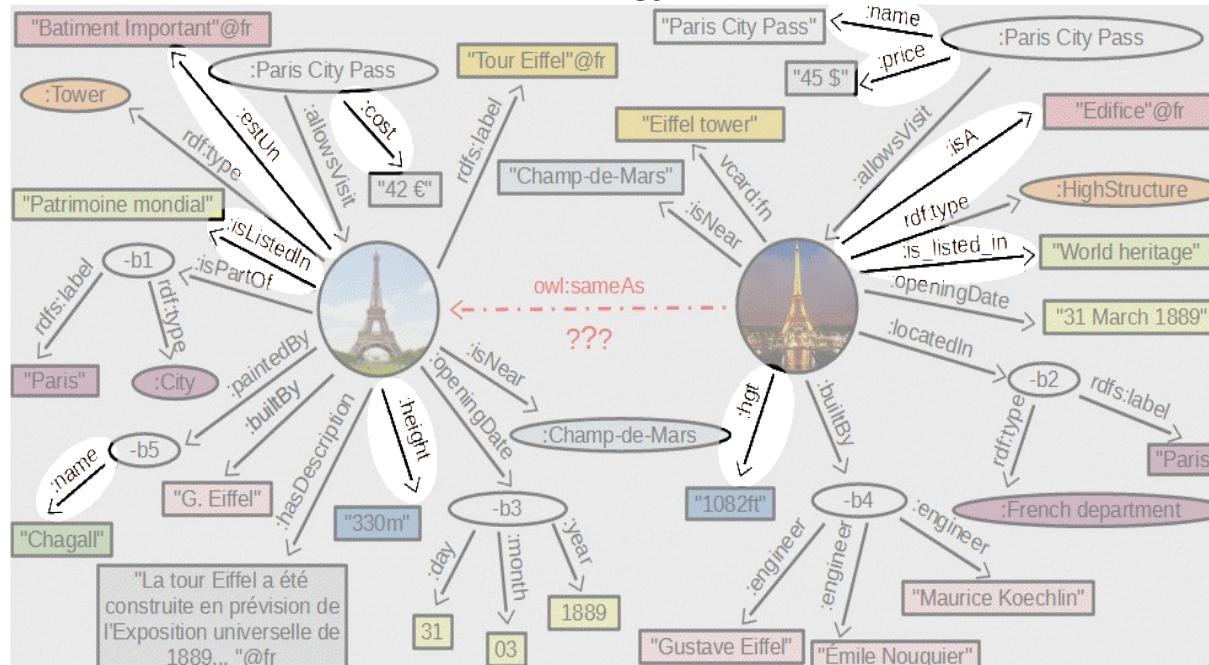
- Any problems encountered when searching for similarity between two characteristic of an entity described in two distinct data sets.



LPT 2.1

LPT 2.1: Predicat terminological problem

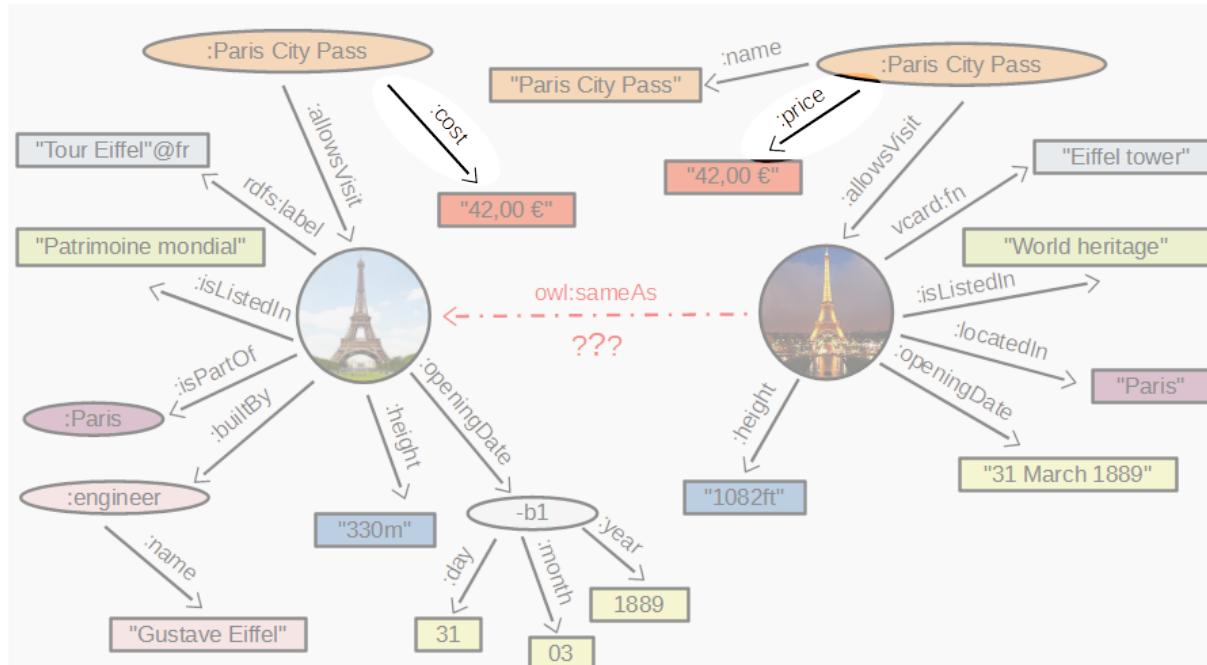
- Any problem encountered when searching for similarity between two features of an entity described in two separate datasets in relation to terminology.



LPT 2.1.1

LPT 2.1.1: Predicat terminological synonymy problem

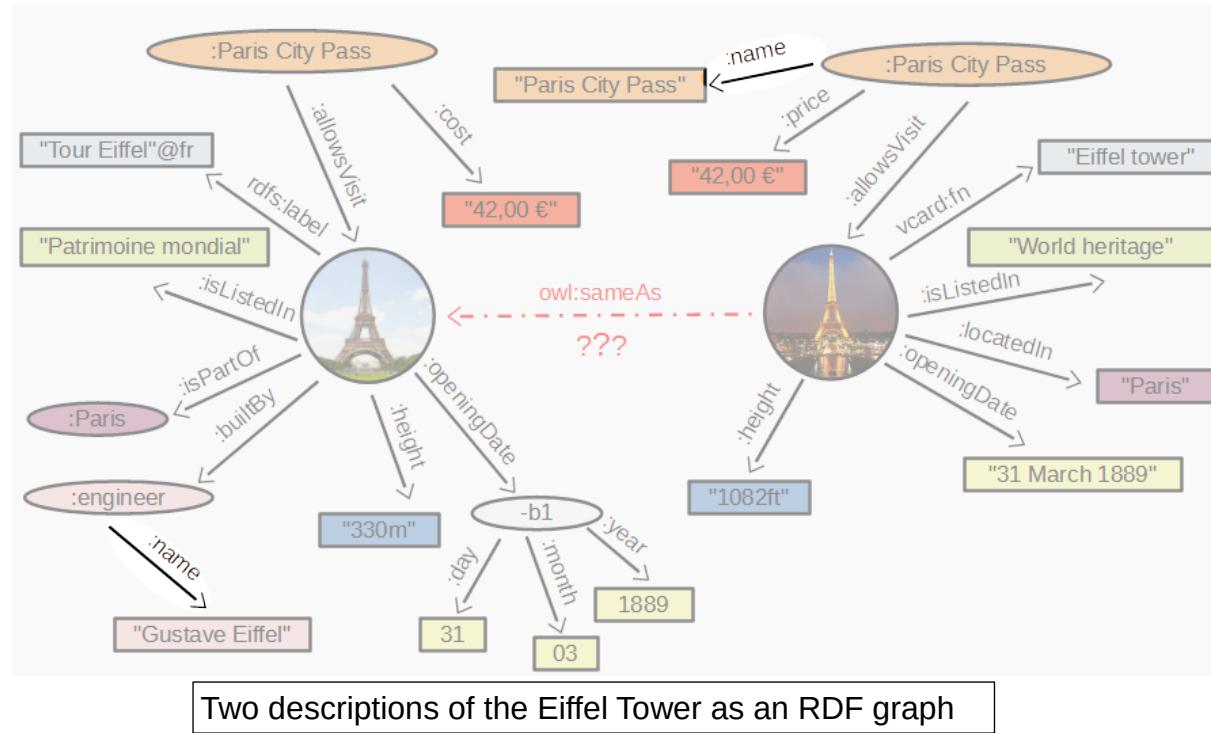
- Any problems encountered when the same characteristic is designed with synonymous terms.



LPT 2.1.2

LPT 2.1.2: Predicat terminological homonymy problem

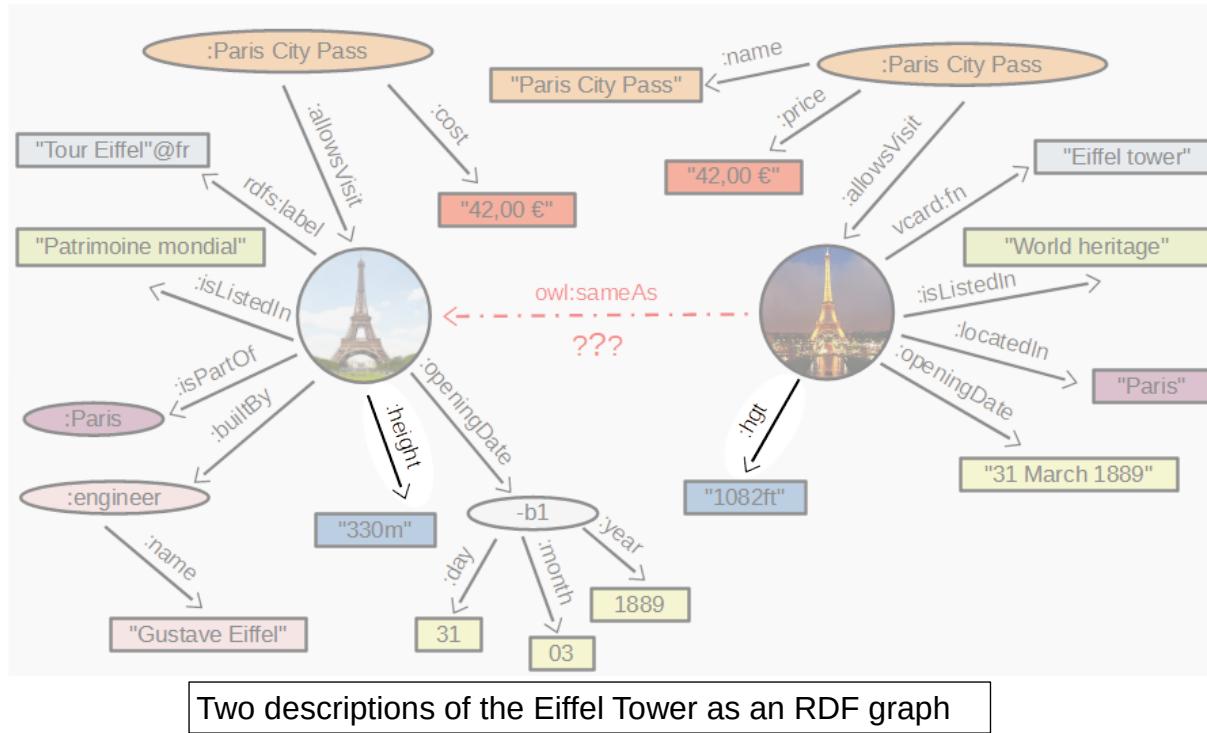
- Any problems encountered when searching for similarity between two characteristic of an entity described in two distinct data sets.



LPT 2.1.3

LPT 2.1.3: Predicat terminological acronymy problem

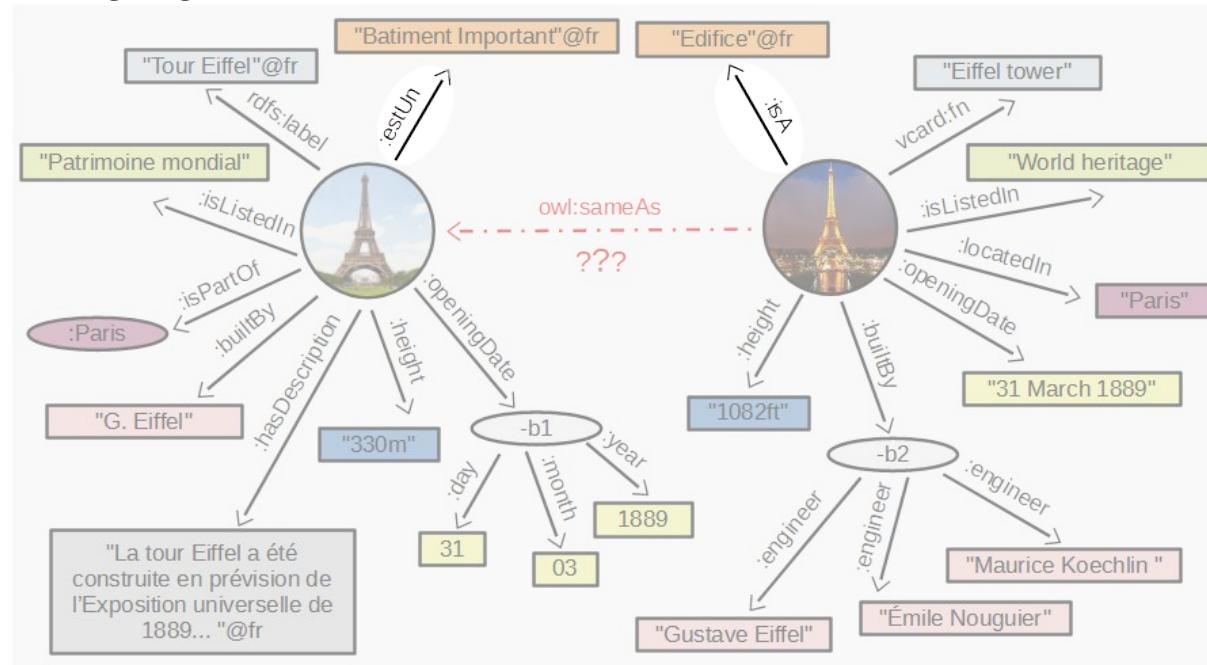
- Any problems encountered when searching for similarity.



LPT 2.1.4

LPT 2.1.4: Predicat terminological multilingual problem

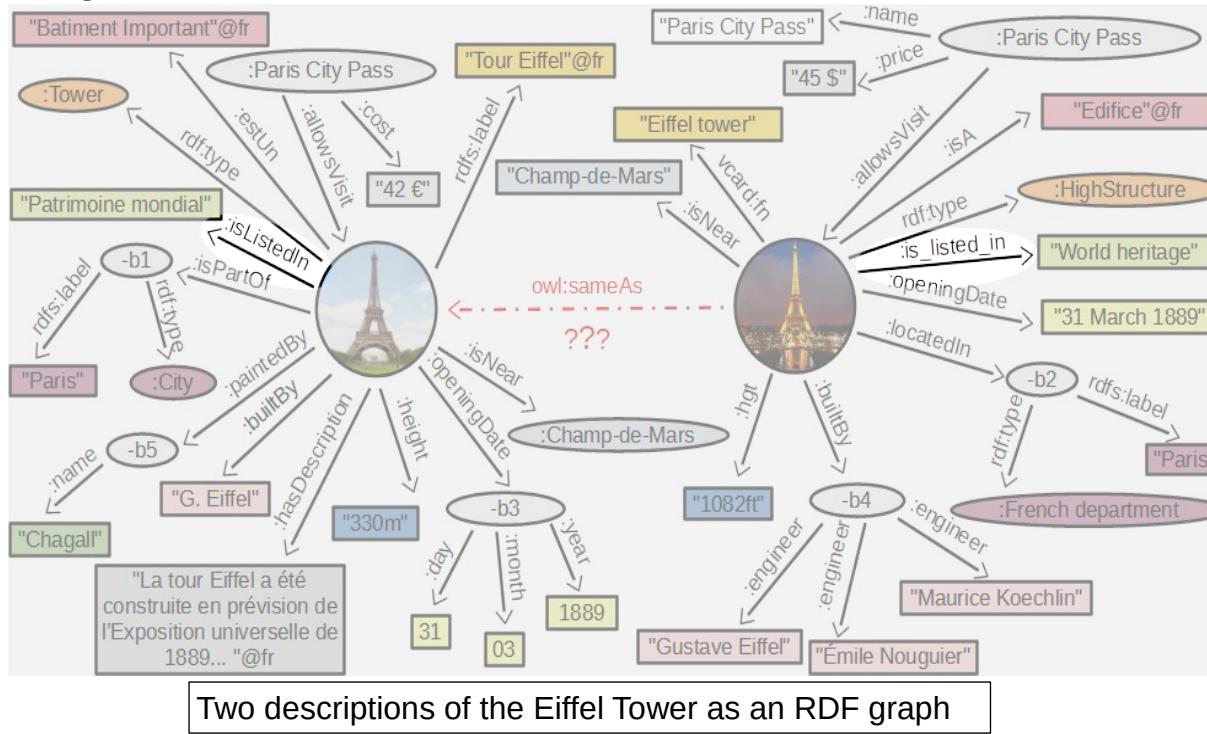
- Any problem encountered when searching for similarity between two characteristics expressed in different natural languages.



LPT 2.1.5

LPT 2.1.5: Predicat terminological data quality problem

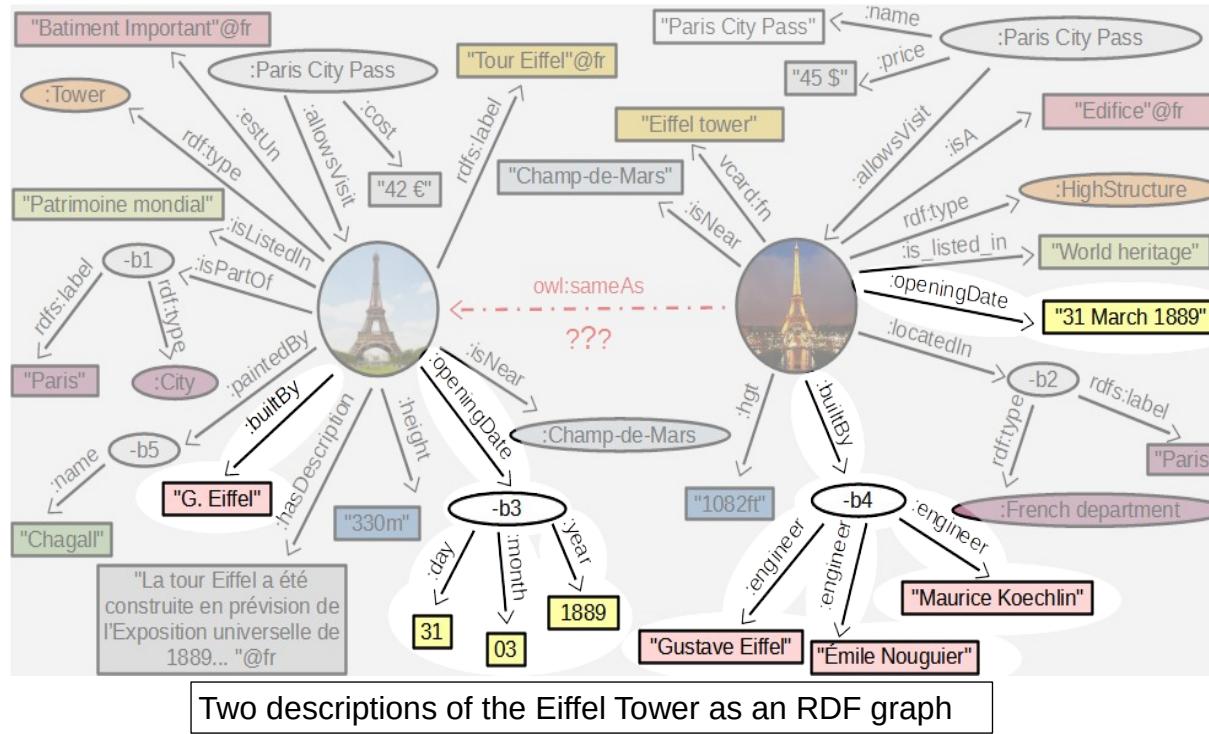
- Any problem encountered when searching for similarity between two features with typographical errors in their designation.



LPT 2.2

LPT 2.2: Predicat structural problem

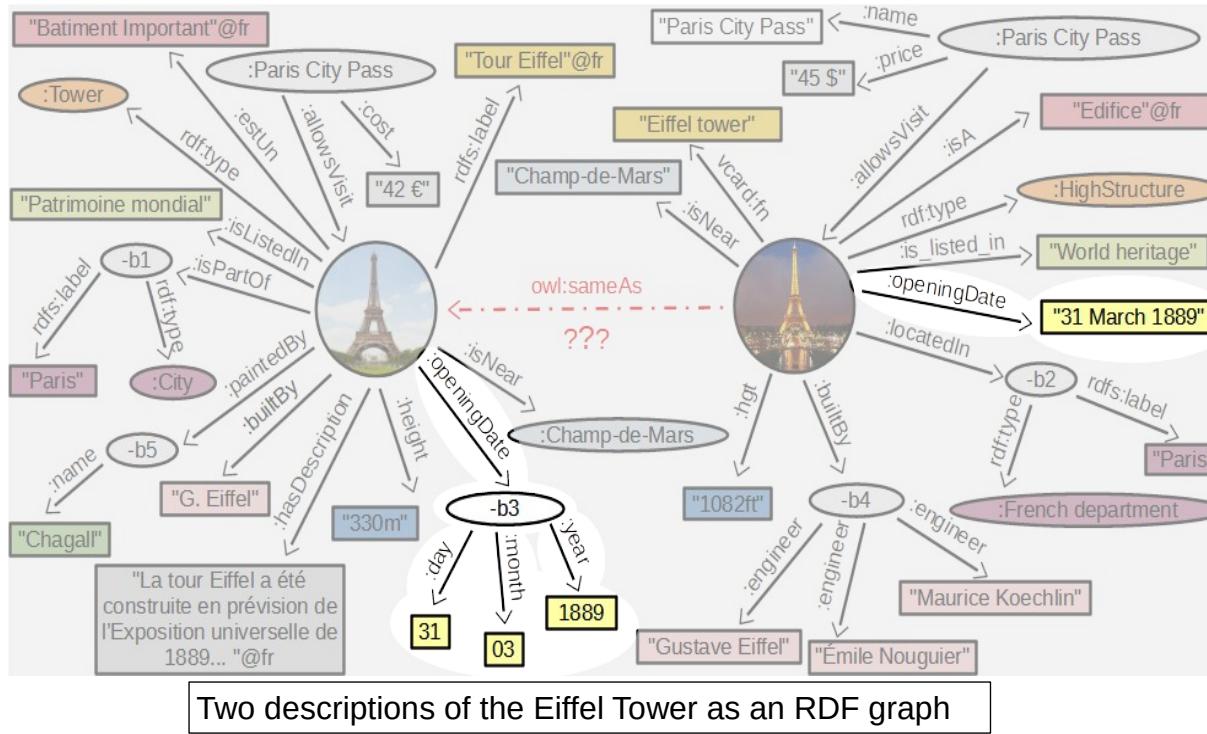
- Any problems encountered when searching for similarity between two characteristic of an entity described in two distinct data sets.



LPT 2.2.1

LPT 2.2.1: Predicat structural intra-aggregation problem

- Any problem encountered when the value of a characteristic is an aggregation of values.



LPT 2.2.2

LPT 2.2.2: Predicat structural inter-aggregation problem

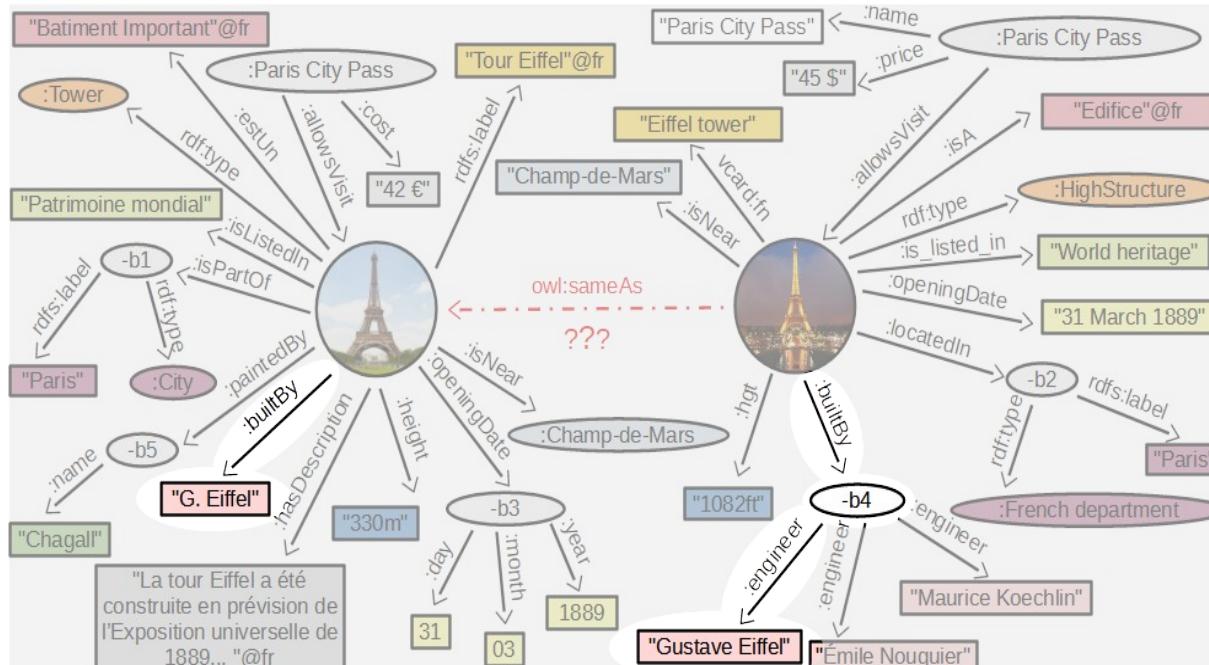
- Any problems encountered when searching for similarity .

Price versus Total price

LPT 2.2.3

LPT 2.2.3: Predicat structural internal path discrepancy problem

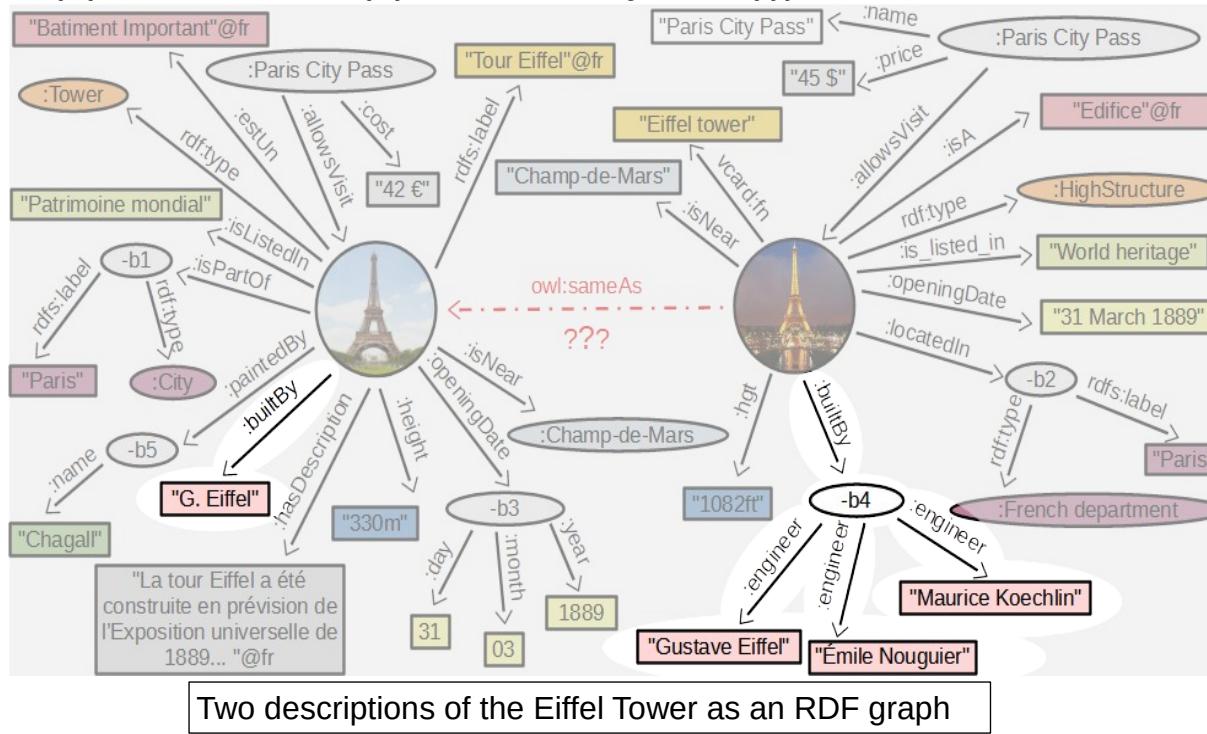
- Any problem encountered when information (value) is not at the same distance from the entity.



LPT 2.2.4

LPT 2.2.4: Predicat structural constraint mismatch problem

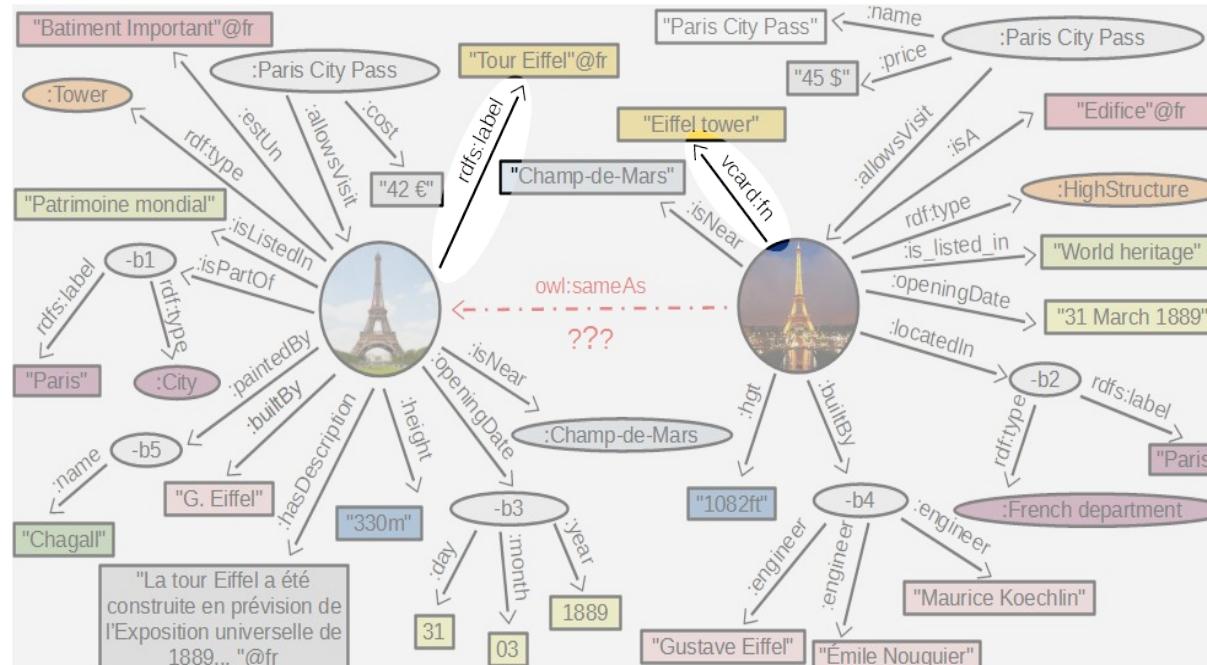
- Any problems encountered when the constraints for the property are not the same (zero-or-more (*), one-or-more (+), zero-or-one (?), and exactly-one (.)).



LPT 2.3

LPT 2.3: Predicat vocabulary problem

- Any problems encountered in the similarity search when the same information is expressed with different vocabularies.



Two descriptions of the Eiffel Tower as an RDF graph

LPT 3

LPT 3: Class problem

- Any problem encountered when searching for similarity between two entities described in two different data graphs in relation to class problems.

LPT 3.1

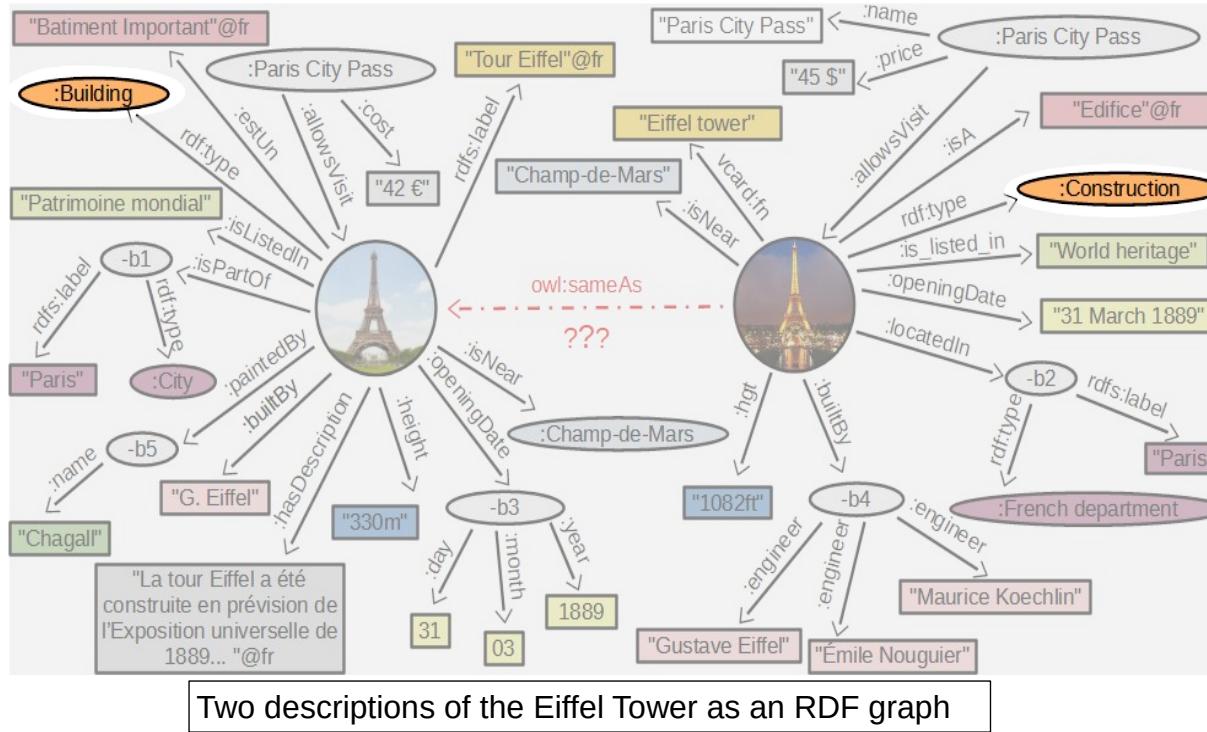
LPT 3.1: Class terminological problem

- Any problem encountered when searching for similarity between two entities described in two different data graphs in relation to class terminology problems.

LPT 3.1.1

LPT 3.1.1: Class terminological synonymy problem

- Any problem encountered when searching for similarity between two classes of membership of an entity described in two distinct data sets when they are synonymous.



LPT 3.1.2

LPT 3.1.2: Class terminological homonymy problem

- Any problem encountered when searching for similarity between two classes of membership of an entity described in two distinct data sets in relation to homonymy problems.

LPT 3.1.3

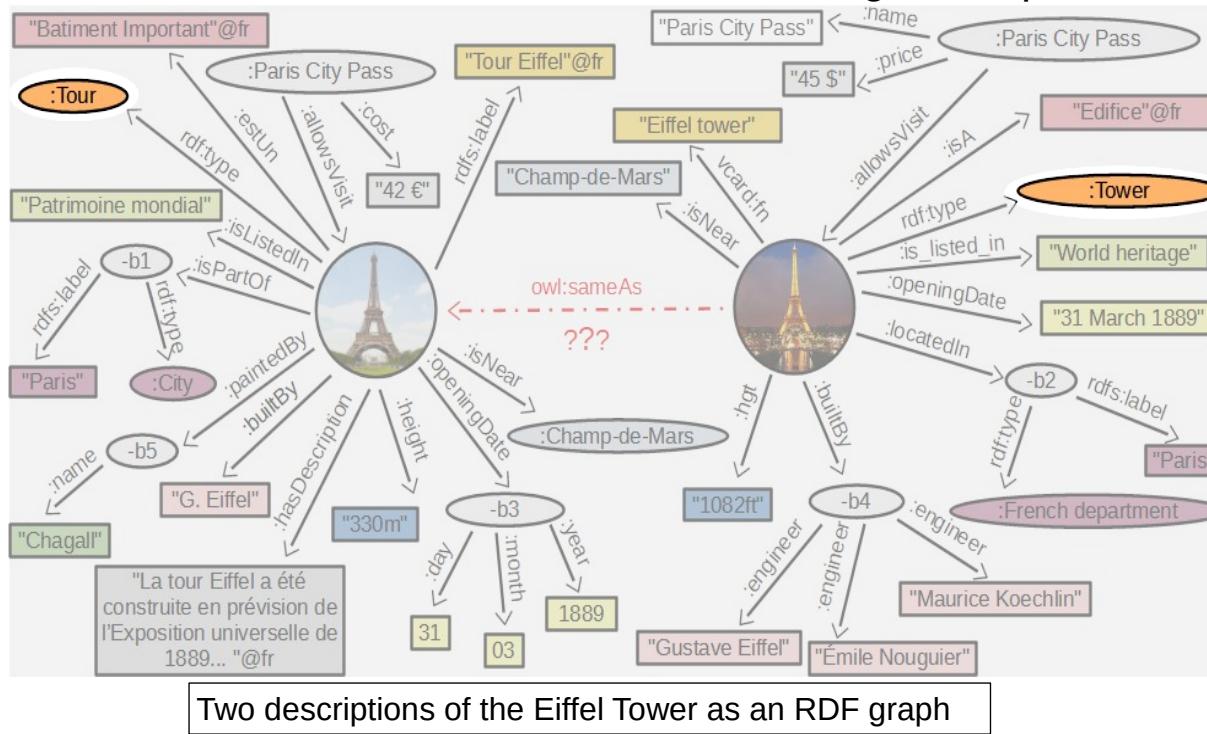
LPT 3.1.3: Class terminological acronymy problem

- Any problem encountered when searching for similarity between two classes of membership of an entity described in two distinct datasets in relation to acronymic problems.

LPT 3.1.4

LPT 3.1.4: Class terminological multilingual problem

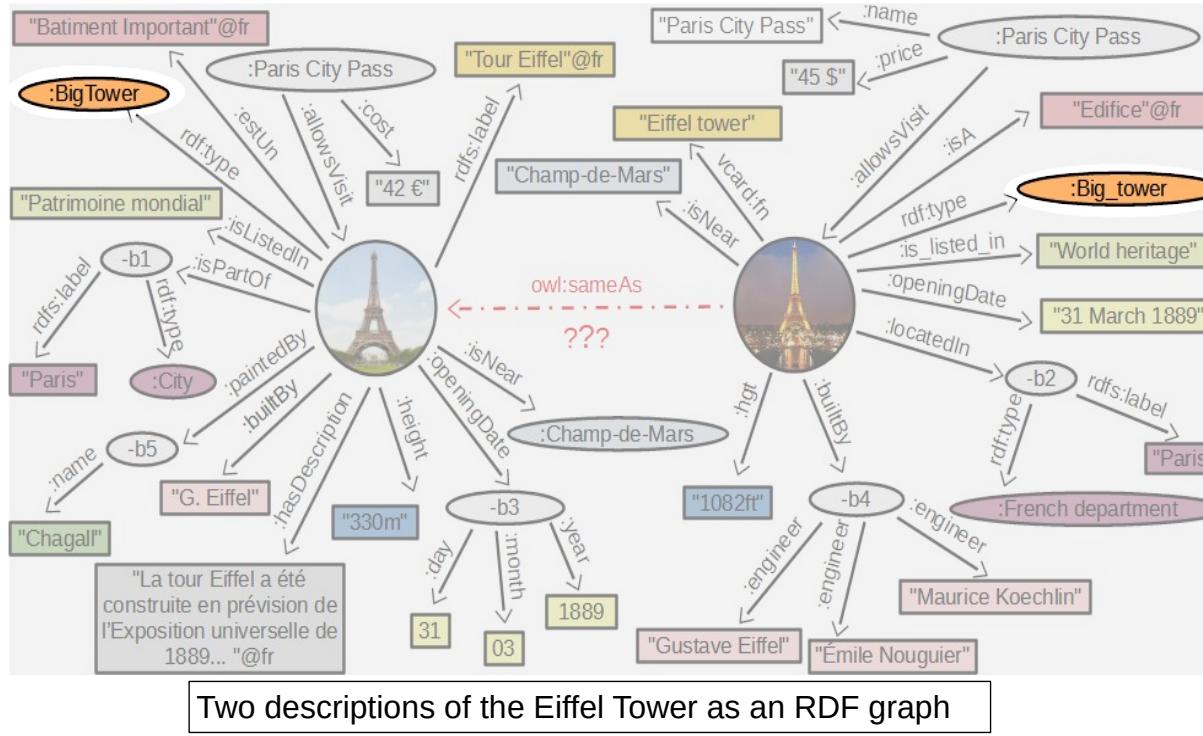
- Any problem encountered when searching for similarity between two classes of membership of an entity described in two distinct datasets in relation to multilingualism problems.



LPT 3.1.5

LPT 3.1.5: Class terminological data quality problem

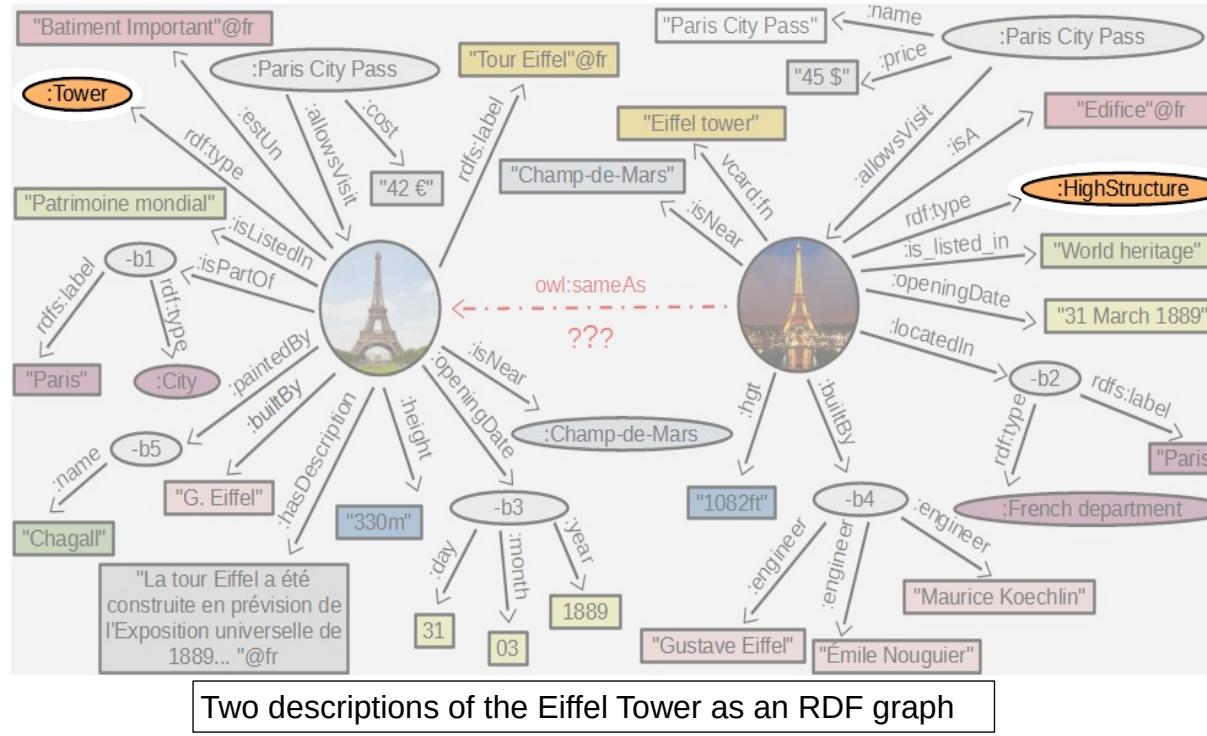
- Any problem encountered when searching for similarity between two classes of membership of an entity described in two distinct datasets in relation to typography problems.



LPT 3.2

LPT 3.2: Class generalisation/specialisation problem

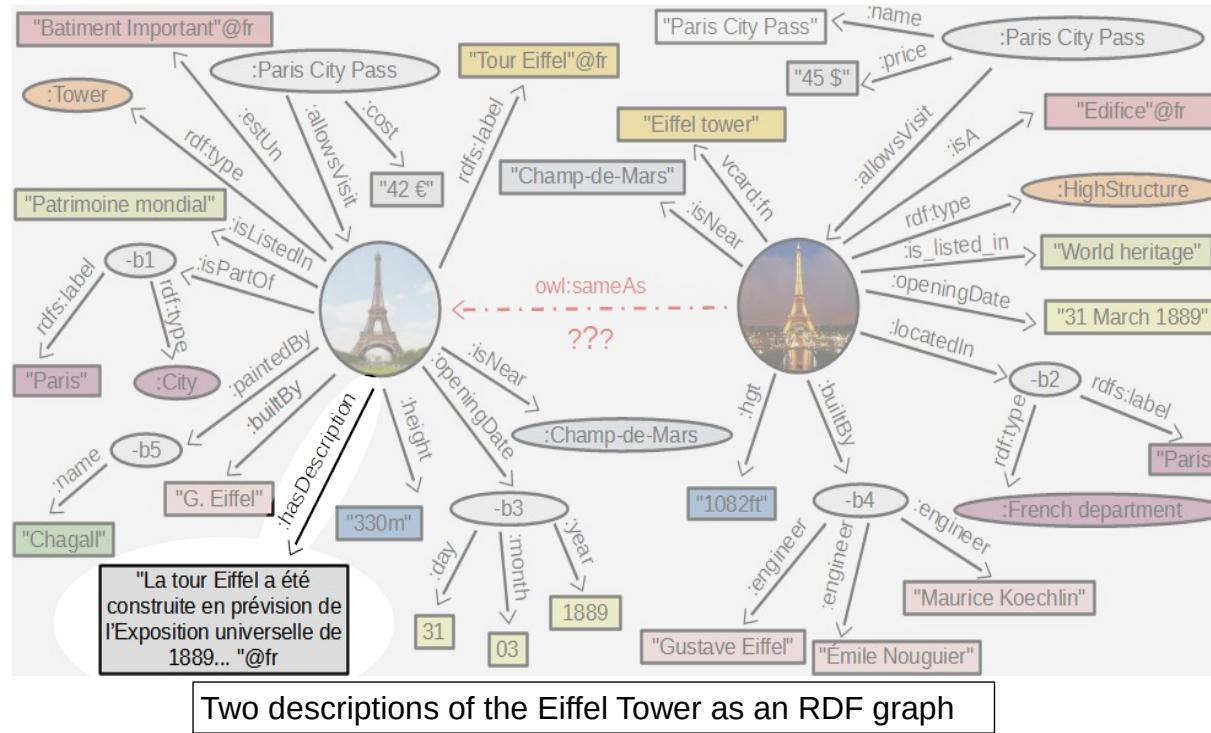
- Any problem encountered when searching for similarity between two classes of membership of an entity described in two distinct data sets in relation to generalisation/specialisation problems.



LPT 4

LPT 4: Subgraph problem

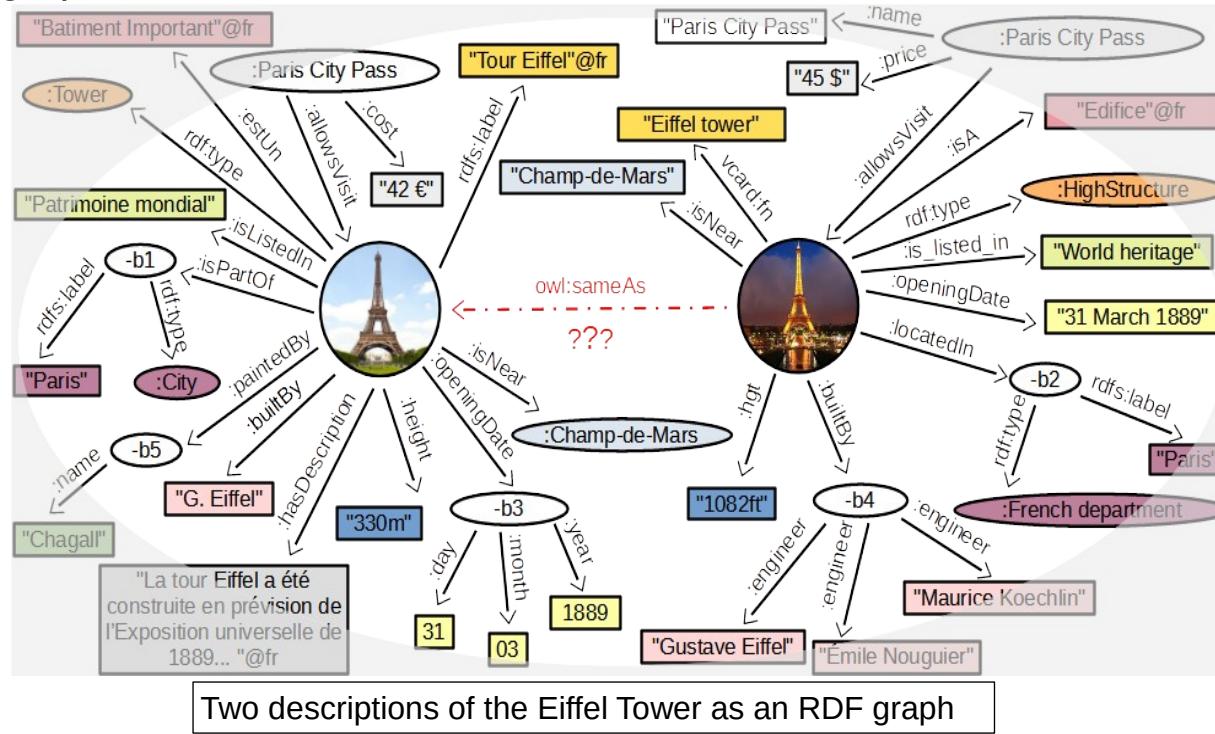
- Any problem encountered when searching for similarity between two groups of features of an entity described in two different data sets.



LPT 5

LPT 5: Graph problem

- Any problem encountered when searching for similarity between two entities described in two different data graphs.



Two descriptions of the Eiffel Tower as an RDF graph

LPT 5.1: Graph syntax problem

- Any problem encountered when searching for similarity between two entities described in two different data graphs who are not expressed in the same ontology language.
- « This obviously happens when comparing, for instance, a directory with a conceptual model. This also happens when two ontologies are modelled by using different knowledge representation formalisms, for instance, OWL and F-logic. »¹
- « For example, to define the class of chairs in RDF Schema (Brickley and Guha, 2000), one uses <rdfs:Class ID="Chair">. In LOOM, the expression (defconcept Chair) is used to define the same class. »²

¹ Ontology matching

J Euzenat, P Shvaiko - 2007 - Springer

² KLEIN, Michel. Combining and relating ontologies:
an analysis of problems and solutions.
In : OIS@ IJCAI. 2001.

LPT 5.2

LPT 5.2: Graph semantic of primitive problem

- Any problem encountered when searching for similarity between two entities described in two different data graphs who use different syntaxes.
- « For example, the OIL RDF Schema syntax (Broekstra et al., 2001) interprets multiple <rdfs:domain> statements as the intersection of the arguments, whereas RDF Schema itself uses union semantics »¹

¹ KLEIN, Michel. Combining and relating ontologies:
an analysis of problems and solutions.
In : OIS@ IJCAI. 2001.

LPT 5.3

LPT 5.3: Graph language expressivity problem

- Any problem encountered when searching for similarity between two entities described in two different data graphs using languages with different expressiveness.
- « For example, some languages have constructs to express negation, others have not. »¹

¹ KLEIN, Michel. Combining and relating ontologies:
an analysis of problems and solutions.
In : OIS@ IJCAI. 2001.

LPT 5.4

LPT 5.4: Graph conceptual problem

- Any problem encountered when searching for similarity between two entities described in two different data graphs in relation to the conceptualisation of models.

LPT 5.4.1

LPT 5.4.1: Graph conceptual coverage problem

- Any problem encountered when searching for similarity between two entities described in two different data graphs where the domain of one only partially covers that of the other.
- « example of an ontology about cars: one ontology might model cars but not trucks. »¹

¹ KLEIN, Michel. Combining and relating ontologies:
an analysis of problems and solutions.
In : OIS@ IJCAI. 2001.

LPT 5.4.2

LPT 5.4.2: Graph conceptual granularity problem

- Any problem encountered when searching for similarity between two entities described in two different data graphs where the levels of detail of the domain are different.
- « Another one might represent trucks but only classify them into a few categories »¹

¹ KLEIN, Michel. Combining and relating ontologies:
an analysis of problems and solutions.
In : OIS@ IJCAI. 2001.

LPT 5.4.3

LPT 5.4.3: Graph conceptual perspective problem

- Any problem encountered when searching for similarity between two entities described in two different data graphs describing the same domain but with a different perspective.
- « also called difference occurs when two ontologies describe the same domain, at the same level of detail, but from a different perspective. This occurs for maps with different purposes: a political map and a geological map do not display the same objects. » ¹

¹ Ontology matching

J Euzenat, P Shvaiko - 2007 - Springer

LPT 5.4.4

LPT 5.4.4: Graph conceptual logical problem

- Any problem encountered when searching for similarity between two entities described in two different data graphs with differences in logical representation.
- « For example, in some languages it is possible to state explicitly that two classes are disjoint (e.g., disjoint A B), whereas it is necessary to use negation in subclass statements (e.g., A subclass-of (NOT B), B subclass-of (NOT A) in other languages. » ¹

¹ KLEIN, Michel. Combining and relating ontologies:
an analysis of problems and solutions.
In : OIS@ IJCAI. 2001.

LPT 5.4.5

LPT 5.4.5: Graph conceptual key heterogeneity problem

- Any problem encountered when searching for similarity between two entities described in two different data graphs in relation to the constitution of identification keys.
- « Key identification algorithms aim to discover discriminative properties on two datasets independently and thus identify potential candidates for link specifications of property-based state-of-the-art tools »¹

¹ Achichi, Manel; Bellahsene, Zohra; Ellefi, Mohamed Ben; Todorov, Konstantin Linking and disambiguating entities across heterogeneous RDF graphs, 2019.

LPT 5.4.6

LPT 5.4.6: Graph conceptual dataset currentness problem

- Any problem encountered when searching for similarity between two entities described in two different data graphs in relation to the dynamics or temporality of the data.
- « The temporal evolution (or the lack thereof) of data and its dynamicity can lead to conceptual issues across datasets »¹

¹ Achichi, Manel; Bellahsene, Zohra; Ellefi, Mohamed Ben; Todorov, Konstantin Linking and disambiguating entities across heterogeneous RDF graphs,2019.

LPT 5.5

LPT 5.5: Graph semiotique problem

- Any problem encountered when searching for similarity between two entities who are interpreted differently by humans.
- Example : Murder knife and kitchen knife

LPT 5.6

LPT 5.6: Graph schematic discrepancy problem

- Any problem encountered when searching for similarity between two entities described in two different data graphs with schema similarities.

LPT 5.7: Graph scalability Problem

- Any problem encountered when searching for similarity between two entities described in two different data graphs is related to the amount of computing resources needed to solve this task on large data sets.
- The processing time for studying the similarity of individuals in two separate datasets is proportional to the size of the two sets.
- « ...besides the qualitative challenges that the classic IM approaches should face, additional quantitative and scalability challenges are introduced. »¹

¹ ASSI, Ali, MCHEICK, Hamid, et DHIFLI, Wajdi. Data linking over RDF knowledge graphs: A survey. *Concurrency and Computation : Practice and Experience*, 2020, vol. 32, no 19, p. e5746.