

Inferring meta-data from triple store populated datasets

. Regarding triple store representations of data as an 'intermediate' format of what could come from any data-source: relational, XML, NoSQL, etc. the intention is to be able to learn as much as possible from the data in raw form without resorting to any external mechanism for using it later in BI applications.

. 'Canonical' translation of tabular, XML, key-value and other forms of data to triple representation. Automated ETL. This should be done following conventions for triple generation.

. Given that there should be no commitment of using any triple related facilities such as schema in RDF or relating ontologies by inference, the meta-data provided should be available from the raw data itself in its triple form. In principle RDFS, nor any other ontology classification method will be used.

. Materialization: Render inferences into triples again for later use.

. Statements form (Quads, SVO, etc.)

. Subject type inference

Given subjects in the statements loaded in the triple store, the model should give enough insight as to provide assertions of subjects belonging to a given (inferred) subject type.

. Object type inference

Object or values should be recognized as belonging to some type because of being referred to as some kind of value of a predicate. Aggregation of this 'mentions' of the value in predicates will be used to assert its inferred type.

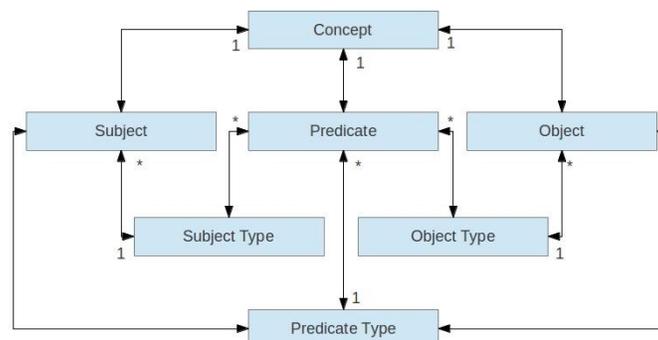
. Predicate (function) type inference

A predicate is to be regarded functionally and, as such, will have domain (subjects) and range (objects). The type of the function is the collection of mappings from subject to object that holds between the triples under consideration.

. Concepts ('reification')

'Reification' goal is to distinguish a 'Concept' as a whole, independently of its references or occurrences in the dataset. For example one could 'reify' some subject, for example a person, and regardless of it being a subject in some statement and object in another.

. Pseudo UML model (triple store data populates subject, predicate and object, the rest is inferred)



Situation recognition & multidimensional addressing

It could be possible, performing the adequate folding in some triple set, to aggregate what could be regarded as abstract 'situations', being performed by 'roles' of some given 'classes'. Here the term 'role' should be used interchangeably with the concept of 'meta class' or the class of the previously mentioned class in the given situation.

Then, at the upper most triple level (see diagram at bottom of page) subjects in some statements will be Actors of the previously mentioned classes and in the given roles. Predicates will aggregated as 'Interactions' or instances of the situation. And finally, objects will be aggregated as 'Occurrences' of the role given the actor in the situation (measures here should be used too because is more appropriate for some kind of situations).

Take, for example, a commercial exchange situation. There will be buyers and sellers (roles). For this example, buyers and sellers will be physical persons or people (the class playing the role). An statement may say that "J sells X", "J price Y", etc. Given the statements in a such form, maybe using meta-data from the techniques mentioned in the previous page or some learning algorithms, some 'folding' could be performed to achieve the arrangement in the picture.

TBD.: This could also be an arrangement for schema less data comprehension. Analysis & dimensional arrangements could be provided adding more meta-data from the inferred situations and using a contextual model for the nodes in the tree. Interactions (sells, price, etc.) should be described in the learned model of the situation as also object roles and actor classes.

TBD.: Example medical examinations, symptoms, diagnose.

TBD.: Example financial markets, variations, flow, price change, etc.

. Pseudo 'fold' model:

