The Typological Database System and the Relation Registry

Menzo Windhouwer (MPI)

Purpose

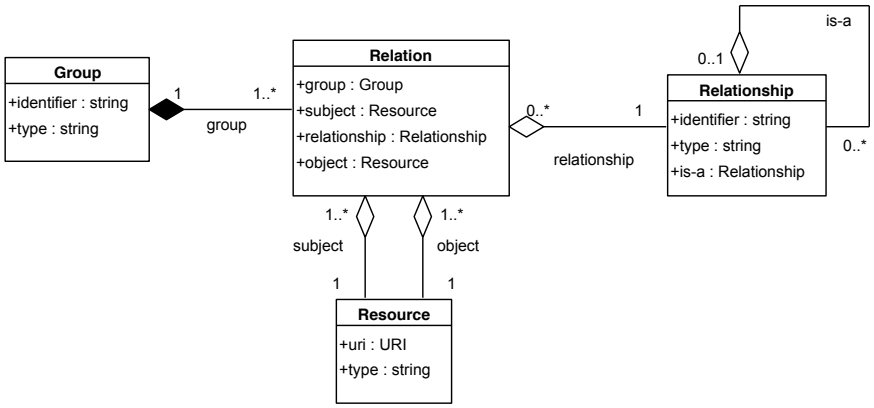
The Typological Database System (TDS; http://languagelink.let.uu.nl/tds/) has a knowledge architecture which uses a hybrid approach, i.e., it has a global ontology and a set of local ones. The concepts in the global ontology will be imported into ISOcat (the Data Category Registry (DCR); http://www.isocat.org/). However, to maintain the full ontology the relationships between the concepts also need to be stored. The idea is that those will be stored in the Relation Registry (RR). Which should mean that the original OWL ontology could/should be retrievable by the combination of the information stored in the DCR and the RR. Also the RR should be used to store (loose) equivalence relationships between TDS concepts and data categories now already in the DCR.

My view on the Relation Registry

I think the RR is basically a triple store, i.e., it contains various types relationships between resources identified by URIs. Although to keep relationships grouped, e.g., to recreate the original TDS global ontology, it may mean it needs to support one of the more extended models, e.g., named graphs (see http://www.w3.org/2004/03/trix/ and the discussion at http://www.jenitennison.com/blog/node/101). However, it may also be enough to maintain the provenance information internally without the need to actually output it.

The need to also reconstruct the original document implies that the original relationships need to be maintained. For example, the TDS uses the propriety tds:equatesWith relationship to indicate that two concepts describe the same linguistic phenomena (but probably from a different theoretical background). This specific relationship should be maintained to be able to reconstruct the original ontology. However it may be placed in a relationship taxonomy, for example under equivalence relationships, so generic algorithms can also utilize the relationship.

This results in the following sketch of an UML diagram for the RR:



Lateron I'll try to give an example instantiation of this model in a set of database tables.

TDS global ontology

For a description of the TDS global ontology see http://www.windhouwer.nl/menzo/professional/papers/E-MELD-2005.pdf. The following sections will describe the concept types and relationships and their mapping to the DCR and/or the RR.

TDS concepts

The TDS global ontology makes the distinction between the following concept types (in the OWL implementation all instantiated as classes):

1. Linguistic objects (Container Data Categories (?)), e.g., Sentence, Morpheme
2. Linguistic properties (Complex and Simple Data Categories), e.g., Basic Word Order, Referential
3. Linguistic relationships (Container Data Categories), e.g., Agreement

TDS Relationships

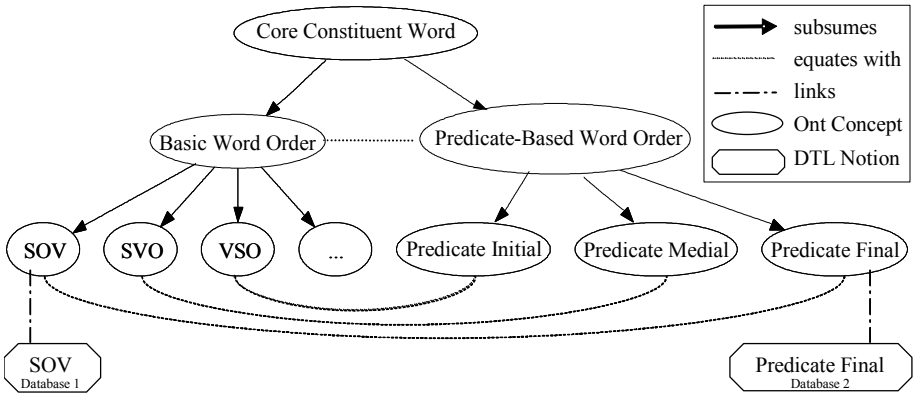
The TDS global ontology currently contains the following relationships:

1. Subsumption (the is-a backbone)
2. Loose synonymy (alternate names; will probably end up in the Name Sections of the Data Category specification, i.e., not a relationship in the RR sense)
3. Related phenomena (the tds:equivalentWith relationship)
4. Meronomy ((in)direct part-of relationships; implemented along the lines of http://www.w3.org/2001/sw/BestPractices/OEP/SimplePartWhole/index.html)
5. Determination (specific role in reified Linguistic relationships)
6. Form-function (specific role in reified Linguistic relationships)
7. Other roles in reified Linguistic relationships

Either existing OWL class relationships statements are used, or OWL object properties are created.

An example

Let's reuse the example in Figure 5 of the E-MELD 2005 paper:



We'll ignore the lowest part containing the DTL Notions, as those are part of the local TDS ontologies. To keep the example simple I'll just reuse the labels as the identifying URIs for the classes. For each class in the UML diagram we'll create a table.

Table 1 Relation

group	subject	relationship	object
TDS global ontology	Basic Word Order	rdfs:subClassOf	Core Constituent Word Order
TDS global ontology	SOV	rdfs:subClassOf	Basic Word Order
TDS global ontology	SVO	rdfs:subClassOf	Basic Word Order
TDS global ontology	VSO	rdfs:subClassOf	Basic Word Order
TDS global ontology	Predicate-Based Word Order	rdfs:subClassOf	Core Constituent Word Order
TDS global ontology	Predicate Initial	rdfs:subClassOf	Predicate-Based Word Order
TDS global ontology	Predicate Medial	rdfs:subClassOf	Predicate-Based Word Order
TDS global ontology	Predicate Final	rdfs:subClassOf	Predicate-Based Word Order
TDS global ontology	Basic Word Order	tds:equatesWith	Predicate-Based Word Order
TDS global ontology	SOV	tds:equatesWith	Predicate Final

TDS global ontology	SVO	tds:equatesWith	Predicate Medial
TDS global ontology	VSO	tds:equatesWith	Predicate Initial

Table 2 Group

identifier	type
TDS global ontology	OWL 1.0

Table 3 Resource

uri	type
Core Constituent Word	OWL class
Basic Word Order	OWL class
SOV	OWL class
SVO	OWL class
VSO	OWL class
Predicate-based Word Order	OWL class
Predicate Initial	OWL class
Predicate Medial	OWL class
Predicate Final	OWL class

Table 4 Relationship

identifier	type	is-a
is-a	generic	
related-to	generic	
rdfs:subClassOf	OWL 1.0 – builtin	is-a

tds:equatesWith	OWL 1.0 – object property	related-to

Notes

Some random notes:

1. Symmetric relationships, e.g., tds:equivalentWith is one, could be encoded as a property of the relationship type but probably we just want to instantiate it twice (which I didn't do in the example above; so for example also the tuple (TDS global ontology, Predicate Initial, tds:equatesWith, VSO) next to the tuple (TDS global ontology, VSO, tds:equatesWith, Predicate Initial) would be created in the Relation table

2. A student assistant created for me a mapping from TDS concepts to existing data categories in the DCR (unfortunately no overlap with this example, i.e., all those concepts should be new to the DCR). These could be easily added to the Relation table. Probably with another group identifier and using the generic relationship types. Actually the mapping states the following related-to sub types: exact, subset and superset. Those additional types maybe useful in the generic set (which we need to identify anyway).

group	subject	relationship	object
TDS – DCR	Core Constituent Word	related-to	http://www.isocat.org/datcat/ISO-DC-1234

3. …