

The Semantics of SPARQL

Enrico Franconi and Sergio Tessaris
Free University of Bozen-Bolzano, Italy

2 November 2005

Definition 1 (RDF graphs)

We call \mathcal{RDF}_T the set of RDF terms, namely the union of all RDF URI references with the set of all literals in their canonical representation and with all the bnode names; the canonical representation of a literal is a chosen representative of all the literals associated to the same value, if the literal is non ill-typed, otherwise it is the literal itself. RDF graphs are defined in the standard way.

Definition 2 (Pattern Solution)

A pattern solution (aka variable substitution) is a partial function from a finite subset, possibly empty, of the query variables alphabet to \mathcal{RDF}_T ; optionally, we may consider pattern solutions mapping to \mathcal{RDF}_T without bnodes.

Definition 3 (Answer Set)

An answer set is a set of pattern solutions. An answer set is non-redundant if it does not contain any redundant tuple.

A tuple $S \in \mathcal{AS}$ is redundant in an answer set \mathcal{AS} if $L(\mathcal{AS})$ – the logical translation of the answer set – is logically equivalent to $L(\mathcal{AS} \setminus \{S\})$ – the logical translation of the answer set diminished by the tuple.

The logic translation $L(\mathcal{AS})$ of an answer set \mathcal{AS} is a closed existential predicate logic formula, where URIs and literals (in their canonical representation) are constants and blank nodes are existential variables, and the body is a conjunction of atomic formulae of some predicate R in correspondence with each solution $S \in \mathcal{AS}$. This correspondence is built as follows:

- the arity of R is equal to the cardinality of the set V^A , which is the union of the active domain of each S ;
- there is a bijective correspondence between the positions of the arguments of R and the query variables in V^A ;
- an atomic formula in the conjunction is built for each S by putting in each position i of the predicate R either the value of S for the variable corresponding to i – if S is defined for it – or a special constant ϵ otherwise.

Definition 4 (RDF Merge)

The RDF merge $\uplus\langle \mathcal{G}_1 \dots \mathcal{G}_n \rangle$ of a sequence of graphs $\langle \mathcal{G}_1 \dots \mathcal{G}_n \rangle$ (i.e., a dataset)

is the ordered merge union of the graphs, where repeated bnodes are substituted with fresh ones, by keeping the names of the bnodes coming first in the sequence order. Note that, w.r.t. the standard definition of RDF merge, if any of the graphs contains variables, then those are not renamed (i.e., variables are treated as URIs).

Definition 5 (*Basic Graph Pattern*)

A basic graph pattern is a finite RDF graph where query variables are allowed to label nodes and edges, in alternative to elements of \mathcal{RDF}_T . Variables may be annotated with a set of elements of \mathcal{RDF}_T to restrict the possible assignments of variables.

Definition 6 (*Answer Set of a BGP query*)

Given an RDF graph \mathcal{G} and a basic graph pattern query \mathcal{Q} , the answer set (resp. non-redundant answer set) of \mathcal{Q} for \mathcal{G} is the maximum set (resp. maximum non-redundant set) of pattern solutions such that for each such \mathcal{S} :

1. the domain of \mathcal{S} is restricted to the query variables in \mathcal{Q} ;
2. the bnodes in the range of \mathcal{S} can only be among those in \mathcal{G} ;
3. if a variable is annotated with a set of elements of \mathcal{RDF}_T then \mathcal{S} must map that variable to one of these elements;
4. $\mathcal{Q}_{[\mathcal{S}]}$ - i.e., the application of the substitution \mathcal{S} to \mathcal{Q} - is a well-typed RDF graph;
5. $\mathcal{G} \models (\mathcal{G} \uplus \mathcal{Q})_{[\mathcal{S}]}$
where \models is a semantically well defined logical implication relation among RDF graphs, namely either simple entailment, or RDF entailment, or RDFS entailment, or OWL-DL entailment, or OWL-Full entailment.

Lemma 1

1. Both the answer set and the non-redundant answer set of a query for a graph are unique.
2. An answer set without substitutions to bnodes is always non-redundant.
3. In the case of simple, RDF, and RDFS entailment, the answer set of any query for a ground graph (i.e., a graph without bnodes) is always non-redundant.
4. In the case of simple entailment, the answer set corresponds to all the substitutions defined as follows:
 - consider all the query variables in \mathcal{Q} as fresh bnodes;
 - consider all subgraph isomorphisms between \mathcal{Q} and \mathcal{G} , where bnodes may be mapped to any node;

- a substitution is the projection over the bnodes in \mathcal{Q} that were query variables of the above isomorphisms.

Definition 7 (SPARQL Algebra – still informal)

A SPARQL query expression constructs an answer set from an RDF graph, by combining basic steps of the form of basic query patterns – constructing answer sets from the RDF graph – with higher level algebraic operators which transform answer sets. Algebraic operators are:

1. *Join*: $AS \times AS \rightarrow AS$
2. *Filter*: $AS \rightarrow AS$
3. *Optional*: $AS \times AS \rightarrow AS$
4. *Union*: $AS \times AS \rightarrow AS$
5. *Select*: $AS \rightarrow AS$

Optionally, a *Construct* operator may at the end transform an answer set in an RDF graph.