Toby Segaran - Programming the Semantic Web - 2009

Tom Rochette <tom.rochette@coreteks.org>

July 24, 2025 — daae079c

0.1 Context

0.2 Learned in this study

0.3 Things to explore

1 Overview

2 Notes

2.1 1. Why Semantics?

- Semantics: symbols can refer to things or concepts, and sequences of symbols convey meaning
- Ways to model data
  - Tabular Data
    - Easy to understand and formulate
    - Fields which contain more than a single piece of information are harder to use/parse
    - The models are rigid, limited and usually not changeable by the users
  - Relational Data
    - Fast and powerful tools for storing large sets of data where the data model is well understood and the usage patterns are fairly predictable
    - Pivot tables are essentially tables that indicate equality between two other tables entries
    - Schema migration is often a huge headache
  - Semantic Relationships
    - Extremely flexible, adaptible to new changes describing the data
    - Gets rid of normalization, degrades performance, columns type cannot be set/constrained

2.2 2. Expressing Meaning

- The triple is the fundamental building block of semantic representations
- It is composed of a subject, a predicate, and an object
- A subject corresponds to an entity - a "thing" for which we have a conceptual class
- Predicates are a property of the entity to which they are attached
- Objects are either entities or literal values such as strings or numbers
- Multiple triples can be tied together by using the same subjects and objects in different triples

2.3 ## 3. Using Semantic Data

2.4 4. Just Enough RDF

2.4.1 The RDF Data Model

- RDF is a language for expressing data models using statements expressed as triples
- Each statement is composed of a subject, a predicate, and an object

2.4.2 Resources

- A resource is simply anything that can be identified with a Universal Resource Identifier (URI)

2.4.3 Blank Nodes

- Blank nodes are graph nodes that represent a subject (or object) for which we would like to make assertions, but have no way to address with a proper URI

2.4.4 Literal Values

- Can have an optional language (e.g., English, French) and type (e.g., integer, boolean, string, float)

2.4.5 RDF Serialization Formats

- Four serialization formats:
  - N-Triples
  - N3
  - RDF/XML
  - RDF in attributes (RDFa)
- N3 uses the letter "a" as a predicate representing the RDF "type" relationship
  - tom a foaf:Person

2.4.6 SPARQL

- SPARQL provides four forms of queries:
  - SELECT
  - CONSTRUCT
  - ASK
  - DESCRIBE

2.5 5. Sources of Semantic Data

- Degree: Number of nodes connected to a given node
- Betweenness centrality: Centrality is defined as the percentage of shortest paths in the graph that pass through a given node
- Clique: A group of nodes that are all connected to one another. The smallest cliques have only two members (the two connected nodes)
- Clustering: Calculated from the fraction of its neighbors that are connected to one another
- Data models within Freebase are called schemas and are broken down along areas of interest called domains. Domains serve only to collect components of a data model
- In Freebase, any object can have one or more types
- Types provide a basic categorization of objects within Freebase and indicate what properties you might expect to find linked to the object
- A domain is also a namespace, and types have keys in domains (e.g., /people/person -> domain = people, type = person)
- Types operate as namespaces for properties (e.g., /people/person/date_of_birth)
- Every object in Freebase is automatically given the type object (/type/object), which provides properties available to all objects, such as name, id, and type
- The type property (/type/object/type) is used to link the object with type definitions
- In a self-similar fashion, types themselves are nothing more than objects that have a /type/object/type link to the root type object /type/type

2.6 6. What Do You Mean, "Ontology"?

- Ontologies allow us to express the formal rules for inference

- Classes are used to define the characteristics of a group of things and to specify their relationships to other classes
- Classes describe groups of entities
- The collection of types that use the property is called the domain of the property
- A property definition may also indicate which types of values this property can take on, representing the range of the property
- When a property does not indicate its domain, we cannot infer anything about the resources it is describing, as it may be used by any type of resource
- If the property define a type as its domain, then we can infer that anything described by that property is of the domain type
- If the property defines several types as its domain, we can infer that the resource described by that property is all of the domain types
- If a property does not specify a range, then we can't infer anything about the value of the property
- If the property specifies one or more types as its range, then a value of the property can be inferred to be all of the types specified by the property range
- The process of making a subject-predicate-object statement in a subject is called reification in RDF
- OWL is broken into three sub-languages of increasing complexity and expressiveness
  - OWL-Lite
  - OWL DL
  - OWL Full
- Data modeling is not a singular activity, is it iterative:
  - Build a model
  - Populate some instances
  - Run some queries
  - Repeat

3 Entities of interest

- rdf:type: The type of a resource. This property specifies that a resource is an instance of a specific RDFS class
- rdf:XMLLiteral: The class of all XML literal values
- rdfs:domain: Specifies that a property has a domain of a specified class. In a triple, the subject will always be an instance of the class specified by the rdfs:domain property
- rdfs:range: Species that a property as a range of a specified class. In a triple, the object will always be an instance of the class specified by the rdfs:range property
- rdfs:subclassOf: Specifies that a class is a subclass of another class, and therefore that all instances of the subclass are also instances of the superclass
- owl:Thing: All classes implicitly subclass this class
- owl:Class: The class of RDF resources that are classes
- owl:ObjectProperty: Indicate that the linked entity is an object property. The class of all properties that have ranges that are instances of owl:Class
- owl:DatatypeProperty: Indicate that the linked entity is a datatype (intrinsic) property (string, integer, float, character, boolean, date, datetime, etc.). The class of all properties that have ranges that are literals, and instances of rdfs:Datatype
- owl:FunctionalProperty/owl:InverseFunctionalProperty
- owl:inverseOf: Indicates that a property is an inverse of the given property (director directed film vs film directedBy director)

4 Definitions

- RDF: Resource Description Framework
- RDFS: Resource Description Framework Schema
- FOAF: Friend Of A Friend

- OWL: Web Ontology Language

5 See also

6 References

- Segaran, Toby, Colin Evans, and Jamie Taylor. Programming the semantic web. " O'Reilly Media, Inc.", 2009.