Knowledge Representation and Reasoning for Semantic Web

Svatopluk Šperka

Faculty of Information Technology, Brno University of Technology

December 31, 2009

Semantic Web

- Contemporary web consists mostly of documents structured for *presentation* purposes, not for *machine processing*.
- Non-human manipulation of informations contained in such documents is fundamentally limited.
- Two possible scenarios:
  1. Develop increasingly sophisticated tools to extract information from hypertext documents by means of AI and computational linguistics.
  2. Publish informations in machine-processable form. That is the idea of *Semantic Web*.
- Endorsed by *W3C* and its standards.

Knowledge

- In our context – what is known; facts and informations.
- *Assertional knowledge* – concerning concrete entities.
- *Terminological knowledge* – concerning generalized concepts and facts.
- *Ontology* – Explicit and formal specification of conceptualization.
- Need to capture knowledge in a form that is processable by machines, i.e. explicitly and formally, again.

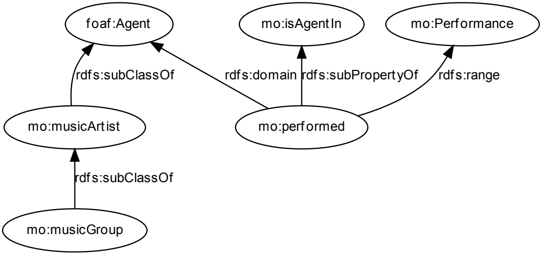
Resource Description Framework

- Data model for structuring information.
- Basic element is *triple* – universal linguistic construct consisting of *subject*, *predicate* and *object*. Each triple represents a *statement*.
- In RDF terminology there are *resources* and *properties* (and statements).
- Additional terms – *literal*, *blank node*.
- Set of statements form labeled directed multigraph.
- Labels are *URI*s – *Uniform Resource Identifiers*.
- Resource can be typed and thus become instance of particular class.

Example of a triple as graph.

RDF Schema

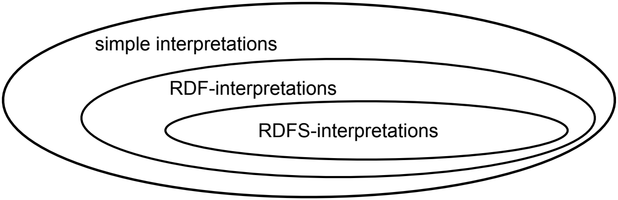
- Light-weight ontology language.
- Following interdependencies in domain are expressible by *RDFS*:
  - class hierarchies using subclassing
  - property hierarchies using subproperty construct
  - property restrictions – its domain and range
- *RDF* and *RDFS* itself is defined in *RDF Schema*.



Excerpt from Music Ontology (http://musicontology.com)

RDF(S) Formal Semantics

- Set of *propositions* $\mathbb{P}$. In our case propositions are triples.
- *Entailment relation*: $\models \subseteq 2^{\mathbb{P}} \times 2^{\mathbb{P}}$ (e.g. $\{p_1, p_2\} \models \{p_3, p_4\}$).
- *Model-Theoretic Semantics*:
  - interpretation I is model of p ($I \models p$), if I satisfies p
  - extendable to set $P \subseteq \mathbb{P}$. $I \models P \Leftrightarrow \forall p \in P : I \models p$
  - $P' \subseteq \mathbb{P}$ is entailed by $P \subseteq \mathbb{P}$ ($P \models P'$) *iff* every interpretation I such that $I \models P$ also $I \models P'$
- Hierarchy of graph interpretations.



Simple Interpretation

- Simple interpretation I given vocabulary V:
  - IR – non-empty set of *resources* (universe of discourse)
  - IP – set of *properties* of I (may overlap with IR)
  - $I_{EXT} : IP \rightarrow 2^{IR \times IR}$. $I_{EXT}(p)$ is *extension* of property p
  - $I_S : V \rightarrow IR \cup IP$
  - I_L – function from typed literals from V into IR
  - LV – subset of IR called *literal values* containing plain literals
- Interpretation function $\cdot^{\mathcal{I}}$:
  - untyped literal: $(\text{"}\mathbf{a}\text{"})^{\mathcal{I}} = \mathbf{a}$
  - untyped literal with language information: $(\text{"}\mathbf{a}\text{"}@\mathbf{t})^{\mathcal{I}} = \langle \mathbf{a}, \mathbf{t} \rangle$
  - typed literal $\mathbf{l}$: $(\mathbf{l})^{\mathcal{I}} = I_L(\mathbf{l})$
  - URI $u \in V$: $u^{\mathcal{I}} = I_S(u)$
- Truth value of grounded triple. s p o.$^{\mathcal{I}} = \mathit{true}$ iff
  - $\mathtt{s}$, $\mathtt{p}$, o are in vocabulary V
  - $(\mathtt{s}^{\mathcal{I}}, \mathsf{o}^{\mathcal{I}}) \in I_{EXT}(\mathtt{p}^{\mathcal{I}})$
- $G^{\mathcal{I}} = \mathit{true} \Leftrightarrow \forall T \in G : T^{\mathcal{I}} = \mathit{true}$
- $G_1 \models_S G_2$ if every simple interpretation that is model of G_1 is model of G_2.

RDF vocabulary V_{RDF}

- `rdf:type rdf:Property rdf:XMLLiteral rdf:nil`
- `rdf:Statement rdf:subject rdf:predicate rdf:object`
- `rdf:List rdf:first rdf:rest`
- `rdf:Seq rdf:Bag rdf:Alt`
- `rdf:value`

RDF-Interpretations

- RDF-interpretation of V is simple interpretation of $V \cup V_{RDF}$ that satisfies:
  - $x \in IP$ iff $(x, \texttt{rdf:Property}^{\mathcal{I}}) \in I_{EXT}(\texttt{rdf:type}^{\mathcal{I}})$
  - proper handling of well-typed and ill-typed XML literals
- Axiomatic triples:
  - `rdf:` $\{\texttt{type}, \texttt{subject}, \texttt{predicate}, \texttt{object}, \texttt{first}, \texttt{rest}, \texttt{value}, \texttt{\_i}\}$ `rdf:type rdf:Property`
  - `rdf:nil rdf:type rdf:List`
- $G_1 \models_{RDF} G_2$ if every RDF-interpretation that is model of G_1 is model of G_2.

RDFS vocabulary V_{RDFS}

- rdfs:domain rdfs:range
- rdfs:Resource rdfs:Literal rdfs:Datatype
- rdfs:Class rdfs:subClassOf rdfs:subPropertyOf
- rdfs:member rdfs:Container rdfs:ContainerMembershipProperty
- rdfs:comment rdfs:seeAlso rdfs:isDefinedBy
- rdfs:label

RDFS-Interpretations

- $I_{CEXT} : IR \rightarrow 2^{IR}. I_{CEXT}(y) = \{x \mid (x, y) \in I_{EXT}(\texttt{rdf:type}^{\mathcal{I}})\}$
- $IC = I_{CEXT}(\texttt{rdfs:Class}^{\mathcal{I}})$
- RDFS-interpretation of V is RDF-interpretation of $V \cup V_{RDFS}$ that satisfies:
  - $IR = I_{CEXT}(\texttt{rdfs:Resource}^{\mathcal{I}})$.
  - $LV = I_{CEXT}(\texttt{rdfs:Literal}^{\mathcal{I}})$.
  - $(x, y) \in I_{EXT}(\texttt{rdfs:domain}^{\mathcal{I}}) \wedge (u, v) \in I_{EXT}(x) \Rightarrow u \in I_{CEXT}(y)$. Analogously for `rdfs:range`.
  - $I_{EXT}(\texttt{subPropertyOf}^{\mathcal{I}})$ is reflexive and transitive on IP. Analogously for `rdfs:subClassOf` and IC.
  - $(x, y) \in I_{EXT}(\texttt{rdfs:subPropertyOf}^{\mathcal{I}}) \Rightarrow x, y \in IP \wedge I_{EXT}(x) \subseteq I_{EXT}(y)$
    Analogously for `rdfs:subClassOf`.
  - $x \in IC \Rightarrow (x, \texttt{rdfs:Resource}^{\mathcal{I}}) \in I_{EXT}(\texttt{rdfs:SubclassOf}^{\mathcal{I}})$.
  - ...

Syntactic Reasoning with Deduction Rules

- Deduction rule has form $\frac{p_1 \cdots p_n}{p}$.
- P' can be derived from P using deduction calculus. $P \vdash P'$.
- *Soundness*: $P \vdash P' \Rightarrow P \models P'$.
- *Completeness*: $P \models P' \Rightarrow P \vdash P'$.
- Soundness and completeness does not guarantee *decidability* (e.g. FOPL).

Simple Entailment

- URIs have no special meaning – all are treated as equal.
- Two deduction rules:
  - $\frac{u\ a\ x.}{u\ a\ \_:n.}\ \mathrm{se1} \qquad \frac{u\ a\ x.}{\_:n\ a\ x.}\ \mathrm{se2}$
- Rules can be safely applied if $\_:n$ is not in graph or has been introduced by *weakening* same URI, literal or blank node.
- **Theorem**: Graph G_1 simply entails graph G_2, if G_1 can be extended to graph G_1' by virtue of the rules *se1*, *se2* such that G_2 is contained in G_1' ($G_2 \subseteq G_1'$).

RDF Entailment

- Special meaning of URIs in vocabulary V_{RDF} has to be reflected in additional deduction rules.
- For axiomatic triples
  - $\overline{u\,a\,x.}\ \mathrm{rdfax}$
- To deduce property type for URIs used as predicates
  - $\dfrac{u\,a\,y.}{a\ \texttt{rdf:type}\ \texttt{rdf:Property}.}\ \mathrm{rdf1}$

Complexity of Entailment

- Can be shown that set of inferable triples cannot become infinite.
- RDFS entailment is decidable.
- Simple, RDF and RDFS entailments are NP-COMPLETE (can be transformed to deciding graph homomorphism).
- Without blank nodes it is in complexity class P.

Web Ontology Language

- Its acronym is *OWL*.
- RDFS can't express:
  - local scope of properties.
  - disjointness of classes.
  - creating classes as combination of others.
  - cardinality restrictions.
  - characteristics of properties.
- W3C recommended standard for ontology modeling.

Description Logic

- Description Logics is family of logics for knowledge representation classified by allowed language constructs.
- Generally subset of Predicate Logic
- Very well studied concerning complexity and decidability.
- Basic (usable) DL is $\mathcal{ALC}$ (Attribute Language with Complement)
  - *ABox* – $C(a), R(a, b)$
  - *TBox* – $C \equiv D, C \sqsubseteq D$ (C, D are concept descriptions)
    $C, D ::= A \mid \top \mid \bot \mid \neg C \mid C \sqcap D \mid C \sqcup D \mid \forall R.C \mid \exists R.C$
- DL can be extended by allowing additional constructs
  - $\mathcal{S}$ – $\mathcal{ALC}$ with role transitivity
  - $\mathcal{H}$ – role hierarchies ($R \sqsubseteq S$)
  - $\mathcal{O}$ – nominals (closed classes with one element)
  - $\mathcal{N}$ – cardinality restrictions ($\leq nR.C$)
  - ...
- More on notation and complexity of DL flavours
  `http://www.cs.manchester.ac.uk/~ezolin/dl`

Model-theoretic Semantics of OWL

- Vocabulary
  - $\mathbf{I}$ – set of symbols for individuals
  - $\mathbf{C}$ – class names
  - $\mathbf{R}$ – roles
- Functions mapping symbols to *domain of interpretation*
  - $I_{\mathbf{I}} : \mathbf{I} \rightarrow \Delta$
  - $I_{\mathbf{C}} : \mathbf{C} \rightarrow 2^{\Delta}$ (class extension)
  - $I_{\mathbf{R}} : \mathbf{R} \rightarrow 2^{\Delta \times \Delta}$ (property extension)
- Interpretation function $\cdot^{\mathcal{I}}$
  - $\top^{\mathcal{I}} = \Delta \qquad \bot^{\mathcal{I}} = \emptyset$
  - $\neg C = \Delta \setminus C \qquad C \sqcap D = C^{\mathcal{I}} \cap D^{\mathcal{I}}$
  - $(\forall R.C)^{\mathcal{I}} = \left\{ x \mid \forall y.(x, y) \in R^{\mathcal{I}} \Rightarrow y \in C^{\mathcal{I}} \right\}$
  - $(\leq nR.C)^{\mathcal{I}} = \left\{ x \mid \left| \left\{ (x, y) \in R^{\mathcal{I}} \mid y \in C^{\mathcal{I}} \right\} \right| \leq n \right\}$
  - ...
- Interpretation $\mathcal{I}$ is model of knowledge base K ($\mathcal{I} \models K$) if axioms of knowledge base hold.
  - $C(a) \in K \Rightarrow a^{\mathcal{I}} \in C^{\mathcal{I}}$
  - ...

OWL Species

- OWL species:
  - *OWL Lite* – $\mathcal{SHIF}(D)$, decidable, ExpTime worst case complexity
  - *OWL DL* – $\mathcal{SHOIN}(D)$, decidable, NExpTime worst case complexity
  - *OWL Full* – undecidable, semantically difficult to understand, unsupported by software tools
- OWL2 species
  - *OWL 2 DL* – $\mathcal{SROIQ}$, decidable, NExpTime, backward compatible with OWL DL
  - *OWL 2 EL* with polynomial time algorithms
  - *OWL 2 QL* for implementation using relational databases, polynomial algorithms for all standard inference types
  - *OWL 2 RL* with polynomial time algorithms using rule-based reasoning
  - *OWL 2 Full* is backward compatible with OWL Full

Conclusion

Expressivity vs. reasoning efficiency

Bibliography

- Pascal Hitzler, Markus Krötzsch and Sebastian Rudolph
  Foundations of Semantic Web Technologies
  CRC Press, 2009
- Grigoris Antoniou, Frank van Harmelen
  A Semantic Web Primer, 2nd ed.
  MIT Press, 2008
- Herman J. ter Horst
  Completeness, decidability and complexity of entailment for RDF Schema and a semantic extension involving the OWL vocabulary
  ISWC 2004
- Jean-François Baget
  RDF Entailment as Graph Homomorphism
  ISWC 2005

Bibliography

- Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, Peter Patel-Schneider (editors)
  The Description Logic Handbook: Theory, Implementation and Applications
  Cambridge University Press, 2003
- **RDF Semantics**
  http://www.w3.org/TR/rdf-mt/