The Semantic Web
How the Web can exploit Knowledge Representation (and the other way round)

Frank van Harmelen
AI Department
Vrije Universiteit Amsterdam
Goal

- General *propaganda* for the Semantic Web
- Convince the old folk there is *real substance* there
- Convince the young folk there is *exciting work* to be done

- No: formal definitions, theorems, proofs, complexity results, benchmarks
Semantic Web

- The vision & politics
- What is required
- Some technology
  - XML, RDF, DAML+OIL
- The underlying logic
- Research directions?
The computer has changed:

- **first**: computer = computing
- **then**: computer = games, text processing and powerpoint-presentations
- **now**: “computer” = entry point to info space
Semantic Web: the vision

WWW is an impressive success:
- amount of available information (1.6 Giga-page)
- number of web-servers (30 million)
- number human users (500 million)

However, we’ve only seen two generations:
1. handwritten HTML
2. database generated pages

The real power will come with the 3rd generation:
3. machine accessible semantics
on the Web

What it's like to be a machine
“Intelligent” things we can’t do today

- **Search engines**
  - concepts, not keywords
  - semantic narrowing/widening of queries

- **Shopbots**
  - semantic interchange, not screenscraping

- **Service description and integration**

- **Navigation**
  - by semantic proximity, not hardwired links

- .....
Semantic Web: the politics

- Focus of $80 million DARPA project DAML
- Focus of €20 million EU action line
  (and much more to come)
- Pushed hard by Tim Berners-Lee / W3C
  (“Weaving the Web”)
- Is central to EU 6th Framework
Semantic Web

- The vision & politics
- **What is required**
- Some technology
  - XML, RDF, DAML+OIL
- The underlying logic
- Research directions?
machine accessible meaning

(What it’s like to be a machine)
Required are:

- **Explicit meta-data** for Web-resources:
  Web pages carry their content on their sleeve

- **Shared domain descriptions**
  as basis for meta-data
  (“ontologies”)

- **Machine-processable Web-content**
Shared content-vocabularies: Ontologies

Formal, explicit specification of a shared conceptualisation of a shared consensual knowledge abstract model of some domain, machine processable concepts, properties, relations, functions.
Shared domain descriptions ("ontologies")

- Classes + class-hierarchy
- instances
- slots/values
- inheritance (multiple? defaults?)
- restrictions on slots (type, cardinality)
- properties of slots (symm., trans., …)
- relations between classes (disjoint, covers)
- reasoning tasks: classification, subsumption
Real life examples

- Lightweight:
  - Yahoo topic hierarchy
  - Open directory (400,000 general categories)

- Heavy weight
  - Cyc, 300,000 axioms

- Very specific
  - SNOMED (200,000 medical concepts)
  - DAML library (180 ontologies)
  - **METAR code**
    (international code for weather conditions at air terminals)
Semantic Web

- The vision & politics
- What is required
- **Some technology**
  - XML, RDF, DAML+OIL
- The underlying logic
- Research directions?
TBL talk: (XML 2000)

- Self-desc. doc.
- Data
- Rules
- Proof
- Ontology vocabulary
- RDF + rdfschema
- XML + NS + xmschema
- Unicode
- URI
- Digital Signature
- Trust
XML: Document = labelled tree

- node = label + attr/values + contents

```
<course date="...">
  <title>...</title>
  <teacher>...</teacher>
  <name>...</name>
  <http>...</http>
  <students>...</students>
</course>
```

- **DTD**: simple grammar to describe legal trees
- **XML Schema**: not so simple grammar for the same

- So:
  **why not use XML to represent ontologies?**
XML ≠

machine accessible meaning

在樂迷們熱心奔走之下，樂音港指揮家林克昌離台之際，可能性又提升了幾分。兩廳院主任李炎、國家音樂廳董事長黃子明日前親訪林克昌，在香港所開會，並提出多場客席邀約。此外，台灣省立交響樂團團長陳維雄也早提「下訂」，邀請林克昌赴台中演譯。從八月十日起至九月止，為期長達一個月。

在台灣許多公家機構中，陳維雄是以實際行動表達對林克昌肯定的業界人士之一，曾多次公開表示對林克昌指揮才華的欣賞，且幾乎每個樂季都邀請林克昌客席演出。

此外，林克昌上個月在俄羅斯與頂尖的「俄羅斯國家管弦樂團」領銜柴可夫斯基晚期三大交響曲以及「羅密歐與茱麗葉」、「格林卡里拉舞曲」、「薩格勒布頌歌」，最後的ＤＡＴ母帶也在兩天內寄回台灣。製作人楊志銘與林克昌執演之後，都對錄音成果——尤其是錄音表現感到相當滿意，楊志銘估計呈現了七分林克昌指揮神韻。

俄羅斯國家管弦樂團首席布爾尼日前也親訪林克昌的指揮藝術有三大特點：一是控制自如的彈性速度；二是強烈的動態對比；三是宛如呼吸般連繫的旋律處理。這些對音響而言都構成很大挑戰。俄羅斯音師雖然採用多軌混音，但定位、場面都有可觀之處。
The semantic pyramid again

- **Self-desc. doc.**
- **Data**
- **Rules**
- **Proof**
- **Logic**
- **Ontology vocabulary**
- **RDF + rdfschema**
- **XML + NS + xmlschema**
- **Unicode**
- **URI**
- **Trust**
- **Digital Signature**
Bluffer’s guide to RDF (1)

- **Object -> Attribute -> Value** triples

- Objects are **web-resources**

- Value is again an Object:
  - Triples can be **linked**
  - Data-model = graph

---

Diagram:

- **pers05** is linked to **ISBN...** via Attribute: Author-of
- **ISBN...** is linked to **MIT** via Attributes: Publ-by
- **ISBN...** is linked to itself via Attributes: Publ-by
- **ISBN...** is linked to **pers05** via Attribute: Author-of
Bluffer’s guide to RDF (2)

- Every identifier is a URL
  - = world-wide unique naming!
- Has XML syntax

Any statement can be an object
- graphs can be nested

NYT claims

pers05

Author-of

ISBN...
What does RDF Schema add?

- Defines **vocabulary** for RDF
- Organizes this vocabulary in a **typed hierarchy**
  - Class, subClassOf, type
  - Property, subPropertyOf
  - domain, range

```
Person
  \_ subClassOf
    \_ Author
      \_ communicatesTo
        \_ domain
          \_ Frank
            \_ communicatesTo
              \_ range
                \_ Reader
                  \_ subClassOf
                    \_ Lynda
```

Frank communicatesTo Lynda
The semantic pyramid again
Beyond RDF: DAML+OIL

DAML+OIL extends RDF Schema to a full-fledged knowledge representation language.

- logical expressions
- data-typing
- cardinality
- quantifiers
DAML+OIL as RDF(S) extension

- class-def
- subclass-of
- slot-def
- subslot-of
- domain
- range

- class-expressions
  - AND, OR, NOT
- slot-constraints
  - has-value, value-type
  - cardinality
- slot-properties
  - trans, symm
DAML+OIL: politics

- Officially required for US DAML programme
- De facto required for EU Semantic Web action line
- Very fast take-up in research community:
  - editors
  - browsers,
  - visualisers
  - inference engines
  - storage & query
  - ...
- Early industrial commitment: Glaxo Smith Klein
- Now almost a W3C recommendation (OWL)
  - March ‘03
Semantic Web

- The vision & politics
- What is required
- Some technology
  - XML, RDF, DAML+OIL
- The underlying logic
- Research directions?
Non-standard model-theory (Pat Hayes)
  • E.g: classes members of themselves

Simple model-theoretic properties:
  • Entailment,
  • skolemisation,
  • (strong) Herbrand property,
  • interpolation theorem

Axiomatisations (Stanford, Essen, Lyon)
### DAML+OIL: Constructors

<table>
<thead>
<tr>
<th>Constructor</th>
<th>Abbreviation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>intersectionOf</td>
<td>$C_1 \land \ldots \land C_n$</td>
<td>Human $\land$ Male</td>
</tr>
<tr>
<td>unionOf</td>
<td>$C_1 \lor \ldots \lor C_n$</td>
<td>Doctor $\lor$ Lawyer</td>
</tr>
<tr>
<td>complementOf</td>
<td>$\neg C$</td>
<td>$\neg$ Male</td>
</tr>
<tr>
<td>oneOf</td>
<td>${x_1 \ldots x_n}$</td>
<td>${\text{john, mary}}$</td>
</tr>
<tr>
<td>toClass</td>
<td>$\forall P.C$</td>
<td>$\forall \text{hasChild.Doctor}$</td>
</tr>
<tr>
<td>hasClass</td>
<td>$\exists P.C$</td>
<td>$\exists \text{hasChild.Lawyer}$</td>
</tr>
<tr>
<td>hasValue</td>
<td>$\exists P.{x}$</td>
<td>$\exists \text{citizenOf.{USA}}$</td>
</tr>
<tr>
<td>minCardinalityQ</td>
<td>$\geq n P.C$</td>
<td>$\geq 2 \text{hasChild.Lawyer}$</td>
</tr>
<tr>
<td>maxCardinalityQ</td>
<td>$\leq n P.C$</td>
<td>$\leq 1 \text{hasChild.Male}$</td>
</tr>
<tr>
<td>cardinalityQ</td>
<td>$= n P.C$</td>
<td>$= 1 \text{hasParent.Female}$</td>
</tr>
</tbody>
</table>

+ XML Schema datatypes:
  - int, string, real, etc
# DAML+OIL: Axioms

<table>
<thead>
<tr>
<th>Axiom</th>
<th>Abbreviation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>subClassOf</td>
<td>$C_1 \sqsubseteq C_2$</td>
<td>Human $\sqsubseteq$ Animal $\land$ Biped</td>
</tr>
<tr>
<td>sameClassAs</td>
<td>$C_1 \doteq C_2$</td>
<td>Man $\doteq$ Human $\land$ Male</td>
</tr>
<tr>
<td>subPropertyOf</td>
<td>$P_1 \sqsubseteq P_2$</td>
<td>hasDaughter $\sqsubseteq$ hasChild</td>
</tr>
<tr>
<td>samePropertyAs</td>
<td>$P_1 \doteq P_2$</td>
<td>cost $\doteq$ price</td>
</tr>
<tr>
<td>sameIndividualAs</td>
<td>$x_1 \doteq x_2$</td>
<td>President_Bush $\doteq$ G_W_Bush</td>
</tr>
<tr>
<td>disjointWith</td>
<td>$C_1 \sqsubseteq \lnot C_2$</td>
<td>Male $\sqsubseteq$ $\lnot$ Female</td>
</tr>
<tr>
<td>differentIndividualFrom</td>
<td>${x_1} \sqsubseteq \lnot {x_2}$</td>
<td>${john} \sqsubseteq \lnot {peter}$</td>
</tr>
<tr>
<td>inverseOf</td>
<td>$P_1 \doteq P_2$</td>
<td>hasChild $\doteq$ hasParent$\lnot$</td>
</tr>
<tr>
<td>transitiveProperty</td>
<td>$P^+ \sqsubseteq P$</td>
<td>ancestor$^+$ $\sqsubseteq$ ancestor</td>
</tr>
<tr>
<td>uniqueProperty</td>
<td>Thing $\sqsubseteq \leq 1 P$</td>
<td>Thing $\sqsubseteq \leq 1$hasMother</td>
</tr>
<tr>
<td>UnambiguousProperty</td>
<td>Thing $\sqsubseteq \leq 1 P$</td>
<td>Thing $\sqsubseteq \leq 1$isMotherOf$\lnot$</td>
</tr>
</tbody>
</table>

* Axioms (mostly) reducible to subClass/PropertyOf
DAML+OIL

- **Standard model theory**  
  (Patel-Schneider, Horrocks, van Harmelen)

- **FOL axiomatisation**  (Fikes, McGuinness)

- **mapping to SHIQ**  (Horrocks)
  - gives decidability result
  - gives theorem-prover (FaCT)

- **axiomatisation is machine-verified**  (Waldinger)
Semantic Web

- The vision, politics, players
- What is required
- Some technology
  - XML, RDF, DAML+OIL
- The underlying logic

Research directions
- short, medium, long
- highly personal
Short: Language extensions

- “Rules” (e.g. role-chaining)
- “Queries” (constructive bindings of $\exists$ vars)
- “Defaults” (non-monotonicity)
Medium: approximate inference

- Deduction = **exact**
- true/false, not: “almost”,
  “yes except a few”
  “not by a long shot”, etc.
- Ontologies will be sloppy (“scraping”)

Use for Semantic Web reasoning:
- **Approximate classification** (search)
- **Approximate ontology mapping** (agents)
- **Approximate pre/post-conditions** (webservices)
Medium: anytime inference

- Current inference = exact, **abrupt**
- nothing ................ yes!
we want **gradual** answers:
- **anytime computation**
  - agent can decide **how good is good enough**
    (human or machine)
- **deadline computation**
  - pay for quality
  - load balancing
Long: How does the SW change KR?

- it's large
- It’s even larger
- no referential integrity
- many authors, distributed authority, trust
- high variety in quality of knowledge
- diverse vocabularies
- decentralised
- high change rate, time-dependent content
- local containment of inconsistencies
- justifications as first order citizens
“The Semantic Web will globalise KR, just as the WWW globalised hypertext”

(TBL)