The Semantic Web

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

Frank van Harmelen

AI Department

Vrije Universiteit Amsterdam

vrije Universiteit

Goal

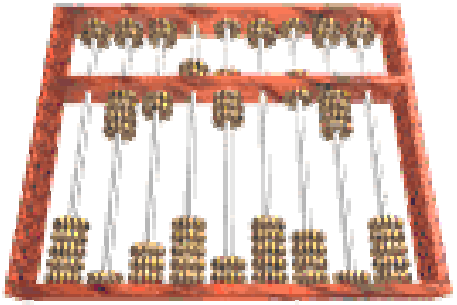
- General propaganda for the Semantic Web
- Convince the old folk there is real substance there
- Convince the young folk there is exciting work to be done

- 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?

Historical perspective



- 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:
  - ❶ handwritten HTML
  - ❷ database generated pages

  } human readers

  The real power will come with the 3rd generation:
  - ❸ machine accessible semantics

What it's like to be a machine on the Web

林克昌 根留台灣 可能增高

在愛戴者熱心奔走之下，華裔名指揮家林克昌根留台灣的可行性又提升了幾分。兩廳院主任李炎、國家音樂廳樂團副團長黃奕明日前親赴林克昌、石聖芳寓所拜會，並提出多場客席邀約。此外，台灣省立交響樂團團長陳澄雄也早早「下訂」，邀請林克昌赴台中霧峰，從八月十日起訓練省交，為期長達一個月。

在台灣諸多公家樂團中，陳澄雄是以實際行動表達對林克昌肯定的樂界人士之一，曾多次公開表示對林克昌指揮才華的欽佩，而且幾乎每個樂季都邀請林克昌客席演出。

此外，林克昌上個月赴俄羅斯與頂尖的「俄羅斯國家管絃樂團」灌錄了柴可夫斯基晚期三大交響曲以及「羅密歐與茱麗葉」、「斯拉夫進行曲」、「義大利隨想曲」，最後的DAT母帶也在前兩天寄回台灣。製作人楊忠衡與林克昌試聽之後，都對錄音效果－尤其音質表現感到相當滿意，楊忠衡估計呈現了七分林克昌指揮神韻。

俄羅斯國家管絃樂團首席布魯尼日前也讚譽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；三是宛如呼吸歌唱的旋律處理。這些對錄音師而言都構成很大挑戰。俄國錄音師雖然採用多軌混音，但定位、場面都有可觀之處。。

“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

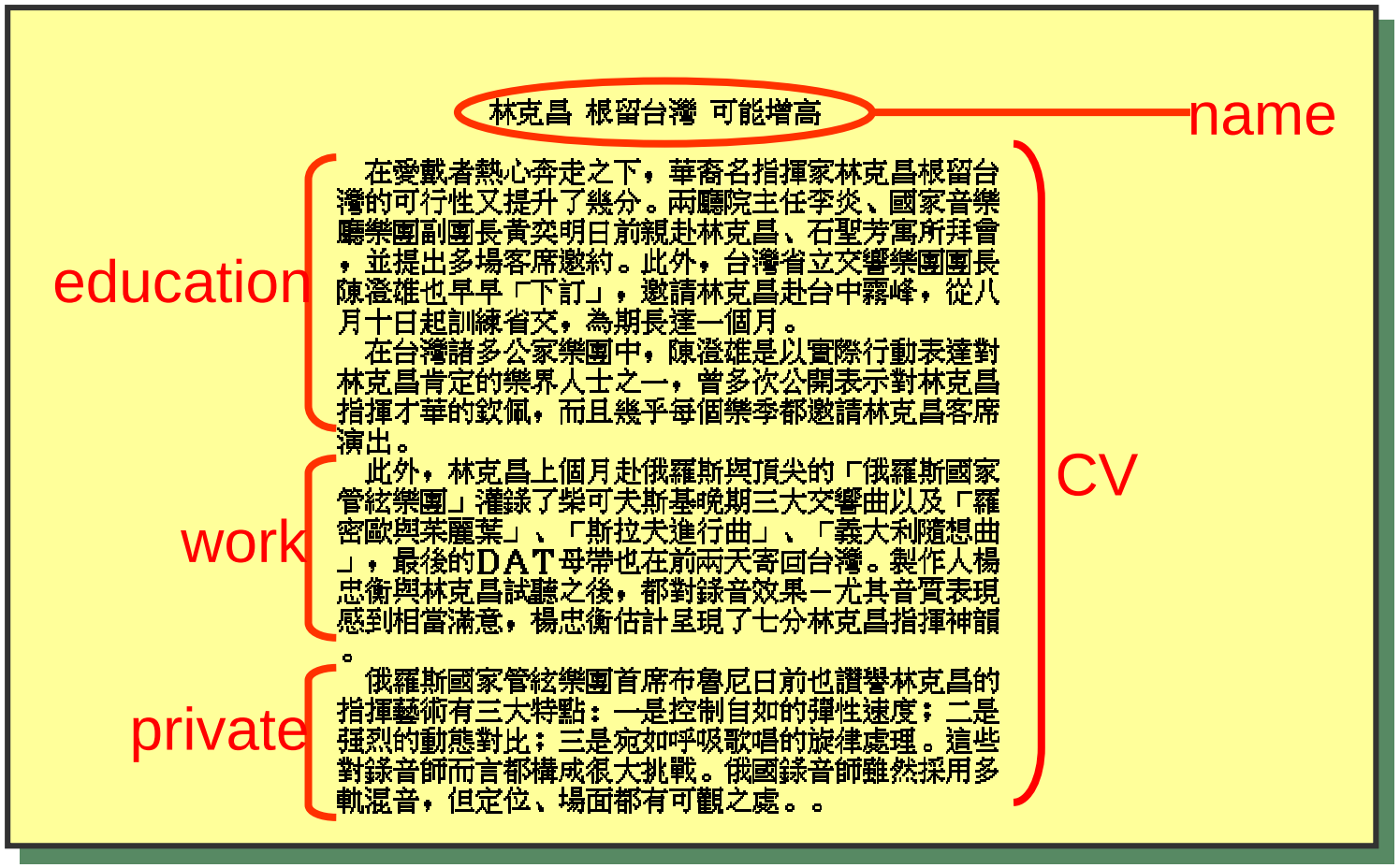
- Focus of $80 million DARPA project DAML
- Focus of €20 million EU action line
  (and much more to come)
- Pushed hard by Tim Berners-Lee / W3C
  (“Weaving the Web”)
- Is central to EU 6th Framework

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)



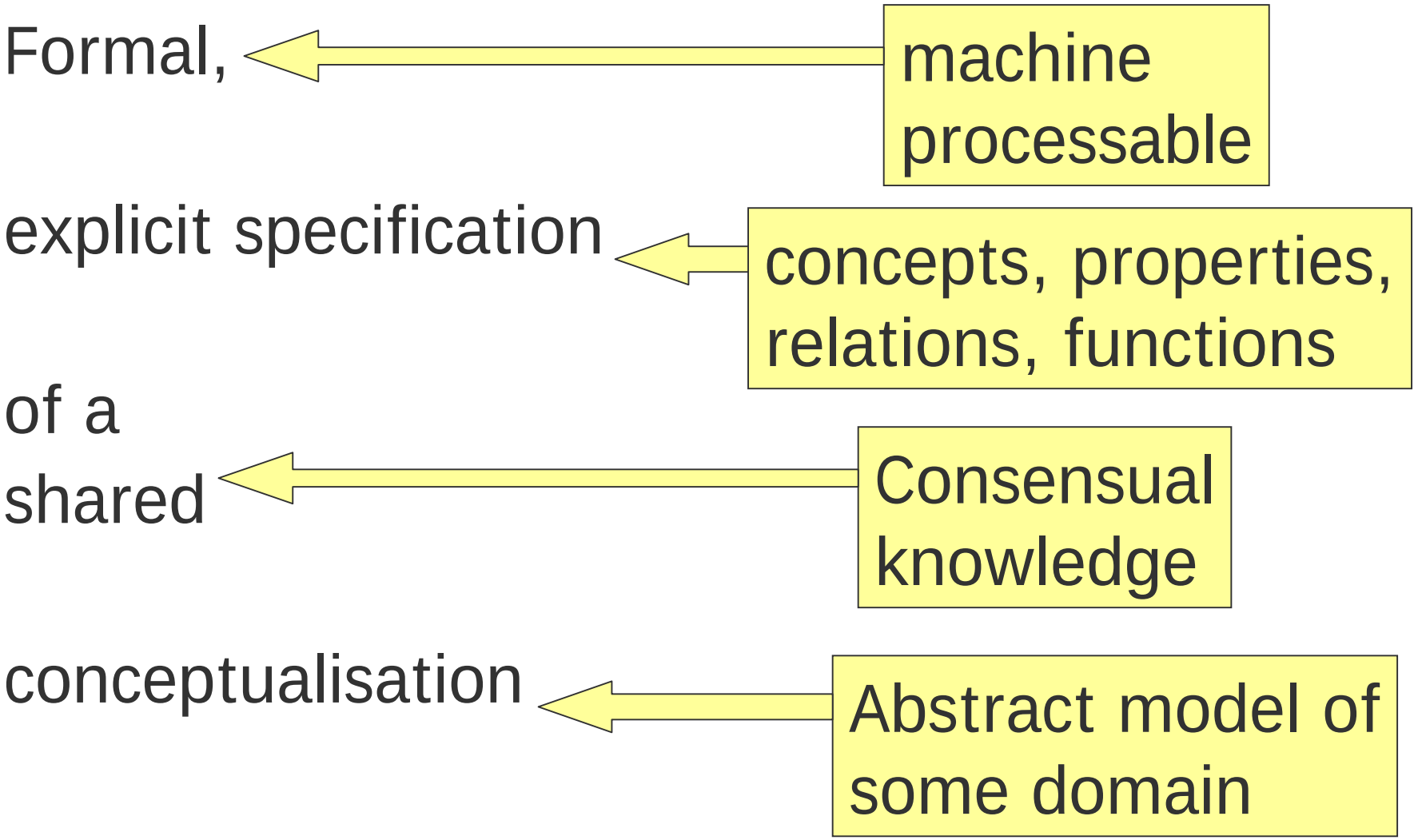
Required are:

- **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



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

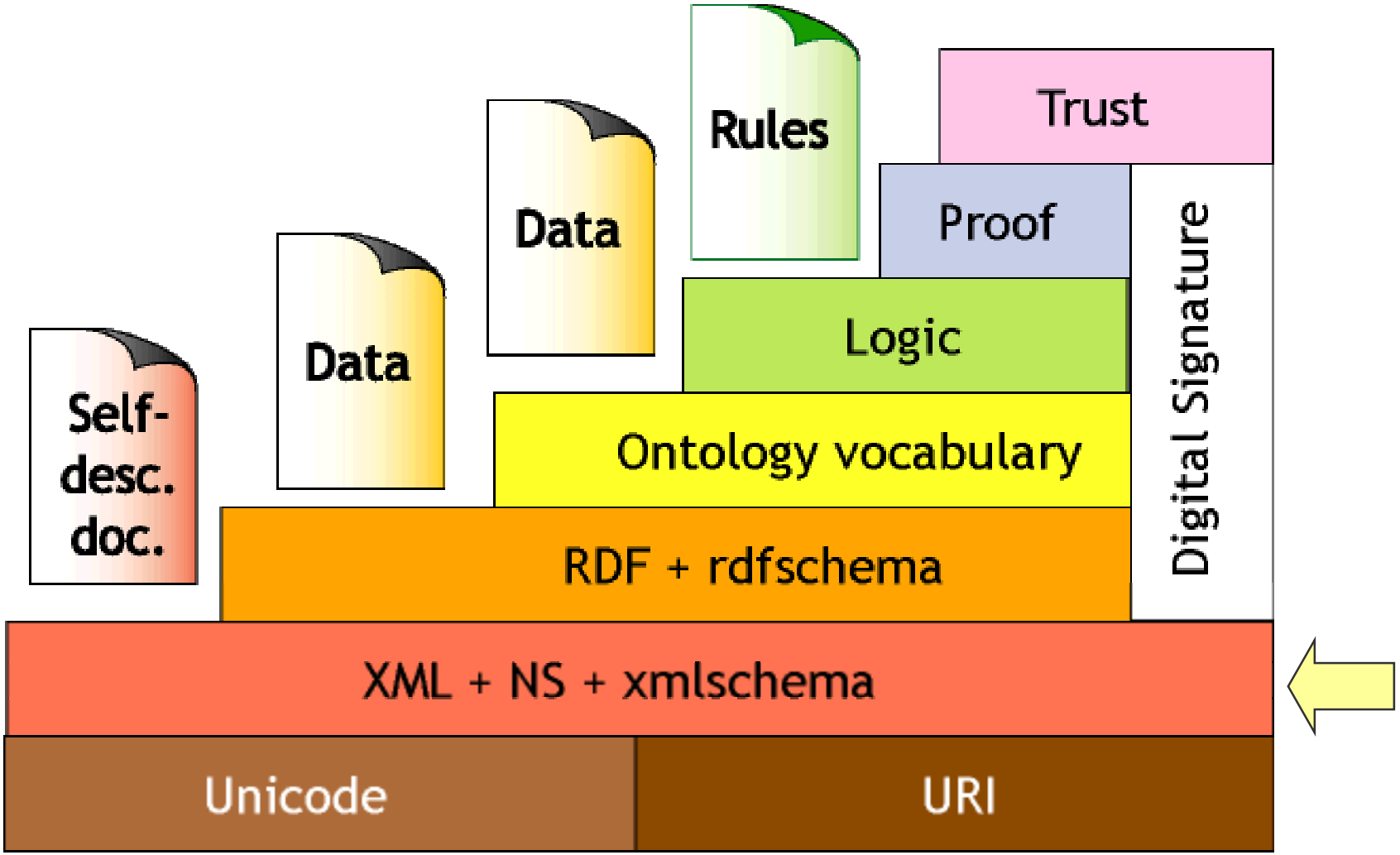
- Lightweight:
  - Yahoo topic hierarchy
  - Open directory (400.000 general categories)
- Heavy weight
  - Cyc, 300.000 axioms

- 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)

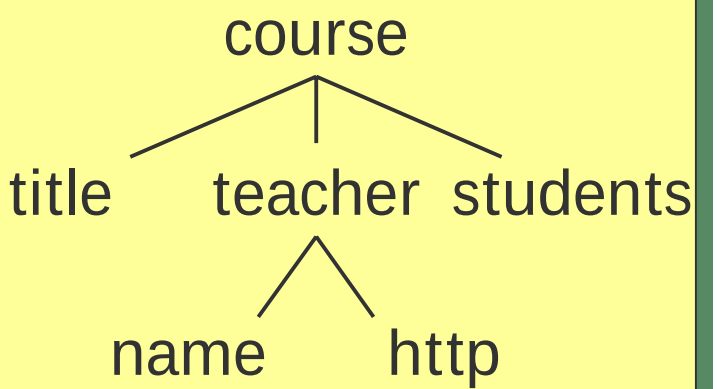


XML: Document = labelled tree

- node = label + attr/values + contents

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

=

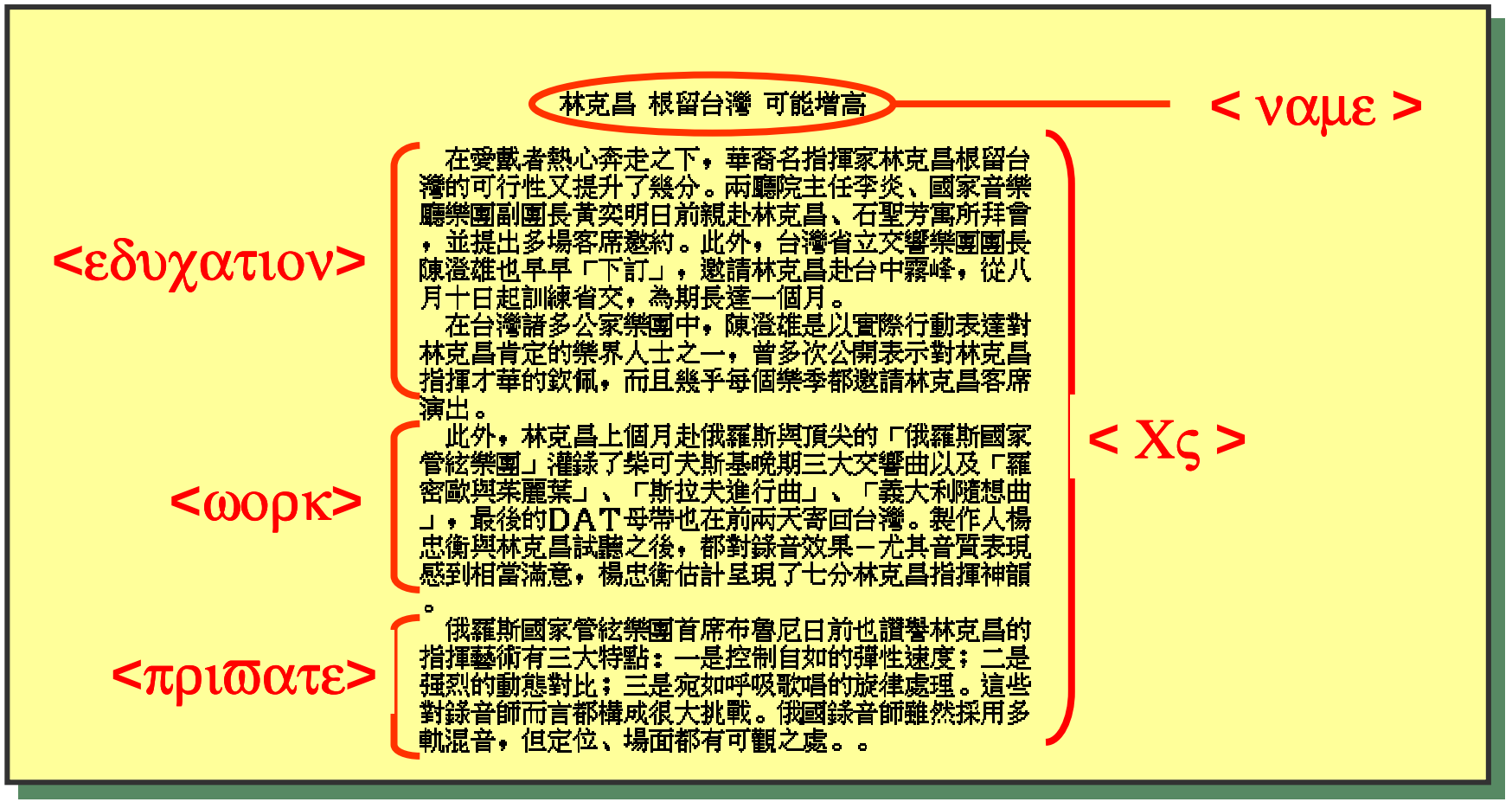


- DTD: simple grammar to describe legal trees
- XML Schema: not so simple grammar for the same

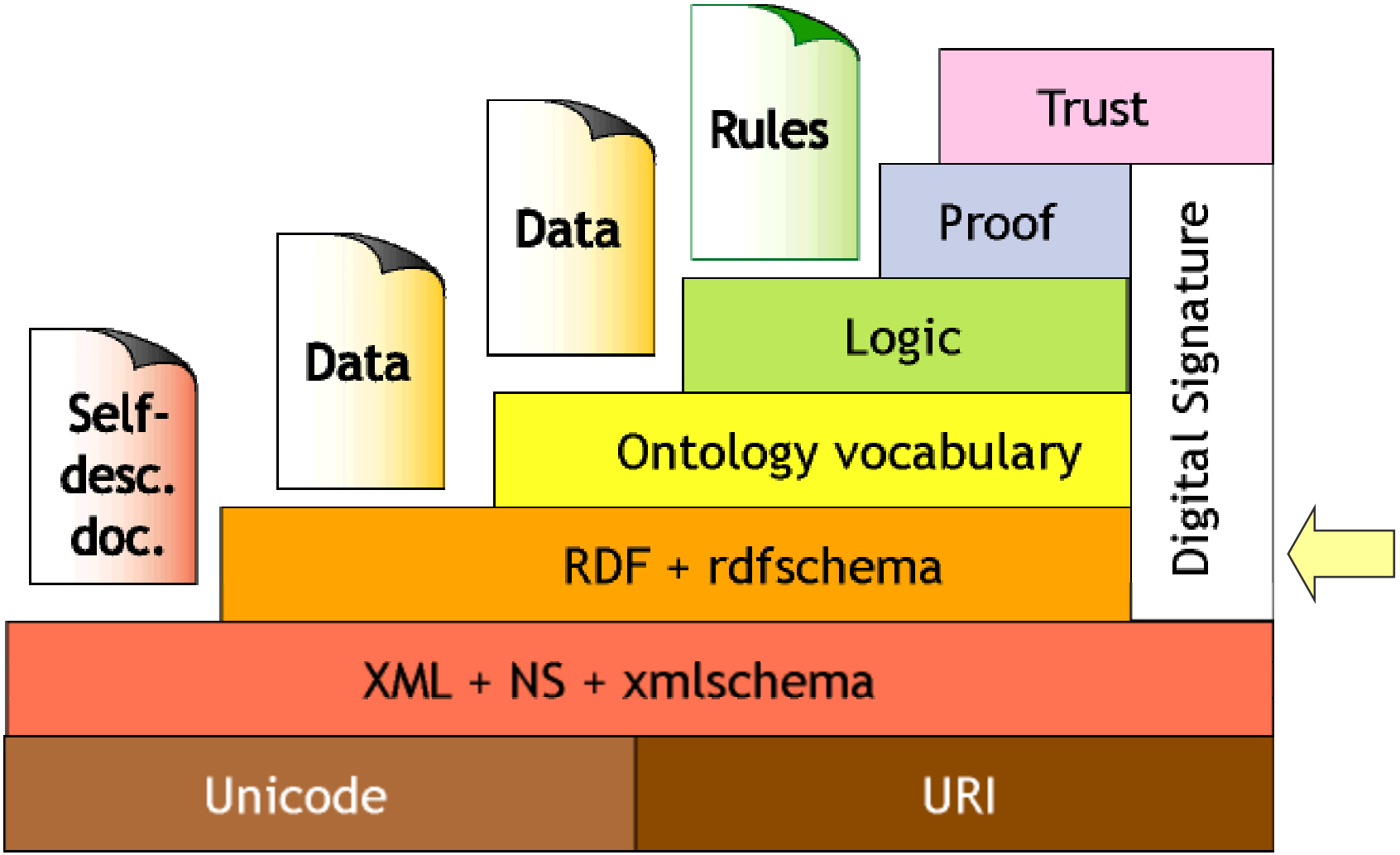
- So:

why not use XML to represent ontologies?

XML ≠
machine accessible meaning



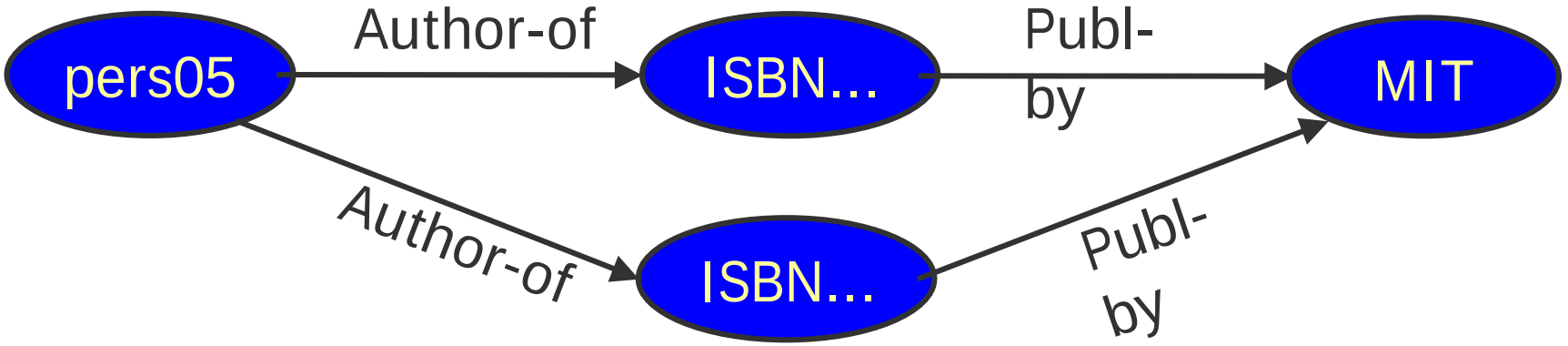
The semantic pyramid again



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

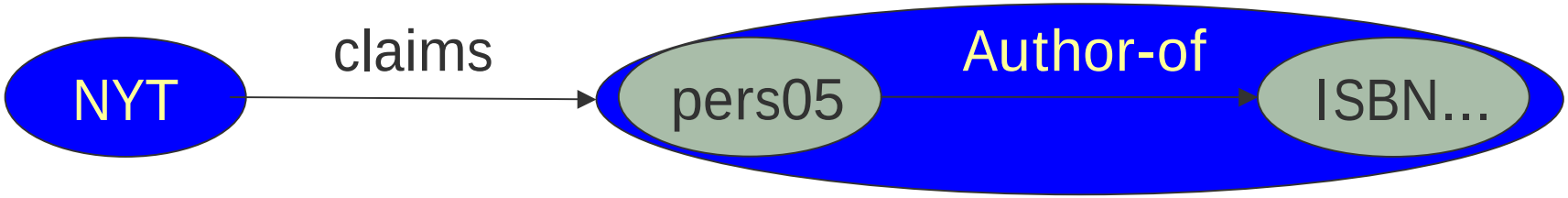


Bluffer's guide to RDF (2)

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

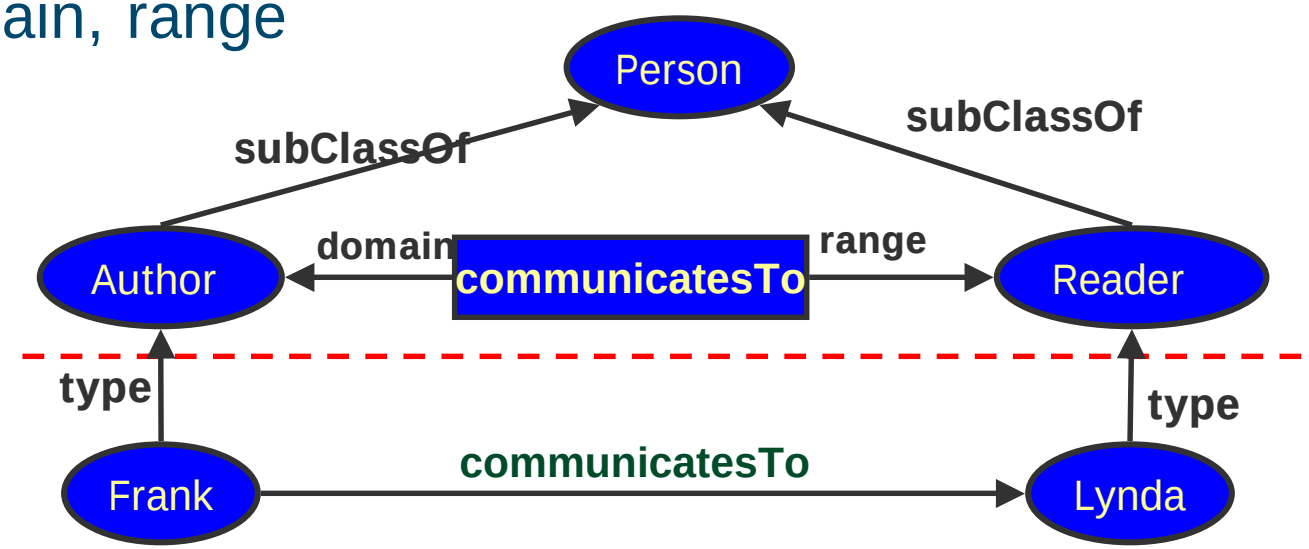
```
<rdf:Description rdf:about="#pers05">
  <authorOf>ISBN...</authorOf>
</rdf:Description>
```

- Any statement can be an object
  - graphs can be **nested**

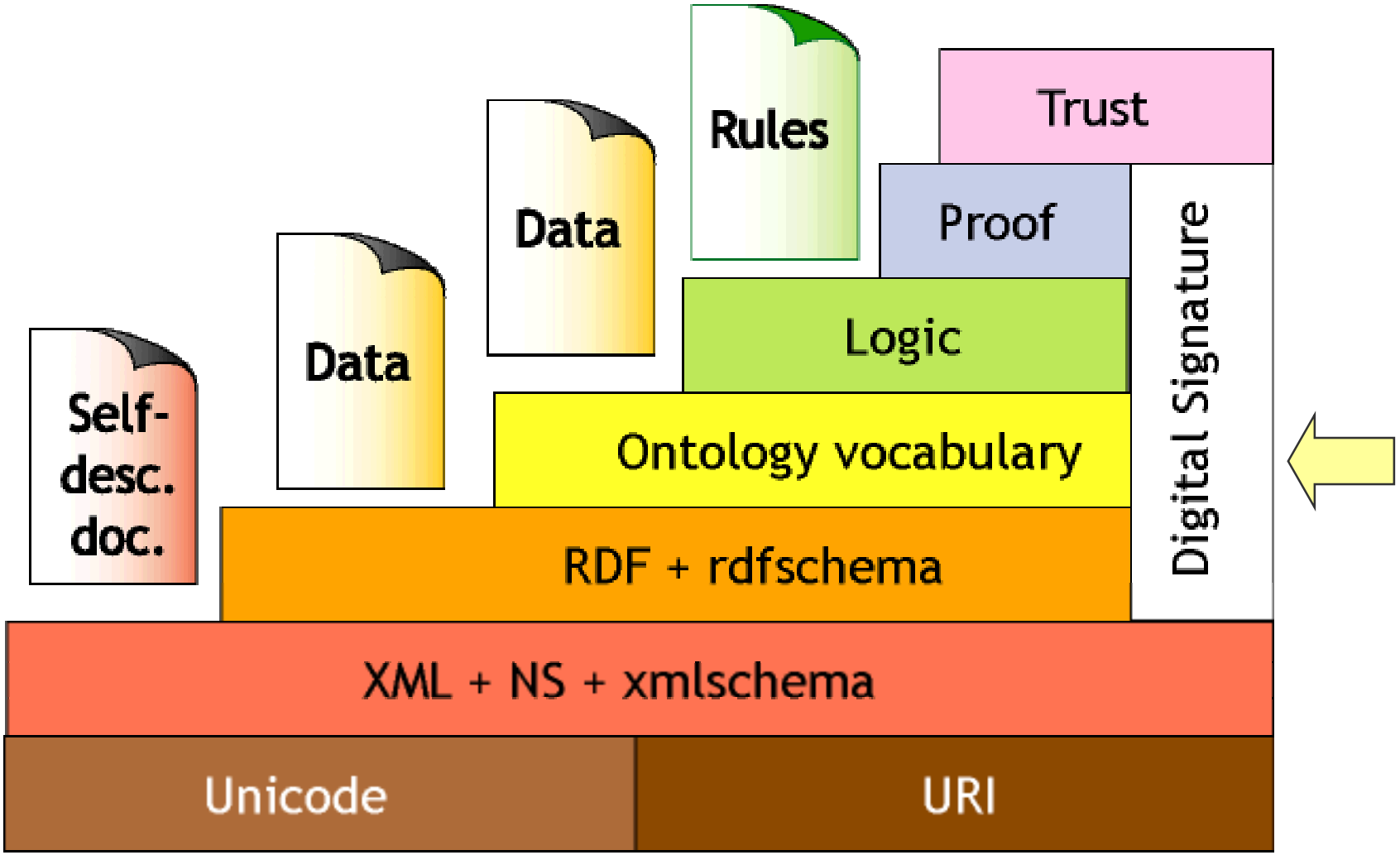


What does RDF Schema add?

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



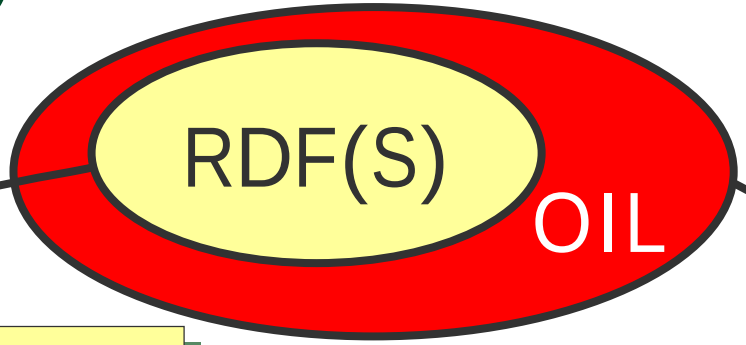
The semantic pyramid again



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

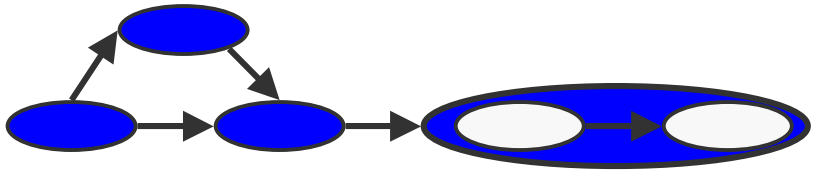
DAML+OIL: politics

- 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?

RDF(S)



- 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

Constructor	Abbreviation	Example
intersectionOf	$C_1 \wedge \ldots \wedge C_n$	Human $\wedge$ Male
unionOf	$C_1 \vee \ldots \vee C_n$	Doctor $\vee$ Lawyer
complementOf	$\neg C$	$\neg$Male
oneOf	$\{x_1 \ldots x_n\}$	{john, mary}
toClass	$\forall P.C$	$\forall$hasChild.Doctor
hasClass	$\exists P.C$	$\exists$hasChild.Lawyer
hasValue	$\exists P.\{x\}$	$\exists$citizenOf.{USA}
minCardinalityQ	$\geqslant n\, P.C$	$\geqslant 2$ hasChild.Lawyer
maxCardinalityQ	$\leqslant n\, P.C$	$\leqslant 1$ hasChild.Male
cardinalityQ	$=n\, P.C$	$=1$ hasParent.Female

+ XML Schema datatypes:

- **`int, string, real`**, etc

DAML+OIL: Axioms

Axiom	Abbreviation	Example
subClassOf	$C_1 \sqsubseteq C_2$	Human $\sqsubseteq$ Animal $\wedge$ Biped
sameClassAs	$C_1 \doteq C_2$	Man $\doteq$ Human $\wedge$ Male
subPropertyOf	$P_1 \sqsubseteq P_2$	hasDaughter $\sqsubseteq$ hasChild
samePropertyAs	$P_1 \doteq P_2$	cost $\doteq$ price
sameIndividualAs	$x_1 \doteq x_2$	President_Bush $\doteq$ G_W_Bush
disjointWith	$C_1 \sqsubseteq \neg C_2$	Male $\sqsubseteq \neg$Female
differentIndividualFrom	$\{x_1\} \sqsubseteq \neg\{x_2\}$	{john} $\sqsubseteq \neg${peter}
inverseOf	$P_1 \doteq P_2^-$	hasChild $\doteq$ hasParent$^-$
transitiveProperty	$P^+ \sqsubseteq P$	ancestor$^+ \sqsubseteq$ ancestor
uniqueProperty	Thing $\sqsubseteq \leqslant 1P$	Thing $\sqsubseteq \leqslant 1$hasMother
UnambiguousProperty	Thing $\sqsubseteq \leqslant 1P^-$	Thing $\sqsubseteq \leqslant 1$isMotherOf$^-$

☞ 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

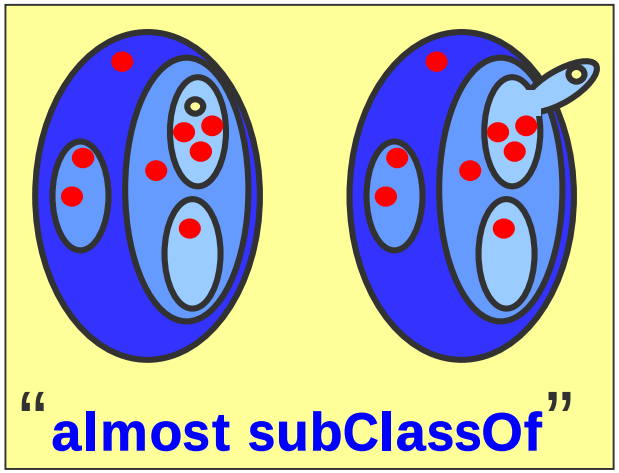
- The vision, politics, players
- What is required
- Some technology
  - XML, RDF, DAML+OIL
- The underlying logic
- **Research directions**
  - **short, medium, long**
  - **highly personal**

Short: Language extensions

- “Rules” (e.g. role-chaining)
- “Queries” (constructive bindings of ∃ vars)
- “Defaults” (non-monotonicity)

Medium: approximate inference

- Deduction = **exact**
- true/false, not: “allmost”,
  “yes except a few”
  “not by a long shot”, etc.

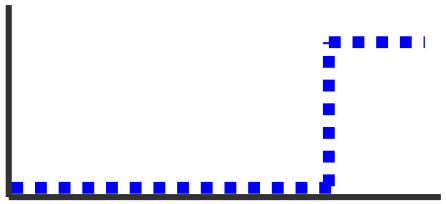


- Ontologies will be sloppy (“scraping”)

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

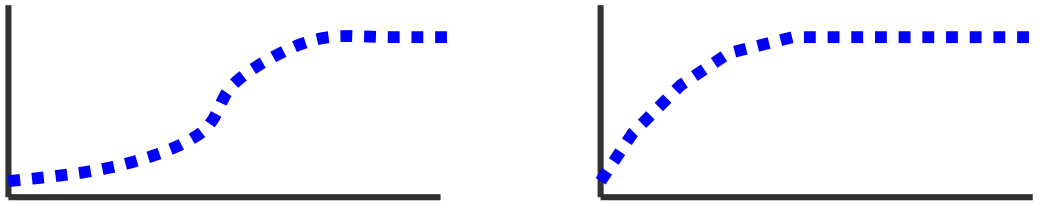
Medium: anytime inference

- 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)