### **OWL: a Reasonable Ontology Language?**

#### Ian Horrocks

<ian.horrocks@comlab.ox.ac.uk> Information Systems Group Oxford University Computing Laboratory





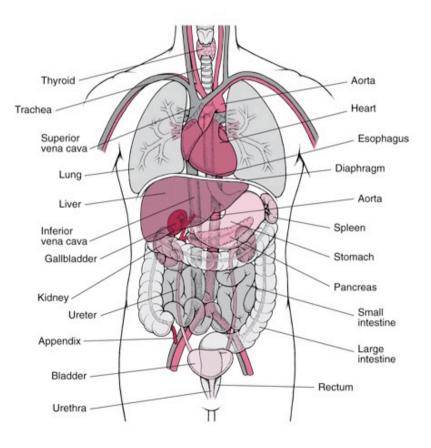


### An explicit specification of a conceptualization



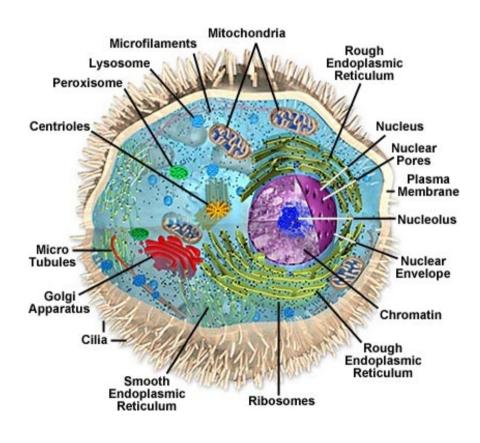


- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy



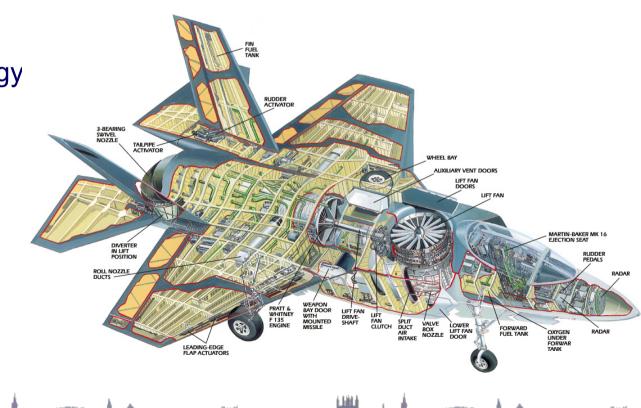


- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology



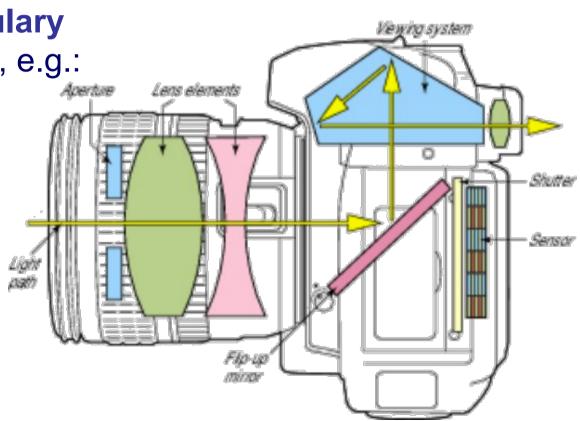


- Introduces vocabulary relevant to domain, e.g.:
  - Anatomy
  - Cellular biology
  - Aerospace



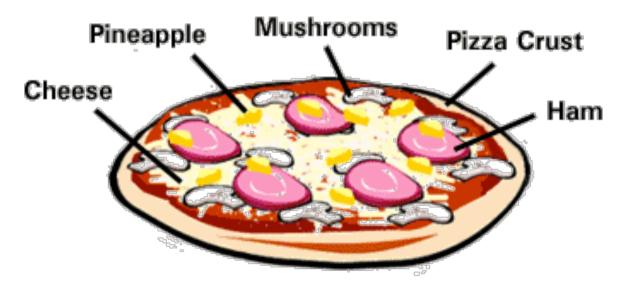


- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology
  - Aerospace
  - Photography





- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology
  - Aerospace
  - Photography
  - Pizzas

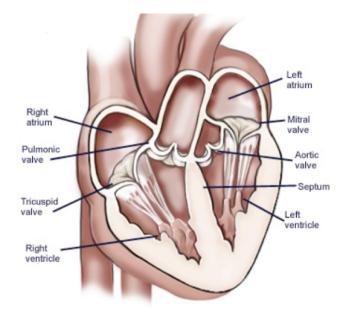




A model of (some aspect of) the world

- Introduces vocabulary
   relevant to domain
- Specifies *relative* **meaning** (aka semantics) of terms

Heart is a muscular organ that is part of the circulatory system



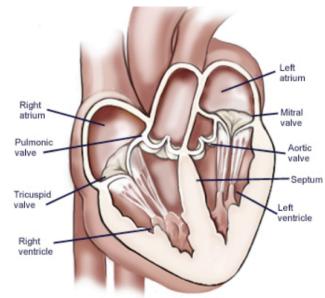
A model of (some aspect of) the world

- Introduces vocabulary
   relevant to domain
- Specifies *relative* **meaning** (aka semantics) of terms

Heart is a muscular organ that is part of the circulatory system

• Formalised e.g. using suitable logic

 $\begin{array}{l} \mathsf{Heart}\sqsubseteq\mathsf{MuscularOrgan}\sqcap\\ \exists \mathsf{isPartOf}.\mathsf{CirculatorySystem} \end{array}$ 





- Coherent **shared view** of domain
  - Help identify and resolve disagreements
- Ontology-based Information Systems
  - User-centric view of data that is independent of logical/physical schema
  - Answers reflect knowledge & data, e.g.:



Now... *that* should clear up a few things around here



- $Q(x) \gets \mathsf{Patient}(x) \land \mathsf{suffersFrom}(x,y) \land \mathsf{VascularDisease}(y)$
- i.e., "Patients suffering from Vascular Disease"



 $Q(x) \leftarrow \mathsf{Patient}(x) \land \mathsf{suffersFrom}(x,y) \land \mathsf{VascularDisease}(y)$ 

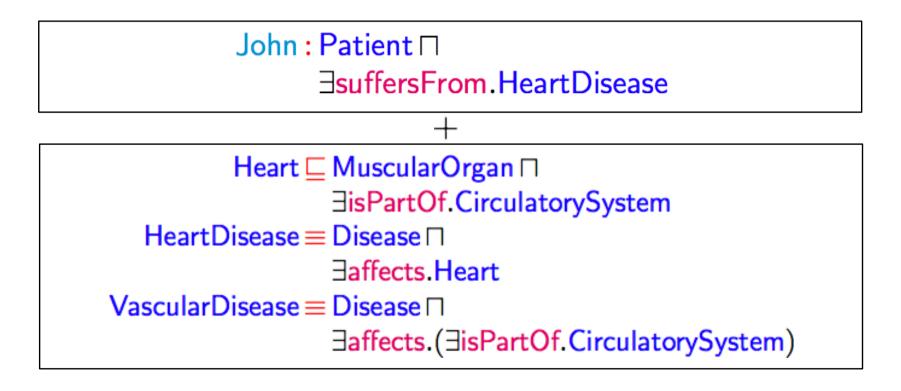
i.e., "Patients suffering from Vascular Disease"

John : Patient □ ∃suffersFrom.HeartDisease



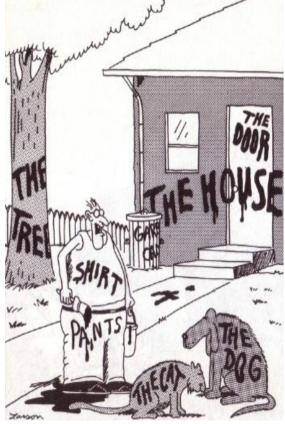
 $Q(x) \gets \mathsf{Patient}(x) \land \mathsf{suffersFrom}(x,y) \land \mathsf{VascularDisease}(y)$ 

i.e., "Patients suffering from Vascular Disease"



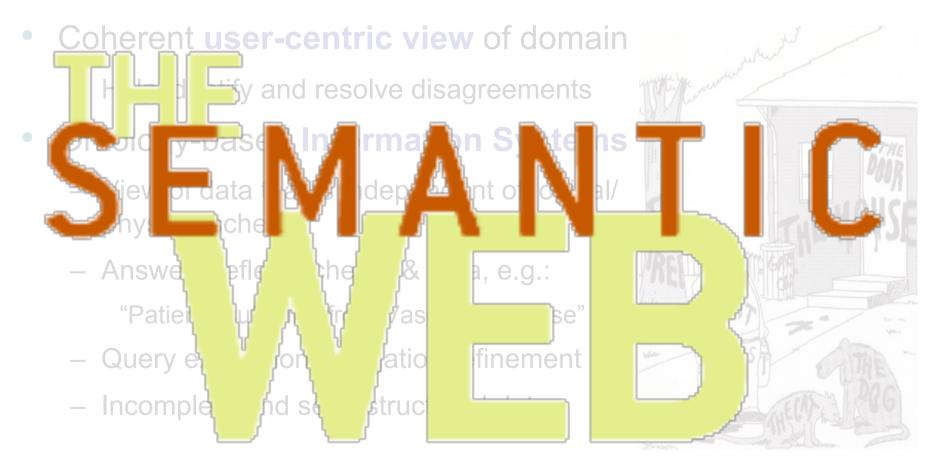
- Coherent **shared view** of domain
  - Help identify and resolve disagreements
- Ontology-based Information Systems
  - User-centric view of data that is independent of logical/physical schema
  - Answers reflect knowledge & data, e.g.:
    "Patients suffering from Vascular Disease"
  - Query expansion/navigation/refinement
  - Incomplete and semi-structured data

### More "intelligent" applications



Now... *that* should clear up a few things around here





### More "intelligent" applications

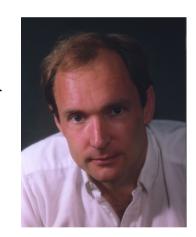
Now... *that* should clear up a few things around here



• According to **TBL** circa 1998:

"... a consistent logical web of data ..." in which

"... information is given well-defined meaning ..."





- According to **TBL** circa 1998:
  - "... a consistent logical web of data ..." in which
  - "... information is given well-defined meaning ..."
- By now has evolved into:

"a platform for distributed applications and sharing (linking) data"



- According to **TBL** circa 1998:
  - "... a **consistent logical web of data** ..." in which "... information is given **well-defined meaning** ..."
- By now has evolved into:
  - "a platform for distributed applications and sharing (linking) data"
  - RDF provides uniform syntactic structure for data
  - OWL provides machine readable schemas (ontologies)

• According to **TBL** circa 1998:

"... a **consistent logical web of data** ..." in which "... information is given **well-defined meaning** ..."

- By now has evolved into:
  - "a platform for distributed applications and sharing (linking) data"
  - RDF provides uniform syntactic structure for data
  - OWL provides machine readable schemas (ontologies)

i.e., a large distributed ontology based information system

- RDF standard first published 1999; revised 2004
- RDF extended to **RDFS**, a primitive ontology language
  - classes and properties; sub/super-classes (and properties); range and domain (of properties)
- But RDFS lacks important features, e.g.:
  - existence/cardinality constraints; transitive/inverse properties;
     localised range and domain constraints, ...
- And RDF(S) has "higher order flavour" with no (later non-standard) formal semantics
  - difficult to understand or to provide reasoning support



• EU On-To-Knowledge project developed OIL



- EU On-To-Knowledge project developed OIL
- DAML program developed DAML-ONT
- Efforts soon merged to produce DAML+OIL
  - Further development carried out by "Joint EU/US Committee"





- EU On-To-Knowledge project developed OIL
- DAML program developed DAML-ONT
- Efforts soon merged to produce **DAML+OIL** 
  - Further development carried out by "Joint EU/US Committee"
- DAML+OIL submitted to WSC as basis for standardisation
  - WebOnt WG developed OWL (2004)
  - OWL WG developed OWL 2 (2009)
- OWL (2) based on SHOIN (SROIQ)
   Description Logics!?

# What are Description Logics (DLs)?

- Fragments of first order logic designed for KR
- Useful computational properties
  - Decidable (essential)
  - Low complexity (desirable)
- Succinct and variable free syntax

 $\begin{array}{l} \mathsf{Heart}\sqsubseteq\mathsf{MuscularOrgan}\sqcap\\ \exists \mathsf{isPartOf}.\mathsf{CirculatorySystem} \end{array}$ 

 $\begin{aligned} \forall x. [\mathsf{Heart}(x) & \to \mathsf{MuscularOrgan}(x) \land \\ & \exists y. [\mathsf{isPartOf}(x, y) \land \\ & \mathsf{CirculatorySystem}(y)]] \end{aligned}$ 

### Can exploit the results of 20+ years of DL research

- Well defined (model theoretic) semantics

Constructor	DL Syntax	Example	FOL Syntax
intersectionOf	$C_1 \sqcap \ldots \sqcap C_n$	Human ⊓ Male	$C_1(x) \wedge \ldots \wedge C_n(x)$
unionOf	$C_1 \sqcup \ldots \sqcup C_n$	Doctor ⊔ Lawyer	$C_1(x) \lor \ldots \lor C_n(x)$
complementOf	$\neg C$	¬Male	$\neg C(x)$
oneOf	$\{x_1\} \sqcup \ldots \sqcup \{x_n\}$	{john} ⊔ {mary}	$x = x_1 \lor \ldots \lor x = x_n$
allValuesFrom	$\forall P.C$	∀hasChild.Doctor	$\forall y. P(x, y) \rightarrow C(y)$
someValuesFrom	$\exists P.C$	∃hasChild.Lawyer	$\exists y. P(x, y) \land C(y)$
maxCardinality	$\leqslant nP$	≤1hasChild	$\exists \leq n y. P(x, y)$
minCardinality	$\geqslant nP$	≥2hasChild	$\exists^{\geqslant n}y.P(x,y)$





Can exploit the results of 20+ years of DL research

- Well defined (model theoretic) semantics
- Formal properties well understood (complexity, decidability)



I can't find an efficient algorithm, but neither can all these famous people.

[Garey & Johnson. Computers and Intractability]

Can exploit the results of 20+ years of DL research

- Well defined (model theoretic) semantics
- Formal properties well understood (complexity, decidability)
- Practical reasoning algorithms

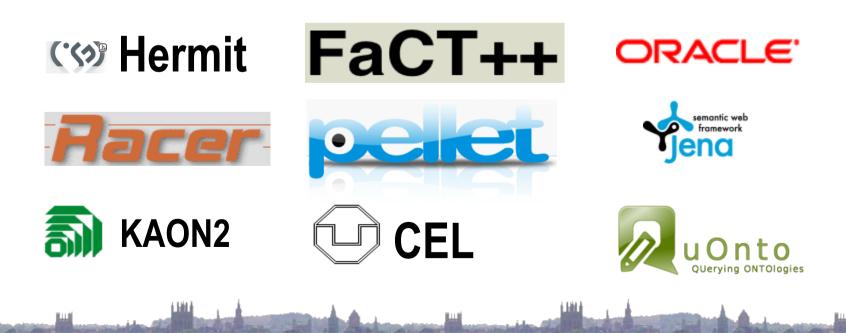
□-rule	if 1. $(C_1 \sqcap C_2) \in \mathcal{L}(v)$ , v is not indirectly blocked, and
	2. $\{C_1, C_2\} \not\subseteq \mathcal{L}(v)$
	then $\mathcal{L}(v) \to \mathcal{L}(v) \cup \{C_1, C_2\}.$
⊔-rule	if 1. $(C_1 \sqcup C_2) \in \mathcal{L}(v)$ , v is not indirectly blocked, and
	2. $\{C_1, C_2\} \cap \mathcal{L}(v) = \emptyset$
	then $\mathcal{L}(v) \to \mathcal{L}(v) \cup \{E\}$ for some $E \in \{C_1, C_2\}$
∃-rule	if 1. $\exists r. C \in \mathcal{L}(v_1), v_1$ is not blocked, and
	2. $v_1$ has no safe r-neighbour $v_2$ with $C \in \mathcal{L}(v_1)$ ,
	then create a new node $v_2$ and an edge $\langle v_1, v_2 \rangle$
	with $\mathcal{L}(v_2) = \{C\}$ and $\mathcal{L}(\langle v_1, v_2 \rangle) = \{r\}.$
∀-rule	if 1. $\forall r.C \in \mathcal{L}(v_1), v_1$ is not indirectly blocked, and
	2. there is an r-neighbour $v_2$ of $v_1$ with $C \notin \mathcal{L}(v_2)$
	then $\mathcal{L}(v_2) \to \mathcal{L}(v_2) \cup \{C\}.$
∀ <sub>+</sub> -rule	if 1. $\forall r.C \in \mathcal{L}(v_1), v_1$ is not indirectly blocked, and
	2. there is some role $r'$ with $Trans(r')$ and $r' \equiv r$
	3. there is an r'-neighbour $v_2$ of $v_1$ with $\forall r'.C \notin \mathcal{L}(v_2)$
	then $\mathcal{L}(v_2) \to \mathcal{L}(v_2) \cup \{ \forall r'.C \}.$
choose-rule	if $1. \leq n r.C \in \mathcal{L}(v_1), v_1$ is not indirectly blocked, and
	2. there is an r-neighbour $v_2$ of $v_1$ with $\{C, \neg C\} \cap \mathcal{L}(v_2) = \emptyset$
	then $\mathcal{L}(v_2) \to \mathcal{L}(v_2) \cup \{E\}$ for some $E \in \{C, \neg C\}$ .
≽-rule	if $1. \ge n r.C \in \mathcal{L}(v)$ , v is not blocked, and
	2. there are not n safe r-neighbours $v_1, \ldots, v_n$ of v
	with $C \in \mathcal{L}(v_i)$ and $v_i \neq v_j$ for $1 \leq i < j \leq n$
A SAME A	a a construction of the second s





Can exploit the results of 20+ years of DL research

- Well defined (model theoretic) semantics
- Formal properties well understood (complexity, decidability)
- Practical reasoning algorithms
- Effective implemented systems







### **Ontologies** before:

Name	Original	de-	primi-	arti-	Σ	de-	primi-
	Language	fined	tive	ficial		fined	tive
		concepts			roles		
CKB	SB-ONE	23	57	58	138	2	46
Companies	BACK	70	45	81	196	1	39
FSS	SB-ONE	34	98	75	207	0	47
Espresso	SB-ONE	0	145	79	224	11	41
Wisber	TURQ	50	81	152	283	6	18
Wines	CLASSIC	50	148	237	435	0	10



### **Ontologies** after:

Swoogle Semantic Web Search Engine				
🔺 🕨 🕂 🐺 http://swoogle.uml	oc.edu/	C Qr Google		
Semantic v	veb search	Want more results? Login		
ontology <u>document</u>	term more >> Swoogle Search	3		
Searchir	ng over 10,000 ontologies			
		11		

#### **Ontologies** after:

#### Welcome to the Protege Ontology Library!

#### **OWL** ontologies

- AIM@SHAPE Ontologies A: Ontologies pertaining to digital shapes. Source: AIM@SHAPE NoE Advanced and Innovative Models And Tools for the development of Semantic-based systems for Handling, Acquiring, and Processing knowledge Embedded in multidimensional digital objects.
- amino-acid.owl 

  A small OWL ontology of amino acids and their properties. Source: Amino Acid Ontology Web site 

  Acid Ontology Web site
- Basic Formal Ontology (BFO) 🗗
- bhakti.owl : An OWL ontology for the transcendental states of consciousness experienced by practitioners of bhakti-yoga, a form of Vedic consciousness engineering.
- Biochemical Ontologies &: Over 30 ontologies for knowledge representation and reasoning across scientific domains. Ontologies are normalized into non-disjoint primitive skeletons and

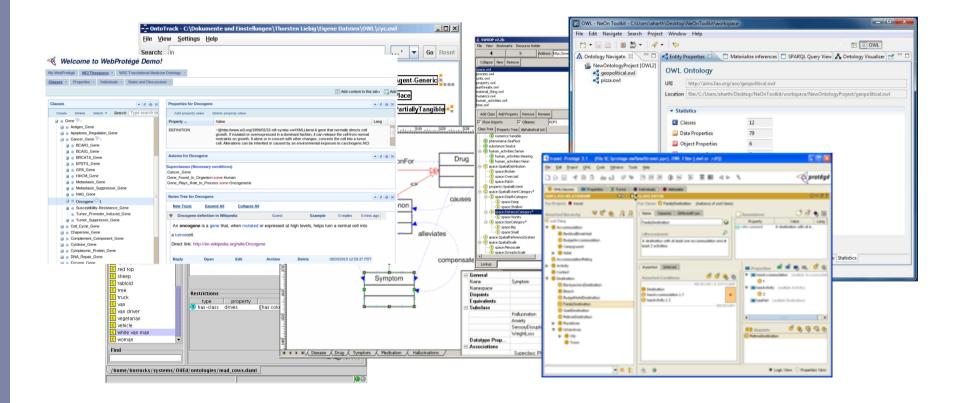


Tools before:

```
> (load-tkb "demo.kb" :verbose T)
                    . . . . . . . . . . . . . . .
> (classify-tkb :mode :stars)
  pppppppppppppppccpcpcccpcppcpcpppcccpcp
  pccccppcpcppcccp
  ጥ
> (direct-supers 'MAN)
  (C[HUMAN] C[MALE])
>
```

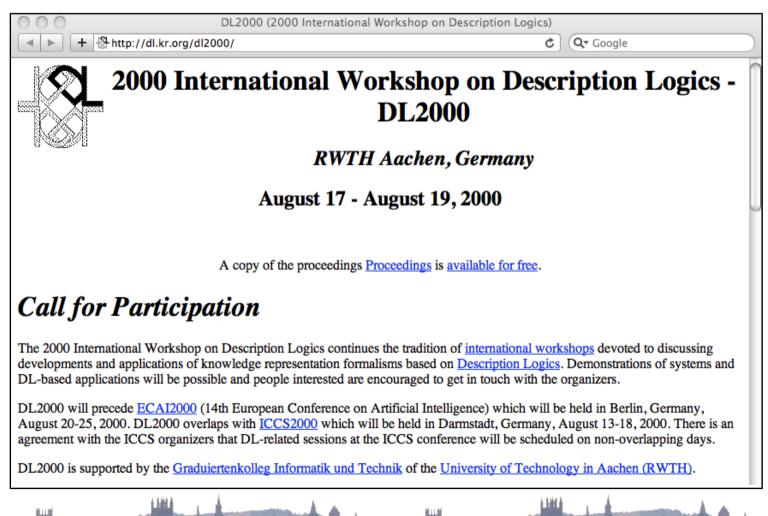


#### Tools after:





#### "Profile" before:





#### "Profile" after:





**Applications** before:

#### **Applications** after:



#### **Applications** after:

- eScience, eCommerce, geography, engineering, defence, ...
- Major impact in healthcare and life sciences
- Mainstream technology supported by, e.g., ORACLE 11g
- Increasing impact in business applications



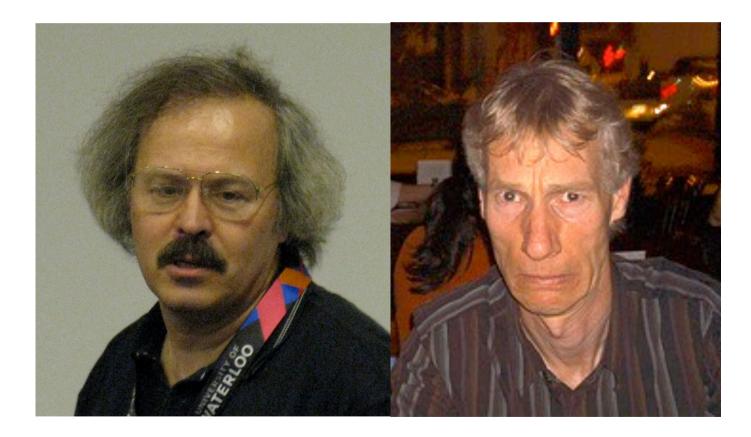


#### Peter and lan before:





#### Peter and Ian after:



# Where We Are Now

- OWL (2) ontology language a W3C standard
- OWL (2) based on **AI research** (in particular DLs)
- Wide range of tools and infrastructure now available
- High profile applications
- Support from mainstream technology vendors

# So everybody's happy?



# So everybody's happy?





## **Of course not!**





It is too complicated, and users will never understand it or be able to use it!

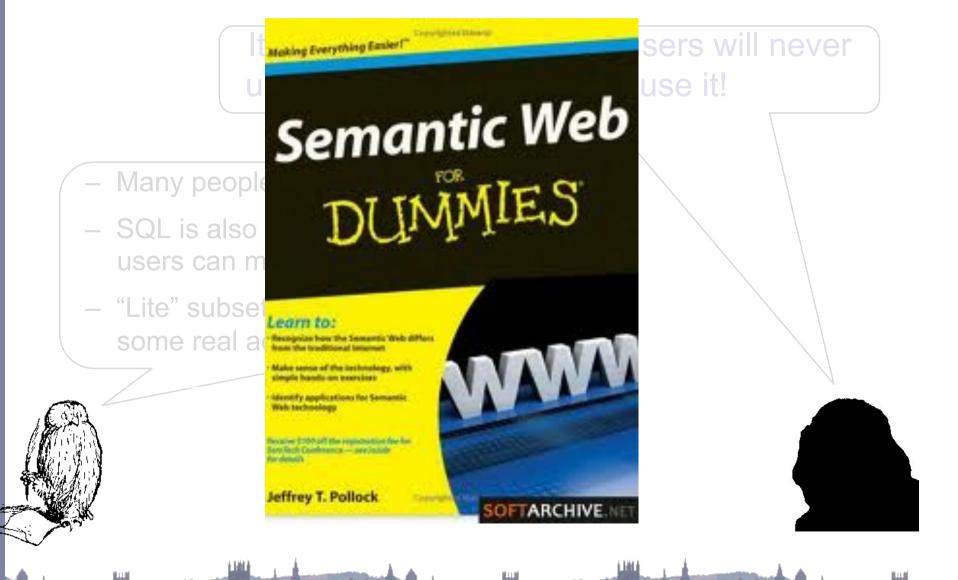


It is too complicated, and users will never understand it or be able to use it!

- Many people are now using it!
- Naive users can manage with a small subset (c.f. SQL, MS-Word, ...)
- "Lite" subsets only useful if they confer some computational advantage











Complexity is too high, and it won't scale!



Complexity is too high, and it won't scale!

– What do we mean by "scale"?

Reasoning with whole web doesn't make sense

– Even so, scalability is a real problem

 $\mathcal{SROIQ}$  satisfiability/subsumption is 2NEXPTIME-complete







Thanks to: Arthur Gordon, Alison Gurlitz, Stephen Lam and Eugene Moy



So is OWL reasoning doomed to failure?

- High complexity doesn't mean that *bad* performance is guaranteed
  - Just that we can't guarantee **good** performance
- Highly optimised implementations (may) work well in practice
- Main problem is relatively low "robustness"
  - Optimisations exploit features of *typical* ontologies
  - Small changes in ontology can lead to large changes in performance – "it worked OK yesterday"
- Large data sets may also be problematical
- Users/applications can choose tractable subsets (profiles) if greater scalability and/or robustness is needed



#### OWL 2 profiles:

#### - **OWL 2 EL**

- polynomial (combined) complexity
- highly effective "one pass" classification algorithms

#### - **OWL 2 RL**

- polynomial (combined) complexity
- convenient rule-extended database implementation

#### – **OWL 2 QL**

- AC<sup>0</sup> (data) complexity (< logspace)
- highly scalable query rewriting implementation





**Rules!** 



- ✓ More natural/intuitive and easy to understand
- Can describe arbitrary relational structures
- ✓ UNA and CWA semantics is more intuitive/appropriate

**Rules!** 

✓ Better scalability



- More natural/intuitive and easy to understand
- Can describe arbitrary relational structures
- ✓ UNA and CWA semantics is more intuitive/appropriate

**Rules!** 

- ✓ Better scalability
- Less natural/intuitive and easy to understand
- X Can't describe unbounded structures
- X UNA and CWA inappropriate in Web setting
- X Poor at dealing with incomplete information



Fuzzy

Logic!



✓ Need to deal with vague concepts, e.g., "tall"

- ✓ Information may also be vague/noisy, e.g., the Web
- ✓ Strictly extends "crisp" languages (1 = true; 0 = false)

**Fuzzy** 

Logic!





- ✓ Need to deal with vague concepts, e.g., "tall"
- ✓ Information may also be vague/noisy, e.g., the Web
- ✓ Strictly extends "crisp" languages (1 = true; 0 = false)
- X Developing ontologies may be more difficult
- X How will fuzzy values be determined/agreed?
- Keasoner implementations still prototypical
- Practicality still an open question

**Fuzzy** 

Logic!



FOL/CL!



Expressive superset of most other languages

✓ FOL reasoners now highly capable

and Specialised reasoners can be used for subsets

FOL/CL!

Undecidability not important

and little different from high complexity

- Expressive superset of most other languages
- ✓ FOL reasoners now highly capable
  - and Specialised reasoners can be used for subsets

FOL/CL!

- Undecidability not important
   and little different from high complete
  - and little different from high complexity
- Reasoners are much less robust
- X Poor at proving non-subsumption (normal case)
- X Difficult to recognise subsets
- Incomplete answers typically used in unsound way
- \* Insert favourite logic/KR-formalism



# **Undecidability -v- High Complexity**



# **Undecidability -v- High Complexity**

- Can think of undecidable as a very high complexity class
  - Result is very low robustness of reasoner performance

Users have to make do with imperfect tests which sometimes fail to yield results" ... "analogous to 404 errors on the Web

# **Undecidability -v- High Complexity**

- Can think of undecidable as a very high complexity class
  - Result is very low robustness of reasoner performance

Users have to make do with imperfect tests which sometimes fail to yield results" ... "analogous to 404 errors on the Web

- But in practice
  - Even SOTA FOL theorem provers are not very effective for non-theorems/non-subsumption
  - Vast majority tests are non-subsumptions, so answer to most tests is "don't know" (almost every link gives a 404 error)
  - Users expect/demand (fast and) complete reasoning; otherwise they simply won't use the reasoner



## **Incompleteness -v- Incorrectness**



# **Incompleteness -v- Incorrectness**

- Applications often treat failure to prove "yes" as "no"
  - and incomplete reasoners often don't even distinguish



#### **Incompleteness -v- Incorrectness**

- Applications often treat failure to prove "yes" as "no"
  - and incomplete reasoners often don't even distinguish



#### **Incompleteness -v- Incorrectness**

- Applications often treat failure to prove "yes" as "no"
  - and incomplete reasoners often don't even distinguish

Isn't this just negation as failure?

- Absolutely not!
  - Failure in NAF means failure of entailment
    - $eg \phi$  is true if  $\phi$  is not entailed
  - It doesn't mean failure of an incomplete reasoner to prove that  $\phi$  is entailed
  - Treating "don't know" as "no" is simply incorrect







I need to express .....,\* which I can't express in OWL

\* Insert favourite expressive feature



I need to express .....,\* which I can't express in OWL

✓ There are many things that can't be expressed in OWL

✓ Some of them would certainly be very useful

\* Insert favourite expressive feature



I need to express .....,\* which I can't express in OWL

✓ There are many things that can't be expressed in OWL

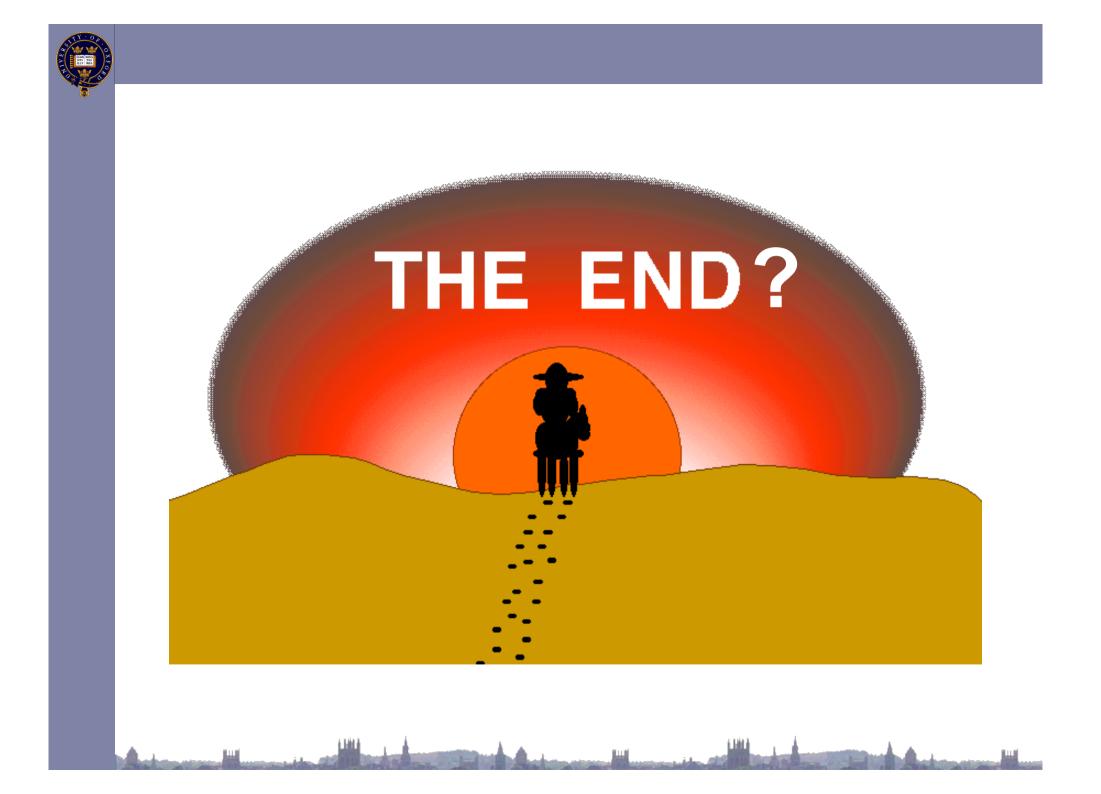
- ✓ Some of them would certainly be very useful
- ✗ It's too complicated
- X It's too complex
- It should have been based on ......

\* Insert favourite expressive feature



#### **Conclusions?**

- There is no "right choice" of ontology language "you pays your money, and you takes your choice"
- Standardisation requires *some* choice
- Claim: OWL was a (not totally un-)reasonable choice:
  - good compromise between expressive power and robust tool performance
  - has allowed for the development of a range of tools, infrastructure and applications that could previously only have been dreamt of





# **Ongoing Research**

- Optimisation/Profiles
  - [Kazakov], [Glimm et al], [Faddoul et al], [Savo et al]
- Query answering
  - [Kontchakov et al], [Konev et al], [Baader et al]
- Diagnosis and repair
  - [Horridge et al], [Peñaloza et al]
- Extensions
  - [Motik et al], [Artale et al]



# **Ongoing Standardisation Efforts**

- Standardised query language
  - SPARQL standard for RDF
  - Currently being extended for OWL, see <u>http://www.w3.org/TR/sparql11-entailment/</u>
- RDF
  - Revision currently being considered, see <a href="http://www.w3.org/2009/12/rdf-ws/">http://www.w3.org/2009/12/rdf-ws/</a>



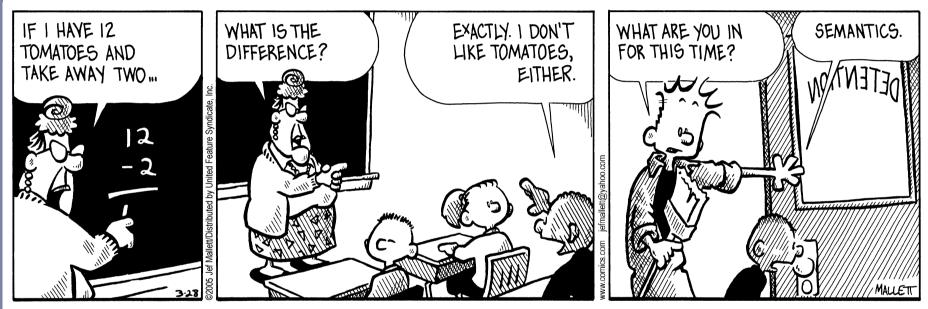


# Thank you for listening





# Thank you for listening



FRAZZ: © Jeff Mallett/Dist. by United Feature Syndicate, Inc.

Any questions?