Semantic Search

Prashanth Kamle (08305006)

Under the guidance of Dr. Pushpak Bhattacharyya
Department of Computer Science
IIT Bombay

Semantic Search

- Is that a yes or a no?

Semantic Search (contd)

1. Understand that “tall” relates to height
2. Fetch the heights of “Taj Mahal” and “Eiffel Tower” from semantically annotated sites on the internet
3. Perform a comparison
4. Return the result

In short, do Logical Inferencing

Outline

- Motivational example
- Ontologies
- Description Logic
- Querying
- Conclusion

Present day search: Keyword based

1. Automobile stereo and radio retail store
2. Automobile engine rebuilding, repair and exchange workshop
3. Car repair and retail shop
4. Jeep repair and retail shop
5. Motor mending and replacement workshop

Query	Results
Automobile	1,2
Automobile retail	1
Car repair	3
Motor repair	
Engine repair	2
Motor exchange	

*Example from Ontoseek(1999) by Guarino et al

Improve: Add structure

No	Business type	Activity	Object	Market area
1	Store	Retail	Radio, Stereo	Automobile
2	Workshop	Rebuilding, repair, exchange	Engine	Automobile
3	Shop	Retail, repair	Car	
4	Shop	Retail, repair	Jeep	
5	Workshop	Replacement, mending	Motor	

No	Business type	Activity	Object	Market area	Result
1	-	-	Automobile	-	
2	-	Retail	Automobile	-	
3	-	Repair	Car	-	3
4	-	Repair	Motor	-	
5	-	Repair	Engine	-	2
6	-	Repair	Motor	-	

Improve: Add structure

No	Business type	Activity	Object	Market area
1	Store	Retail	Radio, Stereo	Automobile
2	Workshop	Rebuilding, repair, exchange	Engine	Automobile
3	Shop	Retail, repair	Car	
4	Shop			
5	Workshop			

Increase in Precision!

No	Business type			arket area	Result
1	-	-	Automobile	-	
2	-	Retail	Automobile	-	
3	-	Repair	Car	-	3
4	-	Repair	Motor	-	
5	-	Repair	Engine	-	2
6	-	Repair	Motor	-	

Improve further: Use Ontology

No	Document	Disambiguated description
1	Automobile stereo and radio retail store	[car, auto, automobile, machine, motorcar], [radio receiver, receiving set, radio set, radio, tuner, wireless], [stereo, stereo system, stereophonic system], [retail, sell retail], [shop, store]
2	Automobile engine rebuilding, repair and exchange workshop	[car, auto, automobile, machine, motorcar], [engine], [rebuilding], [repair, fix, fixing,mending, reparation], [substitution, exchange], [workshop, shop]
3	Car repair and retail shop	[car, auto, automobile, machine, motorcar], [repair, fix, fixing, mending, reparation], [retail, sell retail], [shop, store]
4	Jeep repair and retail shop	[jeep, landrover], [repair, fix, fixing, mending, reparation], [retail, sell retail], [shop,store]
5	Motor mending and replacement workshop	[motor], [repair, fix, fixing, mending, reparation], [replacement, replacing], [workshop,shop]

Use Ontology (contd)

No	Query	Disambiguated query	Result
1	Automobile	[car, auto, automobile, machine, motorcar]	1,2,3,4
2	Automobile retail	[car, auto, automobile, machine, motorcar], [retail, sell, retail]	1,3,4
3	Car repair	[car, auto, automobile, machine, motorcar], [repair, fix, fixing,mending, reparation]	2,3,4
4	Motor repair	[motor], [repair, fix, fixing, mending, reparation]	2,5
5	Engine repair	[locomotive, engine, locomotive engine, railway locomotive], [repair,fix, fixing, mending, reparation]	-
6	Motor exchange	[motor], [substitution, exchange]	2,5

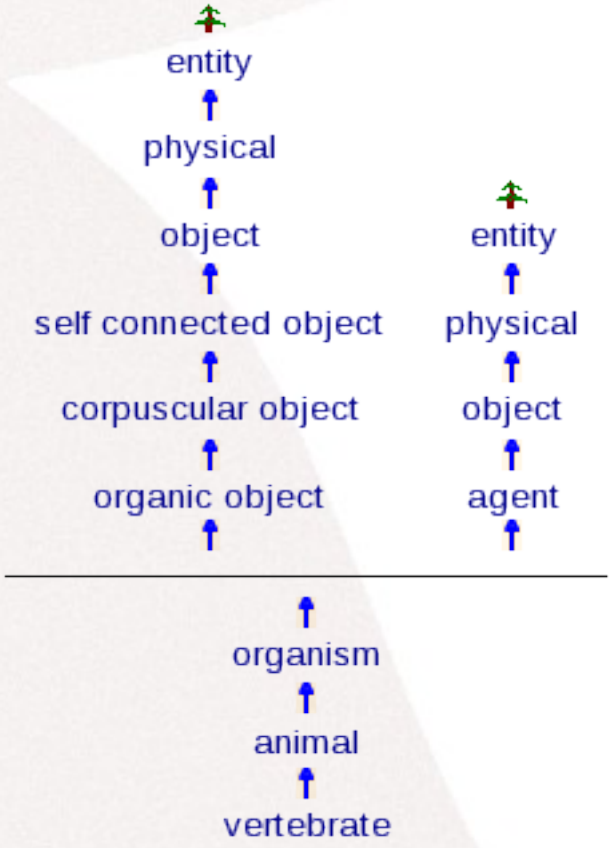
Use Ontology (contd)

No	Query	Disambiguated query	Result
1	Automobile	[car, auto, automobile, machine, motorcar]	1,2,3,4
2	Automobile retail	[car, auto, automobile, machine, motorcar], [retail, sell retail]	1,3,4
3	Car repair	[car, auto, automobile, machine, motorcar], [repair, fix,	2,3,4
4	Motor repair	ration]	2,5
5	Engine repair	railway eparation]	-
6	Motor exchange	[motor], [substitution, exchange]	2,5

Increase in Recall!

Ontology

- Individuals
- Concepts
- Relations
- Roles
- Axioms



invertebrate is disjoint from vertebrate.

If bone is an instance of bone, then there exists vertebrate vert so that bone is a part of vert.

Description Logic

- Description logics (DL) are a family of knowledge representation languages
- Used to represent the concept definitions of an application domain formally
- “Description”
  - refers to concept descriptions used to describe a domain
- “Logic”
  - logic-based semantics which can be given by a translation into first-order predicate logic.

DL – Constituents

Concepts Unary predicates *Eg. Person, Female*

Roles Binary predicates *Eg. hasChild*

Individuals Constants *Eg. Mary, John*

Constructors

- Union $\sqcup$: *Eg.* $Man \sqcup Woman$
- Intersection $\sqcap$: *Eg.* $Person \sqcap Female$
- Restriction Exists $\exists$: *Eg.* $\exists hasChild.Female$
- Restriction ForAll $\forall$: *Eg.* $\forall hasChild.Engineer$
- Negation $\neg$: *Eg.* $\neg Man$
- Number restriction: $\leq k, \geq m$

Axioms Mother $\sqsubseteq$ Parent

DL Constituents (contd)

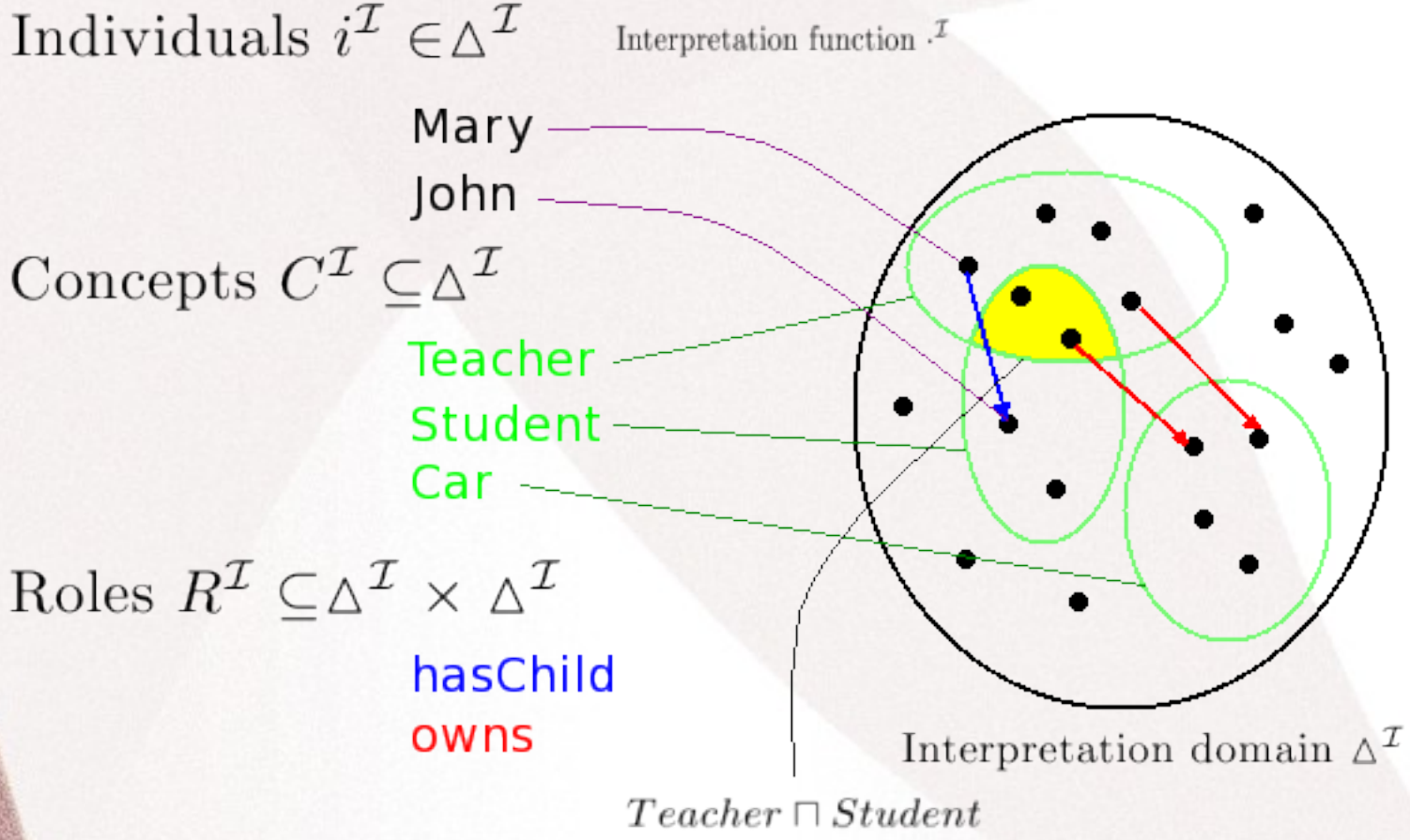
- “A man that is married to a doctor and has at least five children, all of whom are professors”

$$Human \sqcap \neg Female \sqcap \exists married.Doctor \sqcap (\geq 5 child) \sqcap \forall child.Professor$$

DL Interpretation

- An interpretation $\mathcal{I}$ is a tuple$(\triangle^I, \cdot^I)$ where
  - $\triangle^I$ is the domain
  - $\cdot^I$ is a mapping which maps
    - Names of individuals to elements of $\triangle^I$
    - Names of concepts to subsets of $\triangle^I$
    - Names of roles to subsets of $\triangle^I \times \triangle^I$

Concepts, Roles and Interpretation



DL Knowledge base

- TBox

$$Woman \equiv Person \sqcap Female$$
$$Man \equiv Person \sqcap \neg Woman$$
$$Mother \equiv Woman \sqcap \exists hasChild.Person$$
$$Father \equiv Man \sqcap \exists hasChild.Person$$
$$Parent \sqsubseteq Person$$

- ABox

$$\langle PETER \rangle : Father$$
$$\langle MARY \rangle : Mother$$
$$\langle MARY, PETER \rangle : hasChild$$
$$\langle PETER, HARRY \rangle : hasChild$$

Inferencing on the KB

- Satisfiability: Is there some interpretation that satisfies axioms in TBox?
- Subsumption: Is concept A more general than concept B?
- Equivalence: Are concept A and concept B the same?
- Instance check: Can assertion α be entailed by the ABox?
- Retrieval: Which individuals satisfy concept C?

Inferencing on the KB

- Satisfiability: Is there some interpretation that satisfies axioms in TBox?
- Subsumption: Is concept A more general than co
- Equivalence: Are concepts A and B the same?
- Instance check: Can assertion α be entailed by the ABox?
- Retrieval: Which individuals satisfy concept C?

All of these can be reduced to checking satisfiability

Tableaux Inferencing Algorithm

(1) Convert description to Negation Normal form

(2) For any existential restriction, introduce a new individual as role filler such that it satisfies the constraints expressed by the restriction.

(3) Use value restrictions in interaction with already defined role relationships to impose new constraints on individuals

(4) For disjunctive constraints, try both possibilities in successive attempts. Backtrack if you reach an obvious contradiction

(5) If an at-most number restriction is violated then the algorithm must identify different role fillers

Tableaux Inferencing algorithm (example)

$$(\exists teaches.Biology) \sqcap (\exists teaches.Statistics) \sqsubseteq (\exists teaches.(\underline{Biology \sqcap Statistics}))$$

Bio-statistics

$$(\exists R.A) \sqcap (\exists R.B) \sqsubseteq (\exists R.(A \sqcap B))$$

Tableaux Inferencing algorithm (example)

- Check whether $(\exists R.A) \sqcap (\exists R.B) \sqsubseteq (\exists R.(A \sqcap B))$
- If $C = (\exists R.A) \sqcap (\exists R.B) \sqcap \neg(\exists R.(A \sqcap B))$ is unsatisfiable, then $(\exists R.A) \sqcap (\exists R.B) \sqsubseteq (\exists R.(A \sqcap B))$
- Move the negations as far inside as possible.

$$C = (\exists R.A) \sqcap (\exists R.B) \sqcap \forall R.(\neg A \sqcup \neg B)$$

  C is now in negation normal form.

- Now, we try to construct an interpretation $\mathcal{I}$ such that $C^{\mathcal{I}} \neq \phi$. This means there must exist an individual in $\Delta^{\mathcal{I}}$ that is an element of $C^{\mathcal{I}}$. So, we construct an individual $b \in C^{\mathcal{I}}$.
- Since C is the conjunction of 3 concepts, b must satisfy $b \in (\exists R.A)^{\mathcal{I}}, b \in (\exists R.B)^{\mathcal{I}}$ and $b \in (\forall R.(\neg A \sqcup \neg B))^{\mathcal{I}}$.

Tableaux Inferencing algorithm (contd)

- From $b \in (\exists R.A)^{\mathcal{I}}$, we can see that there must exist an individual c such that $(b, c) \in R^{\mathcal{I}}$ and $c \in A^{\mathcal{I}}$. Similarly, $b \in (\exists R.B)^{\mathcal{I}}$ implies that there must exist an individual d with $(b, d) \in R^{\mathcal{I}}$ and $d \in B^{\mathcal{I}}$.
- Since b must also satisfy $\forall R.(\neg A \sqcup \neg B)$, and c, d were introduced as fillers of b for R, we get 2 more constraints $c \in (\neg A \sqcup \neg B)^{\mathcal{I}}$ and $d \in (\neg A \sqcup \neg B)^{\mathcal{I}}$.
- Now, $c \in (\neg A \sqcup \neg B)^{\mathcal{I}}$ means $c \in (\neg A)^{\mathcal{I}}$ or $c \in (\neg B)^{\mathcal{I}}$. $c \in (\neg A)^{\mathcal{I}}$ clashes with the constraint $c \in A^{\mathcal{I}}$, implying that this choice leads to an obvious contradiction. Hence, we must choose $c \in (\neg B)^{\mathcal{I}}$. Similarly, we must choose $d \in (\neg A)^{\mathcal{I}}$ in order to satisfy the constrait $d \in (\neg A \sqcup \neg B)^{\mathcal{I}}$ without contradicting $d \in B^{\mathcal{I}}$.

Tableaux Inferencing algorithm (contd)

- Now, since we have satisfied all constraints without encountering an obvious contradiction, we can conclude that C is satisfiable.
- We have generated an interpretation $\mathcal{I}$ as proof of this fact: $\Delta^{\mathcal{I}} = \{b, c, d\}$; $R^{\mathcal{I}} = \{(b, c), (b, d)\}$; $A^{\mathcal{I}} = \{c\}$ and $B^{\mathcal{I}} = \{d\}$. This means that, for this interpretation, $b \in C^{\mathcal{I}}$ i.e. $b \in ((\exists R.A) \sqcap (\exists R.B))^{\mathcal{I}}$, but $b \notin (\exists R.(A \sqcap B))^{\mathcal{I}}$. This shows that $(\exists R.A) \sqcap (\exists R.B)$ is not subsumed by $\exists R.(A \sqcap B)$.

$$(\exists R.A) \sqcap (\exists R.B) \not\sqsubseteq (\exists R.(A \sqcap B))$$

DL Family

- $\mathcal{ALC}$ - Concepts constructed using $\sqcup, \sqcap, \exists, \forall, \neg$, but roles are atomic.
- $\mathcal{S}$ is an abbreviation for $\mathcal{ALC}$ with transitive roles.
- $\mathcal{H}$ for role hierarchy. Eg: $hasDaughter \sqsubseteq hasChild$
- $\mathcal{O}$ for nominals. Eg: $\{Mary, Hohn\}$
- $\mathcal{I}$ for inverse roles. Eg: $isChildOf \equiv hasChild^{-1}$
- $\mathcal{N}$ for cardinality restrictions. Eg: $\geq 2 hasChild$
- $\mathcal{F}$ restricts cardinality to be 0 or 1.
- $\mathcal{Q}$ for qualified number restrictions. Eg: $\geq 2 hasChild.Professor$
- $\mathcal{R}$ for role inclusion and role disjointness.
- $(\mathcal{D})$ to denote use of datatype properties, data values or data types.

For example, $\mathcal{SHOIQ} = \mathcal{ALC}$ with transitive roles $+\mathcal{H} + \mathcal{O} + \mathcal{I} + \mathcal{Q}$

OWL – Web Ontology Language

- A language based on RDF, RDFS and XML to represent Ontologies
- A W3C standard

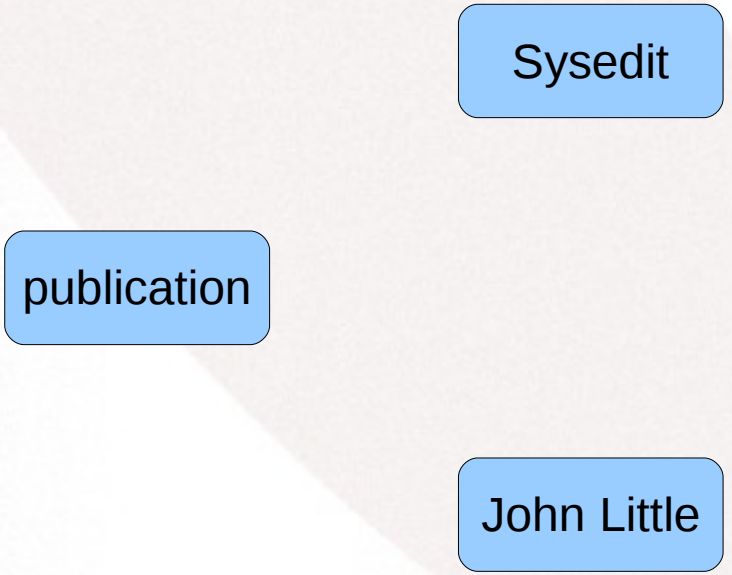
```
-<rdfs:Class rdf:ID="Animal">
   <rdfs:subClassOf rdf:resource="#Organism"/>
  -<rdfs:comment>
     An Organism with eukaryotic Cells, and lacking stiff cell walls, plastids, and photosynthet
   </rdfs:comment>
 </rdfs:Class>


-<owl:ObjectProperty rdf:ID="wears">
  -<rdfs:comment>
     (wears ?AGENT ?CLOTHING) means that ?AGENT is wearing the item of Clothing
   </rdfs:comment>
   <rdf:type rdf:resource="#BinaryPredicate"/>
   <rdfs:domain rdf:resource="#Animal"/>
   <rdfs:range rdf:resource="#Clothing"/>
 </owl:ObjectProperty>
```

OWL Species

- OWL DL
  - Uses SHOIN(D)
  - Non deterministic exponential time reasoning
- OWL Lite
  - Uses SHIF(D)
  - Deterministic exponential time reasoning
- OWL Full
  - Goes well outside DL framework
  - Reasoning undecidable

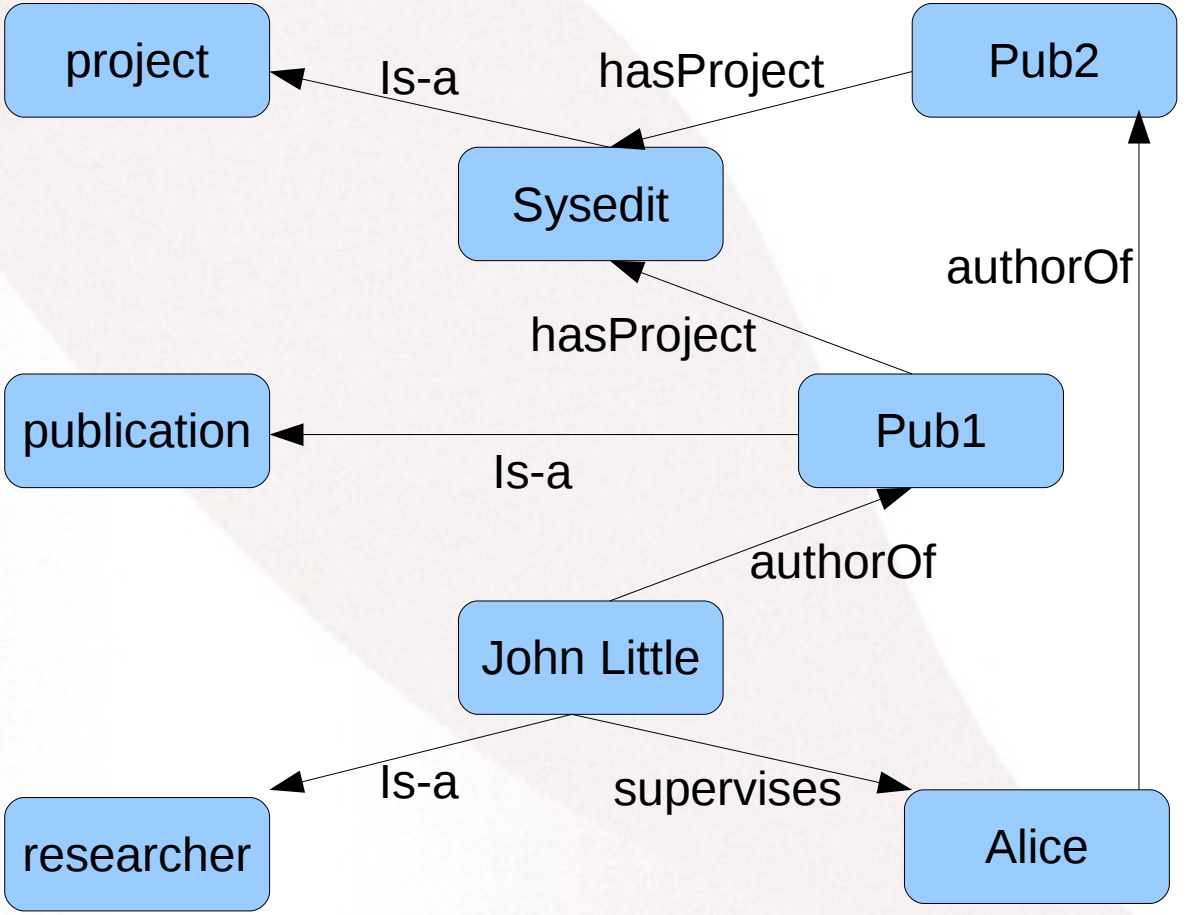
Querying

- Q = “John Little Sysedit publications”



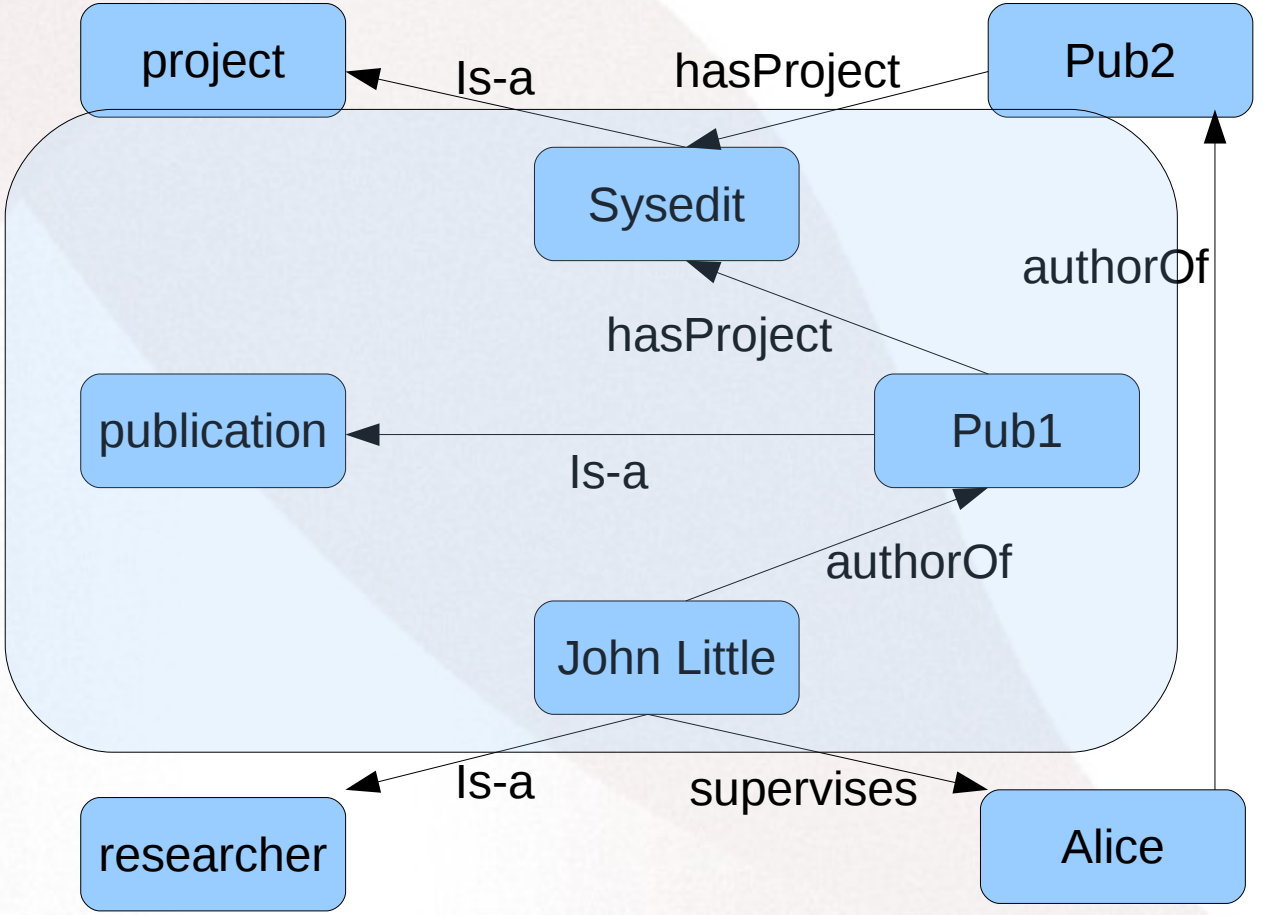
Querying

- Q = “John Little Sysedit publications”



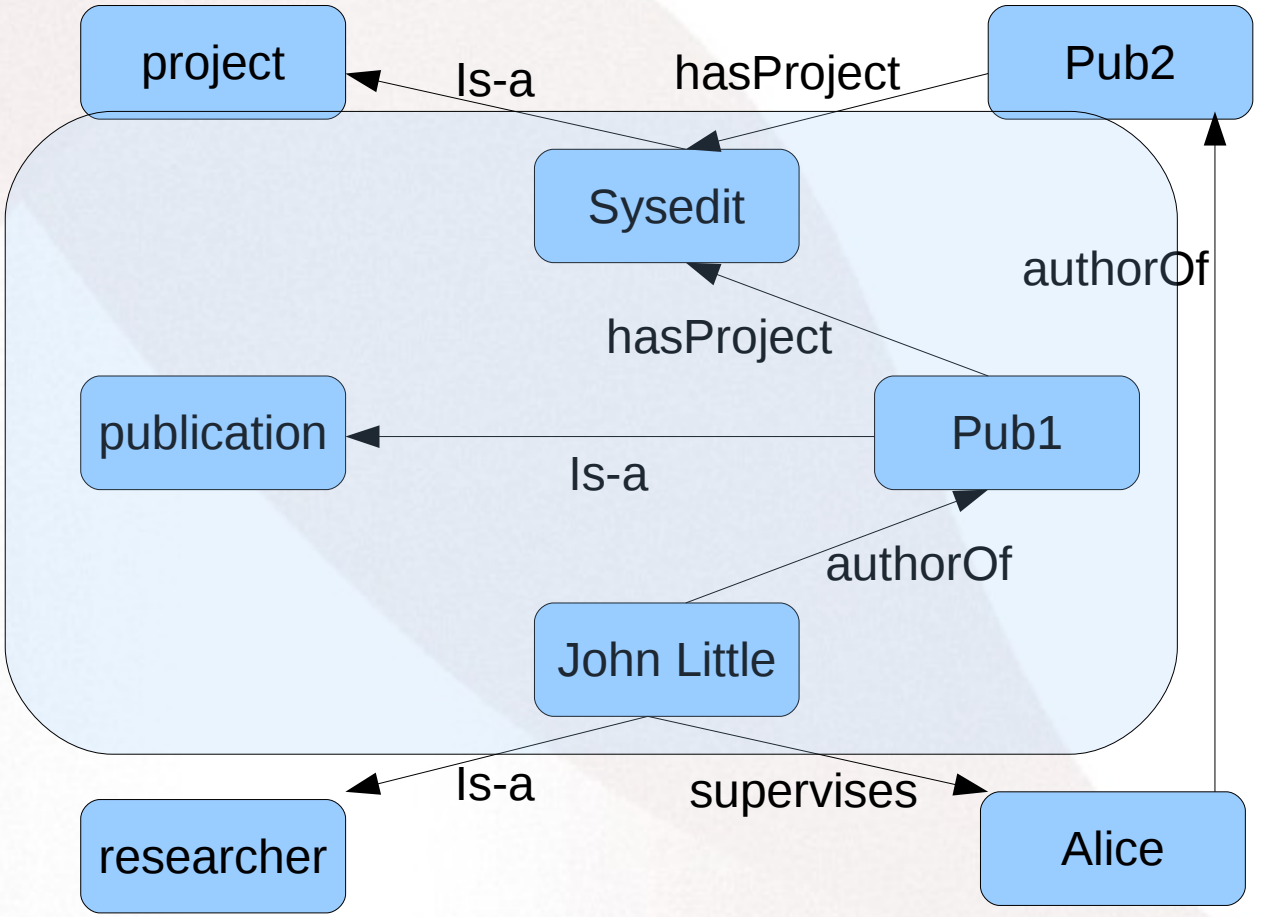
Querying

- Q = “John Little Sysedit publications”



Querying

$$Q = \langle x, JohnLittle \rangle : name \wedge \langle x, y \rangle : author \wedge \langle y, z \rangle : hasProject$$
$$\wedge \langle z, Sysedit \rangle : name \wedge \langle y : publication \rangle$$



Conclusions

- Semantic web is a hot topic of research
- Semantic annotation of documents using ontologies and inferencing gives better search results
- Need of the hour
  - Standards
  - Annotation tools
  - Efficient large scale inferencing engines
  - World wide acceptance and use

References

[1] Grigoris Antoniou and Frank van Harmelen. Web ontology language: OWL. In Handbook on Ontologies in Information Systems, pages 76–92. Springer-Verlag, 2003.

[2] Franz Baader and Werner Nutt. Basic description logics. In The description logic handbook: theory, implementation, and applications, pages 43–95, New York, NY, USA, 2003. Cambridge University Press.

[3] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web: Scientific american. Scientific American, May 2001.

[4] Ian Horrocks, Peter F. Patel-schneider, and Frank Van Harmelen. From SHIQ and RDF to OWL: The making of a web ontology language. Journal of Web Semantics, 1:7–26, 2003.

References

[5] Ian Niles and Adam Pease. Towards a standard upper ontology. In FOIS ’01: Proceedings of the international conference on Formal Ontology in Information Systems, pages 2–9, New York, NY, USA, 2001. ACM.

[6] Thanh Tran, Philipp Cimiano, Sebastian Rudolph, and Rudi Studer. Ontology-based interpretation of keywords for semantic search. In Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007), Busan, South Korea, volume 4825 of LNCS, pages 519–532, Berlin, Heidelberg, November 2007. Springer Verlag.

[7] Wikipedia. Ontology — Wikipedia, the free encyclopedia, 2009. [Online; accessed 10-April-2009].

[8] Jidi Zhao. Introduction to description logic and ontology languages, 2008. Institute of Computer Technology, TU Vienna.