

Semantic Web: an Introduction

Jian-hua Yeh (葉建華)

真理大學資訊科學系助理教授

au4290@email.au.edu.tw

Textbook

- **A Semantic Web Primer**
- **By Grigoris Antoniou, Frank van Harmelen**
- 天瓏圖書公司

History of the Semantic Web

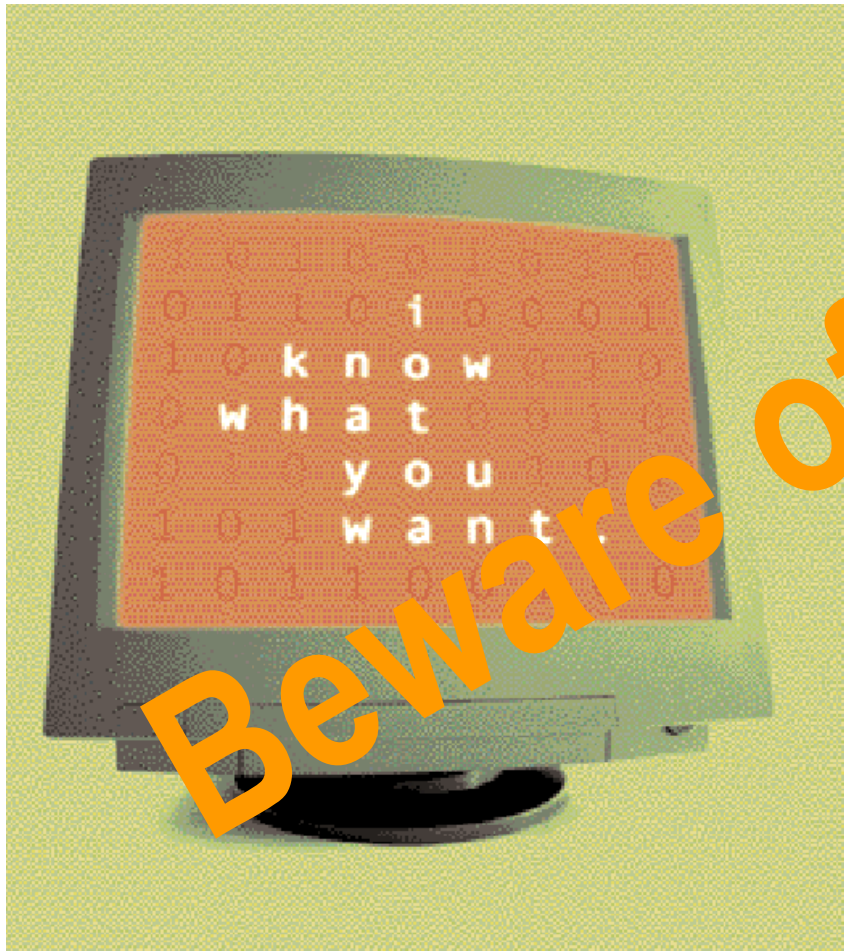
- Web was “invented” by **Tim Berners-Lee** (amongst others), a physicist working at CERN
- TBL’s original vision of the Web was much more ambitious than the reality of the existing (syntactic) Web:



“... a goal of the Web was that, if the interaction between person and hypertext could be so intuitive that the **machine-readable** information space gave an accurate representation of the state of people's thoughts, interactions, and work patterns, then **machine analysis** could become a very powerful management tool, seeing patterns in our work and facilitating our working together through the typical problems which beset the management of large organizations.”

- TBL (and others) have since been working towards realising this vision, which has become known as the **Semantic Web**
 - E.g., article in May 2001 issue of Scientific American...

Scientific American, May 2001:



THE SEMANTIC WEB

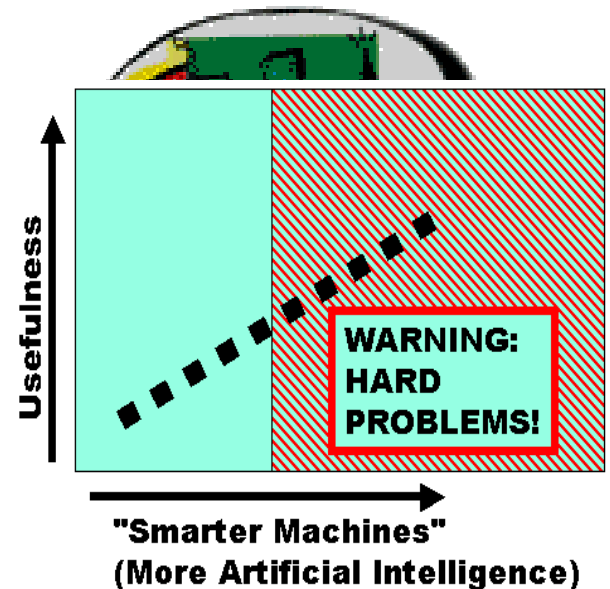
A new form of Web content
that is meaningful to computers
will unleash a revolution of new abilities

by
TIM BERNERS-LEE,
JAMES HENDLER and
ORA LASSILA

PHOTO CREDIT HERE

Beware of the Hype

- Hype seems to suggest that Semantic Web means: “semantics + web = AI”
 - “A new form of Web content that is meaningful to computers will unleash revolution of new abilities”
- More realistic to think of it as meaning “semantics + web + AI = more useful web”
 - Realising the complete “vision” is too hard for now (probably)
 - But we can make a start by adding **semantic annotation** to web resources



Where we are Today: the Syntactic Web

http://www2002.org

WWW 2002

THE ELEVENTH INTERNATIONAL WORLD WIDE WEB CONFERENCE

Shanahan Waikiki Hotel
Honolulu, Hawaii, USA
7-11 May 2002

1 LOCATION. 5 DAYS. LEARN. INTERACT.

Registered participants coming from:
Australia - Canada - Chile - Denmark - France - Germany - China - Hong Kong - India - Ireland - Japan - Malta - New Zealand - The Netherlands - Norway - Singapore - Switzerland - The United States - Vietnam - Zambia

On 7-11 May 2002, Honolulu, Hawaii will provide the backdrop for The Eleventh International World Wide Web Conference. This prestigious series of the International World Wide Web Conference Committee (IW3C2) attracts participants from around the world, and provides a public forum for the World Wide Web Consortium (W3C) through the annual W3C track.

The conference is being organized by the International World Wide Web Conference Committee (IW3C2), the University of Hawaii and the Pacific Telecommunications Council (PTC).

FEATURED SPEAKERS (CONFIRMED)

Tim Berners-Lee, inventor of the World Wide Web and Director of the W3C, who now holds the 3Com Founders chair at the Laboratory for Computer Science (LCS) at the Massachusetts Institute of Technology (MIT).

Richard A. DeMillo, vice president and chief technology officer for Hewlett-Packard Company.

Jan Foster, guru of "Grid Computing", associate professor at MIT.

W3C Award Prize Winner, 2001

Tim Berners-Lee - Homepage

Contents

Short bio
Before you mail me
Address
Table of contents
Speaking engagements
Press interviews

See also

Longer bio
Slides from some talks
Design Issues: web architecture
World Wide Web Consortium
Frequently Asked Questions
Working the Web

Tim Berners-Lee

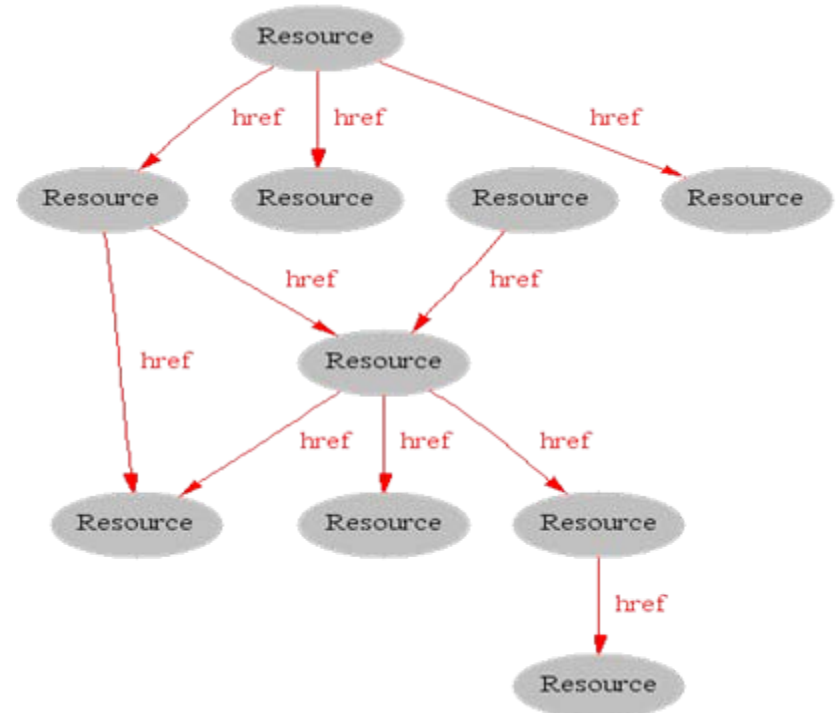
Working the Web by Tim Berners-Lee with Mark Fischetti, Warper San Francisco: Hardback: ISBN0060515861, Abridged audio cassette abridged ISBN006049521256 and various other languages.

Bio

A graduate of Oxford University, England, Tim now holds the 3Com Founders chair at the Laboratory for Computer Science and Artificial Intelligence (LCS) at the Massachusetts Institute of Technology (MIT). He directs the World Wide Web Consortium, an open forum of companies and organizations with the mission to lead the Web to its full potential.

With a background of system design in real-time communications and text processing software development, in 1989 he invented the World Wide Web, an internet-based hypermedia initiative for global information sharing, while working at CERN, the European Particle Physics Laboratory. He wrote the first web client (browser-editor) and server in 1990.

Before coming to CERN, Tim worked with Image Computer Systems, of Farnborough, Dorset, England and before that at



[Hendler & Miller 02]

The Syntactic Web is...

- **A hypermedia, a digital library**
 - A library of documents called (web pages) interconnected by a hypermedia of links
- **A database, an application platform**
 - A common portal to applications accessible through web pages, and presenting their results as web pages
- **A platform for multimedia**
 - BBC Radio 4 anywhere in the world! Terminator 3 trailers!
- **A naming scheme**
 - Unique identity for those documents

A place where computers do the presentation (easy) and people do the linking and interpreting (hard).

Why not get computers to do more of the hard work?

Hard Work using the Syntactic Web...

Find images of Peter Patel-Schneider, Frank van Harmelen and Alan Rector...



Rev. Alan M. Gates, Associate Rector of the Church of the Holy Spirit, Lake Forest, Illinois

Impossible (?) using the Syntactic Web...

- **Complex queries involving background knowledge**
 - Find information about “animals that use sonar but are not either bats or dolphins”, e.g., Barn Owl
- **Locating information in various repositories**
 - Travel enquiries
 - Prices of goods
 - Results of human experiments
- **Finding and using “resources”**
 - Visualise surface between two proteins
- **Delegating complex tasks to the web “agents”**
 - Book me a holiday somewhere warm, not too far away, and that doesn't speak French or English



What is the Problem?

- Consider a typical web page:



The screenshot shows the homepage for the WWW 2002 conference. At the top, there's a header with the URL <http://www2002.org> and the text "WWW 2002 THE ELEVENTH INTERNATIONAL WORLD WIDE WEB CONFERENCE". Below this, it specifies the location: "Sheraton Waikiki Hotel, Honolulu, Hawaii, USA" and the dates "7-11 May 2002". A tagline reads "1 LOCATION. 5 DAYS. LEARN. INTERACT.".

On the left, a vertical navigation menu lists various links: "Conference Proceedings", "Call for Participation Program", "Registration Information", "Hotel Accommodation", "Conference Committee", "Sponsorship/Exhibition Opportunities", "Volunteer Information", "Information about Hawaii", and "Previous & Future WWW Conferences".

The main content area features a section titled "Registered participants coming from:" followed by a list of countries: "Australia · Canada · Chile · Denmark · France · Germany · Ghana · Hong Kong · India · Italy · Ireland · Japan · Malta · New Zealand · The Netherlands · Norway · Singapore · Switzerland · The United States · Vietnam · Zambia". Below this is a "REGISTER NOW" button.

A paragraph of text states: "On 7-11 May 2002, Honolulu, Hawaii will provide the backdrop for The Eleventh International World Wide Web Conference. This prestigious series of the International World Wide Web Conference Committee (IW³C²) attracts participants from around the world, and it provides a public forum for the World Wide Web Consortium (W3C) through the annual W3C track."

Below this, it mentions the organizers: "The conference is being organized by the International World Wide Web Conference Committee (IW³C²), the University of Hawaii and the Pacific Telecommunications Council (PTC)."

A section titled "FEATURED SPEAKERS (CONFIRMED)" displays two speakers: "Tim Berners-Lee, inventor of the World Wide Web and Director of the W3C who now holds the 3Com Founders chair at the Laboratory for Computer Science (LCS) at the Massachusetts Institute of Technology (MIT)." and "Richard A. DeMillo, vice president and chief technology officer for Hewlett-Packard Company.".

At the bottom, there are logos for various sponsors and a mention of "McArthur Prize Winner".

- Markup consists of:
 - rendering information (e.g., font size and colour)
 - Hyper-links to related content
- Semantic content is accessible to humans but not (easily) to computers...

What information can we see...

WWW2002

The eleventh international world wide web conference

Sheraton waikiki hotel

Honolulu, hawaii, USA

7-11 may 2002

1 location 5 days learn interact

Registered participants coming from

**australia, canada, chile denmark, france, germany, ghana, hong kong, india,
ireland, italy, japan, malta, new zealand, the netherlands, norway,
singapore, switzerland, the united kingdom, the united states, vietnam,
zaire**

Register now

**On the 7th May Honolulu will provide the backdrop of the eleventh
international world wide web conference. This prestigious event ...**

Speakers confirmed

Tim berners-lee

Tim is the well known inventor of the Web, ...

Ian Foster

Ian is the pioneer of the Grid, the next generation internet ...

[illegible]

<name>       

❄️🌀🍷 🍷●🍷❖🍷■◆🌀 ✠■◆🍷◻■🌀◆✠◻■🌀● ◆◻◻●♂ ◆✠♂🍷 ◆🍷♂🍷◻■</name>

<location>💧🌊🐍📱🌀🔹📱 🔹🌀⌘&⌘&⌘ 🌊📱🔹🐍●

👉🏻◻️◼️◻️●◼️◼️📦 🌊🌀◼️🌀)()(📦 ✚️💧✌️</location>

<date>            **</date>**

<slogan>  ●               **</slogan>**

[illegible]

၁. ၂. ၃. ၄. ၅. ၆. ၇. ၈. ၉. ၁၀. ၁၁. ၁၂. ၁၃. ၁၄. ၁၅. ၁၆. ၁၇. ၁၈. ၁၉. ၂၀. ၂၁. ၂၂. ၂၃. ၂၄. ၂၅. ၂၆. ၂၇. ၂၈. ၂၉. ၃၀. ၃၁. ၃၂. ၃၃. ၃၄. ၃၅. ၃၆. ၃၇. ၃၈. ၃၉. ၄၀. ၄၁. ၄၂. ၄၃. ၄၄. ၄၅. ၄၆. ၄၇. ၄၈. ၄၉. ၅၀. ၅၁. ၅၂. ၅၃. ၅၄. ၅၅. ၅၆. ၅၇. ၅၈. ၅၉. ၆၀. ၆၁. ၆၂. ၆၃. ၆၄. ၆၅. ၆၆. ၆၇. ၆၈. ၆၉. ၇၀. ၇၁. ၇၂. ၇၃. ၇၄. ၇၅. ၇၆. ၇၇. ၇၈. ၇၉. ၈၀. ၈၁. ၈၂. ၈၃. ၈၄. ၈၅. ၈၆. ၈၇. ၈၈. ၈၉. ၉၀. ၉၁. ၉၂. ၉၃. ၉၄. ၉၅. ၉၆. ၉၇. ၉၈. ၉၉. ၁၀၀.

<introduction> ☀️ 🌊 ⚡️ ♦️ 💎 📦 📦 ♦️



































🔥📐🌀🔗📐🔥 🌀📐🔗🔗📐🔥🌀

<speaker>*✕○ ㄹㄴ□■ㄴ□◆ㄹ●ㄴㄴ</speaker>

<bio> ❄️ ⚔️ ☉ ⚔️ ♦️ ⚡️ ♍️ ♦️ ♍️ ●● & 🟩🟩♦️🟩 ⚔️🟩⚡️♍️🟩♦️🟩🟩 🟩♂️ ⚡️ ⚡️ ♍️
 ⚡️ ♍️ ♂️ 📺 </bio>...

But What About...



❄️🌀🍷 🍷●🍷❖🍷■◆🌀 ✠■◆🍷□■🌀◆✠□■🌀● ◆□□●⚡ ◆✠⚡🍷 ◆🍷🌀🍷□■</conf>

<place>💧🌀♐️◻️🌀🔹◻️◼️ 🔹🌀⌘&⌘&⌘ 🔹◻️🔹♐️●





















<date>         **</date>**

<slogan> ●◻ྐྱ◻✦✦◻■ འུའུའུའུ◻✦ ●ལྱུའུའུ◻■ ✦■◻✦ལྱུའུའུ◻✦ **</slogan>**

[illegible]

<introduction>☀️🐍⌘⬢🐍📦 📦⬢

<speaker>❄️🏹🌐 🌀📊🗲️📊🔍📖🎧📈📈</speaker>

<bio>❄️🌞🌕 🌕💎 💎🌊🐍 💎🐍🔴🟡 &🟩🟩💎🟩 🌕🟩❖🐍🟩💎🟩🟩 🟩🔼 💎🌊🐍 ❖🐍🐍🚪...

Need to Add “Semantics”

- **External agreement** on meaning of annotations
 - E.g., Dublin Core
 - Agree on the meaning of a set of annotation tags
 - Problems with this approach
 - Inflexible
 - Limited number of things can be expressed
- Use **Ontologies** to specify meaning of annotations
 - Ontologies provide a vocabulary of terms
 - New terms can be formed by combining existing ones
 - Meaning (**semantics**) of such terms is formally specified
 - Can also specify relationships between terms in multiple ontologies

Ontology: Origins and History

Ontology in Philosophy

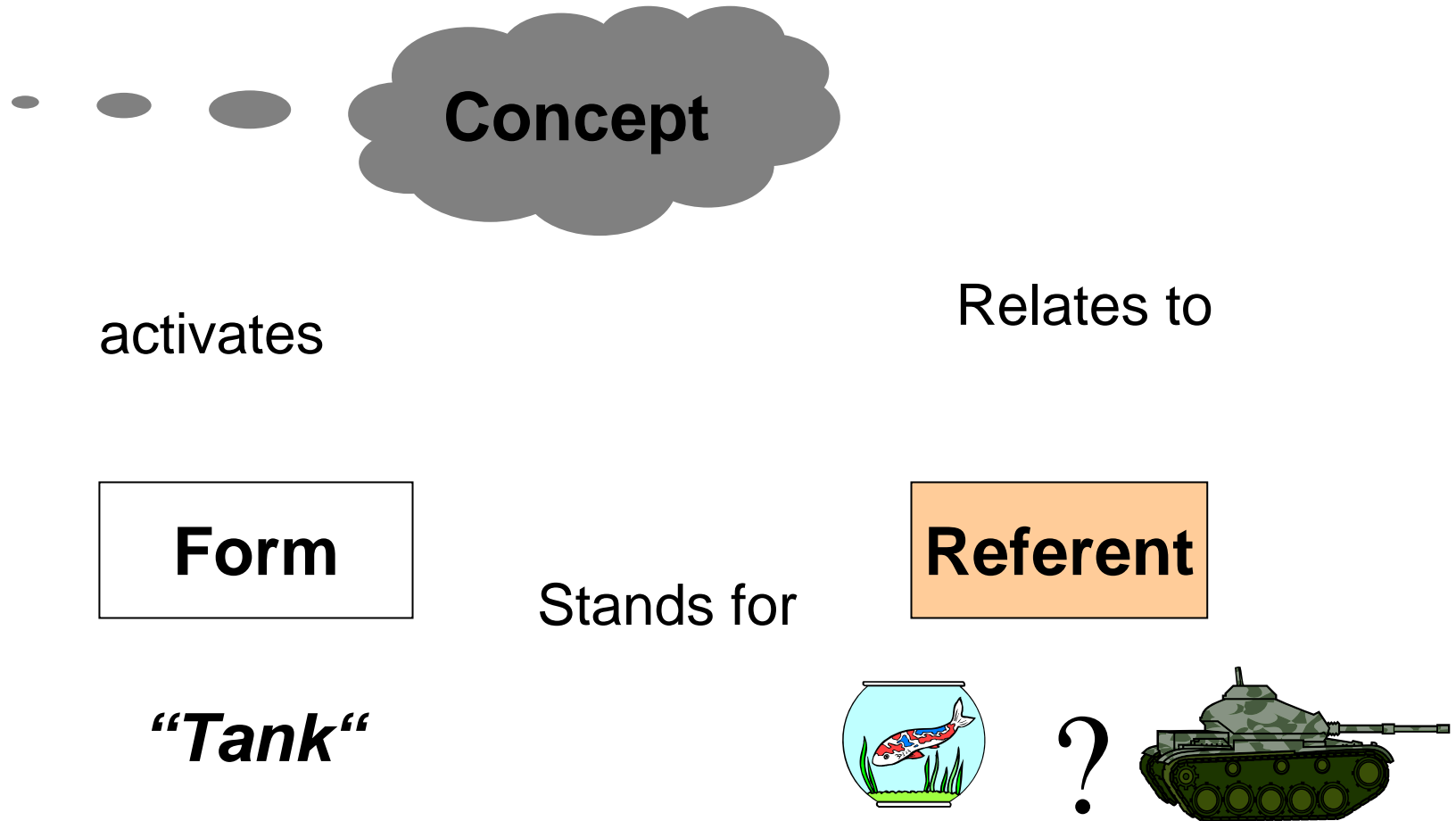
a philosophical discipline—a branch of philosophy that deals with the nature and the organisation of reality

- **Science of Being (Aristotle, Metaphysics, IV, 1)**
- **Tries to answer the questions:**

What characterizes being?

Eventually, what is being?

Ontology in Linguistics



[Ogden, Richards, 1923]

Ontology in Computer Science

- **An ontology is an engineering artifact:**
 - It is constituted by a specific vocabulary used to describe a certain reality, plus
 - a set of explicit assumptions regarding the intended meaning of the vocabulary.
- **Thus, an ontology describes a formal specification of a certain domain:**
 - Shared understanding of a domain of interest
 - Formal and machine manipulable model of a domain of interest

“An explicit specification of a conceptualisation” [Gruber93]

Structure of an Ontology

Ontologies typically have two distinct components:

- **Names for important concepts in the domain**
 - **Elephant** is a concept whose members are a kind of animal
 - **Herbivore** is a concept whose members are exactly those animals who eat only plants or parts of plants
 - **Adult_Elephant** is a concept whose members are exactly those elephants whose age is greater than 20 years
- **Background knowledge/constraints on the domain**
 - **Adult_Elephants** weigh at least 2,000 kg
 - All **Elephants** are either **African_Elephants** or **Indian_Elephants**
 - No individual can be both a **Herbivore** and a **Carnivore**

A Semantic Web — First Steps

Make web resources more accessible to automated processes

- **Extend existing rendering markup with **semantic markup****
 - Metadata annotations that describe content/function of web accessible resources
- **Use Ontologies to provide **vocabulary** for annotations**
 - “Formal specification” is accessible to machines
- **A prerequisite is a standard web ontology language**
 - Need to agree common **syntax** before we can share semantics
 - Syntactic web based on **standards** such as **HTTP** and **HTML**

Ontology Design and Deployment

- **Given key role of ontologies in the Semantic Web, it will be essential to provide **tools** and **services** to help users:**
 - **Design and maintain high quality ontologies, e.g.:**
 - **Meaningful** — all named classes can have instances
 - **Correct** — captured intuitions of domain experts
 - **Minimally redundant** — no unintended synonyms
 - **Richly axiomatised** — (sufficiently) detailed descriptions
 - **Store (large numbers) of **instances** of ontology classes, e.g.:**
 - Annotations from web pages
 - **Answer **queries** over ontology classes and instances, e.g.:**
 - Find more general/specific classes
 - Retrieve annotations/pages matching a given description
 - **Integrate** and align multiple ontologies

Example Ontology

