What is the Semantic Web?

The Semantic Web aims to make information on the World Wide Web accessible to computers.

- Not only parsable by computers (i.e., XML), but also understandable (in some sense) by computers.
- Prior agreements between humans are not needed to provide meaning (as is the case for XML).
- Human guidance is not always needed.

Scientific American, May 2001:

PHOTO CREDIT HERE

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

Scientific American, May 2001:

PHOTO CREDIT HERE

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

- **Realising the complete "vision" is too hard for now (probably)**
- **But we can make a start by adding semantic annotation to web resources**

What is Knowledge Representation?

Knowledge Representation aims to make information accessible to computers.

- Not only parsable by computers (i.e., databases) but also understandable (in some sense) by computers.
- Prior agreements between humans are not needed to provide meaning (as is the case for databases).
- Human guidance is not always needed.

Why the Semantic Web matches Knowledge Representation

- Semantic Web is not just data
  - Divergent interpretations of reality
  - Missing information (and not just null values)
- Doesn‘t match assumptions of databases or object-oriented programming
  - Open world (no closed-world assumption)
  - Objects change status over time
- Semantic Web is a representation system

Differences between Knowledge Representation and Databases

- Database (relational, object-oriented, semi-structured) assumptions
  - All relevant information is known (not there implies not true)
  - All information is definite
    - No disjunction, as in John‘s friend is either Susan or Bill
  - Objects have a single minimal class/type
    - Can‘t have John is both a student and an employee (unless there is a student-employee class/type)
- Knowledge Representation assumptions
  - Relevant information may be missing
    - E.g., no information about John‘s friends doesn‘t mean that he doesn‘t have any
  - Indefinite information may be present
    - E.g., John has a friend, but who it is is not known
  - Multiple (and changing types) allowed

The Semantic Web and/vs Knowledge Representation

- The Semantic Web is both an opportunity and a challenge for Knowledge Representation
  - An opportunity because the Semantic Web is (or will be) a source of information with very similar goals to those of Knowledge Representation
  - A challenge because some of the characteristics of the Semantic Web violate some of the assumptions that have generally been held in Knowledge Representation
- Knowledge Representation is both a resource and a cautionary tale for the Semantic Web
  - Knowledge Representation techniques can be utilized in the Semantic Web
  - Problems encountered in Knowledge Representation have already plauged the Semantic Web.

Why Knowledge Representation is a resource for the Semantic Web

- Knowledge Representation provides formal rigor
  - Meaning of information-bearing constructs are formally determined
    - Needed for computers to process the constructs (compare with formal syntax, also required for computer processing)
- Knowledge Representation is concerned with reasoning
  - Determining what follows from a collection of information
  - Provides an account of what can be (reliably) done by a computer
- Knowledge Representation systems are becoming quite powerful and reliable

Why Knowledge Representation is a cautionary tale for the Semantic Web

- Early Knowledge Representation was not formal
  - Lead to many interminable debates about the meaning of constructs
  - Current work in Knowledge Representation is very concerned with the formal meaning of constructs
- Computing with representations is difficult
  - Many problems are intractable or even undecidable
  - Lead to a retreat to simpler languages
  - Current systems are quite capable, even on expressive languages
    - Heavily optimized code
    - Computers are much more powerful

Why the Semantic Web is an opportunity for Knowledge Representation

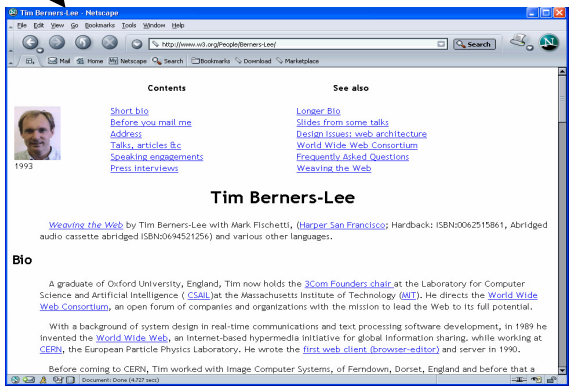
- The World Wide Web forms a (very) large source of information (albeit in awkward formats)
- The Semantic Web aims to transform much of this information into a form compatible with the goals of Knowledge Representation
- The Semantic Web also will contain services that can be controlled by other computers (and reasoned about)
  - A potential solution to the Grounding Problem in Knowledge Representation

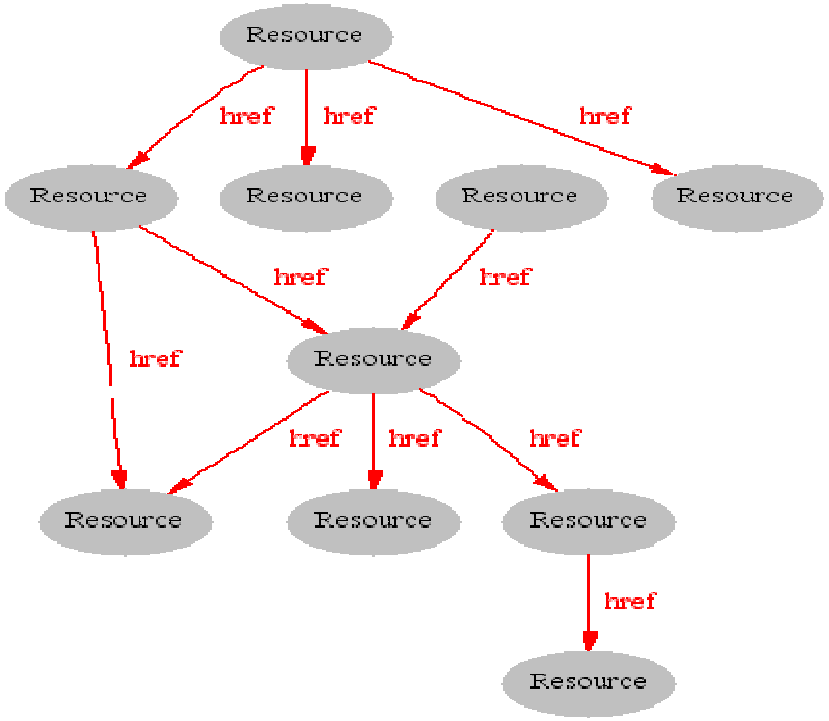
Why the Semantic Web is a challenge for Knowledge Representation

- Very large amounts of information will be part of the Semantic Web
  - Can overwhelm formal reasoning methods
- The Semantic Web contains differing interpretations of reality
  - Which one(s) to choose?
- Recovering from inconsistencies
  - How to determine how inconsistencies arise
  - How to determine who to trust

Why is the Semantic Web a good idea?

Where we are Today: the Syntactic Web





[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?

[Goble 03]

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”
- Locating information in data repositories
  - Travel enquiries
  - Prices of goods and services
  - Results of human genome experiments
- Finding and using “web services”
  - Visualise surface interactions between two proteins
- Delegating complex tasks to web “agents”
  - Book me a holiday next weekend somewhere warm, not too far away, and where they speak French or English

What is the Problem?

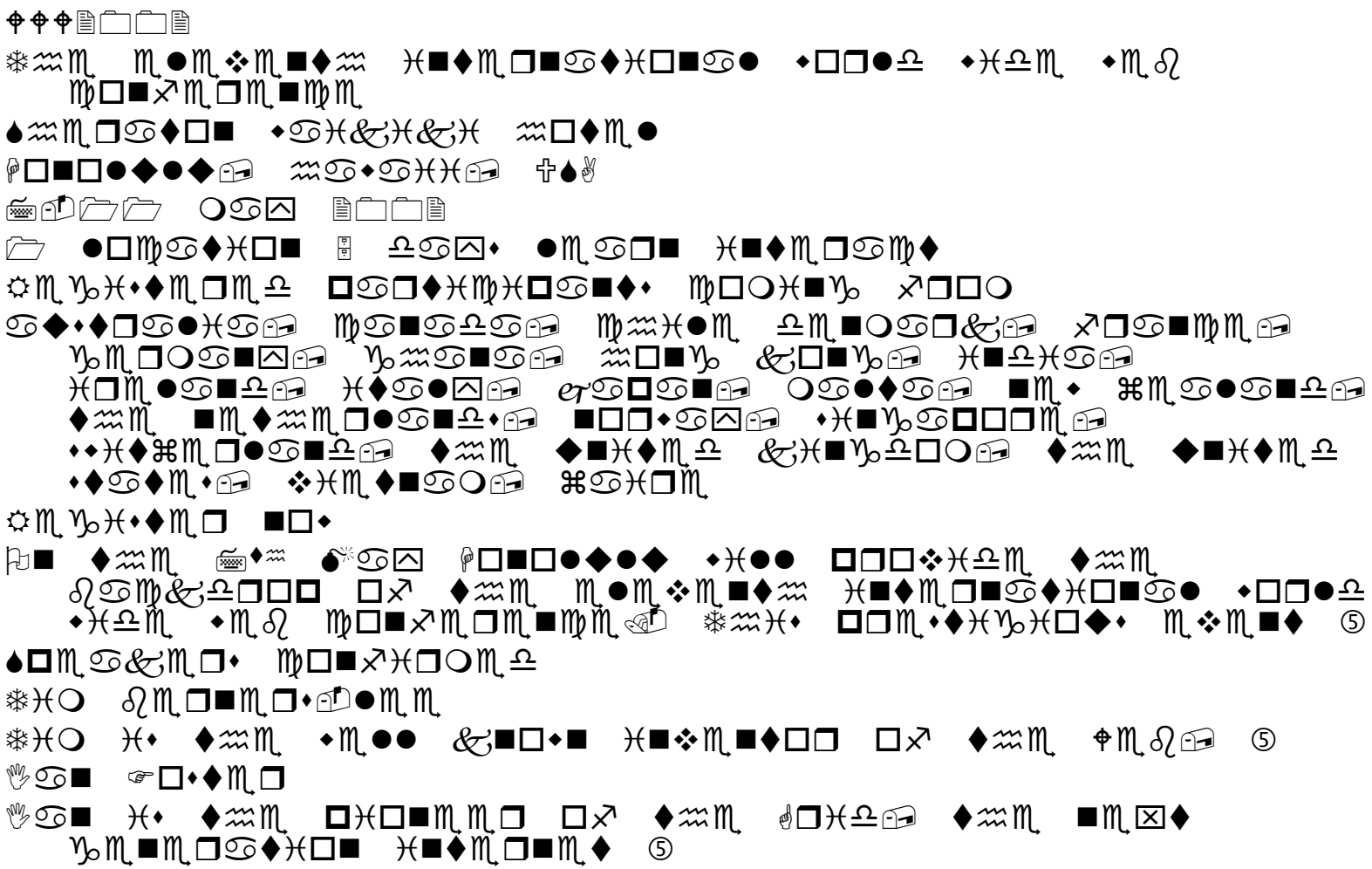
- Consider a typical web page:

- 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 …

What information can a machine see…



Solution: XML markup with “meaningful” tags?

<name>⌖⌖⌖🗎🗀🗀🗎
❄♒♏ ♏●♏❖♏■♦♒ ♓■♦♏❒■♋♦♓□■♋● ⬩□❒●♎ ⬩♓♎♏ ⬩♏♌♍□■</name>

<location>♦♒♏❒♋♦□■ ⬩♋♓&♓&♓ ♒□♦♏●
☝□■□●◆●◆🖃 ♒♋⬩♋♓♓🖃 ✞♦✌</location>

<date>🖮🖃🗁🗁 ❍♋🖂 🗎🗀🗀🗎</date>

<slogan>🗁 ●□♍♋♦♓□■ 🯅 ♎♋🖂⬩ ●♏♋❒■ ♓■♦♏❒♋♍♦</slogan>

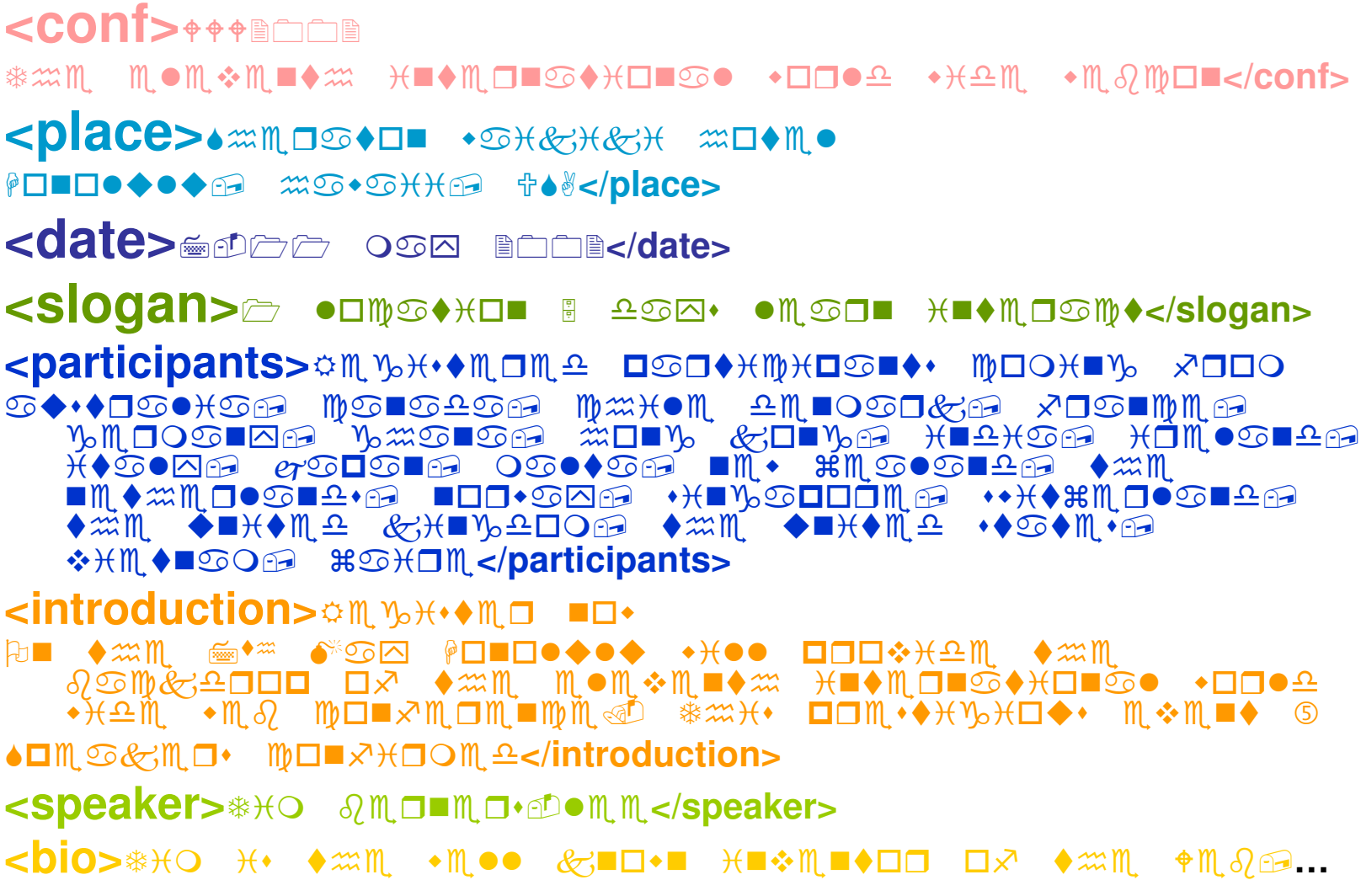
<participants>☼♏♑♓⬩♦♏❒♏♎ □♋❒♦♓♍♓□♋■♦⬩ ♍□❍♓■♑ ↗❒□❍
♋◆⬩♦❒♋●♓♋🖃 ♍♋■♋♎♋🖃 ♍♒♓●♏ ♎♏■❍♋❒&🖃 ↗❒♋■♍♏🖃
♑♏❒❍♋■🖂🖃 ♑♒♋■♋🖃 ♒□■♑ &□■♑🖃 ♓■♎♓♋🖃 ♓❒♏●♋■♎🖃
♓♦♋●🖂🖃 ♂♋□♋■🖃 ❍♋●♦♋🖃 ■♏⬩ ⌘♏♋●♋■♎🖃 ♦♒♏
■♏♦♒♏❒●♋■♎⬩🖃 ■□❒⬩♋🖂🖃 ⬩♓■♑♋□□❒♏🖃 ⬩⬩♓♦⌘♏❒●♋■♎🖃
♦♒♏ ◆■♓♦♏♎ &♓■♑♎□❍🖃 ♦♒♏ ◆■♓♦♏♎ ⬩♦♋♦♏⬩🖃
❖♓♏♦■♋❍🖃 ⌘♋♓❒♏</participants>

<introduction>☼♏♑♓⬩♦♏❒ ■□⬩
🏳■ ♦♒♏ 🖮⬩♒ 💣♋🖂 ☝□■□●◆●◆ ⬩♓●● □❒□❖♓♎♏ ♦♒♏
♌♋♍&♎❒□□ □↗ ♦♒♏ ♏●♏❖♏■♦♒ ♓■♦♏❒■♋♦♓□■♋● ⬩□❒●♎
⬩♓♎♏ ⬩♏♌ ♍□■↗♏❒♏■♍♏📣 ❄♒♓⬩ □□♏⬩♦♓♑♓□◆⬩ ♏❖♏■♦ ⑤
♦□♏♋&♏❒⬩ ♍□■↗♓❒❍♏♎</introduction>

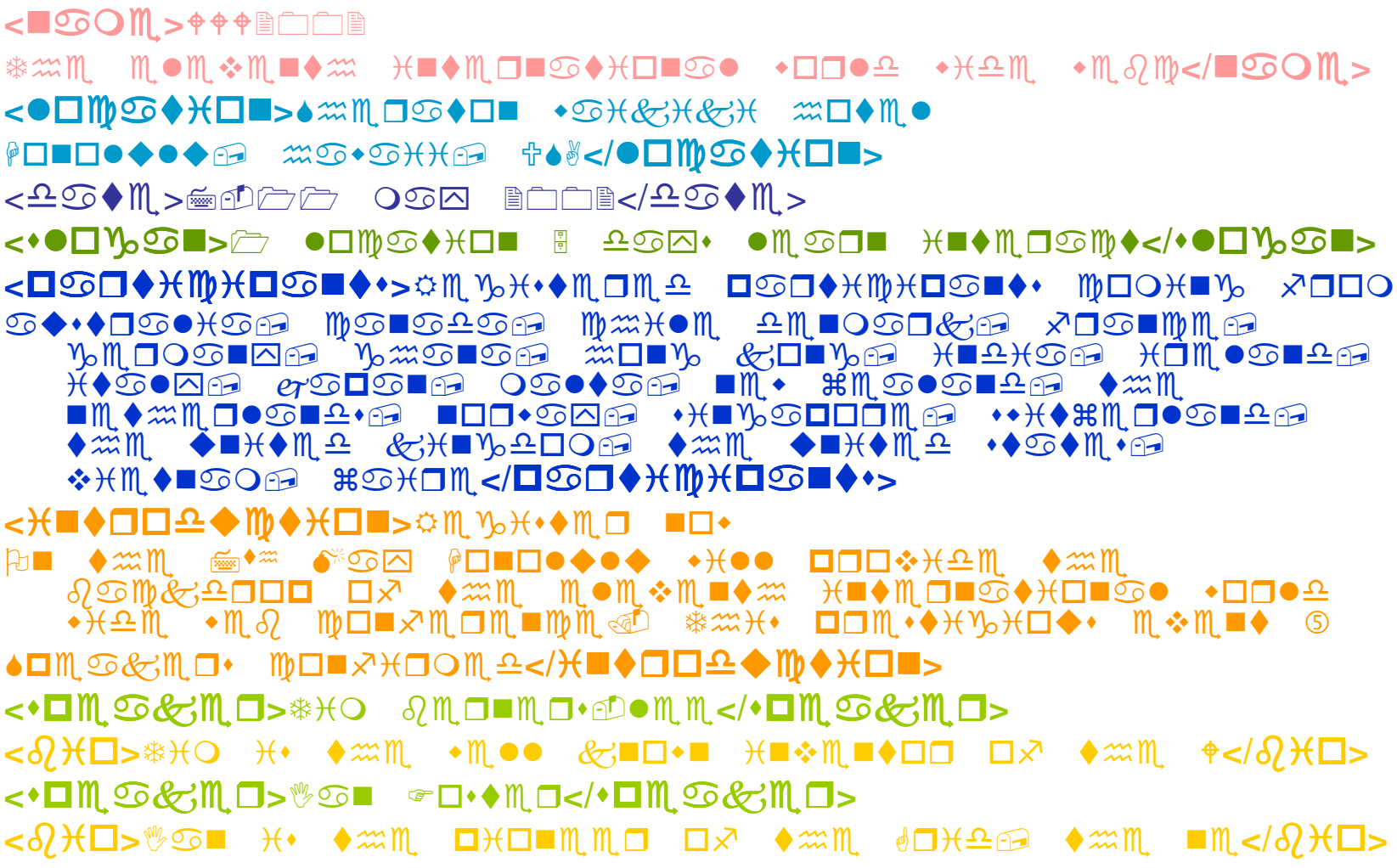
<speaker>❄♓❍ ♌♏❒■♏❒⬩🖃●♏♏</speaker>

<bio>❄♓❍ ♓⬩ ♦♒♏ ⬩♏●● &■□⬩■ ♓■❖♏■♦□❒ □↗ ♦♒♏
⌖♏♌🖃</bio>…

But What About...



Machine sees…



Need to Add "Semantics"

Two very different possible approaches:

1. External agreement on meaning of annotations
   - Agree on the meaning of a set of annotation tags, e.g., Dublin core
   - Problems with this approach
     - Inflexible
     - Limited number of things can be expressed
2. Use on-line 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

Semantic Web takes second approach

Characteristics of the Semantic Web

- Part of the Web
  - Uses Web addressing (URIs)
  - Adheres to Web philosophy
  - Connected to the rest of the Web (e.g, services)
- *Large* part of the Web
  - Very many connected documents
  - Diverse, conflicting
- Semantic
  - Contains ontological information about meaning of objects

What are Ontologies?

Ontology: Origins and History

Ontology in Philosophy

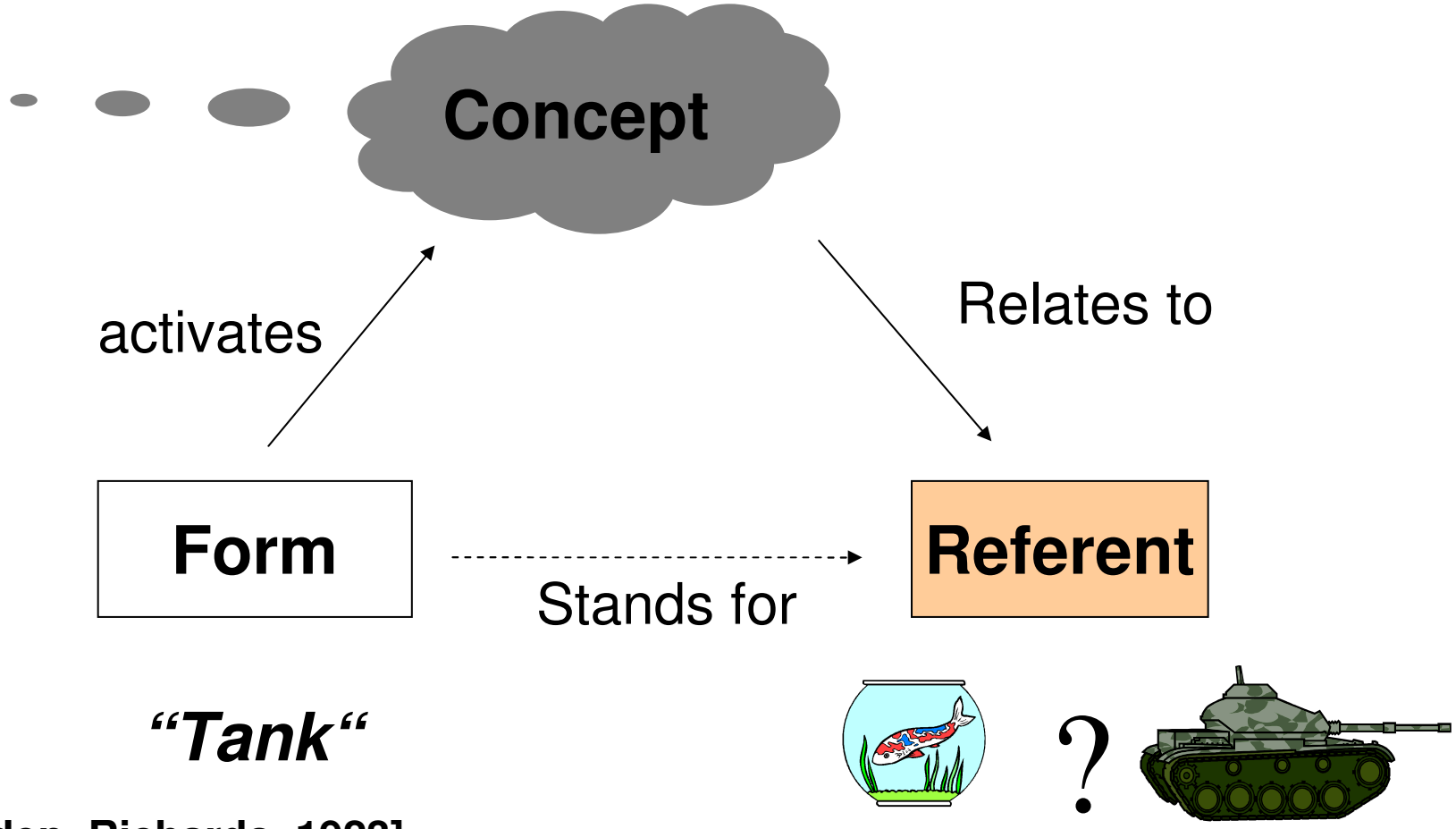
a philosophical discipline—a branch of philosophy that deals with the nature and the organisation of reality

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

  What characterizes being?

  Eventually, what is being?

Ontology in Linguistics



"Tank"

[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

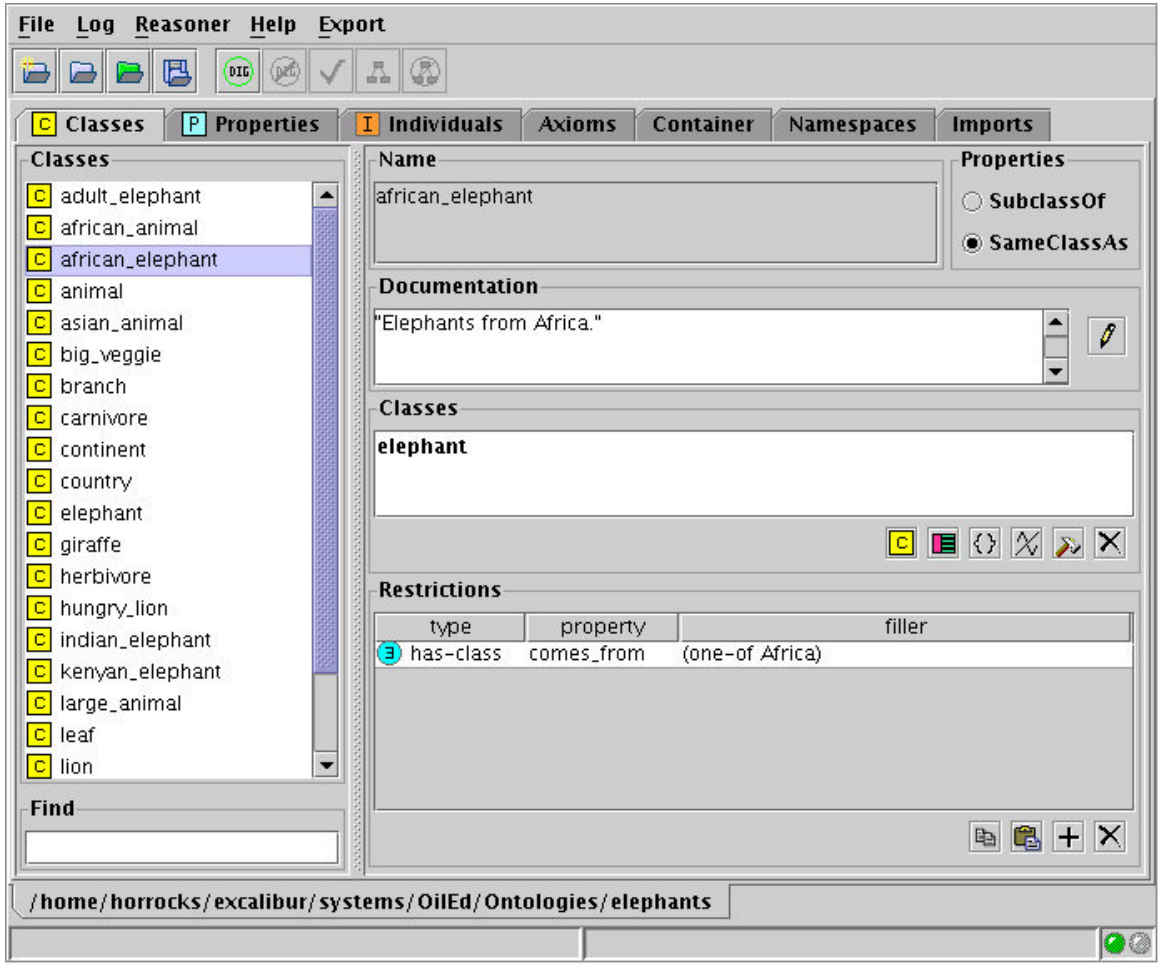
"An explicit specification of a conceptualisation" [Gruber93]

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

Example Ontology



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

Ontology Languages

- Wide variety of languages for “Explicit Specification”
  - Graphical notations
    - Semantic networks
    - Topic Maps (see http://www.topicmaps.org/)
    - UML
    - RDF
  - Logic based
    - Description Logics (e.g., OIL, DAML+OIL, OWL)
    - Rules (e.g., RuleML, LP/Prolog)
    - First Order Logic (e.g., KIF)
    - Conceptual graphs
    - (Syntactically) higher order logics (e.g., LBase)
    - Non-classical logics (e.g., Flogic, Non-Mon, modalities)
  - Probabilistic/fuzzy
- Degree of formality varies widely
  - Increased formality makes languages more amenable to machine processing (e.g., automated reasoning)

Many ontology languages use "object oriented" model based on

- Objects/Instances/Individuals
  - Elements of the domain of discourse
- Types/Classes/Concepts
  - Sets of objects sharing certain characteristics
- Relations/Properties/Roles
  - Sets of pairs (tuples) of objects
- Such languages are/can be:
  - Well understood
  - Formally specified
  - (Relatively) easy to use
  - Amenable to machine processing

Description Logics are a family of such ontology languages