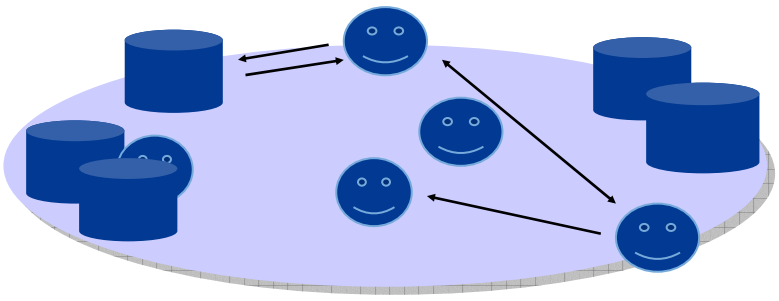
Software Design and Development in Distributed and (Very) Heterogeneous Environments – the Case of Semantic Web

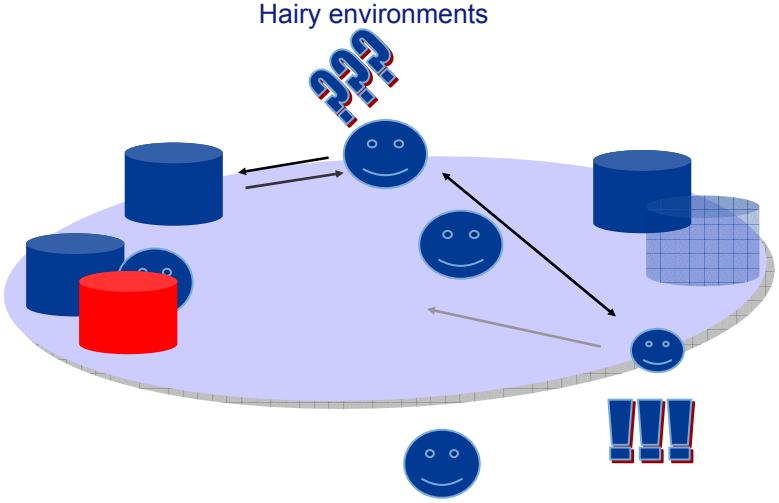
Santtu Toivonen
VTT Information Technology

Ohjelmistojen suunnittelumenetelmät ja –työkalut
22.11.2004

Distributed environments



Hairy environments



22.11.2004

Hairy environments: some concepts

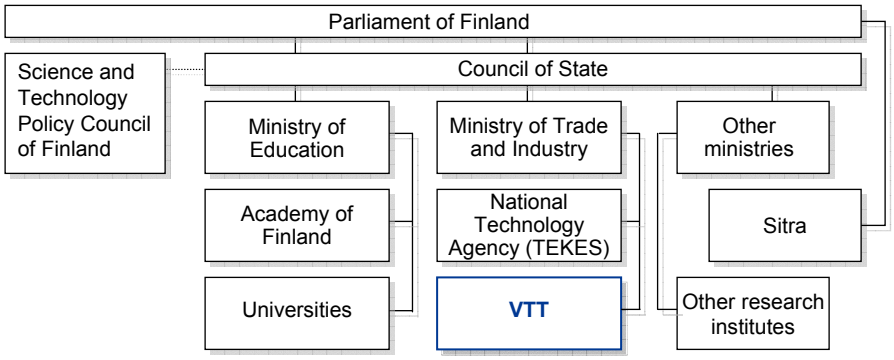
- (Semantic) Web
- **Ontology**
- Agent
  - Software agent
  - Human agent
- Web Services
- Reasoning
- Context
- Mobility
- Communication
- Interoperability, collaboration
- Openness
- Security, privacy, trust
- (Pick your favorite AI/CS problem)

22.11.2004

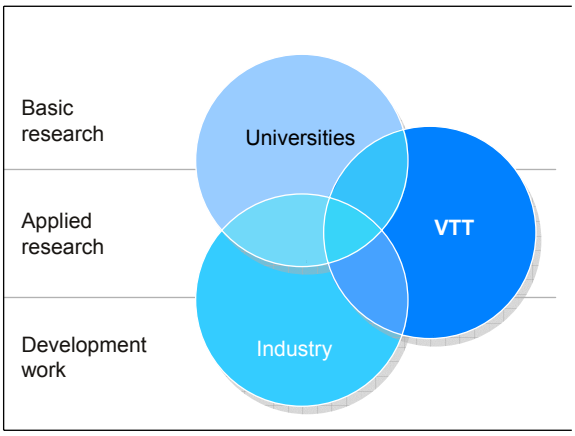
How did I end up in this mess?

- Education
  - First theoretical philosophy major, then cognitive science
  - Master's thesis on the definition of software agent
- Work
  - Sonera Research
  - VTT Information Technology
- Interests
  - Agent communication
  - Ontologies for agents
  - Distribution of cognition
  - Philosophy of mind

VTT in a nutshell



VTT in a nutshell (contd.)



VTT in a nutshell (contd.)

- < 3000 employees in VTT
- Six Research Institutes in VTT
  - VTT Electronics
  - **VTT Information Technology**
  - VTT Industrial Systems
  - VTT Processes
  - VTT Biotechnology
  - VTT Building and Transport
- > 400 employees in VTT Information Technology
  - Microelectronics
  - Microsensing
  - Telecommunications
  - **Networks**
  - Information Systems
  - Media
  - Human Interaction Technologies (in Tampere)

VTT in a nutshell (contd.)

- From the Networks web site: “We research Internet and Web technologies. We generate development environments required by Internet applications and service platforms. Moreover, we provide services for device manufacturers, Web operators, service providers and content producers.”
- Five research groups in Networks
  - **Service Platforms**
  - Multimedia
  - Telecommunications Software
  - Mobility (→ Security)
  - Wireless Systems

VTT in a nutshell (contd.)

- The role of VTT has changed in its 60+ years
  - From tailored tests (“VTT Testaa”) to longer term projects
- Main ways of doing research
  1. Jointly funded projects
     - **TEKES**
     - EU
  2. Direct assignments
  3. Self-financed projects (from government, Ministry of Trade and Industry

The burden of pleasing everybody

- In a typical TEKES funded research project the funding comes from
  - One or more research partners (VTT, universities)
  - Two or more industry partners
  - TEKES itself

→ All have their own objectives
→ Everybody's interests have to be taken into account
  - Resarch partners: theoretical (for example graduate studies)
  - Industry partners: ROI
  - TEKES: percentually enough industry funding, international and national research cooperation, contentually challenging, etc.

Past and present projects enabling my work in this area

- Kontti (Context-aware Services for Mobile Users)
  - 2002 - 2003
  - http://www.vtt.fi/tte/projects/kontti/
  - http://www.vtt.fi/inf/pdf/publications/2004/P539.pdf
- PAX (Process Descriptions for Agents in XML)
  - 2003 - 2004
  - http://www.vtt.fi/tte/proj/pax/
- DYNAMOS (Dynamic Composition and Sharing of Context-aware Mobile Services)
  - 2004 →
  - http://www.vtt.fi/tte/proj/dynamos/

Doing software research vs. developing software products

- Time span
- **Goals**
- Funding
- Testing
- Target audience, customers
- Colleagues
- **Success factors**, merits
- Project phases
- IPRs, patents, ownership questions in general

→ *Impact on the software design, development, and tools!*

What is the Semantic Web?

- "The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation."
  - Tim Berners-Lee, James Hendler, and Ora Lassila (2001): The Semantic Web. In *Scientific American*, May 2001. http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21
  - Extension
  - HCI
  - Machine-understandability

Need for ontologies (or something alike)

WWWeb is a distributed system
SWeb is an extension of the WWWeb

SWeb is a distributed system

- Parts of a distributed system need to interoperate (otherwise they are not in a same system?)
- In SWeb the interoperation is defined in the “meaning level” or “knowledge level”* (semantics, machine-understandability)
- This presupposes mutual understanding and shared knowledge, which is achieved via **ontologies**

* Allen Newell (1982): The Knowledge Level. *Artificial Intelligence* 18(1): 87-127.

What is an ontology?

- Webster:
  - Main Entry: on·tol·o·gy
  - Pronunciation: än-'tä-l&-jE
  - Function: *noun*
  - Etymology: New Latin *ontologia*, from *ont-* + *-logia* -logy
  - Date: circa 1721
  - **1** : a branch of metaphysics concerned with the nature and relations of being
  - **2** : a particular theory about the nature of being or the kinds of existents

http://www.m-w.com/home.htm

What is an ontology (contd.)?

- Tom Gruber:
  - Short answer: An ontology is a specification of a conceptualization.
  - […] an ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents.

http://www-ksl.stanford.edu/kst/what-is-an-ontology.html

What is an ontology (contd.)?

- Nicola Guarino:
  - […] in AI, an ontology refers to an engineering artifact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words.

http://citeseer.nj.nec.com/guarino98formal.html

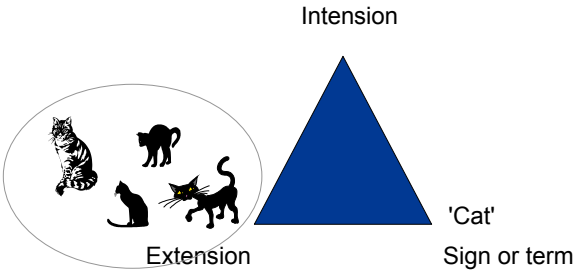
Ontologies vs. data models

- No strict line in between, but ontologies are
  - more general
  - more reusable
  - intended for multiple purposes, goals, and users
  - more easily shareable
  - take stand on semantics of concepts (as opposed to mere structure and integrity)

What is a concept?

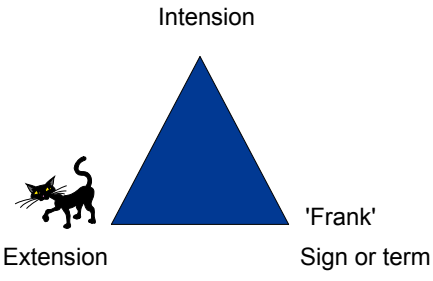
- Concepts (among other things)
  - are in general language-independent (words 'cat' and 'kissa' denote the same concept)
  - are mental or logical representations of reality
  - are related to other concepts
  - do not need symbols but hold them for means of communication
- A concept has
  - **intension** or meaning
  - **extension**, i.e. the set of objects that the concept refers to
- On the difference between intension and extension, consider phrases "Evening star" and "Morning star" that have different meanings (intension) yet both refer to planet Venus (extension)
- *Ontology is mainly concerned with intension*

"Semiotic Triangle" (or one version of it)



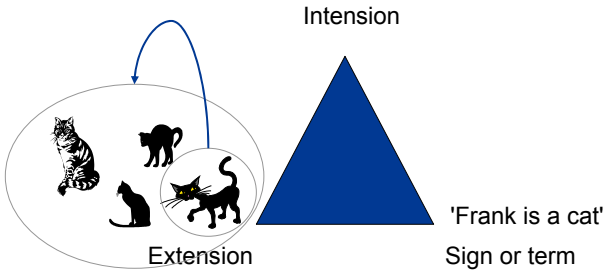
22.11.2004

"Semiotic Triangle" (contd.)

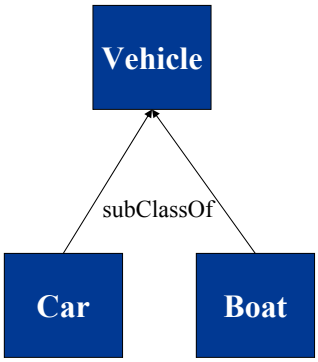


22.11.2004

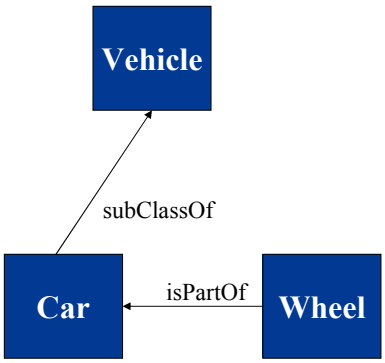
"Semiotic Triangle" (contd.)



A simple example ontology

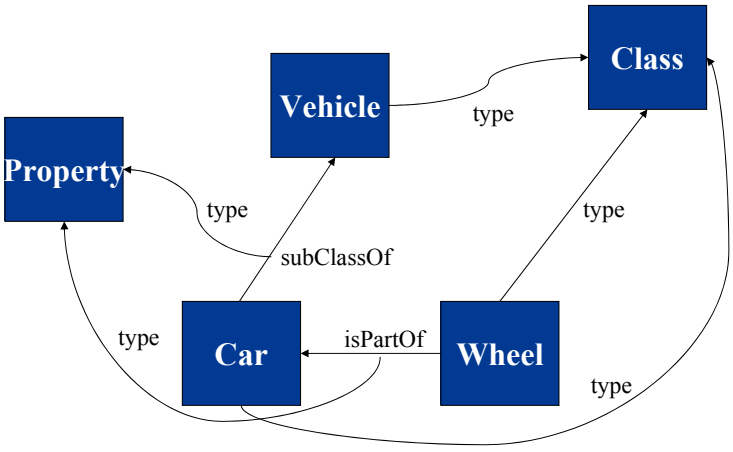


Almost as simple

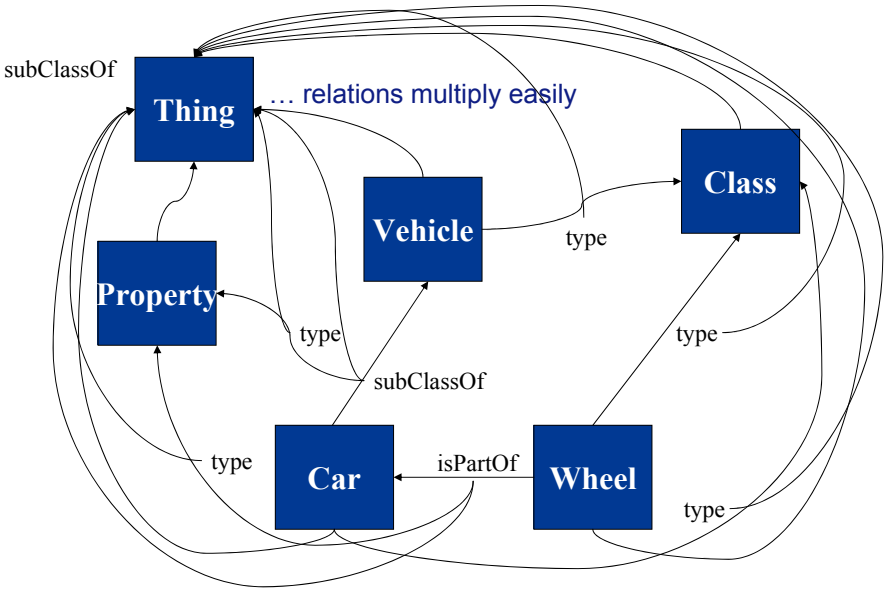


22.11.2004

Gets a bit tricky...



22.11.2004



22.11.2004

Important concepts wrt. ontology

- Concepts
  - abstract or concrete, elementary or composite, real or fictitious
  - can have attributes (properties)
- Taxonomy (typically at least subclass-relation)
- Other relations
  - part of, connected to, on top of, …
  - unary vs. binary vs. n-ary relations
- Axioms (sentences that are always true)
  - axioms typically built using logic (first or higher order)
- Instances, Individuals, Facts, Claims
  - extension
- Production rules and inference mechanisms

22.11.2004

Ontology representation languages in the Semantic Web

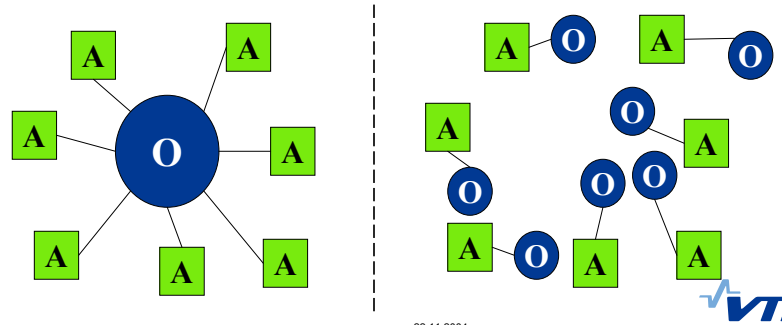
- RDF (Resource Description Framework)
  - Triples: Subject, Predicate, Object
- RDFS (RDF Schema)
  - Vocabulary definitions for RDF (cf. DTD and XML Schema for XML)
  - subClassOf, type, domain, range, ...
- DAML (DARPA Agent Markup Language)
- OIL (Ontology Inference Layer or Ontology Interchange Language)
- OWL (Web Ontology Language)
- DAML, OIL → DAML+OIL → OWL

OWL

- OWL builds on the work of DAML+OIL
  - DAML+OIL an extension of RDF Schema
  - W3C Recommendation since 10.2.2004
- Inconsistencies with RDF(S) and DAML+OIL → problems with the layered approach presented earlier → three versions of OWL:
  - OWL Full
    - Union of OWL and RDF(S)
  - OWL DL (Description Logics)
    - Restricted to DL/FOL fragment (≈ DAML+OIL)
  - OWL Lite
    - Subset of OWL DL

Size and scope of an ontology

- Two extremes (the reality something in between):
  - One huge ontology O that captures "everything"
  - One (small) ontology for each specific application A



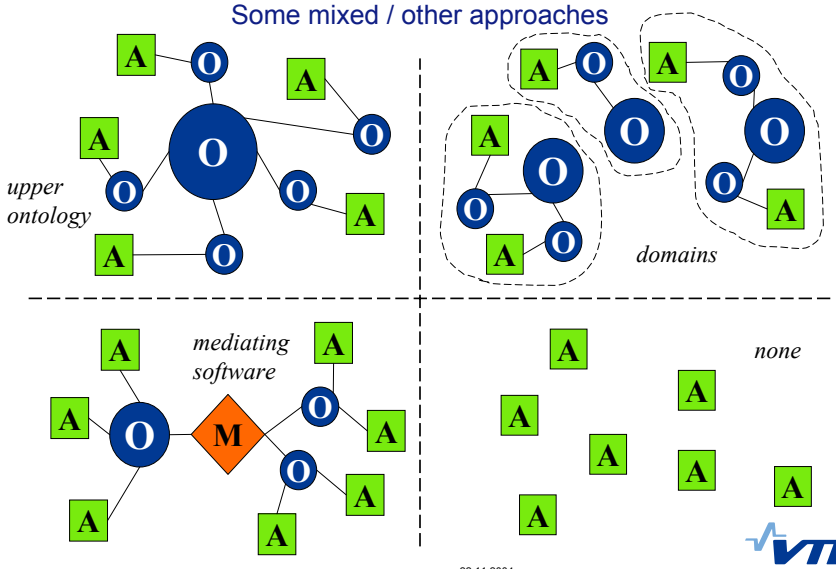
"One large ontology" approach

- Benefits
  - few or no internal inconsistencies
  - for an application developer easier to find
  - uniform documentation
  - less overlapping work!
- Drawbacks
  - who maintains it?
  - who is responsible?
  - heavy and slow to use (both for human user and for application)
  - difficult to take into account everybody's opinions and wishes
- Example: Cyc

"Several small ontologies" approach

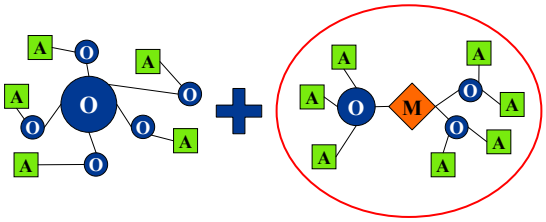
- Benefits
  - ontologies fit well the application demands
  - faster to use
  - easier to form complete picture of an ontology (fewer concepts and interrelations)
- Drawbacks
  - different ontologies do not fit together without either
    - central coordination body or
    - ontology mapping software
  - overlapping work - same concepts defined in multiple ontologies, either in the same way or (even worse!) differently

Some mixed / other approaches



A suitable approach for the Semantic Web?

- "Domains"-approach might not work
- ISWC 2004 Proceedings is full of ontology alignment software, concept matching, web service integration, etc.
- So…



Ontology management

- Individuals and instances change over time, whereas ontologies are intended to be relatively stable and static
- but…
- → ontology **management** is needed
- Issues on interdependencies
  - upper ontology ← domain ontologies
  - domain ontology ⇔ domain ontology
  - ontology ← applications
- Approaches for cross-usage of concepts defined in multiple ontologies
  - **integration** (take existing ontologies and build a new one)
  - **merge** (unify existing ontologies)
  - **mapping software** ("runtime" approach) M

Ontology management (contd.)

- Ontology **versioning** an important aspect
- Five ways for an ontology to change
  1. Changes in the conceptual hierarchy (logical changes)
  2. Changes in the documentation of the concepts (non-logical changes)
  3. Renaming concepts (identifier changes)
  4. Defining new concepts
  5. Deleting concepts
- **Prospective** usage: using data sources that conform with the previous version of the ontology via the new version
- **Retrospective** usage: using data sources that conform with the new version of the ontology via the previous version

Ontology management (contd.)

- Compatibility types
  1. Complete compatible revisions
     - the semantics of the ontology is not changed, for example, **syntactic** changes or updates of natural language descriptions; this type of change is compatible in both prospective and retrospective use
  2. Backward compatible revisions
     - the semantics of the ontology are changed in such a way that the interpretation of data via the new ontology is the same as when using the previous version of the ontology, for example, the addition of an independent class; this type of change is compatible in **prospective use**
  3. Upward compatible revisions
     - the semantics of the ontology is changed is such a way that an older version can be used to interpret newer data sources correctly, for example, the removal of an independent class; this revision is compatible in **retrospective use**
  4. Incompatible revisions
     - the semantics of the ontology is changed in such a way that the interpretation of old data sources is invalid, for example, changing the place in the hierarchy of a class; this type of change is incompatible in both prospective use and retrospective use

What is needed in addition to ontologies?

- Ontologies are the building blocks of the SWeb, but shared meanings only enable interoperability, nothing else
- Need for collaboration between various parties in the SWeb
  - Pragmatics vs. semantics (even if a meaning is shared, it can be used in various ways → need for defining conversation patterns or interaction protocols*)
  - Game theories, maximizing utility
  - Trust, security, privacy, certificates, etc.
  - HCI & usability research, policy definitions

* S. Toivonen and H. Helin. Representing Interaction Protocols in DAML. In van Elst, L., Diagnum, V., and Abecker, A. (Eds.): *Agent-Mediated Knowledge Management: Selected Papers from AAAI 2003 Spring Symposium, volume 2926 of Lecture Notes in Artificial Intelligence*, Berlin, Germany, January 2004, pp. 310--321.

What is needed in addition to ontologies (contd.)?

- The Internet is expanding beyond PCs with wired connections
  - Entails the Web(s)
- Mobility, limited devices and varying connection types
- Context-awareness
  - Location and time
  - User activity
  - User's state of mind
  - Social context
  - Environment (e.g., weather conditions)
  - Device characteristics
  - Network QoS

What is needed in addition to ontologies (contd.)?

- Tools and applications!
  - Critical mass needed
- Jena (HP Labs Bristol)
  - http://www.hpl.hp.com/semweb/jena2.htm
  - Java API for SWeb applications
  - APIs for RDF and OWL, RDQL, in-memory and persistent storage
  - Open source
- Protégé (Stanford University)
  - http://protege.stanford.edu
  - Ontology editor
  - Older than SWeb, plugin-based development
  - Open source

Further information

- Contact:
  - santtu.toivonen@vtt.fi
  - http://www.vtt.fi/tte/staff/tos/cv.html
- Some links
  - ISWC 2004: http://iswc2004.semanticweb.org
  - Annotea: http://www.w3.org/2001/Annotea/
  - FOAF: http://www.foaf-project.org/
  - Sesame: http://www.openrdf.org/
  - MuseoSuomi: http://museosuomi.cs.helsinki.fi/
  - Bibster: http://bibster.semanticweb.org/
  - Haystack:: http://haystack.lcs.mit.edu/
  - Task Computing: http://www.taskcomputing.org/
  - SIR: http://www.profium.com/
  - Swoogle: http://pear.cs.umbc.edu/swoogle/index.php