

Esipuhe

Tämä dokumentti sisältää kaksi esitystä/alustusta Tampereen teknillisen yliopiston (TTY) Hypermedialaboratorion järjestämään syksyn 2006 Mindtrek-tutkijatapahtumaan:

- 1.Sovelluskehitystä vai käsiteanalyysia: ontologioiden filosofiset perusteet ja tekniikka
- 2.Semanttisesta Webistä asiantuntijajärjestelmiin

Esitysten tarkoituksena on lyhyesti esitellä ja tehdä ymmärrettäväksi ontologioiden taustaa, (teknisiä) ideoita ja sovelluksia tietoverkkojen tutussa kontekstissa (*jos ja kun myös yksityiskohdat kiinnostavat, hae sopivan hakukoneen avulla verkosta lisätietoja tekstissä esiintyvien avainsanojen avulla*)

Ossi Nykänen
ossi.nykanen@tut.fi (ossi@w3.org)



#1. Sovelluskehitystä vai käsiteanalyysia: ontologioiden filosofiset perusteet ja tekniikka

Mindtrek-tutkijaseminaari "Ontologiat ja loogiset kuvailutiedot Web-
sovelluksissa digitaalisten ekosysteemien kontekstissa"
10.11.2006 Tampere

Ossi Nykänen
ossi.nykanen@tut.fi

TTY/ Hypermedialaboratorio
<http://matriisi.ee.tut.fi/hypermedia/>



Johdanto

Suomen kielen sivistyssanakirjasta (Nurmi, Rekiaro & Rekiaro 1994) voimme lukea seuraavaa:

ontologia *fil.* **1** oppi olevaisen perusominaisuudesta **2** metafysiikka

Onko tästä tiedosta meille mitään hyötyä?

Eryteisesti: Miten termi (käsite) ontologia eroaa esim. termistä (käsitteestä) **oologia**?

Ilmeisestikin meidän tulisi paitsi tietää miten ko. termiä käytetään kielessä, myös mitä se tarkoittaa käsitteenä suhteessa muihin ymmärtämiimme käsitteisiin ja mitä kaikkea tällä tiedolla voisi ylipäänsä tehdä

Seuraavan alustuksen tavoitteena on päivittää henkilökohtaista ontologiaamme käsitteen ontologia osalta (tietotekniikan kontekstissa)...



OSA 1/2: Ontologiat historian hämärässä

(Luvassa on myös osa 2/2: Tietotekniikkaa & sovellettuja ontologioita)

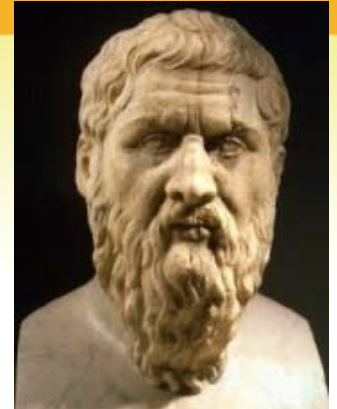


Länsimaisen ontologiakäsitteen lyhyt historia

...

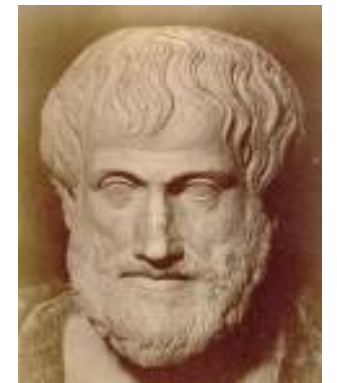
Platon 427-347 eKr.

- "Ideaoppi" (näennäinen-todellinen, ideat tosiolevana)
- "Universaalin" käsite (tieto kohdistuu universaaleihin, ei yksittäisiin olioihin)



Aristoteles 384-322 eKr.

- Platonin ideaopin kritiikki → Aristoteleen "kokonaisfilosofia"
- Teos "Metafysiikka" (toimitettu jälkeempään luentomuistiinpanojen perusteella n. 100 jKr(?); erittäin vaikeaselkoisena pidetty teos)
- Ideoiden suhde aistittaviin olioihin → "tieteellinen" lähestymistapa ja erit. syyn analyysi



...

Aristoteleen kategoriat (Categoriae)

”Jokainen malliesimerkillinen väitelause koostuu subjektista ja predikaatista. Subjekti on ilmaus joka nimeää sen josta on puhe. Predikaatti puolestaan ilmaisee mitä subjektista sanotaan”

Aristoteleen mukaan kaikki väitelauseiden termit voidaan luokitella kymmeneen kategoriaan:

1. Substanssi eli olio (esim. hevonen)
2. Kvantiteetti (esim. neljä)
3. Kvaliteetti (esim. ruskea, vauhkoontunut)
4. Relatiivi (esim. pienempi kuin, ladon sisällä [Huom. ml. tieto ja havainto!])
5. Paikka (esim. kylä)
6. Aika (esim. tänään)
7. Asento (esim. satulolituna)
8. Jollakin oleminen (esim. isännän omistuksessa)
9. Tekeminen (esim. syömässä heinää)
10. Tekemisen kohteena oleminen (esim. syötävä heinä)

Huom. (Kategoriat näyttävät koskevan sekä **kieltä**, **ajattelua** että **todellisuutta** ja A:n logiikka oli rajoittunutta (**S-O-P**) [Organon])



Ontologian käsite

Termi "ontologia" (latinankielinen termi *ontologia*) lienee ensi kertaa löydettävissä kirjallisuudesta vuodelta 1606

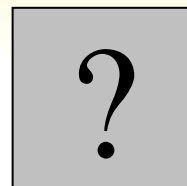
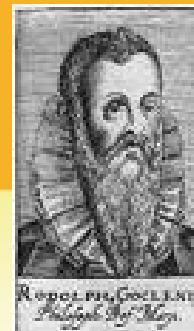
- Rudolf **Göckel** (1547-1628, lat. Rudolf Goclenius), saksalaisen Marburging yliopiston logiikan professori
- Jacob **Lorhard** (1561-1613, lat. Jacobus Lorhardus), sveitsiläisen St. Gallen yliopiston professori

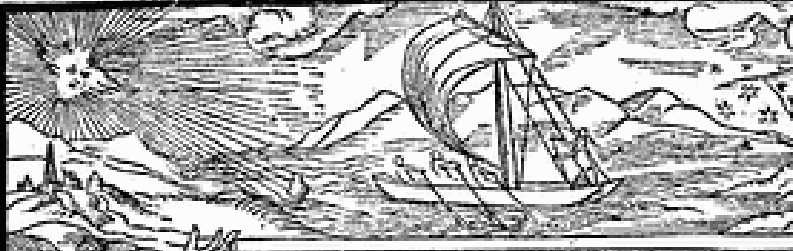
Johannes Clauberg (1647):

- 'Since the science which is about God calls itself *Theosophy* or *Theology*, it would seem fitting to call **Ontosophy** or **Ontology** that science which does not deal with this and that being, as distinct from the others owing to its special name or properties, but with being in general.'

Christian Wolff (1730):

- "*the science of the most general and abstract predicates of anything*"





OGDOAS
SCHOLASTICA

CONTINENS
Diagraphen Typicam
Arithm.

- Grammatices } Latinae.
Graecae.
- Logices.
- Rhetorices.
- Astronomicas.
- Ethicas.
- Physicas.
- Metaphysicas, seu Ontologia.

Ex Praesentium hujus temporis viciorum locutionibus,
Pro Doctrinae & virtutum studio juvenutis
CONFECTA

Jacobo Lorhardo, Gymnasij Sangal-
lensis Rectore, & in Ecclesia
Christi sacro:

APVD GEORGIUM STRAVB
Sangallii ANNO. 1606.



METAPHYSICAE SEV ONTOLO-
giae Diagraphæ.

Λόγος: Intelligibile dicitur omne, quod
intellectu percipi ac comprehēdi potest.

Metaphysicæ
(.qua est Πρῶ-
τη ἢ ἄλλη,
quatenus ab
homine natu-
ralitatis
lumine sine
ullo materiae
conceptu est
intelligibile.)
partes sunt
duae, Altera

Vniversa-
lis, qua cō-
sistit in
Intelligi-
bilium &
Entium
tūm

Generalissi-
mis distribu-
tionibus. In-
telligibilis
autem notetur

Distributio: Intelligibile est vel

Nihil: hoc simpliciter non est
aliquid.

Aliquid:
quod sim-
pliciter
non est
nihil. Est
g. vel

Positivum,
quod ponit
seu affir-
mat quid-
piam. Est g.
vel

Essen-
tia.
Vide
A.
Ens.
Vide
B.

Negativū: Vide B.B.

Communitissimis attributis: Vide C.

Particularis: Vide E.E.

“The first book on Ontology
Jacob Lorhard (1561-1609) –
Ogdoas Scholastica (1606)”

(<http://www.formalontology.it/>)

Metafysiikka ja ontologia

1600-luvulla termejä metafysiikka ja ontologia käytettiin nykykäsitteiden mukaan aluksi synonyymeina (~olevaisen luokittelu)

1700-luvulla termi ontologia muodostui ”tieteelliseksi instrumentiksi” (kaikentyypisen) tiedon käsittelyyn, osin vastareaktionä Aristoteleen perinnön ”painolastille”

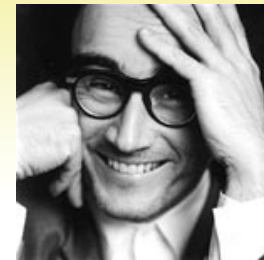
Ontologiakäsitys on sittemmin historiallisesti kehittynyt käsi kädessä tieteen ja tekniikan kanssa



Termit nykyään: metafysiikka (suomeksi)

Metafysiikka

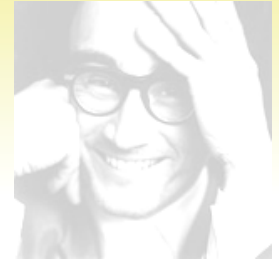
- *"Filosofian osa-alue joka tutkii **olevaa** sellaisenaan. Oppi todellisuuden ylemmistä periaatteista, rakenteesta ja koostumuksesta"* (Saarinen 1985)
- *"(Yleensä) metafysiikalla tarkoitetaan maailman tai olevaisen **alkuperää** ja **perusluonnetta** koskevia yleisiä oppeja kuten materialismi, spiritualismi, dualismi, jne. ...toisaalta ne voidaan omaksua pelkästään metodologisina periaatteina"* (tai metafysiikan käsite voidaan ymmärtää laajemmassa, *empiirisesti testattavassa* merkityksessä) (Niiniluoto 1984)



Termit nykyään: ontologia (suomeksi)

Ontologia

- *”Olemassaolo-oppi; filosofian osa-alue joka tutkii **olemisen käsitettä** ja perimmäistä luonnetta. Samaistetaan joskus metafysiikan kanssa.”* (Saarinen 1985)
- *”Kysymys **yleiskäsitteiden** olemassaolosta kuuluu ontologiaksi kutsuttuun filosofian osaan, jossa tarkastellaan todellisuuden peruselementtejä ja rakennetta.”* (Niiniluoto 1984)



(Pyhä) Yksinkertaistus

Metafysiikka tarkastelee *olevaa* sinänsä

Ontologia tarkastelee käsitteitä (ja kieltä) jolla olevaa on mahdollista *kuvaila loogisesti* (→**kuvailutieto**)

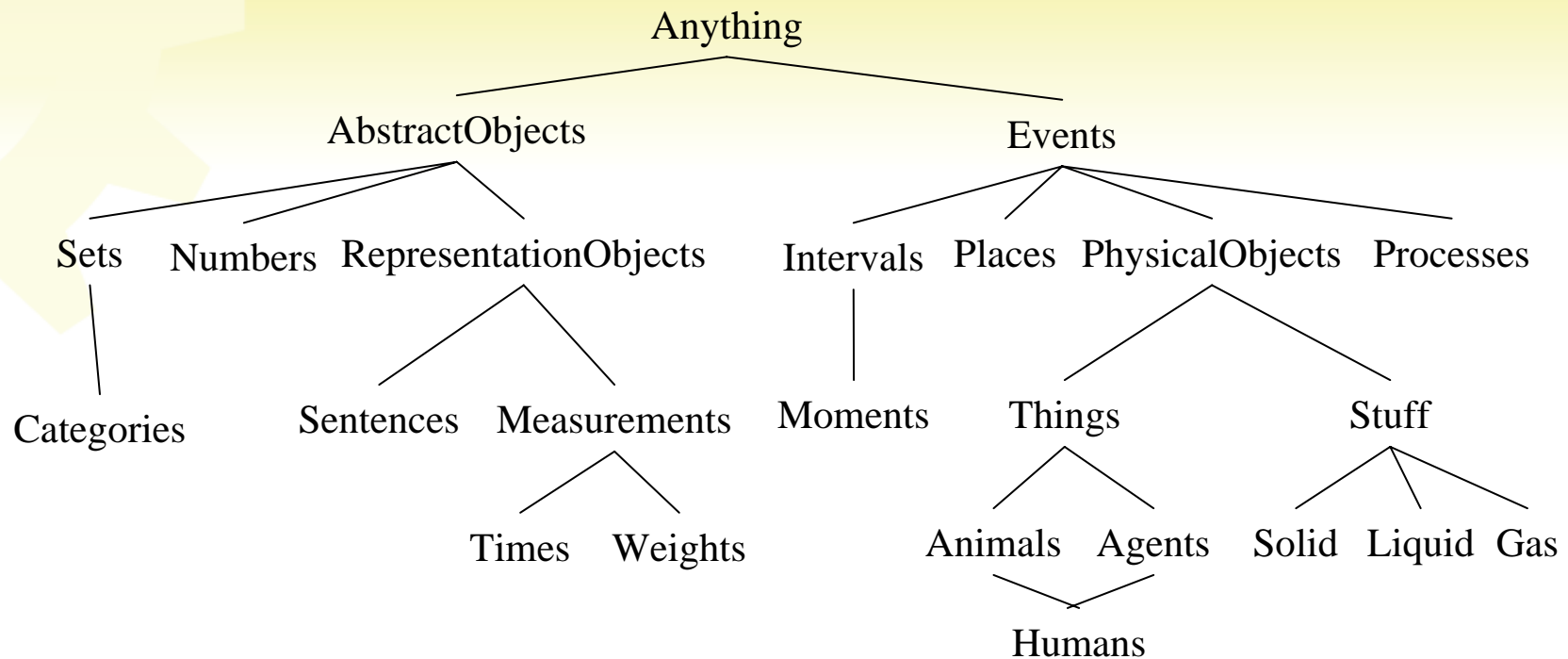
...eli metafysiikka tarvitsee ontologioita (ainakin kommunikoinnin näkökulmasta) mutta ontologioiden ei tarvitse olla ”metafysiikkaa”



OSA 2/2: Tietotekniikkaa & sovellettuja ontologioita



”Klassinen” esimerkki: top-level ontology of the world (vain osa näkyvillä) (Russel & Norvig 1995)



”Useimmat” käytännön ontologiat ovat sovellusalueeltaan huomattavasti rajatumpia!

(Tiedon) mallinnus ja metafysiikka

Tieteen historia opettaa että **hyvä malli on käyttökelpoinen abstraktio** kuvailtavasta ilmiöstä

Kun tavoitteena on tiedon soveltaminen, esiintyvät tietomallit poikkeuksetta suhteessa **tekemiseen** (mukaan lukien kommunikointi)

Esimerkkejä

- Fyysikko ei yleensä usko että kaava $v(t) = v_0 + at$ mallintaa putoavan kiven loppunopeuden *realistisesti*, vaan **sopivalla tarkkuudella**
- Matemaatikko ei yleensä pohdi onko lukua $\sqrt{2}$ *olemassa*, vaan miten se **toimii** osana laskusääntöjä
- Informaatikko ei yleensä usko *määrittelevänsä* runoutta sijoittaessaan Juice Leskisen kirjan *Äeti* kirjastoluokkaan 82.2 (Suomenkieliset runot), vaan helpottaa kirjan **löytämistä**



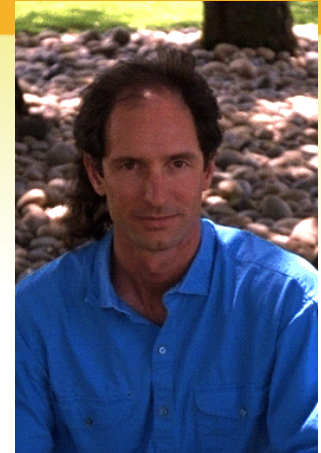
Ontologiat tietotekniikassa (1991-*)

Suosittu (informaali) määritelmä (tiedon jakamisen kontekstissa):

- *"Ontology is a **specification** of a **conceptualisation**."* (Gruber 1993)

Sittemmin yleiseen käyttöön on vakiintunut hieman laajennettu määritelmä

- *"Ontology is a formal, explicit specification of a shared conceptualisation"*



Mitä ja kenelle?

Esitettyjä sovellusalueita:

- Knowledge engineering, artificial intelligence, computer science, knowledge management, natural language processing, e-commerce, intelligent integration information, information retrieval, database design and integration, bio-informatics, education, Semantic Web (Gómes-Pérez, Fernández-López & Corcho 2004):

Korkean tason käyttötapauksia

- Tiedon ulkoistaminen ja reflektio
- Yhteensopivuus deklarativisen tiedon tasolla
- Korkean tason uudelleenkäyttö
- Semanttinen haku
- Sovellusten ja tehtävien rajat ylittävä päättely (yhdessä ongelmanratkaisumenetelmien kanssa (Problem Solving Methods, PSM))



Ontologioiden mallintaminen ja soveltaminen

Ontologia esiintyy joko suhteessa **luonnolliseen kieleen** tai erityiseen **mallinnuskieleen** (joskaan sen ei välttämättä tarvitse olla ”erityinen ontologiakieli”)

Ontologioiden käytännöllisiin sovelluksiin liittyy aina yhteisöllisen sitoutumisen vaihe:

- *”Ontological commitment is a function that links the terms of the ontology vocabulary with a conceptualisation.”* (Guarino, 1998)
- Vrt. formaalisen teorian tulkinta

Hyvä ontologia ei sano (maailmasta) enempää kuin on välttämätöntä

Ontologiatyön merkittävimmät käytännön haasteet liittyvät lähes poikkeuksetta (”ei-tekniseen”) määrittelyyn ja yhteisölliseen sitoutumiseen (prosessit ja käytännöt)



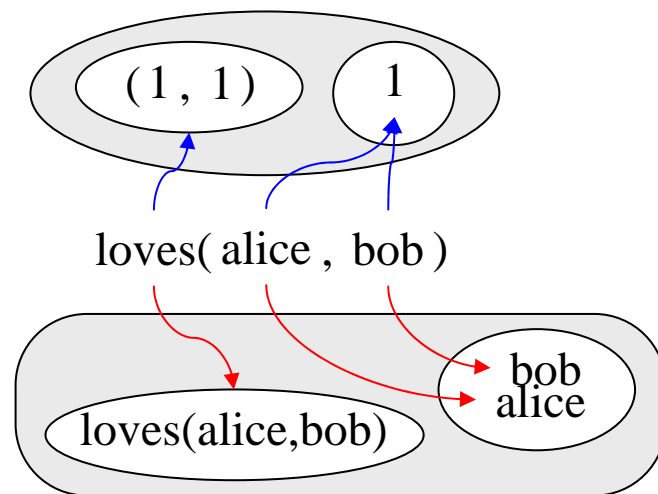
Semantiikka vai "semantiikka"?

Syntaksilla tarkoitetaan sovitun kielen koodaamisessa tarvittavia rakenteita (merkitö, termit, lauseet, "laatikot", "ovaalit", "nuolet", jne.)

Semantiikalla tarkoitetaan *formaalisten kielten teknisessä kielenkäytössä kuvausta* jossa sovitun kielen kukin (mielekäs) termi/lause/jne. **viittaa** (täsmällisesti) jonkin *tunnetuksi oletetun sovellusalueen* käsitteeseen

Semantiikan käsitteeseen liittyy lukuisia väärinkäsityksiä, mm. (Harel & Rumpe 2004):

- Semantics is the metamodel (in UML)
- Semantics is the semantic domain
- Semantics is the context conditions
- Semantics is dealing with behaviour
- Semantics is being executable
- Semantics is the behaviour of a system
- Semantics is the meaning of individual constructs
- Semantics means looking mathematical
- Semantics is _____ (e.g. given by message-passing)



Formaalisten kielten näkökulmasta ontologia voidaan ahtaasti ymmärtää "vain" formaalin teorian **sanastona**; tai "laajimmillaan" kokonaisuena **formaalisena teoriana** (joskin **toisinaan** tulkintafunktio "ajatellaan määritellyksi vain" ns. [kirjoittajan ajatteleman] **luonnollisen tulkinnan** kautta)

Erityyppisiä ontologioita

Tietosisällön (ja käytön) mukaan ontologioita voidaan jaotella esim. seuraavasti (Gómes-Pérez, Fernández-López & Corcho 2004):

1. Knowledge representation (KR) ontologies
2. General or common ontologies
3. Top-level or Upper-level ontologies
4. Domain ontologies
5. Task ontologies
6. Domain-Task ontologies
7. Method ontologies
8. Application ontologies

Sovellus voi tietenkin yhdistellä erityyppisiä ontologiakomponentteja (esim. hyödyntää yleistä päättelyontologiaa sovelluskohtaisen tehtävän suorittamiseen)



Lopuksi

Ok; ontologiat sitten ovat tällaisia: *Entäs sitten?*

Emme jää (tähän) munimaan käsiteanalyysin tasolle vaan jatkamme aiheesta myöhemmin iltapäivällä, kahvitauon jälkeen...

Ai niin:

oologia = linnunmunien tutkimus, linnunmunatiede



#2. Semanttisesta Webistä asiantuntijajärjestelmiin

Mindtrek-tutkijaseminaari "Ontologiat ja loogiset kuvailutiedot Web-
sovelluksissa digitaalisten ekosysteemien kontekstissa"
10.11.2006 Tampere

Ossi Nykänen
ossi.nykanen@tut.fi

TTY/ Hypermedialaboratorio
<http://matriisi.ee.tut.fi/hypermedia/>



Johdanto

Semanttinen Web (SW) voidaan nähdä teknologiana erilaisten (Web-käyttöön suunniteltujen) tietointensiivisten sovellusten toteuttamiseen

Seuraavaksi

- luomme yleiskatsauksen tunnettuihin ontologioihin,
- tarkastelemme SW-perustekniikkaa ontologioiden näkökulmasta ja
- pohjustamme tietojenkäsittelyn periaatteita tiedon tasolla

On hyvä panna merkille että integroituihin sovelluksiin liittyy melko paljon **avoimia kysymyksiä** (joista vain osa on ”teknisiä”)



OSA 1/3: Pikakatsaus erilaisiin ontologioihin



Tunnettuja ”isoja” ontologioita (Gómes-Pérez, Fernández-López & Corcho 2004)

KR ontologies

- Frame Ontology (FO)
- Open Knowledge Base Connectivity (OKBC) Ontology
- RDF & RDF Schema
- Ontology Inference Layer (OIL)
- DAML+OIL
- OWL

Top-level ontologies

- Top-level ontology of universals
- Sowa's top-level ontology
- Cyc's upper ontology
- Standard upper ontology (SUO)

Linguistic ontologies

- WordNet
- EuroWordNet
- Generalized Upper Model (GUM)
- Mikrokosmos Ontology
- SENSUS

Domain ontologies/

Medical ontologies

- GALEN
- Unified Medical Language System (UMLS)
- ON9

E-commerce ontologies

- United Nations Standard Products and Services Codes (UNSPSC)
- North American Industry Classification System (NAICS)
- Standard Classification of Transported Goods (SCTG)
- E-cl@ss
- RosettaNet

Engineering ontologies

- EngMath
- PhySys

Enterprise ontologies

- Enterprise Ontology
- Toronto Virtual Enterprise (TOVE)

Chemistry ontologies

- Chemicals
- Ions
- Environmental pollutants

Knowledge management ontologies

- *Various types of information/domain/enterprise ontologies*
- Knowledge Acquisition (KA) ontologies (*organization, project, person, publication, event, research-topic, research-product*)
- Esperanto/R&D projects

Edellisten lisäksi ontologioita voidaan tietenkin ”löytää” lähes kaikkialta missä tietoa kuvataan



OSA 2/3: Semanttisen Webin ontologiat



Web-ontologioiden kuvailukieli Web Ontology Language (OWL)

Web-ontologioiden teknisenä perustana pidetään yleensä W3C:n **OWL** –suosituksia (Patel-Schneider, Hayes, Horrocks, etc. 2004)

- Taustalla Resource Description Framework (RDF) -tekniikka
- OWL on itsessään **KR-tyyppinen ontologia**
- ...mutta sen ensisijaisena tarkoituksena on tarjota **väline** esim. sovellusontologioiden määrittelyyn/kuvaamiseen
- Ks. <http://www.w3.org/2004/OWL/>

OWL-kontekstissa ontologia määritellään seuraavasti:

- *"An ontology formally defines a common set of terms that are used to describe and represent a domain. Ontologies can be used by automated tools to power advanced services such as more accurate web search, intelligent software agents and knowledge management."* (Heflin 2004)



OWL-kielen suunnittelukriteerit (Heflin 2004)

OWL-kielen suunnittelulle asetettiin seuraavia tavoitteita:

1. Shared ontologies
2. Ontology evolution
3. Ontology interoperability
4. Inconsistency detection
5. Balance of expressivity and scalability
6. Ease of use
7. Compatibility with other standards
8. Internationalisation

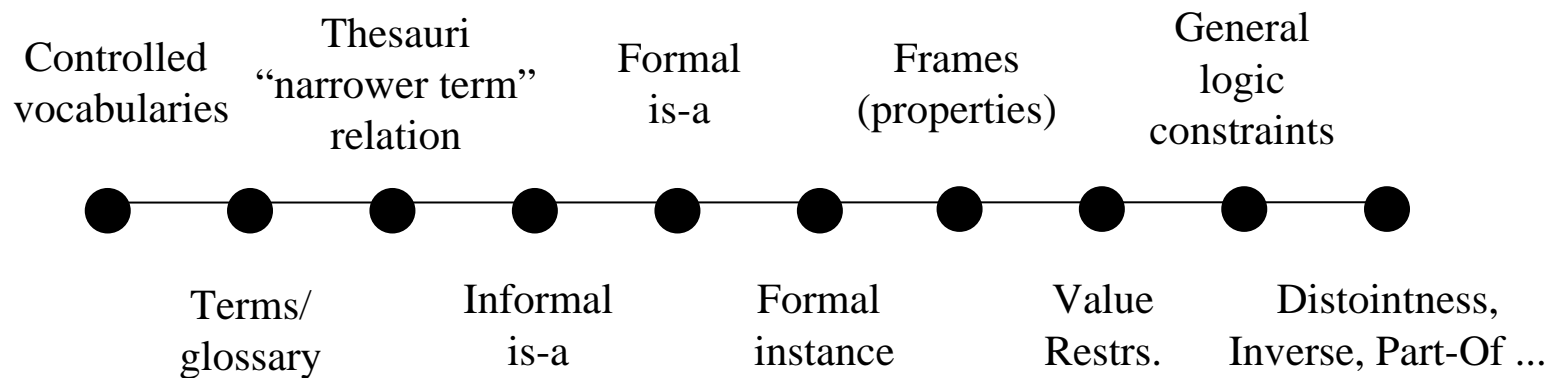
Keskeisiksi käyttötapauksiksi asetettiin vastaavasti:

1. Web portal
2. Multimedia collections
3. Corporate web site management
4. Design documentation
5. Agents and services
6. Ubiquitous computing



”Ontologioiden ravintoketju”

Lassila ja McGuinness (2001) ovat esittäneet erään ontologioiden ilmaisuvoimaan perustuvan erityyppisten ontologioiden kategorisoinnin:



OSA 3/3: Kohti asiantuntijajärjestelmiä



Asiantuntijajärjestelmä?

Asiantuntijajärjestelmällä (expert system) tarkoitetaan ideatasolla (Steels 1987):

- *tietokoneohjelmaa joka on suunniteltu auttamaan (ihmistä) jonkun todellisen, rajoitetun ja vaikean ongelman ratkaisussa*
- *...jonka päättely jäljittelee asiantuntijan päättelyä*
- *...jolla on sisäinen kuva tilastaan ja toiminnastaan*
- *...joka on suoraan käyttäjän käytettävissä*

Sama perusidea voidaan yleistää esim. erilaisiin avustaja- ja agenttisovelluksiin; nykyään suosittuja ovat esim. erilaiset **päätöksenteon tukijärjestelmät** (decision-support system)

- Periaatteellinen toteutustapa: ”logiikkaohjelmointi” (logic programming, LP)
- Toimiva lähestymistapa (esim. teollisuudessa) on valvoa jonkin prosessin etenemistä **sääntöjärjestelmän** tuella (esim. XCON)

Knowledge engineering (1980) → ontology engineering (2004) → ?



Tiedon esittäminen & päättely

Pelkkä kuvailutieto ei siis yksin riitä sovelluksissa; työn tueksi tarvitaan myös kättä pidempää, ”ongelmanratkaisuvälineitä”

Tekoälyn kontekstissa ratkaisua haetaan usein käsitteen **esitys- ja päättelyjärjestelmä** (vapaa suomennos) takaa:

- A **representation and reasoning system** (RRS) is [a system] composed of a language to communicate with a computer, a way to assign meanings to the language, and procedures to compute answers given input in the language (Poole 1998)
- Tunnettuja esimerkkejä: Assembler, ..., **Prolog**, ..., suomen kieli

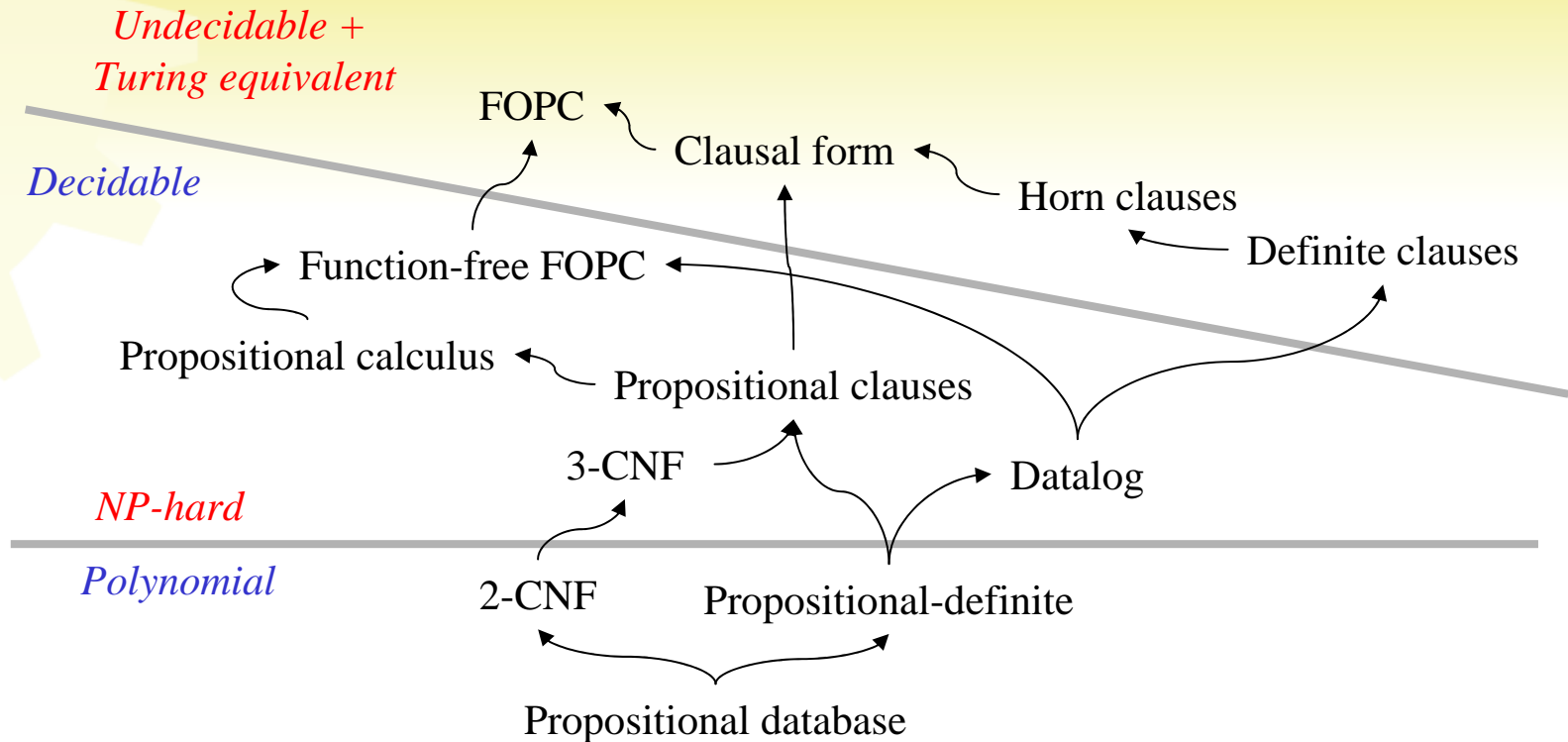


Semanttisen Webin yhteydessä pelkistetyksi työnjaoksi ajatellaan usein seuraavaa:

- RDF/OWL-ontologiat esittävät (staattisen) sovellustiedon (”**tietämyskanta**”)
- Päättely tms. hoidetaan **Semantic Web services** –teknologian välityksellä



Eräs logiikoiden luokittelu (Poole, 1998)



Alilogiikoiden jaottelu ("patologioiden kautta") ei toki suoraan osoita eri menetelmien käyttökelpoisuutta

Haasteita

Semanttisen Webin päättelyä vaativissa sovelluksissa ei *periaatteessa* ole paljoakaan Web-sovelluksille ominaista

Käytännöllisiä haasteita kuitenkin tarjoavat mm.:

- RDF- ja OWL-kielten takaa löytyvän deskriptiivisen logiikan ”epästandardi” perusluonne (ja RRS-järjestelmien eroavaisuudet)
- Sovellusten, tekniikoiden ja teorioiden monimutkaisuus

Sovellusten tulisi myös toimia myös käytännössä – rajoittunut logiikan näkökulma ei riitä

- Esim. ['Edinburgh'] Prolog on [yksinään käytettynä] ”huono” ympäristö *sovellusten* kehittämiseen (Kifer 2005)
- Teoria vs. käytännöt? Aliohjelmakirjastot? Ymmärrettävyys?



Sääntökielten kehitystyö

W3C on käynnistänyt standardin sääntökielen **Rule Interchange Format** kehitystyön 2006

- Ks. <http://www.w3.org/2005/rules/wg/charter>

Kaksiosainen prosessi:

1. Ensimmäisessä vaiheessa (2006-2007) määritellään laajennettavissa oleva sääntökieli (järjestelmä) joka oleellisesti perustuu Hornin logiikkaan
2. Toisessa vaiheessa (2007-2009[?]) sääntöjärjestelmää erikoistetaan (\rightarrow LP) ja laajennetaan (\rightarrow FOL)

Tehtävä on hyvin haasteellinen (!)



Lopuksi: Web tietämuskantana

Mistä tietämuskannan tieto sitten voi olla peräisin?

Vaihtoehtoja:

1. Kaikki tieto on suoraan joko ”käsien syötettyä” tai ”ohjelmallisesti tulostettua” (vrt. *Propositional database*)
2. Syötetyllä tiedolla on (lisäksi) deduktiivinen tulkinta (logiikka)
3. Syötetyllä tiedolla on (lisäksi) induktiivisia tulkintoja (esim. tilastollinen päättely)

Mutta: sekä deduktioon että induktioon liittyy paitsi käytännöllisiä, myös teoreettisia rajoituksia

- Esim. NP & ratkeamattomat ongelmat
- Kaikenlainen (kone)oppiminen edellyttää aina suunnittelijan valitsemaa suuntaamaa (bias) eikä yksikään (luokitteleva) oppimisalgoritmi pärjää arvausta paremmin mv. sovellusalueella (vrt. *Conservation law for generalisation performance* [Schaffer 1994])



Kiitos!

*“Formal ontology [fil.] is the systematic, formal, axiomatic development of the **logic** of all forms and modes of being.*

*It studies the formal **properties** and the **classifications** of the entities of the world (physical objects, events, etc.), and the **categories that model the world** (concept, property, etc.)”*

(Cocchiarella 1991)

