



Semanttinen Web - mitä XML:n jälkeen?

Prof. Eero Hyvönen
Helsingin yliopisto ja
Helsinki Institute for Information Technology
(HIIT)

27-Mar-02

1

Sisältö

- ☞ Semantic Web -visio
- ☞ Tasot ja standardit
- ☞ Sovellusalueita



27-Mar-02

Mitä hyötyä on Webistä?

☞ Keskeistä WWW:n tarjoamat palvelut

- ◆ Viestintä (email, puhe, kuva, video,...)
- ◆ Tiedonhaku (hakukoneet, portaalit,...)
- ◆ Toimenpiteiden suorittaminen
 - Sähköinen liiketoiminta
 - Sähköinen asiointi
 - Ym.

- Siksi esimerkiksi

- ◆ Valtavat 3G-panostukset
 - Suomessa Nokia, Sonera, Elisa, ...
- ◆ dot.com-kupla ja sen pubkeaminen



27-Mar-02

Miten hyötypalveluja Webiin saadaan?

- ☞ Sisältö tärkeää
 - ◆ materiaalin digitointi
- ☞ Palvelun äly tärkeää
 - ◆ sisällön hyödyntäminen
- ☞ Käyttömukavuus ihmiselle tärkeää
- ☞ Ansaintologiikan toimittava
 - ◆ money makes the world go round



27-Mar-02

Kehityksen este Webissä?

- ☞ WWW-palvelu ≈ kone auttaa ihmistä
 - ◆ Edellyttää sisältöjen koneellista "ymmärtämistä"
 - ☞ WWW:n sisällöt ovat ihmislukijaa varten
 - ◆ HTML, PDF, JPEG ...
 - ☞ Kone ei ymmärrä WWW:n sisältöjä
 - ◆ Hakukoneet, ostoagentit, verkkomönkijät jne.
 - ◆ Periaatteessa kaikki WWW-sovellukset
- Perustavaa laatua oleva ristiriita



27-Mar-02

5

Miten Webistä tulee älykkäämpi?

- ☞ Älykkäämmät sovellukset
 - Sisältö pysyy samana
 - Koneesta tehdään ihmismäisempi
- ☞ Älykkäämmiin esitetty sisältö
 - Sisältö helpommin ymmärrettäväksi
 - Kone pysyy tyhempänä
- ☞ Käytännössä molempia tapoja tarvitaan
 - Yhä älykkäämmät järjestelmät käsittelevät yhä älykkäämmiin esitettyjä tietoja



27-Mar-02

6

Ratkaisumalli 1: Älykkäämmät sovellukset

- ☞ Tiedon haku (information retrieval)
- ☞ Tiedon ekstrahointi (information extraction)
- ☞ Tiedon louhinta (data mining)
- ☞ Verkon louhinta (web mining)



27-Mar-02

Ratkaisumalli 1:n ongelmia

- ☞ Kielen automaattisen tulkinnan vaikeus
 - vapaamuotoisuus
 - sisällön semantiikka
- ☞ Ei-tekstuaaliset sisällöt
 - kuva, ääni, musiikki, video, ohjelmisto
 - Miten tulkita algoritmisesti?
- ☞ Tulkintaan ei riitä itse dokumentti
 - Tarvitaan konteksti, common sense



27-Mar-02

Ratkaisumalli 2: Älykkäämmän esitetty sisältö

- ☞ Semantic Webin lähtökohta
 - Talletetaan tieto niin, että tyhmempikin sen ymmärtää!
 - Ihminen tulee konetta vastaan
 - Kone voi auttaa ihmistä itsensä auttamisessa
- ☞ Kiihkeä kehitystyö käynnistynyt



27-Mar-02

Webin sukupolvia

- ☞ 1G WWW:
 - ◆ HTML-sivut ihmisen tulkittavaksi
- ☞ 2G WWW:
 - ◆ XML-sivut ihmisen/koneen tulkittavaksi
- ☞ 3G WWW:
 - ◆ Merkitykset ihmisen/koneen käytettäväksi
 - => Uusi perusta älykkäille WWW-palveluille
- ☞ Kansainvälinen yhteistyö (W3C ym....)



27-Mar-02

Semantic Web: Technology push

Trust level	Planning CPR, SPAR, PDDL,.... Processes BPML, WPD, PSL,.... Services UDDI, WSDL, DAML-S,.... Transactions XML/EDI, KQML,.... Communication TCP/IP, HTTP, SOAP,....
Digital signature, annotations,....	
Logic level	
KIF, RuleML,....	
Ontology level	
WordNet, RosettaNet, DAML+OIL,....	
Meta data level	
RDF, RDFS, Topic Maps,....	
Structure level	
XML, XML DTD/ Schema, XSL,....	
Internet level	
Unicode, URI,....	

27-Mar-02

11

XML ja "Semantic" Web?

- ☞ Semantiikka: "tutkii miten symbolit viittaavat toisiinsa olioihin"
 - Intensio: käsitteen merkitys
 - Extensio: käsitteen viittamat oliot
- ☞ XML on syntaktinen, ei semanttinen



27-Mar-02

12

Miksi XML ei ole "semanttinen"?

- <OSOITE>
 - <NIMI>Onni Ohjelmoija</NIMI>
 - <PUHELIN> 123 456 </PUHELIN>
- </OSOITE>
- <OSOITE>
 - <NIMI>Onni Ohjelmoija</NIMI>
 - <PUHELIN> 123 456 </PUHELIN>
- </OSOITE>

☞ Semantiikka on vain nahkakansissa, ei peltikuorissa



27-Mar-02

Semanttiset metakuvaukset

- ☞ Idea: rakenteelle on annettava merkitys (semantiikka) toisella tasolla
- ☞ Tärkeimmät Semantic Web -standardit
 - W3C: RDF (S), DAML+OIL, ...
 - ISO: Topic Map, XTM



27-Mar-02

RDF Model

- ☞ Idea: kuvataan WWW resursseja (URIs)
- ☞ Olio-Attribuutti-Arvo kolmikot
 - "Helsinki"-"is located in"-"Finland"
 - <http://www.helsinki.fi, Location, http://www.finland.fi>
- ☞ Reifikaatio
 - Hierarkkisuus



27-Mar-02

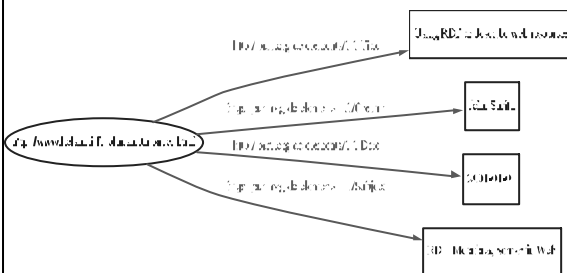
RDF Syntax

- ☞ Yleensä XML-perustainen (muukin käy!)
- ☞ "The WWW-paper "Using RDF ..." was written by John Smith, concerns RDF, Metadata and Semantic Web", and was created on Jan 1, 2000"

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:DC="http://purl.org/dc/elements/1.1/*">
  <rdf:Description rdf:about="http://www.helsinki.fi/john.smith/">
    <DC:Title>Using RDF to describe web resources</DC:Title>
    <DC:Creator>John Smith</DC:Creator>
    <DC:Date>2001-01-01</DC:Date>
    <DC:Subject>RDF, Metadata, Semantic Web</DC:Subject>
  </rdf:Description>
</rdf:RDF>
```

27-Mar-02

Vastaava RDF malli (verkko)



27-Mar-02

RDF Schema

- ☞ Tapa määritellä sanastoja, joilla RDF-kuvaukset annetaan
- ☞ Olio-ohjelmoinnista tuttu luokka-ajattelu
- ☞ Core classes
 - Resource, Class, Property



27-Mar-02

Miksi RDF(S) on tärkeä?

- ☞ Oleellisin innovaatio XML:ään verrattuna
 - ◆ XML:n tietomalli on *puu*
 - ◆ RDF:n tietomalli on yleisempi *verkko*
 - ◆ Ulkoiset metakuvaukset joustavasti
 - ◆ Voidaan esim. yhdistää useita eri verkkoa
- ☞ Olioajattelu merkkaukieliin
- ☞ Yksinkertainen yleinen "frame work"
 - ◆ Ei sidottu sovellusalaan tai -tapaan



27-Mar-02

Topic Map

- ☞ Ideoitu jo ennen WWW:tä 90-luvun alussa!
- ☞ Kirjan hakemisto lähtökohtana
 - O'Reillyn kirjojen hakemistojen yhdistäminen
- ☞ TM = tietämystason verkko (metadata) kuvattavista WWW-resursseista
 - Semanttinen, verkkomainen hakemisto
- ☞ ISO:n standardi



27-Mar-02

Ontologiat

- ☞ Ontologia on formaali, eksplisiittinen määrittely yhteisestä käsitteistöstä (Gruber, 1993)
 - ◆ Formaali: jämpä
 - ◆ Eksplisiittinen: konekin ymmärtää
 - ◆ Yhteinen: kommunikaatio mahdollista
- ☞ Kuvaa sovellusmaailmassa olevat oliot
- ☞ Ensimmäinen edellytys sille, että ihmiset ja koneet voivat ymmärtää toisiaan



27-Mar-02

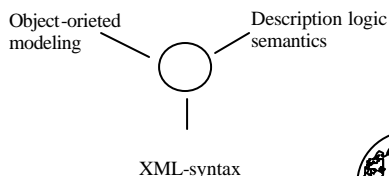
Ontologiat käytännössä

- Yleisiä hierarkkisia sanastoja
 - ◆ Esim. YSA, WordNet
- Yleisiä maailmanmalleja
 - ◆ Esim. CYC
- Spesifejä ammattitermistöjä
 - ◆ Esim. RosettaNet
- Luokittelujärjestelmiä
 - ◆ Esim. tuotteet/palvelut UN/SPSC
 - ◆ Kirjastot UDK,
- Loogisia terminologisia malleja
 - ◆ Ohjelmistoissa



27-Mar-02

Mitä uutta?



27-Mar-02

23

Esimerkki: OIL ontologiakieli

- ☞ Ontology container
 - Yleistä metatietoa ontologiasta (tekijä, kieli jne)
- ☞ Ontology definitions
 - Relaatioiden määrittelyt
 - ◆ Esim. käänteisrelaatio, transitivisuus
 - Luokka hierarkian määrittelyt



27-Mar-02

24

Container

ontology-container
title "African animals"
creator "Ian Horrocks"
subject "animal, food, vegetarians"
description "A didactic example ontology describing African animals"
description.release "1.01"
publisher "I. Horrocks"
type "ontology"
format "pseudo-xml"
format "pdf"
identifier "http://www.cs.vu.nl/~dieter/oil/TR/oil.pdf"
source "http://www.africa.com/nature/animals.html"
language "OIL"
language "en-uk"
relation.hasPart "http://www.ontosRus.com/animals/jungle.ontology"



27-Mar-02

Ontology definitions

ontology-definitions

slot-def eats

inverse is-eaten-by

slot-def has-part

inverse is-part-of properties transitive



27-Mar-02

```
class-def animal
class-def plant
  subclass-of NOT animal
class-def tree
  subclass-of plant
class-def branch
  slot-constraint is-part-of has-value tree
class-def leaf
  slot-constraint is-part-of has-value branch
class-def defined carnivore
  subclass-of animal
  slot-constraint eats value-type animal
class-def defined herbivore
  subclass-of animal
  slot-constraint eats
  value-type plant OR (slot-constraint is-part-of has-value plant)
class-def herbivore
  subclass-of NOT carnivore
class-def giraffe
  subclass-of animal
  slot-constraint eats
  value-type leaf
class-def lion
  subclass-of animal
  slot-constraint eats value-type herbivore
class-def tasty-plant
  subclass-of plant
  slot-constraint eaten-by has-value herbivore carnivore
```



27-Mar-02

Kuvausten käyttö

- ☞ Voidaan tuottaa RDFS-muoto tietokoneelle
- ☞ Konsistenssi voidaan tarkistaa
- ☞ Voidaan vastata loogisiin kysymyksiin
 - Päätelykoneella FaCT
- ☞ Dokumenttien loogisten metakuvausten ymmärtäminen



27-Mar-02

Ontologin haasteita

- ☞ Terminologinen konsensus
 - Eri tarpeiden ja mieltymisten sovittaminen
- ☞ Ontologioiden koko-ongelmat
 - Riittävätkö resurssit, skaalautuvatko välineet?
- ☞ Ontologioiden muuttuminen
 - Esim. Tsekkoslovakia valtio-ontologias
- ☞ Ontologioiden yhdistäminen
 - Esim. tuote/palvelukatalogit



27-Mar-02

Tasot 5-6: Logiikka ja luottamus

☞ Logiikkakatos

- ◆ Päätely ja toiminta ontologian kuvaamiin oloihin liittyen
- ◆ Todistukset/perustelut (proof)
- ◆ Esim. RuleML -saantömerkkuskieli

☞ Luottamus

- ◆ Resurssien kuvailu/arvostelu (annotointi)
 - Kenen suosituksiin itse kukin luottaisi?
- ◆ Digitaaliset allekirjoitukset (digital signatures)
 - Tahojen luotettava tunnistaminen
- ◆ Salaus: tiedot julki vain oikeille tahoille
- ◆ Tietosuojakysymykset (esim. P3P)



27-Mar-02

30

Tärkeimpiä sovellusalueita

- ☞ Kuvausstandardit
- ☞ Informaation haku (information retrieval)
- ☞ Tietämyksen hallinta (knowledge management)
- ☞ Sähköinen liiketoiminta, Web Services
- ☞ Profilointi ja kustomointi



27-Mar-02

Kuvausstandardit

- ☞ Dublin Core (DC)
- ☞ PRISM
- ☞ vCard
- ☞ Platform for Privacy Preference (P3P)



27-Mar-02

Informaation haku

- ☞ Seuraavan polven hakurobotit
 - Metatietojen hyödyntäminen
- ☞ Älykkäät hakemistot
- ☞ Matchmaker-sovellukset
- ☞ Semanttiset portaalit

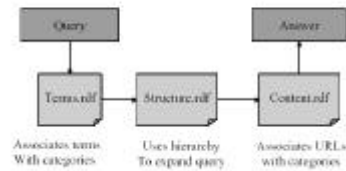


27-Mar-02

Open Directory Project dmoz

- ☞ RDF-based search

Searching the Open Directory



27-Mar-02

(Fensel, 2002)



27-Mar-02

35

Tietämyksen hallinta (knowledge management)

- ☞ Ongelmia
 - Dokumenttien monimuotoisuus
 - Maapalloistuminen -> sisältöjen hajautus
 - Tietämyskannan komplisoituminen
- ☞ SemWeb-tekniikat antavat uusia työkaluja
 - Liima heterogeenisten hajautettujen dokumenttien hallintaan



27-Mar-02

36

Sähköinen kaupankäynti Web Services

- Miten tarjoan oman tuotteen/palvelun kansainvälisille markkinoille?
 - ◆ Vision globaalit rekisterit ja sanastot käytettävissä (UDDI, RosettaNet,...)
- Miten hoidan transaktiot kumppaneiden kanssa?
 - ◆ Esim. tarjouspyyntöön vastaaminen
 - ◆ EDI-XML, SOAP, WSDL
- Toimintakehykset: UDDI, ebXML, ...



27-Mar-02

- ☞ Semantic Webin mahdollisuuksia
 - Sisältöjen rikastaminen
 - Eri standardien yhdistäminen
 - ◆ Esim. toimialaportaalit
 - Liiketoimintaprosessien automatisointi
 - ◆ Semanttiset kuvaukset



27-Mar-02

Profilointi ja kustomointi

- ☞ Ihmiset ja palvelut (P3P)
 - Omat preferenssit, tietosuojat,...
- ☞ Laitteet (CC/PP, FIPA Device Ontology)
 - Esim. MV-kännykälle ei värikuvia
- ☞ Dokumentit (transcoding)
 - Sisältöjen kustomointi eri laitteille ja tarpeisiin



27-Mar-02

Semantic Web ja Suomi

- ☞ Juna lähti jo
 - ◆ XML-standardointihankkeet, 90-luvun loppu
 - ◆ W3C Semantic Web Activity, 2001/kevät
 - ◆ EU:n OntoWeb 2001/kesä
 - ◆ Semantic Web Kick-Off in Finland 2001/syksy
- ☞ Tärkeä ala jatkossa monessa mielessä
 - ◆ Teollinen intressi
 - ◆ Tekninen mahdollisuus
 - ◆ Tieteellinen haaste
 - ◆ Kansallinen intressi



27-Mar-02

Lisätietoja

- ☞ W3C:n Semantic Web Activity
 - ◆ www.w3.org/2001/SW/
- ☞ Tutkimusmaailman portaali
 - ◆ www.semanticweb.org
- ☞ EU:n yhteistyöverkosto
 - www.ontoweb.org
- ☞ Semantic Web in Finland
 - www.cs.helsinki.fi/u/eahyovone/stes/semanticweb/



27-Mar-02

41