

Semantic Web - kohti uutta merkitysten Internetiä

Prof. Eero Hyvönen
Helsingin yliopisto
Helsinki Institute for Information
Technology

1-marras-01

1

Visio tulevaisuuden Webistä

- **Mitä hyötyä on Webistä?**
 - Keskeistä WWW:n tarjoamat palvelut
 - Vrt. 3G, Sonera, Elisa, dot.com-kupla
- **Miten hyötypalveluja Webiin saadaan?**
 - Sisältö tärkeää (materiaalin digitointi)
 - Palvelun äly tärkeää (sisällön hyödyntäminen)
 - Käyttömukavuus tärkeää ihmiselle
- **Keskeinen edellytys?**
 - Sisältöjen koneellinen "ymmärtäminen"

1-marras-01

2

Vision este Webissä

- **WWW:n sisällöt ovat ihmislukijaa varten**
 - HTML, PDF, JPEG ...
- **WWW:tä käyttävät kuitenkin myös koneet**
 - Hakukoneet, ostoagentit, verkkomönkijät jne.
 - Periaatteessa kaikki WWW-sovellukset
 - Yhä enenevässä määrin
- => Perustavaa laatua oleva ristiriita

1-marras-01

3

Ratkaisumalli 1: Älykkäämmät sovellukset

- **Tiedon haku (information retrieval)**
 - Tiedon etsintä "surffailamalla"
 - Assosiativisuus
 - "Lost in hyperspace"-ongelma
 - Hakurobotit
 - Avainsanojen indeksointi
 - Verkon viittausten hyödyntäminen (Google)

1-marras-01

4

- **Tiedon ekstrahointi (information extraction)**
 - Merkityshahmojen tunnistaminen vapaamuotoisesta aineistosta
 - Esim. kiinnostavien uutisten tunnistus
- **Tiedon louhinta (data mining)**
 - Klusterointi, sääntöjen oppiminen ...

1-marras-01

5

- **Verkon louhinta (web mining)**
 - Sisällöt: Content mining
 - Käyttö: Usage mining

1-marras-01

6

Ratkaisumalli 1:n ongelmat

- 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

1-marras-01 7

Ratkaisumalli 2: Semantic Web

- Talletetaan tieto niin, että tyhmempikin sen ymmärtää!
- Ihminen tulee konetta vastaan
- Kone voi auttaa ihmistä itsensä auttamisessa

1-marras-01 8

Semantic Web: Technology push

Luottamus (trust, proof)	Palvelut, prosessit XML/edi, ebXML, UDDI, WSDL,...
Päättele DAML+OIL, RuleML, KIF...	
Ontologiat WordNet, RosettaNet, ...	
Metakuvaukset RDF, RDF(S), Topic Maps,...	
Rakenteen kuvaus XML, XML Schema, XSL, ...	
Internet URI, Unicode...	Kommunikointi SOAP, KQML, ...

1-marras-01 9

Taso 1: Internet

- Ytimenä URI (Universal Resource Identifier)
 - Esim. <http://www.cs.helsinki.fi/>
 - WWW = URI-en verkko
- Unicode-merkistö
 - Kaikki maailman kielet
- Protokollat
 - HTTP, FTP, TCP/IP

1-marras-01 10

Taso 2: Sisältöjen rakenne

- eXtensible Markup Language (XML)
- Meta-kieli sovelluskohtaisille kielille
 - <OSOITE>
<NIMI>Onni Ohjelmoija</NIMI>
<PUHELIN> 123 456 </PUHELIN>
</OSOITE>
- Rakenteen ja ilmasun erottaminen
 - Muunnokset ja ilmasu: XSL

1-marras-01 11

XML ja "Semantic" Web?

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

1-marras-01 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 peltikurossa**

1-marras-01

13

Taso 3: Semanttiset metakuvaukset

- ⊛ **Idea: rakenteelle on annettava merkitys (semantiikka) toisella tasolla**
- ⊛ **Tärkeimmät Semantic Web -standardit**
 - **W3C:**
 - **RDF (S)**
 - **ISO:**
 - **Topic Maps, XTM**

1-marras-01

14

RDF(S)

- ⊛ **RDF Resource Description Framework (1999)**
 - **Yleinen verkkoresurssien kuvaamiskieli**
- ⊛ **RDF Schema (2000)**
 - **RDF-terminologian määrittely**
 - **Käsittehierarkiat (Class/subClass/type)**

1-marras-01

15

Topic Maps

- ⊛ **Ideoitu jo ennen WWW:tä 90-luvun alussa!**
- ⊛ **Kirjan hakemisto lähtökohtana**
- ⊛ **TAO: Topic - Association - Occurrence**
- ⊛ **Semanttinen surffailu Webissä**

1-marras-01

16

RDF(S) ja XTM: semantiikan idea

- ⊛ **Kieliä WWW-resurssien kuvaamiseen**
 - **Kielen muodot viittaavat alla oleviin resursseihin ja jopa itseensä (reifikaatio)**
- ⊛ **Tietomalli ei ole sidottu XML-esitysmuotoon**
 - **Sarjallistamistapa (serialization) periaatteessa yhdentekevä**
 - **Kielet voidaan esittää muunkinlaisella syntaksilla**

1-marras-01

17

Taso 4: Ontologiat

- ⊛ **Ontologia on formaali, eksplisiittinen määrittely yhteisestä käsitteistöstä (Gruber, 1993)**
- ⊛ **Kuvaa sovellusmaailmassa olevat oliot**
- ⊛ **Ensimmäinen edellytys sille, että ihmiset ja koneet voivat ymmärtää toisiaan**

1-marras-01

18

Ontologiat käytännössä

- Yleisiä hierarkkisia sanastoja (esim. YSA, WordNet)
- Yleisiä maailmanmalleja (esim. CYC)
- Spesifejä ammattitermistöjä (esim. RosettaNet)
- Luokittelujärjestelmiä (esim. UDK)
- Loogisia terminologisia malleja

1-marras-01

19

Ontologin välineitä

- **Esityskielet**
 - Esimerkiksi DAML+OIL
- **Ontologiaeditorit**
 - Esimerkiksi Protégé, OilEd

1-marras-01

20

Ontologin haasteita

- **Terminologinen konsensus**
 - Eri tarpeiden ja mieltymisten sovittaminen
- **Ontologioiden koko-ongelmat**
 - Riittävätkö resurssit, skaalautuvatko välineet?
- **Ontologioiden muuttuminen**
 - Esim. Tsekkoslovakia valtio-ontologiassa
- **Ontologioiden yhdistäminen**

1-marras-01

21

Tasot 5-6: Logiikka ja luottamus

- **Logiikkataso**
 - Päättely ja toiminta ontologian kuvaamiin olioihin liittyen
 - Todistukset/perustelut (proof)
 - Esim. RuleML-saantömerkkäuskieli
- **Luottamus**
 - Resurssien kuvailu/arvostelu (annotointi)
 - Digitaaliset allekirjoitukset (digital signatures)

1-marras-01

22

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, XML, ...)**

1-marras-01

23

Sovellusalueita

- **Informaation haku**
 - Seuraavan polven hakurobotit
 - Älykkäät hakemistot
 - Matchmaker-sovellukset
 - Semanttiset portaalit

1-marras-01

24

Tietämyksen hallinta (knowledge management)

- **Ongelmia**
 - Dokumenttien monimuotoisuus
 - Maapalloistuminen -> sisältöjen hajautus
 - Tietämiskannan komplisoituminen
- **SW-tekniikat antavat uusia työkaluja**

1-marras-01 25

Verkkokauppa (B2C)

- Uusi markkinointi- ja jakelukanava
- Uudet sähköiset liiketoimintamallit

Sähköinen liiketoiminta (B2B)

- Transaktiot: EDI/XML
- Toimintakehykset: UDDI, ebXML, WSDL, RosettaNet

Profilointi

- Ihmiset, palvelut, laitteet (CC/PP)

1-marras-01 26

Semantic Web ja Suomi

- **Kansainvälinen juna lähti jo**
 - XML-standardointihankkeet, 90-luvun loppu
 - W3C Semantic Web Activity, 2001/kevät
 - EU:n OntoWeb 2001/kesä
- **Tärkeä ala, kick-off tarpeen Suomessa**
 - Teollinen intressi
 - Tekninen mahdollisuus
 - Tieteellinen haaste
 - Kansallinen intressi

1-marras-01 27

Semantic Web Kick-Off in Finland

- **Johdanto, visio (E. Hyvönen)**
- **W3C:n aktiviteetit (M.-R. Koivunen)**
- **Suomalainen yrityscase (J. Saarela)**
- **Teknologiakatsaus (Helsingin yliopisto)**
- **Suomalaiset hankkeet**

1-marras-01 28

Tervetuloa verkottumaan semanttisesti!

1-marras-01 29