

# Mikä on semanttinen web?

Airi Salminen

Jyväskylän yliopisto

<http://www.cs.jyu.fi/~airi/>

Semanttinen web ja funktionaalinen luettelointi  
seminaari 3.5.2006

# Sisältö

1. Webin ongelmia
2. Semanttinen web: määrittelyjä
3. Keskeisiä ominaisuuksia
4. Metatieto semanttisessa webissä
5. Semanttisen webin tekniikoita
6. Mahdollisuuksia ja haasteita
7. Yhteenveto

- ▶ *Epäyhtenäisyys*: heterogeenisten sisältöjen määrät kasvavat helposti hallitsemattomiksi jopa organisaatioiden intranet-ratkaisuissa.
- ▶ *Kommunikointikuilut*: järjestelmät ja palvelut eivät kommunikoi keskenään paljonkaan edes yhden organisaation sisällä.
- ▶ *Epäluotettavuus*: sisällön ja palvelujen luotettavuus epävarmaa erityisesti Internetissä.

- ▶ HTML-kieli on yhdenmukainen ratkaisu, mutta se kertoo lähinnä dokumenttirakenteista (otsikko, kappale, lista, taulukko, ... ), kirjasintyypeistä (kursivoitu, ...) ja linkityksistä, ei dokumentin sisällön merkityksestä.
- ▶ Sisältöjä muilla esitysmuodoilla paljon, esim. teksturidokumentteja, kuvia, äänitteitä, video-, animaatio- ja kalvoesityksiä.
- ▶ Sisältöjen merkitys ei yleensä ohjelmistojen ymmärrettävissä.

- ▶ Internetin käyttäjien ja tietovarantojen huima kasvu on aktivoinut myös mitä erilaisimpien agenttiohjelmien ja vertaispalvelujen (peer-to-peer) kehittämisen; niiden kontrolloiminen ja niihin liittyvän tietoturvan varmistaminen on ongelmallista.
- ▶ Internet-teknologia rakennettiin rajatulle yhteisölle, sitä ei suunniteltu kaikkien ihmisten yhteiseksi kommunikointitekniikaksi; luottamusta pidettiin itsestään selvyytenä.

An extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.

Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001.

## 2. Semanttinen web: määrittelyjä

The **Semantic Web** provides a common framework that allows **data** to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (**RDF**), which integrates a variety of applications using XML for syntax and URIs for naming.

W3C Semantic Web Activity, <http://www.w3.org/2001/sw/>

XML = Extensible Markup Language

RDF = Resource Description Framework

URI = Uniform Resource Identifier

### 3. Keskeisiä ominaisuuksia

- ▶ Koostuu primääriresursseista ja metatietoresursseista
- ▶ Resurssien merkitykseen, käyttöön ja luotettavuuteen liittyviä metatietoja
- ▶ Metatiedot standardoidussa muodossa, sekä ihmisten että ohjelmien luettavissa (esitystapoina XML, RDF ja muut W3C:n määrittelemät esitystavat)
- ▶ Automaattista päättelyä merkityksestä ja luotettavuudesta
- ▶ Kehitystä koordinoi W3C

Semanttinen web on metatiedolla rikastettu web



Tietoa joka auttaa käyttämään,  
ymmärtämään, prosessoimaan ja  
yhdistämään webin tietoresursseja

### julkaisutiedot

- jokaiseen julkaisuun liittyy Dublin Core –tiedot: creator, subject, ...
- julkaisun kirjoittaja (creator) identifioidaan yksikäsitteisellä URI-tunnisteella
- julkaisun asiasanat annetaan yleisestä suomalaisesta ontologiasta (YSO)

### kirjailijatiedot

- tietoja suomalaisista kirjoittajista, heidän suhteista toisiinsa
- jokainen kirjoittaja tunnistetaan URI-tunnisteella

### YSO

- suomen kielen asiasanat ja niiden väliset suhteet

Tämä kaikki RDF-muodossa, tietoja voidaan yhdistellä

### Vaihtoehtoja metatiedoille

kuvaa jotain tietoresurssia Internetissä	kuvaa Internetin ulkopuolista asiaa
kohteena resurssien luokka (esim. dokumenttiluokka)	kohteena ilmentymä (esim. dokumentti)
kuvaa kohteen sisältöä	kuvaa kohteen kontekstia
ihmisen tuottama	tietokoneohjelma tuottama
upotettu (esim. HTML-dokumenttiin)	ulkoinen
keskistetty metatietovaranto	hajautettu metatietovaranto
tuotettu yhdessä kohteen luomisen kanssa	tuotettu ennen tai jälkeen
käyttäjänä ihminen	käyttäjänä tietokoneohjelma

### Ontologia

- ▶ Yhteisesti sovittu käsitteistö (käsitteet, niiden väliset suhteet)
- ▶ Tietokoneen tulkittavaksi tarkoitettu
- ▶ Mahdollisesti päättelysääntöjä

### Semanttisen webin pelkistetty rakenne



### Tekniikoita sisältöjen identifioimiseen, esittämiseen ja määrittelyyn

- Merkkikoodit ja tietoresurssien osoittaminen: Unicode, URI
- Rakenteisten dokumenttien esittäminen: XML, XML Namespaces
- Rakenne- ja sisältörajoitteiden määrittelemineen: DTD, XML Schema

### Tekniikoita metatiedon esittämiseen ja määrittelyyn

- metatiedon esittäminen: RDF
- metatietoesitystapojen määrittely (ontologian määrittely): RDF Schema
- ontologian määrittely: Web Ontology Language (OWL)

Dublin Core –tiedot RDF-muodossa:

[http://www.lib.helsinki.fi/dublin\\_core/rdfsyntax.html](http://www.lib.helsinki.fi/dublin_core/rdfsyntax.html)

Dublin Core –termistön määrittely RDF-skeemalla:

<http://dublincore.org/schemas/rdfs/>



### Tekniikoita luottamuksen hallintaan

- digitaalinen allekirjoitus: XML-Signature
- tietosuojamenettelytavat: P3P, APPEL
- salakirjoittaminen: XML Encryption
- annotointien jakaminen: Annotea

### Tekniikoita ohjelmien väliseen kommunikointiin

- asiakaskoneen valmiuksien ilmaiseminen palvelinkoneelle: CC/PP
- yhteyskäytäntö informaation välittämiseksi hajautetussa ympäristössä: SOAP
- web-palveluiden kuvaaminen: WSDL

<b>sovellusohjelmat</b>	
<b>primääritietovarannot</b>	<b>DTD:t, XML Schema - määrittelyt, RDF Schema -määrittelyt, RDF-määrittelyt, ontologiat, annotoinnit, ...</b>
<b>URI, Unicode, XML, XML Namespaces, XML Schema, RDF, RDF Schema, XTM, XML-Signature, OWL, Annotea, ...</b>	

### Yhteistyön vahvistaminen

- ▶ Monipuolinen ohjelmistojen yhteistyö ja kommunikointi
- ▶ Ihmisten monipuolinen yhteistyö yhteisöissä, missä on yhdenmukainen käsitteistö, käsitys sisältöjen merkityksestä ja yhteiset arvot
- ▶ Hajallaan olevat yhteisöt voivat työskennellä toisistaan riippumattomasti webin yhteisen tietämyksen lisäämiseksi ja jatkuvaksi kehittämiseksi

### Moniarvoiset Internetin tietosisällöt

- ▶ Ei kasvurajoja
- ▶ Ei keskitettyä kontrollia
- ▶ Epätäydellisyys hyväksytään
- ▶ Ristiriitaisuudet mahdollisia
- ▶ Kaikkiin kysymyksiin ei saada vastauksia

### Entistä joustavampi laajennettavuus ja yhdisteltävyys

- ▶ Uutta tietoa voidaan lisätä niin, ettei vanhaa tarvitse muuttaa
- ▶ Erilaisia sanastoja ja erilaisia rakenteita voidaan yhdistellä
- ▶ Paikallisen ja globaalin tiedon joustava yhdistäminen

### Luottamuksen lisääminen

- ▶ Tietojen tuottaja voi määritellä käyttöoikeudet erilaisille kuluttajille, niin ihmisille kuin myös ohjelmistoille
- ▶ Tietojen kuluttaja voi varmistaa tiedon alkuperän ja sisällön tai palvelun luotettavuuden
- ▶ Palvelun käyttäjä voi varmistaa, mihin hänen tietojansa käytetään
- ▶ Tekniikat hajautettuun luottamuksen hallintaan

### Sovellusalueita:

- web-portaalit (esim. kirjastoportaali)
- organisaation tai yhteisön tietämyksen hallinta (informaatiopalvelut)
- elektroninen kaupankäynti
- agenttiohjelmat (esim. tiedonhakuagentti)



- Metatiedon standardoiminen ongelmallista; vaatii paljon työtä ja yhteistyökykyä
- Käsitteistä ja niiden välisistä suhteista sopiminen voi olla erittäin työlästä.
- Ontologioiden pysyvyys ja ylläpito; luojaisten käsitteistö ei ole välttämättä yhdenmukainen käyttäjien käsitteistön kanssa
- Luottamukseen ja tietoturvaan liittyvät ongelmat erittäin vaikeita
- Vaatii eri alueiden asiantuntijoiden (kohdealue, tietohallinto, tietopalvelut, tekniikka) halua ja kykyä yhteistyöhön

- Teknologia semanttisen webin rakentamiseen on olemassa
- Standardointityö erittäin haastavaa
- Automaatio lisääntyy kirjastopalveluissa (hakuagentit, kysymyksiin vastaavat järjestelmät, päättelyt)
- Kirjastoasiantuntijoiden vahva asiantuntemus metatietojen määrittelystä tarpeellista semanttisen webin rakentamisessa

- Semanttisen webin yhteisöportaali  
<http://www.SemanticWeb.org>
- W3C:n semanttisen webin sivusto  
<http://www.w3.org/2001/sw/>
- XML-kieliperheportaali  
<http://www.cs.jyu.fi/~airi/xmlfamily.html>
- Johdatus XML-kieleen  
<http://www.cs.jyu.fi/~airi/opetus/xml/xml-kieli/xml-jakso1.pdf>