

# Semanttinen web - lyhyt johdatus

Airi Salminen  
Jyväskylän yliopisto  
<http://www.cs.jyu.fi/~airi/>

Huhtikuu 2002

# Sisältö

1. Webin historia
2. World Wide Web
3. Tavoitteita semanttiselle webille
4. Webin tieto ja metatieto
5. Semanttisen webin tekniikoita
6. RDF-metatietomalli
7. Ontologiat ja ontologiakielet
8. Semanttisen webin sovellusalueita
9. Yhteenveto

Kurssin kotisivu: <http://www.cs.jyu.fi/~airi/opetus/SemanttinenWeb.html>

# 1. Webin historia

1960-luku: Verkottumisen alku

1970-luku: Internetin synty

1980-luku: Internet käytäntöön

1990-luku: Webin synty

2000-luku: Kohti semanttista webiä

# 1. Webin historia

- 1957 Neuvostoliitto laukaisee avaruuteen Sputnikin
- 1958 ARPA (myöhemmin DARPA = Defense Advanced Research Projects Agency) perustetaan

## 1960-luku: Verkottumisen alku

- 1962 Licklider (MIT) esittää vision globaalisti toisiinsa kytketyistä tietokoneista
- 1969 ARPANET-verkko saa alkunsa, 4 konetta vuoden lopussa verkossa; käyttö tiedostojen siirtoa ja ohjelmien etäkäyttöä; RFC-käytäntö (Request for Comments) Internetin spesifikaatioiden kehittämiseksi saa alkunsa.

# 1. Webin historia

## 1970-luku: Internetin synty

- 1970 ALOHAnet ja muitakin verkkoja alkaa kehittyä
- 1971 Uusi verkkosovellus: sähköposti
- 1972 "Internetworking architecture" -ajatus *verkkojen verkosta* esitetään DARPA:n piirissä
- 1973 TCP/IP (Transmission Control Protocol/Internet Protocol) -yhteyskäytäntö syntyy
- 1979 noin 200 tietokoneella oma IP-osoite

# 1. Webin historia

## 1980-luku: Internet käytäntöön

- 1983 ARPANET siirtyy TCP/IP-yhteyskäytäntöön
- 1984 Nykyinen nimeämiskäytäntö (esim. [www.cs.jyu.fi](http://www.cs.jyu.fi)) otetaan käyttöön
- 1985 Amerikkalainen yliopistojen tutkimusverkko NSFNET perustetaan
- 1989 NSFNET:ssä mukana AU, CA, DE, DK, FI, FR, IL, IS, IT, JP, MX, NL, NO, NZ, PR, SE, UK

# 1. Webin historia

## 1990-luku: Webin synty

- 1991 World Wide Web (WWW) esitellään Cernissä;  
Internet Society perustetaan
- 1992 Internetiin kytkettyjen koneiden määrä > miljoona
- 1993 Media, yritykset ja julkishallinto alkavat herätä
- 1994 W3C perustetaan
- 1996 PICS (Platform for Internet Content Selection)  
hyväksytään W3C:ssä
- 1998 XML-spesifikaatio hyväksytään W3C:ssä,  
Dublin Core -metatietosuositus esitellään RFC:nä
- 1999 RDF-spesifikaatio hyväksytään W3C:ssä

# 1. Webin historia

## 2000-luku: Kohti semanttista webiä

- 2000 Internetiin kytkettyjen koneiden määrä ylittää sata miljoonaa
- 2001 Semantic Web Activity käynnistyy W3C:ssä
- 2006 Miljardi vuorovaikutteista laitetta kytkettynä Internetiin?



## 2. World Wide Web

World Wide Web (WWW) syntyi Cernin hajallaan toimivien tutkijoiden ja tutkimusprojektien yhteistyön ja tietämyksen hallinnan tukemiseen. (ks. Berners-Lee et al., 1994)

### WWW:n keskeiset tekniikat

- Resurssien osoittaminen: URI (Universal Resource identifier)
- Rakenteisten hypertekstidokumenttien esittäminen: HTML (Hypertext Markup Language)
- Hypertekstidokumenttien siirtäminen: HTTP (Hypertext Transfer Protocol)

## 2. World Wide Web

### Ongelmia

- WWW:n perusratkaisu oli niin onnistunut, että webin sisältöjen määrä alkoi kasvaa vaikeasti hallittavaksi.
- WWW:n sisällöstä valtaosa on maailman eri kielillä kirjoitettua luonnollista kieltä; sisällön merkitys ei yleensä ohjelmistojen ymmärrettävissä.
- XML tarjoaa yhdenmukaisen esitystavan dokumenteille sekä tekniikan määrittellä termit (elementtien ja attribuuttien nimet) ja rakenteet tietyille dokumenttijoukolle; samallekin sovellusalueelle kehitetään kuitenkin lukuisasti toisistaan poikkeavia dokumenttityyppejä.

## 2. World Wide Web

### Ongelmia

- Internetin käyttäjien ja tietovarastojen huima kasvu on aktivoanut myös mitä erilaisimpien agenttiohjelmien ja vertaispalvelujen (peer-to-peer) kehittämisen; niiden kontrolloiminen ja niihin liittyvän tietoturvan varmistaminen on ongelmallista.
- Sisällöntuottajien määrän kasvaessa sisällöstä on syntynyt mitä kirjavinta; kaikki ei ole sopivaa kaikille eikä kaikki ole tarpeen kaikille.

### 3. Tavoitteita semanttiselle webille

#### Kontrolloimattoman kasvun mahdollisuus

- Webin kasvua ei edelleenkään haluta rajoittaa.
- Ei keskitettyä kontrollia.
- Epätäydellisyys hyväksytään.
- Ristiriitaisuudet mahdollisia.
- Kaikkiin kysymyksiin ei saada vastauksia.

### 3. Tavoitteita semanttiselle webille

#### Monipuolista metatietoa

- Nykyistä huomattavasti enemmän webin tietovarantojen merkitykseen, käyttöön, käyttötapoihin ja luotettavuuteen liittyvää metatietoa.
- Metatietojen kieli huomattavasti homogeenisempaa ja kontrolloidumpaa kuin dokumenttisisältöjen.

### 3. Tavoitteita semanttiselle webille

#### Metatiedot koneiden ymmärrettävään muotoon

- Metatiedot formaalissa standardoidussa muodossa niin, että sekä ihmiset että erilaiset ohjelmistot pystyvät niitä hyödyntämään.
- Automaattisia päättelysääntöjä.

### 3. Tavoitteita semanttiselle webille

#### Entistä joustavampi laajennettavuus ja yhdisteltävyys

- Hajallaan olevien yhteisöjen täytyy voida työskennellä toisistaan riippumattomasti webin yhteisen tietämyksen lisäämiseksi ja jatkuvaksi kehittämiseksi.
- Uutta tietoa täytyy voida lisätä niin, ettei vanhaa tarvitse muuttaa.
- Erilaisia sanastoja ja erilaisia rakenteita täytyy voida yhdistellä.
- Tietoalkioiden monipuoliset linkitysmahdollisuudet.
- Paikallisen ja globaalin tiedon joustava yhdistäminen: eri järjestelmistä peräisin olevaa eri muotoista paikallista tietoa täytyy voida yhdistää globaalissa tietoverkossa.

# 3. Tavoitteita semanttiselle webille

## Luottamuksen lisääminen

- Tietojen tuottajan voitava määritellä käyttöoikeudet erilaisille kuluttajille, niin ihmisille kuin myös ohjelmistoille.
- Tietojen kuluttajan voitava varmistaa tiedon alkuperä ja ja sisällön tai palvelun luotettavuus.
- Palvelun käyttäjän voitava varmistaa, mihin hänen tietojaan käytetään.
- Juridiset, kaupalliset ja yhteiskunnalliset näkökohdat voitava huomioida.
- Tekniikat hajautettuun luottamuksen hallintaan.



## 4. Webin tieto ja metatieto

**metatieto = tietoa tiedosta**

**esimerkiksi**

- **dokumentista**
- **dokumenttikokoelmasta**
- **dokumentin osasta**
- **HTML-sivusta**
- **tietokannasta**
- **ohjelmasta**
- **museoesineestä**
- **henkilöstä**

# 4. Webin tieto ja metatieto

## Vaihtoehtoja metatiedoille

<b>liittyy webin tietoon</b>	<b>liittyy webin ulkopuoliseen asiaan</b>
<b>upotettu</b>	<b>ulkoinen</b>
<b>keskitetty</b>	<b>hajautettu</b>
<b>ihmisen tuottama</b>	<b>automaattisesti tuotettu</b>

# 4. Webin tieto ja metatieto

## Metatieto voi kuvata esimerkiksi

- sisällön rakennetta
- sisällön merkitystä
- sisällön kieltä
- kontekstia (tekijä, organisaatio, prosessi, ohjelma)
- versiota
- tallennusmuotoa
- tallennusmediaa
- säilytysaikaa
- käyttöoikeuksia

# 4. Webin tieto ja metatieto

## Esimerkki metatiedosta HTML-dokumentissa

```
<head>
  <title>Airi Salminen</title>
  <link rel="schema.DC" href="http://purl.org/dc/elements/1.1/">
  <meta name="DC.Title" content="Airi Salminen">
  <meta name="DC.Creator" content="Airi Salminen">
  <meta name="DC.Subject" content="Airi Salminen">
  <meta name="DC.Description" content="Representative publications,
    research interest, projects, courses, and contact information">
  <meta name="DC.Date.Updated" content="2002-04-10">
  <meta name="DC.Type" content="Text.Home page">
  <meta name="DC.Format" content="html">
  <meta name="DC.Identifier" content="http://www.cs.jyu.fi/~airi/index.html">
  <meta name="DC.Language" content="en">
  <meta name="DC.Relation" content="http://www.cs.jyu.fi/~airi/docman.html">
  <meta name="DC.Relation"
    content="http://www.cs.jyu.fi/~airi/publications.html">
  <meta name="DC.Rights" content="Copyright 2002 Airi Salminen">
</head>
```

# 4. Webin tieto ja metatieto

## Lisäyksiä suomenkielistä hakua varten:

```
<head>
```

```
...
```

```
<meta name="DC.Description" content="Representative publications,  
research interest, projects, courses, and contact information">
```

```
<meta name="DC.Description" content="keskeiset julkaisut, tutkimusintressit,  
projektit, opetus, yhteystiedot">
```

```
...
```

```
</head>
```

**Suomenkielistä kuvailua varten käytettävissä on suomalainen versio Dublin Core -metatietomäärityksistä. Ks. JHS 143, Asiakirjojen kuvailuformaatti, Yleinen Suomalainen Asiasanasto.**

## 4. Webin tieto ja metatieto

### XML-kieli metatiedon esittämisessä

- Elementin merkkkaus (markup) sisältää metatietoa elementin sisällöstä (yleensä sen merkityksestä)
- Dokumentin merkkkaus sisältää tietoa dokumentin rakenteesta
- Dokumentin DTD sisältää tietoa dokumentin rakenteesta
- XML-dokumentin sisältö voi koostua johonkin asiaan liittyvistä metatiedoista, esimerkiksi julkaisujen bibliografisista tiedoista.

## 4. Webin tieto ja metatieto

```
<?xml version = "1.0"?>
<runo tekija = "Murasaki Shikibu" tekija_syntyntyt = "974">
<info_linkki xmlns:xlink="http://www.w3.org/1999/xlink"
  xlink:type="simple"
  xlink:href=
    "http://digital.library.upenn.edu/women/omori/court/murasaki.html">
  Lis&auml;tietoa runoilijasta
</info_linkki>
<sae>
<rivi>This life of ours would not cause you sorrow</rivi>
<rivi>if you thought of it as like </rivi>
<rivi>the mountain cherry blossoms</rivi>
<rivi>which bloom and fade in a day. </rivi>
</sae>
</runo>
```

**Note:** The text of the rivi elements is taken from <http://www.slip.net/~knabb/rexroth/translations/japanese.htm>, containing Kenneth Rexroth's translations of Japanese poetry

## 4. Webin tieto ja metatieto

**This life of ours would not cause you sorrow  
if you thought of it as like  
the mountain cherry blossoms  
which bloom and fade in a day.**

[Lisätietoa  
runoilijasta](#)



## 4. Webin tieto ja metatieto

**This life of ours would not cause you sorrow  
if you thought of it as like  
the mountain cherry blossoms  
which bloom and fade in a day.**

[Lisätietoa  
runoilijasta](#)

**Suomen kieltä ymmärtävälle yllä olevan tekstin XML-merkattu muoto kertoo esimerkiksi, että**

- **teksti muodostuu runosta ja se puolestaan yhdestä säkeestä**
- **runon kirjoittaja on Murasaki Shikibu**
- **runon kirjoittaja on syntynyt vuonna 974**
- **teksti "Lisätietoa runoilijasta" on linkki osoitteeseen**

**<http://digital.library.upenn.edu/women/omori/court/murasaki.html>**

## 4. Webin tieto ja metatieto

Jos runoon liitetään DTD, myös siinä on paljon metatietoa.

```
<!DOCTYPE runo [  
<!ELEMENT runo (info_linkki? nimi?, sae+)>  
<!ATTLIST runo  
    tekija CDATA #REQUIRED  
    tekija_syntynyt CDATA #IMPLIED>  
<!ELEMENT otsikko (#PCDATA) >  
<!ELEMENT info_linkki (#PCDATA) >  
<!ATTLIST info_linkki  
    xmlns:xlink CDATA #FIXED "http://www.w3.org/1999/xlink"  
    xlink:type CDATA #FIXED "simple"  
    xlink:href CDATA #REQUIRED >  
<!ELEMENT sae (rivi+) >  
<!ELEMENT rivi (#PCDATA) >]
```

# 4. Webin tieto ja metatieto

**DTD määrittelee kaikkien DTD:tä noudattavien dokumenttien**

- **sanaston: runo, info\_linkki, nimi, sae, rivi tekija, tekija\_syntynyt, xlink:type, xlink:href**
- **elementtirakenteen**
- **sallitut attribuutit**

## 4. Webin tieto ja metatieto

XML:ää käytetään lukuisissa yhteyksissä nimenomaisesti metatiedon esittämiseen, esimerkiksi

- **Dublin Coren XML-muoto, esimerkiksi Open Archives Metadata Harvesting Protocol**
- **BiblioML: UNIMARC-muotoisten bibliografisten tietojen esittämiseen**

**XML muodostaakin perustan, jonka päälle myös semanttinen web rakentuu.**

# 5. Semanttisen webin tekniikoita

## Semanttisen webin pelkistetty rakenne



# 5. Semanttisen webin tekniikoita

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

# 5. Semanttisen webin tekniikoita

## Tekniikoita metatiedon esittämiseen ja määrittelyyn

- metatiedon esittäminen: RDF
- metatietoesitystapojen määrittely: RDF Schema
- aihekartat: Topic Maps (XTM)
- ontologian määrittely: DAML+OIL

# 5. Semanttisen webin tekniikoita

## Tekniikoita luottamuksen hallintaan

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



# 5. Semanttisen webin tekniikoita

sovellusohjelmat	
primääritietovarannot	DTD:t, XML Schema - määrittelyt, RDF Schema -määrittelyt, RDF-määrittelyt, ontologiat, aihekartat, annotoinnit, ...
URI, Unicode, XML, XML Namespaces, XML Schema, RDF, RDF Schema, XTM, XML-Signature, DAML+OIL, Annotea, ...	

## 6. RDF-metatietomalli

RDF = Resource Description Framework  
malli resurssien kuvaamiseen webissä

RDF-spesifikaatio: <http://www.w3.org/TR/REC-rdf-syntax/>

**resurssi** on mitä tahansa, mitä voidaan osoittaa internetissä; osoittaminen tapahtuu URI:llä

esim. tiedosto, palvelu, sivusto, tiedoston osa  
kirja, henkilö, yritys

# 6. RDF-metatietomalli

## Esimerkkejä resursseista

**resurssi**

**URI**

**kurssin kotisivu**

**<http://www.cs.jyu.fi/~airi/opetus/SemanttinenWeb.html>**

**TKTL**

**<http://cs.jyu.fi>**

**Airi Salminen**

**<http://cs.jyu.fi/henkilot/asalminen>**

**A.S.:n kotisivu**

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

## 6. RDF-metatietomalli

RDF-kuvauksessa resursseihin liitetään joukko ominaisuuksia. Ominaisuuden liittäminen resurssiin voidaan ilmaista kolmikolla:

(ominaisuus, resurssi, ominaisuuden arvo)

Esim.

(language, <http://www.cs.jyu.fi/~airi/opetus/SemanttinenWeb.html>, "fi")

## 6. RDF-metatietomalli

```
(dc:Creator,  
http://www.cs.jyu.fi/~airi/opetus/SemanttinenWeb.html,  
"Airi Salminen")  
(dc:Language,  
http://www.cs.jyu.fi/~airi/opetus/SemanttinenWeb.html,  
"fi")
```

### XML-syntaksi:

```
<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 about =  
    "http://www.cs.jyu.fi/~airi/opetus/SemanttinenWeb.html">  
    <dc:Creator>Airi Salminen</dc:Creator>  
    <dc:Language>fi</dc:Language>  
  </rdf:Description>  
</rdf:RDF>
```

## 6. RDF-metatietomalli

XML-syntaksissa täytyy ilmaista, missä nimiavaruudessa ominaisuuksien nimet on sovittu. Esimerkiksi

<http://purl.org/dc/elements/1.1/>

on URI-tunnus Dublin Coren nimistölle versiossa 1.1.

```
<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 about =
    "http://www.cs.jyu.fi/~airi/opetus/SemanttinenWeb.html">
    <dc:Creator>Airi Salminen</dc:Creator>
    <dc>Date>2002-04-05</dc>Date>
  </rdf:Description>
</rdf:RDF>
```

## 6. RDF-metatietomalli

Kurssin kotisivuun voitaisiin liittää seuraava kuvaus ulkoisena tiedostona:

```
<?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
        about="http://www.cs.jyu.fi/~airi/opetus/SemanttinenWeb.html">
<dc:title>Semanttinen web - Jyväskylän yliopisto</dc:title>
<dc:creator>Airi Salminen</dc:creator>
<dc:description>Home page of a course about the semantic web.
</dc:description>
<dc:description>Kotisivu kevään 2002 kurssille
TJT D63 Digitaalisen median erityiskysymyksiä, jonka teemana on
semanttinen web.</dc:description>
<dc:date>2002-04-05</dc:date>
<dc:format>text/html</dc:format> <dc:language>fi</dc:language>
<dc:publisher>Jyväskylän yliopisto</dc:publisher>
</rdf:Description> </rdf:RDF>
```

## 6. RDF-metatietomalli

**RDF Schema (RDFS)** -määrittelykielen avulla on mahdollista määritellä jotain tiettyä sovellusaluetta varten:

- mitä ominaisuuksia resursseihin voidaan liittää
- millaisia arvoja ominaisuuksilla voi olla

**RDFS-spesifikaatio on vasta Candidate**

**Recommendation -vaiheessa, ei W3C-suositus!**

**<http://www.w3.org/TR/2000/CR-rdf-schema-20000327/>**



## 6. RDF-metatietomalli

RDFS-määrittelyt ovat RDF-määrittelyjä ja ne voidaan kirjoittaa RDF:n XML-syntaksia käyttäen.

RDFS:llä on mahdollista määritellä luokkahierarkioita.  
Esim.

luokan **Person** määrittely:

<http://www.w3.org/TR/2000/CR-rdf-schema-20000327/#s7.1>

# 7. Ontologiat ja ontologiakielet

**Ontologia**

=

**yhteisesti sovittu, formaalisti esitetty, tietokoneen  
tulkittavaksi tarkoitettu käsitteistö**

# 7. Ontologiat ja ontologiakielet

## Esimerkkejä ontologioista:

- **YSA - Yleinen Suomalainen Asiasanasto**  
<http://vesa.lib.helsinki.fi/>
- **UNSPSC - Universal Standard Products and Services Classification**  
<http://eccma.org/unspsc/>
- **WordNet - Lexical database for the English language**  
<http://www.cogsci.princeton.edu/~wn/>
- **SHOE-ontologioita**  
<http://www.cs.umd.edu/projects/plus/SHOE/onts/index.html>

# 7. Ontologiat ja ontologiakielet

## Ontologiakieliä:

- **määrittelykieli ontologioiden määrittelyyn**
- **annotointikieli tietyn ontologian käsitteistön liittämiseen esimerkiksi html-sivuun**
- **kyselykieli annotoitujen dokumenttien hakemiseen**

# 7. Ontologiat ja ontologiakielet

## Ontologiakieliä:

- **RDF(S) = RDF + RDF Schema**
- **XOL (Ontology Exchange Language)**  
<http://www.ai.sri.com/pkarp/xol/>
- **SHOE**  
<http://www.cs.umd.edu/projects/plus/SHOE/>
- **OIL (Ontology Inference Layer)**  
<http://www.ontoknowledge.org/oil/>
- **DAML+OIL**  
<http://www.daml.org/2001/03/daml+oil-index>

## 8. Semanttisen webin sovellusalueita

- web-portaalit
- organisaation tai yhteisön tietämyksen hallinta
- elektroninen kaupankäynti
- agenttiohjelmat

# 8. Semanttisen webin sovellusalueita

## Web-portaalit

- tehokkaan portaalin rakentamiseksi tarvitaan käsitteistön yhdenmukaistamista
- yhdenmukaistaminen voidaan tehdä palveluihin liitetyillä metatiedoilla

# 8. Semanttisen webin sovellusalueita

## Tietämyksen hallinta

- Evidence-based Medical Information Service: lääketieteellistä tietoa saksankielisille lääkäreille  
<http://www.evimed.ch/>
- organisaatiomuisti ontologiakielellä
- julkishallinnon tietämyksen hallinta

e-GIF (electronic Government Interoperability Framework)

<http://www.govtalk.gov.uk/>

Suomalainen julkishallinnon metatietojen standardointi

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



# 8. Semanttisen webin sovellusalueita

## Elektroninen kaupankäynti

- Palvelujen yhteensovittaminen [TBG01], sisällönkuvailuun tarvitaan ontologioita
- UDDI - Universal Description, Discovery and Integration of Business for the Web  
<http://www.uddi.org/>
- RosettaNet: organisaatio, joka standardoi IT-alan kaupankäyntiprosesseja  
<http://www.rosettanet.org>

# 8. Semanttisen webin sovellusalueita

## Agenttiohjelmat

- Electric elves [CGK01]: työn organisointia auttavat agentit
- ostoagentit
- tiedonetsijäagentit, esimerkiksi WebWatcher

## 9. Yhteenveto

- **Globaalissa tietoverkossa käytetään dokumenteissa mitä moninaisimpia kieliä, samallakin sovellusalueella.**
- **Kielten rikkautta dokumenttisisällöissä ei haluta rajoittaa.**
- **Tarvitaan metataso, jolla kieli on huomattavasti kontorolloidumpaa kuin dokumenttisisällöissä.**

## 9. Yhteenveto

- Metatason määrittelyiden täytyy olla niin formaaleja, että ohjelmat pystyvät käyttämään määrittelyjä.
- Jotta sisältökielien rikkautta voitaisiin hyödyntää, metatasojakin täytyy määritellä eri kielille.
- Webin käytön tehostaminen edellyttää luottamuksen entistä parempaa hallintaa.