

Web ja semanttinen web organisaatioissa

Airi Salminen

Jyväskylän yliopisto

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

18.11.2003

Tampereen yliopisto

Sisältö

1. Webin käyttö organisaatioissa
2. Webin ongelmia
3. Osaratkaisuja: XML ja RDF
4. Semanttinen web: määrittelyjä
5. Tavoitteita semanttiselle webille
6. Webin tieto ja metatieto
7. Semanttisen webin tekniikoita
8. Mahdollisuuksia ja haasteita organisaatioissa
9. Yhteenveto

1. Webin käyttö organisaatioissa

World Wide Web syntyi Cernin hajallaan toimivien tutkijoiden ja tutkimusprojektien yhteistyön ja tietämyksen hallinnan tukemiseen.

WWW:n keskeiset tekniikat

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

1. Webin käyttö organisaatioissa

HTML

- ▶ Internetissä jaettavaksi tarkoitettujen dokumenttien esittämiseen standardoidussa muodossa
- ▶ kielessä mekanismi hypertekstilinkkien esittämiseen
- ▶ kieli on vapaasti kaikkien käytettävissä
- ▶ ymmärrettävissä sekä ihmisille että tietokoneelle
- ▶ HTML-selaajaohjelmat (web-selaajia, esim. Netscape, MS Internet Explorer, Opera) helposti saatavissa käyttöön; myös julkisohjelmia

1. Webin käyttö organisaatioissa

Web-teknologiaa on hyödynnetty organisaatioissa kolmella tasolla:

- ▶ Internet
- ▶ Extranet
- ▶ Intranet

1. Webin käyttö organisaatioissa

Internet

- ▶ Hajautettu dokumenttivarasto (HTML-dokumentit, teksti, kuvat, musiikki, video)
- ▶ julkisohjelmavarasto (HTML-selaajat, ...)
- ▶ palvelukanava (uutispalvelut, kauppapaikat, ...)
- ▶ jakelukanava (tiedottamiseen, markkinointiin, palvelujen tarjoamiseen, myymiseen)
- ▶ kommunikointikanava (sähköposti, uutisryhmät)

1. Webin käyttö organisaatioissa

Extranet

- ▶ organisaatioverkoston keskinäiseen tiedon, tietämyksen, dokumenttien, työprosessien hallintaan
- ▶ web-teknologia käytössä
- ▶ yhteys Internetiin
- ▶ käyttörajoitukset organisaatioverkoston sisällä

1. Webin käyttö organisaatioissa

Intranet

- ▶ organisaation sisäiseen tiedon, tietämyksen, dokumenttien, työprosessien hallintaan
- ▶ web-teknologia käytössä
- ▶ yhteys Internetiin
- ▶ käyttörajoitukset organisaation resursseihin

2. Webin ongelmia

- ▶ *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.

2. Webin ongelmia

- ▶ 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.

2. Webin ongelmia

- ▶ Internetin käyttäjien ja tietovarastojen 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.

XML = Extensible Markup Language

T. Bray, J. Paoli, C. M. Sperberg-McQueen, and E. Maler (Eds.), Extensible Markup Language (XML) 1.0 (Second Edition), W3C Recommendation 6 October 2000, <http://www.w3.org/TR/REC-xml>

RDF = Resource Description Framework

O. Lassila, & R. R. Swick (Eds.), Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation, Feb. 1999. <http://www.w3.org/TR/REC-rdf-syntax/>

3. Osaratkaisuja: XML ja RDF

XML

- ▶ Säännöt sovellusaluekohtaisten dokumenttityyppien määrittelyyn ja esittämiseen.
- ▶ Voidaan määritellä dokumenttien merkkauksessa käytettävässä nimistö ja dokumenttirakenteet.
- ▶ Nimistöön voidaan sisällyttää semantiikkaa.
- ▶ Maailman kaikilla kielillä kirjoitetuille sisällöille.

3. Osaratkaisuja: XML ja RDF

XML

- ▶ XML hyväksyttiin W3C:n suositukseksi kuten HTML:kin.
- ▶ Vapaasti käytettävissä, sekä ohjelmistokehittäjien että sisällöntuottajien.
- ▶ Pohjautuu jo vuonna 1986 standardoituun SGML-merkkaukieleen.

3. Osaratkaisuja: XML ja RDF

Dokumenttityyppimäärittely (DTD, Document Type Definition) runoille.

```
<!DOCTYPE runo [  
<!ELEMENT runo (info_linkki? otsikko?, sae+)>  
<!ATTLIST runo  
    tekija CDATA #REQUIRED  
    tekija_syntynyt CDATA #OMITTED>  
<!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) >]
```

3. Osaratkaisuja: XML ja RDF

```
<?xml version = "1.0"?>
<runo tekija = "Murasaki Shikibu" tekija_syntynyt = "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>
```

Huom: Elementtien rivi sisältö on otettu sivustosta <http://www.bopsecrets.org/rexroth/translations/japanese.htm>, missä on Kenneth Rexrothin japanilaisten runojen käännöksiä englanniksi.

3. Osaratkaisuja: XML ja RDF

**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](#)

3. Osaratkaisuja: XML ja RDF

Edellä näytettyyn tekstiin liittyy paljon metatietoa, joka on tallennettu sen XML-muotoon:

- ▶ teksti on 1-säkeinen, 4-rivinen runo
- ▶ runon tekijä on Murasaki Shikibu, joka on syntynyt vuonna 974
- ▶ runosta lähtee tyyppiä *simple* oleva linkki
- ▶ linkin kohde on osoitteessa
<http://digital.library.upenn.edu/women/omori/court/murasaki.html>

3. Osaratkaisuja: XML ja RDF

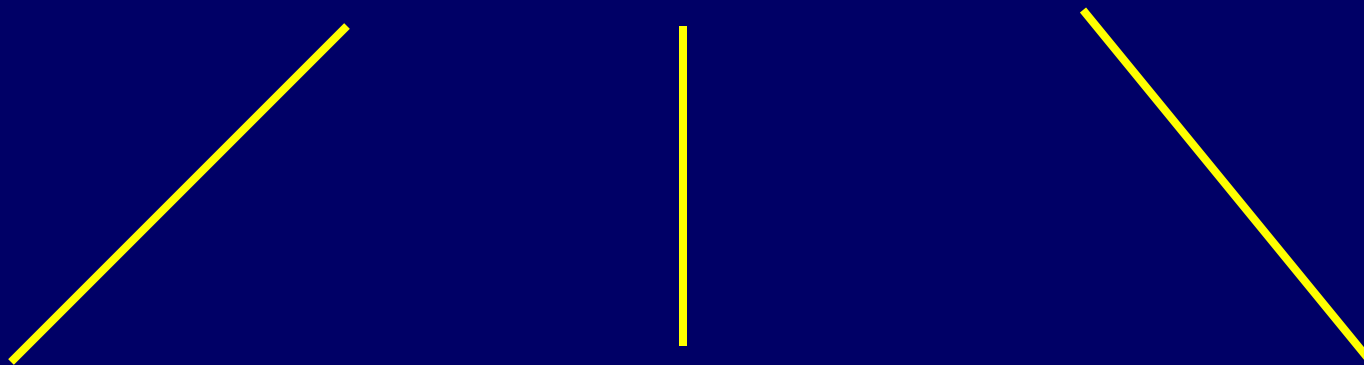
RDF

- ▶ Malli metatietojen liittämiseksi Internetin resursseihin.
- ▶ *Resurssi* on mitä tahansa, mitä voidaan osoittaa internetissä; osoittaminen tapahtuu URI:llä
- ▶ Esimerkkejä resursseista: tiedosto, palvelu, sivusto, tiedoston osa, kirja, henkilö, yritys.

3. Osaratkaisuja: XML ja RDF

RDF-kuvaus koostuu ominaisuuskuvauksista:

(ominaisuus, resurssi, ominaisuuden arvo)



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

3. Osaratkaisuja: XML ja RDF

RDF-kuvauksille on määritelty XML-syntaksi

RDF-kuvaus kolmikkoina:

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

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

3. Osaratkaisuja: XML ja RDF

- ▶ XML ja RDF auttavat organisaatioiden tiedon hallintaa tukevien ratkaisujen kehittämisessä.
- ▶ Käyttöönotto vaatii tietosisältöjen, niihin liitettävien metatietojen ja niiden käsittelytapojen standardointia.
- ▶ Tekniikoiden käyttöönotto yhden organisaation sisällä ei kuitenkaan riitä, tarvitaan organisaatioiden ja yhteisöjen välistä kommunikointia ja resurssien jakamista tukevia ratkaisuja.

3. Osaratkaisuja: XML ja RDF

- ▶ Eri yhteisöt määrittelevät toisistaan poikkeavia XML-kieliä samallekin sovellusalueelle.
- ▶ Eri yhteisöt määrittelevät toisistaan poikkeavia RDF-malleja samallekin sovellusalueelle.
- ▶ Luonnollista kieltä olevien sisältöjen merkityksen ymmärtäminen ohjelmien avulla ongelmallista.
- ▶ Resurssien yhdistäminen ongelmallista.

4. Semanttinen web: määrittelyjä

The abstract representation of data on the World Wide Web, based on the RDF standards and other standards to be defined. It is being developed by the W3C, in collaboration with a large number of researchers and industrial partners.

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

4. Semanttinen web: määrittelyjä

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.

<http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>

5. Tavoitteita semanttiselle webille

Moniarvoiset Internetin tietosisällöt

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

5. 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.
- ▶ Paikallisen ja globaalin tiedon joustava yhdistäminen.

5. 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.

5. Tavoitteita semanttiselle webille

Keskeisiä ominaisuuksia:

- ▶ Koostuu primääriresursseista ja metatietoresursseista.
- ▶ Resurssien merkitykseen, käyttöön ja luotettavuuteen liittyviä metatietoja.
- ▶ Metatiedot standardoidussa muodossa, sekä ihmisten että ohjelmien luettavissa.

5. Tavoitteita semanttiselle webille

Keskeisiä ominaisuuksia:

- ▶ Kehitystä koordinoi W3C.
- ▶ XML sekä primääri- että metatietoresurssien esitysmuoto.
- ▶ RDF metatietojen esitysmuoto.
- ▶ Automaattista päättelyä merkityksestä ja luotettavuudesta.

5. Tavoitteita semanttiselle webille

Keskeisiä ominaisuuksia:

- ▶ Monipuolinen ohjelmistojen yhteistyö ja kommunikointi mahdollista.
- ▶ Mahdollistaa ihmisten monipuolisen yhteistyön yhteisöissä, missä on yhdenmukainen käsitys sisältöjen merkityksestä ja yhteiset arvot.

5. Tavoitteita semanttiselle webille

Keskeisiä ominaisuuksia:

- ▶ Vaatii ihmisten monipuolista yhteistyötä.
- ▶ Vaatii halua ja kykyä sopia yhteisön käsitteistöstä ja luotettavista web-toimintatavoista.
- ▶ Vaatii eri alueiden asiantuntijoiden (kohdealue, tietohallinto, tietopalvelut, tekniikka) halua ja kykyä yhteistyöhön.

6. Webin tieto ja metatieto

metatieto

**tietoa joka auttaa käyttämään,
ymmärtämään, prosessoimaan ja
yhdistämään webin tietoresursseja**

6. Webin tieto ja metatieto

Metatieto voi kuvata esimerkiksi dokumentista

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

6. Webin tieto ja metatieto

Vaihtoehtoja metatiedoille

kuvaa webin resurssia	kuvaa webin ulkopuolista asiaa
upotettu	ulkoinen
keskitetty	hajautettu
ihmisen tuottama	automaattisesti tuotettu

6. 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.Description" content="Home page of Professor Airi
    Salminen, her representative publications, research interest, projects,
    courses, and contact information">
  <meta name="DC.Date.Updated" content="2002-12-10">
  <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">
</head>
```

XML-kieli metatiedon esittämisessä

- **Merkkaus (markup) sisältää metatietoa dokumentin rakenteesta ja elementtien merkityksestä.**
- **Dokumenttityypin määritys (DTD) sisältää tietoa dokumentin rakenteesta.**
- **XML-dokumentin sisältö voi koostua johonkin asiaan liittyvistä metatiedoista, esimerkiksi julkaisujen bibliografisista tiedoista.**

Ontologia

- ▶ Yhteisön tietovarantojen merkitykseen liittyvää metatietoa.
- ▶ Yhteisesti sovittu käsitteistö (käsitteet, niiden väliset suhteet).
- ▶ Tietokoneen tulkittavaksi tarkoitettu.
- ▶ Mahdollisesti päättelysääntöjä.

6. Webin tieto ja metatieto

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>

6. Webin tieto ja metatieto

Ontologiakieliä:

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

7. Semanttisen webin tekniikoita

Semanttisen webin pelkistetty rakenne



7. 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

7. Semanttisen webin tekniikoita

Tekniikoita metatiedon esittämiseen ja määrittelyyn

- metatiedon esittäminen: RDF
- metatietoesitystapojen määrittely: RDF Schema
- aihekartat: Topic Maps (XTM)
- ontologian määrittely: Web Ontology Language (OWL)

7. Semanttisen webin tekniikoita

Tekniikoita luottamuksen hallintaan

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

7. 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, OWL, Annotea, ...	

8. Mahdollisuuksia ja haasteita organisaatioissa

Sovellusalueita:

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

8. Mahdollisuuksia ja haasteita organisaatioissa

Web-portaalit

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

8. Mahdollisuuksia ja haasteita organisaatioissa

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

8. Mahdollisuuksia ja haasteita organisaatioissa

Elektroninen kaupankäynti

- Palvelujen yhteensovittaminen, 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. Mahdollisuuksia ja haasteita organisaatioissa

Agenttiohjelmat

- työn organisointia auttavat agentit
- ostoagentit
- tiedonetsijäagentit, esimerkiksi WebWatcher

8. Mahdollisuuksia ja haasteita organisaatioissa

Haasteita

- Metatiedon standardoiminen yhteisössä ongelmallista; vaatii paljon työtä.
- Luottamukseen ja tietoturvaan liittyvät ongelmat erittäin vaikeita.
- Yhteistyön käynnistäminen ja ylläpitäminen eri alueiden asiantuntijoiden kesken vaikeaa.

9. Yhteenveto

- Globaalissa tietoverkossa käytetään dokumenteissa mitä moninaisimpia kieliä, samallakin sovellusalueella.
- Kielien 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 ja eri yhteisöille.
- Webin käytön tehostaminen edellyttää luottamuksen entistä parempaa hallintaa.

9. Yhteenveto

Vaikutuksia organisaation tietopalveluun

- Automaatio lisääntyy (hakuagentit, kysymyksiin vastaavat järjestelmät, päättelyt).
- Luottamuksen hallinnan merkitys kasvaa.
- Työ suuntautuu entistä enemmän tiedon hakemisesta suunnitteluun ja yhteistyöhön eri alueiden asiantuntijoiden kesken.
- Tietopalvelujen asiantuntemus tärkeää semanttisen webin kehittämisessä.

Lähteitä

- Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H.F., Secret, A., The World-Wide Web. *Communications of the ACM* 37, 8, August 1994, 76-82.
- Berners-Lee, T., Hendler, J., & Lassila, O., The Semantic Web. *Scientific American*, May 2001.
<http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>
- *Communications of the ACM*, Vol. 45, No. 2, 2002, Special Section on the Ontologies.
- Hyvönen, E., Semantic Web - kohti uutta merkitysten Internetiä, esitelmä Semantic Web Kick-Off in Finland -tilaisuudessa 2.11.2001.
<http://www.cs.helsinki.fi/u/eahyvone/stes/semanticweb/SemanticWebVisio.PDF>

Lähteitä

- Kobayashi, M. & Takeda, K., Information retrieval on the Web. *ACM Computing Surveys* 32, 2 (June 2000), 144-173.
- Salminen, A., XML Family of Languages. Overview and Classification of W3C Specifications. <http://www.cs.jyu.fi/~airi/xmlfamily.html>.

Lisää lähteitä:

Salminen, A., Kevään 2002 semanttinen web -kurssin kotisivu, Jyväskylän yliopisto.
<http://www.cs.jyu.fi/~airi/opetus/SemanttinenWeb.html>