

Semanttinen Web

Ossi Nykänen

Tampereen teknillinen yliopisto (TTY),
Digitaalisen median instituutti (DMI),
Hypermedialaboratorio
W3C Suomen toimisto



- Hyvin lyhyt versio:
 - Semanttinen Web (SW) on järjestelmä jonka päämääränä on mallintaa ja esittää maailmanlaajuisesti semanttista kuvailutietoa tietokoneohjelmien ymmärtämässä muodossa. SW:n ytimen määrittelee joukko W3C:n asettamia suosituksia. W3C eli World Wide Web Consortium asettaa SW:n ohella muitakin Web-perustekniikoita.
- Pidempi versio, otsikoita:
 - W3C ja universaalit Web-standardit
 - Miksi Semanttinen Web? (Ja miten?)
 - Perustekniikkaa ja ideoita
 - Lopuksi
- Esityksen tavoite on esitellä ja tehdä ymmärrettäväksi Semanttisen Webin perusidea, tekniikan yksityiskohdat ohittaen

**World Wide Web Consortium
(W3C) ja universaalit
Web-standardit**

- W3C kehittää yhteensopivia teknologioita ja siten *ohjaa Webin kehittymistä täyteen mittaansa*
 - ...asettamalla teknisiä **suosituksia** (esim. HTML, XML, WAI)
- 3 päätstoa, 15 aluetstoa, n. 360 jäsenorganisaatiota



Jäseneksi? <http://www.w3c.tut.fi/joining.html>



Applications

Interaction

XHTML, CSS, SVG, SMIL, CDF, XForms, MathML, InkML, ...

Mobile Web

XHTML Basic, Mobile SVG, SMIL Mobile, XForms Basic, CC/PP, CDF, ...

Voice

VoiceXML, SRGS, SSML, CCXML, EMMA, ...

Web Services

SOAP, XOP, WSDL, WS-CDL, Addressing, ...

Semantic Web

OWL, SKOS, ...

Privacy, Security

P3P, APPEL, XML Sig, XML Enc, XKMS, ...

Web Accessibility, Internationalization, Device Independence, Quality Assurance

XML, Namespaces, Schemas, XQuery/XPath, XSLT, DOM, XML Base, XPointer, ...; RDF/XML, SPARQL, ...

XML Infoset; RDF Graph

Web Architectural Principles

URI/IRI, HTTP

The Web

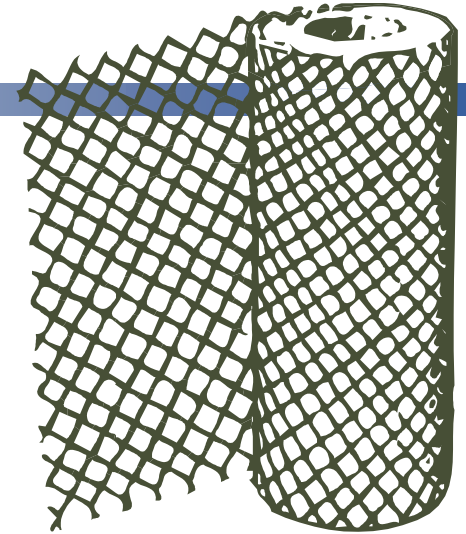
Internet

Miksi Semanttinen Web? (Ja miten?)

Mitä tietokoneet eivät osaa tehdä (hyvin)?

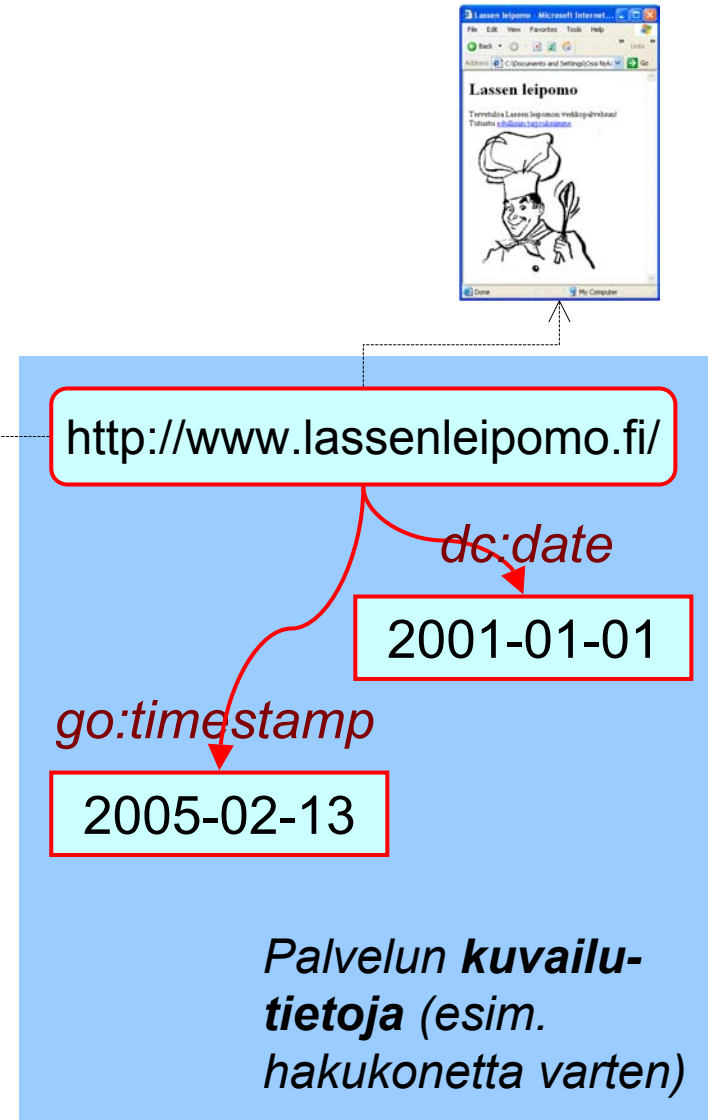
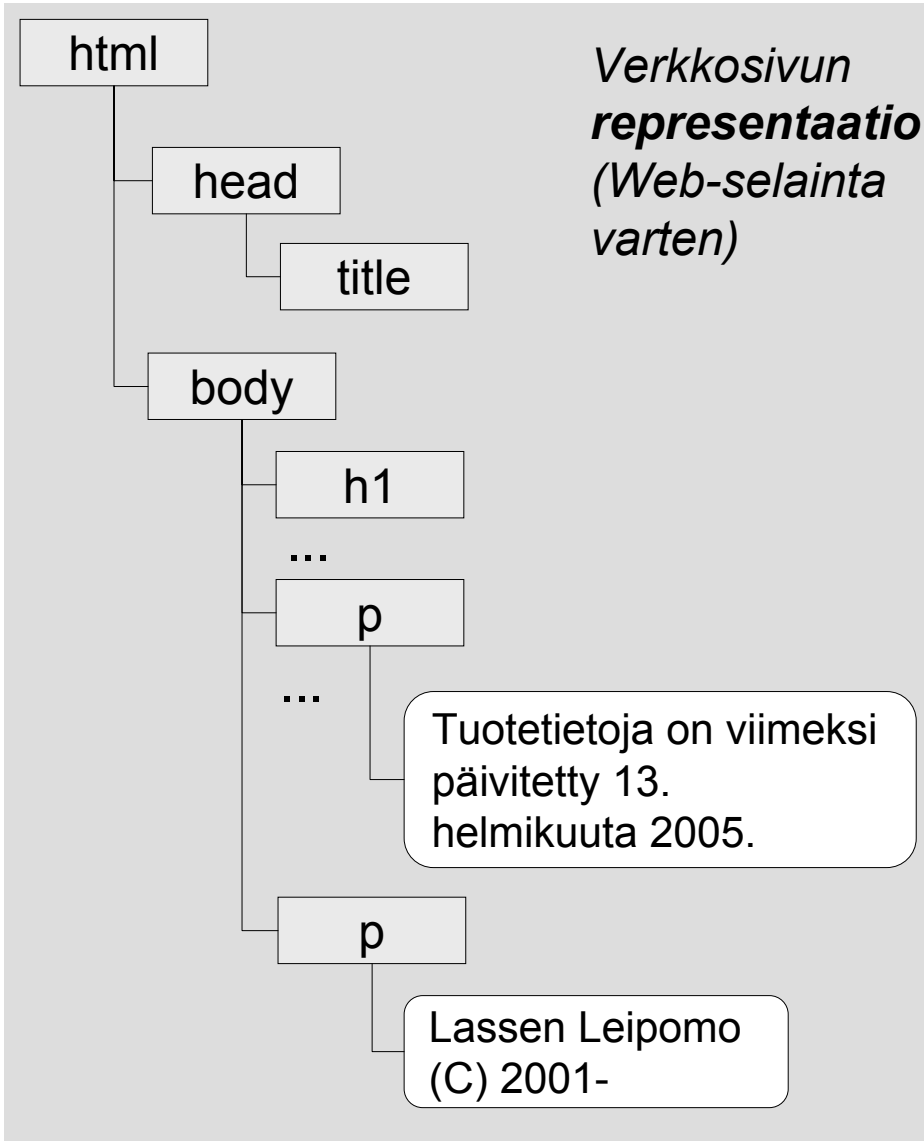
- **Jäsentää** luonnollista tekstiä
 - "Edustamamme Lassen Leipomo Oy:n (Tampere) puhelinnumero on 123 456."
- **Tulkita** kontekstuaalista tietoa
 - "Edellä mainittu yritys harjoittaa leipomotoimintaa Tampereella."
- **Yhdistellä** eri tavoin kuvattua tietoa
 - "Se-ja-se Oy:n puhelinnumero on (03) 123 456."
 - "The phone number of the Acme Inc. is +48 3 789 000."
- Yo. pulmat hankaloittavat esim.
 - tiedon yhdistämistä ja hakemista
 - tietojenkäsittelyn automatisointia
 - tiedon havainnollistamista, automatisoitua päättelyä, jne.





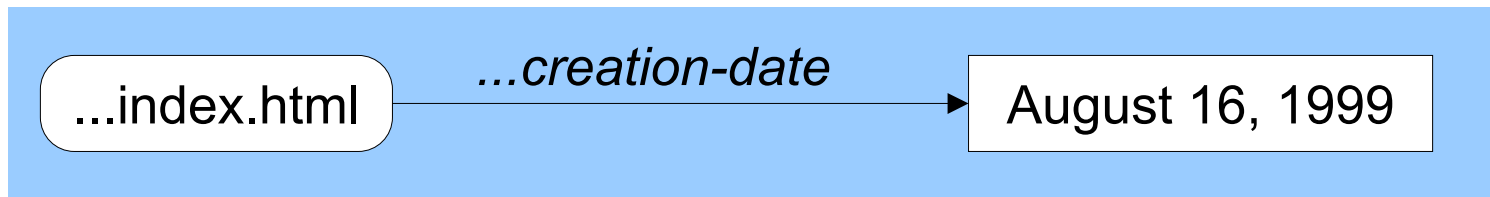
- Jäsennetään ja kirjataan tieto sellaisessa muodossa että myös tietokoneohjelmat osaavat tulkita sitä
- Käytännössä (Web-kontekstissa)
 - sovitaan **täsmällinen lauserakenne** semanttisen tiedon välittämiseen
"Lassen leipomon puhelinnumero on **+358 3 123 456**."
 - sovitaan että lauseissa esiintyvät **nimet** (**subjektit**, **predikaatit** ja **objektit**) voidaan koodata pitkien URI-nimien avulla; näin *hajautettu nimien kehitystyö* on mahdollista
 - sovitaan **skeemojen** ja **käsitemallien** esittämisen pelisäännöistä (luokka, aliluokka, predikaatti, transitiivinen predikaatti, ...)
- Teknisesti kyse on kuvailu- tai metatiedosta
...Semanttinen Web

Ero esim. HTML-julkaisuun verrattuna:



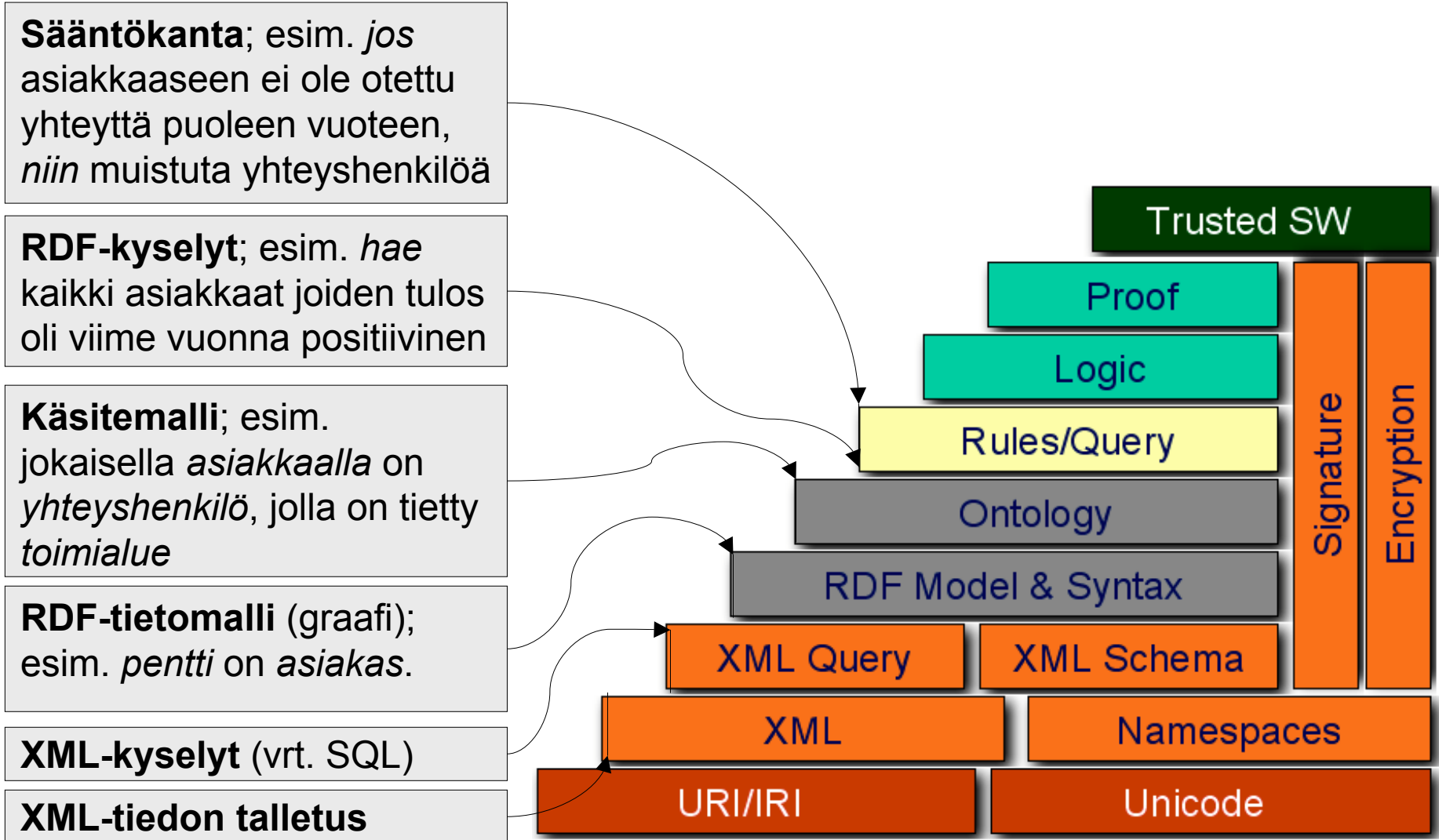
Perustekniikkaa ja ideoita

- **Resource Description Framework (RDF)** määrittelee miten kuvailutietoa mallinnetaan ja julkaistaan

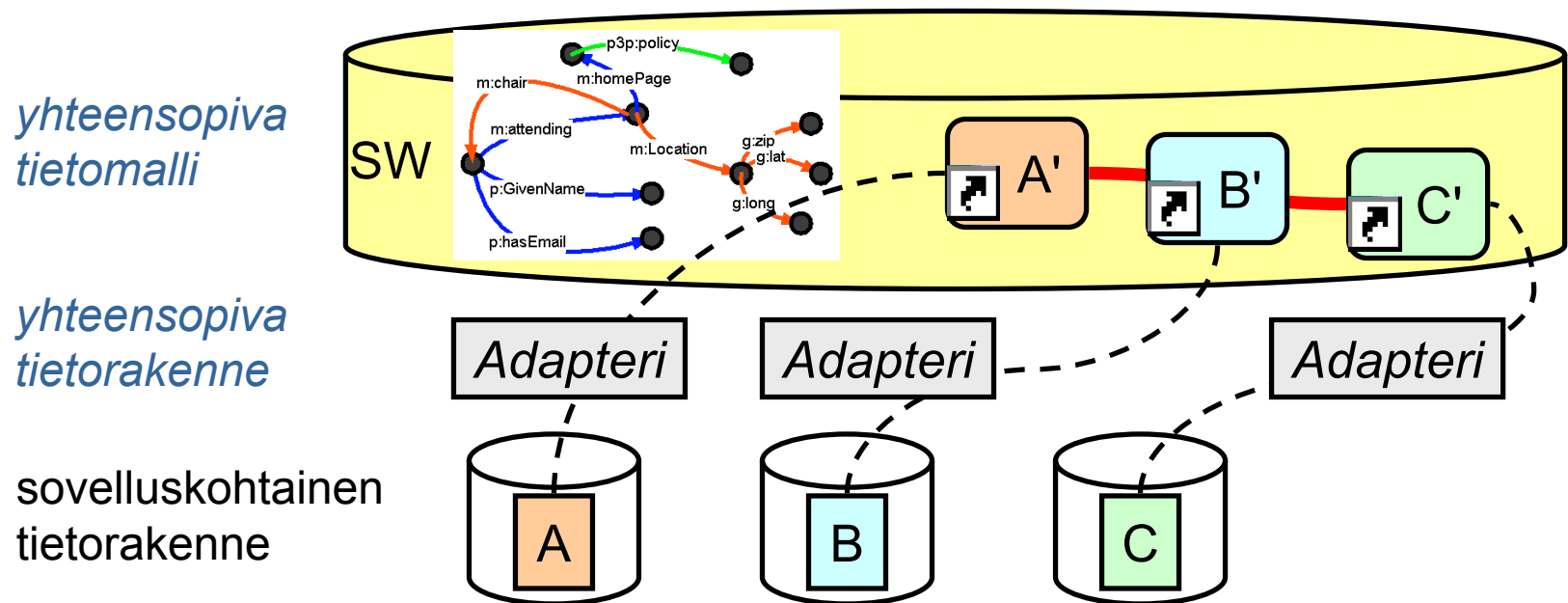


```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf=
  "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:exterms="http://www.example.org/terms/">
  <rdf:Description
    rdf:about="http://www.example.org/index.html">
    <exterms:creation-date>August 16, 1999</exterms:creation-date>
  </rdf:Description>
</rdf:RDF>
```

- XML-syntaksia tärkeämpää on ymmärtää mallin idea



- Kyse on siis pitkälti **"universaalin tietokannan"** määrittelystä



- Idea ei siis ole "siirtää tiedostoja Semanttiseen Webiin", vaan **kuvata jo olemassa olevia tietoja** – ts. kyse ei ole fyysisestä vaan *loogisesta rakenteesta*

- **DC** on sovittu standardi esim. verkkosivujen kuvailemisessa käytettävistä predikaateista (attribuutit)
- Sisältää seuraavat sanat (ja std-tulkinnan näille):
 - **Title**: A name given to the resource.
 - **Creator**: An entity primarily responsible for making the content of the resource.
 - **Subject**: The topic of the content of the resource.
 - **Description**: An account of the content of the resource.
 - **Publisher**: An entity responsible for making the resource available
 - **Contributor**: An entity responsible for making contributions to the content of the resource.
 - **Date**: A date associated with an event in the life cycle of the resource.
 - **Type**: The nature or genre of the content of the resource.
 - **Format**: The physical or digital manifestation of the resource.
 - **Identifier**: An unambiguous reference to the resource within a given context.
 - **Source**: A reference to a resource from which the present resource is derived.
 - **Language**: A language of the intellectual content of the resource.
 - **Relation**: A reference to a related resource.
 - **Coverage**: The extent or scope of the content of the resource.
 - **Rights**: Information about rights held in and over the resource.

(RSS = RDF Site Summary)

Syndikointi



<http://www.w3c.tut.fi/home.rss>

otsikko: **uutisia, joulukuu**
 uutisotsikko1: ○
 uutisotsikko2: ○ ...
 uutisotsikko3: ○ ...
 ...

otsikko: **XForms**
 1.2 Mobile
 teksti: **W3C**
 julkaisi uuden
 version ...



<http://www.oreillynet.com/meerkat/>

- Sovitaan yhteinen tapa julkaista **listamuotoisia tiedotteita** (uutisia, ruokalistoja, työpaikkoja, päivityksiä...)
- ...tiedon hakeminen ja yhdistely, filterointi, ...

Lopuksi

- **World Wide Web Consortium** (W3C) kehittää Web-standardeja
- **Semanttinen Webin** perusidea on asettaa universaali tietomalli (meta)tietojen yhtenäisen kuvailun ja abstraktin käsittelyn tueksi
- **Kuvailukehikon** ohella tarvitaan myös yhteisesti sovittuja **sanastoja** ja **käsitemalleja**

Jäikö joku W3C-asia mietityttämään?

Allekirjoittaneen tavoittaa helposti:

<http://www.w3c.tut.fi>

Ossi@w3.org



- W3C
 - <http://www.w3.org/> (*kotisivu*)
 - <http://www.w3c.tut.fi/> (*W3C Suomen toimiston kotisivu*)
- W3C-aktiviteetit
 - <http://www.w3.org/Consortium/Activities> (*yleiskuva*)
 - <http://www.w3.org/2001/sw> (*Semanttinen Web*)
- W3C:n suositukset, raportit ja teknistä tietoa
 - <http://www.w3c.org/TR/> (*kaikki tekniset dokumentit*)
 - <http://www.w3c.tut.fi/report-archive.html> (*yleisesityksiä suomeksi*)
- Jäseneksi!
 - <http://www.w3.org/Consortium/Prospectus>
 - <http://www.w3c.tut.fi/joining.html>

- Keskeiset SW-spesifikaatiot ovat saaneet W3C-suosituksen (standardin) statuksen vuonna 2004
 - 10 February 2004: OWL Web Ontology Language Overview
 - 10 February 2004: OWL Web Ontology Language Guide
 - 10 February 2004: OWL Web Ontology Language Reference
 - 10 February 2004: OWL Web Ontology Language Semantics and Abstract Syntax
 - 10 February 2004: OWL Web Ontology Language Test Cases
 - 10 February 2004: Resource Description Framework (RDF): Concepts and Abstract Syntax
 - 10 February 2004: RDF Semantics
 - **10 February 2004: RDF Primer**
 - 10 February 2004: RDF Vocabulary Description Language 1.0: RDF Schema
 - 10 February 2004: RDF/XML Syntax Specification (Revised)
 - 10 February 2004: RDF Test Cases
 - 10 February 2004: OWL Web Ontology Language Use Cases and Requirements
- Standardointi on vielä kesken mm. kysely- ja sääntökielten osalta
 - ks. <http://www.w3c.org/TR/> ja <http://www.w3c.org/2001/sw/>