

Internet jolla on merkitystä

Semanttinen Web –visio, kehitys ja tulevaisuus

Ora Lassila

Nokia Services

USA

ICT 2010 avauspuheenvuoro

Keskiviikko 14.4. klo 8:30

Olen puhunut tästä jo aiemmin..



- Onko sanoma muuttunut?
- Kannattaako kuunnella?

Tästä aion puhua tänään

1. Semanttinen Web – alkuperäinen visio ja analyysi
2. Viimeaikainen kehitys – standardit, ym.
3. Uudet haasteet ja vaatimukset
4. Mobiilipalvelut ja Semanttinen Web

Semanttinen Web

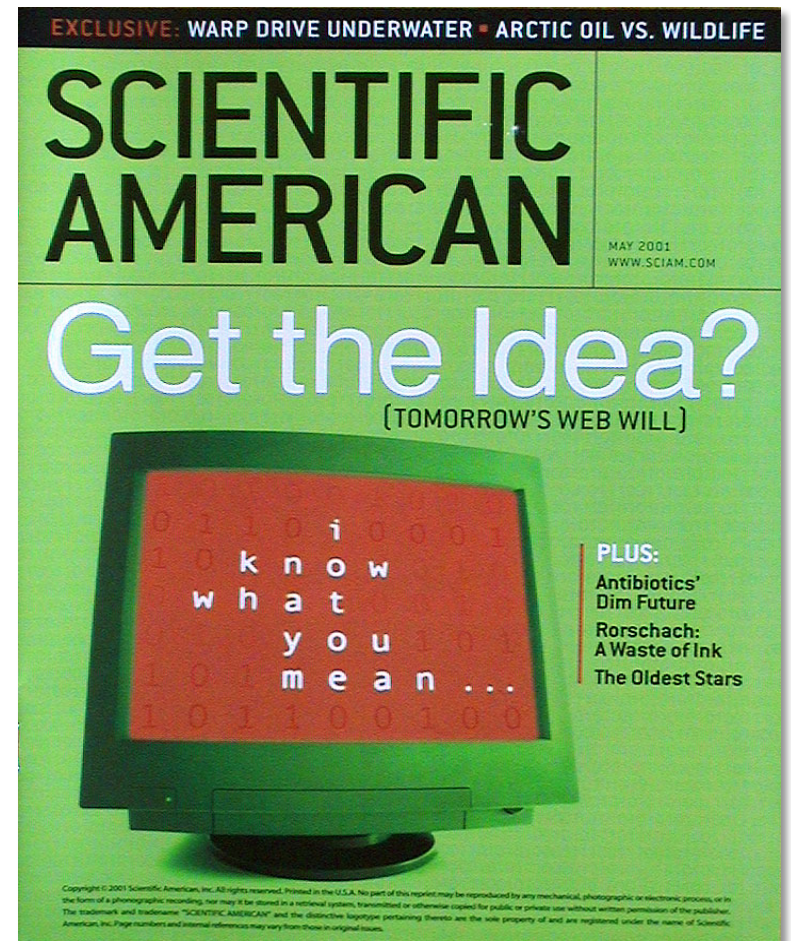
– alkuperäinen visio

Havaintoja

- Modernit informaatiojärjestelmät (esim. nettipalvelut) päättävät monia asioita “puolestasi”:
 - kuinka informaatio **esitetään**
 - kuinka informaatiota voi käyttää (esim. yhdistellä)
 - mitkä toiminnot automatisoidaan
- Ennalta päätetty esitys vaikeuttaa palvelujen tarjoamista
 - vaikkapa näkövammaisille
 - ”epästandardin” päätelaitteen kautta (esim. mobiililaite)
- Eri tietolähteiden yhdistely on vaikeaa
 - tämän automatisointi on erittäin vaikeaa
 - ennalta aavistamattomien tilanteiden automatisointi on **käytännössä mahdotonta**

Semanttinen Web...?

- Semanttinen Web on visio Internetin tulevaisuudesta
- Internetin sisältö on nyt tarkoitettu lähinnä ihmisille
 - tehtävien automatisointi on vaikeaa, edellyttää ihmisen tulkintaa
- Mitä tarvitaan?
 - tietämyksen esittämistä, ontologioita, tekoälyä



[Berners-Lee, Hendler, Lassila 2001]

Semanttinen Web...?

- ~~• Semanttinen Web on visio Internetin tulevaisuudesta~~
- ~~• Internetin sisältö on nyt tarkoitettu lähinnä ihmisille
 - ~~• tehtävien automatisointi on vaikeaa, edellyttää ihmisen tulkintaa~~~~
- ~~• Mitä tarvitaan?
 - ~~• tietämyksen esittämistä, ontologioita, tekoälyä~~~~
- Visio **tietojenkäsittelyn** tulevaisuudesta
- Tietotekniikkaa käytetään nyt työkaluna (vertaa esimerkiksi vasaraan)
 - koneet voisivat suorittaa tehtäviä **käyttäjien puolesta**
- Mitä tarvitaan?
 - tietämyksen esittämistä, ontologioita, tekoälyä

Mikä on "ontologia"...

- Rikas tietomalli joka määrittelee
 - *käsitteitä* ja niiden *attribuutteja*
 - käsitteiden välisiä *suhteita*
 - (mahdollisesti myös käsitteiden *ilmentymiä*)
- Filosofiasa ontologiat liittyvät metafysiikkaan
- Tietojenkäsittelyopissa ontologiat ovat olleet tietämyksen esittämisen perusta
 - ontologioihin voi liittyä päättelysääntöjä joilla olemassa olevasta tiedosta voi päätellä lisää

**Esimerkki
päättelystä:**

*"Musti" on koira,
koirat ovat nisäkkäitä*

⇒ *kaikki mitä tiedämme nisäkkäistä
pätee myös Mustiin*

Alkuperäisen vision “dekonstruktio”

- Visio perustuu seuraaviin olettamuksiin:
 - tarvitaan ontologioita ja loogista päättelyä
 - tarvitaan **paljon** dataa – tyypillisesti monista eri lähteistä
- Agentit tai muut ohjelmistojärjestelmät
 - noutavat dataa eri lähteistä
 - yhdistelevät ja tulkitsevat dataa (ontologioiden avulla)
 - vetävät käyttäjää hyödyttäviä johtopäätöksiä
- Käyttäjälle voidaan antaa suurempi kontrolli
 - siihen miten dataa voi käyttää ja yhdistellä
 - siihen miten data esitetään ja miten sen kanssa vuorovaikutetaan

Runsaasti kritiikkiä..

Eikö XML ole jo ratkaisu tähän ongelmaan?

Formaatit eivät ole olennainen osa ratkaisua. Yhteinen formaatti on hyödyllinen, mutta todelliset ongelmat ovat muualla.

Emmekö voisi vain ohjelmoida kaikkia tarvittavia toimintoja?

Kaikkien tulevaisuuden tarpeiden ja käyttötilanteiden ennalta aavistaminen ei liene realistinen vaihtoehto.

Mistä saadaan riittävästi dataa? Miten voidaan varmistaa oikeudet dataan?

Maailmassa on runsaasti dataa, jopa sellaista dataa jonka formaali semantiikka tunnetaan. Tietoturvakysymykset ovat vaikeampia...

Voidaanko olettaa että jostakin löytyy kaiken kattava ontologia?

Tätä ei tarvita. Visio on rakennettu sen varaan että "annetaan kaikkien kukkien kukkia".

Haiskahtaa tekoälyltä...

Olet oikeassa!

Vaikea kysymys:

- Tarkasti määritellylle ongelmalle löydetään tyypillisesti tarkasti määritelty ratkaisu joka ei edellytä Semanttisen Webin tekniikoiden käyttöä
- Miksi sitten Semanttinen Web on kiinnostava alue?

Semanttinen Web voi olla ratkaisu
niihin ongelmiin joita ei vielä ole
tarkasti määritelty!

W3C:n standardit “rakennuspalikat”

- **RDF: Resource Description Framework**
 - rakenteinen tietomalli
+ yksinkertainen ontologiakieli
- **OWL: Web Ontology Language**
 - RDF:ää ilmaisuvoimaisempi ontologiakieli
- **SPARQL**
 - kyselykieli RDF:ää varten
- **RIF: Rule Interchange Format – (työ vielä käynnissä)**
 - Semanttisen Webin ontologiakielten kanssa yhteensopiva sääntökieli

Nokia osallistui merkittävällä panoksella RDF:n ja OWL:in määrittelyyn

Semanttinen Web ja Suomi

- Nokia
 - Nokia Research Center
 - Forum Nokia, Nokia Services
- TKK/Aalto: *Semantic Computing Research Group*
 - FinnONTO-ohjelma, ...
- Jyväskylä: *Industrial Ontologies Group*
- Profium Oy
- ...



Forum.Nokia.com
Driving mobile innovation



 profium

NOKIA
Connecting People

Semanttinen Web

– viimeaikainen kehitys

Uusia W3C-standardeja

- **RDFa**
 - tapa sisällyttää RDF:ää XHTML-dokumentteihin
- **GRDDL: Gleaning Resource Descriptions from Dialects of Languages**
 - muiden formaattien muuttaminen RDF:ksi
 - perustuu XSLT:n käyttöön
- **POWDER: Protocol for Web Description Resources**
 - miten yhdistetään semanttista metadataa nettiresursseihin
- **SKOS: Simple Knowledge Organization System**
 - RDF-skeema taksonomioiden ja tesaurusten määrittelyyn
- Viime aikoina on julkaistu useita ontologioita ja skeemoja joista on tullut *de facto* -standardeja

Semanttinen Web vai "*semanttinen web*"...?

- Tekoälyn termein: "neat" vai "scruffy"...?
- Microformaatit: askel kohti todellista Semanttista Webiä?
 - GRDDL ja XSLT pelastavat tilanteen
 - tärkeitä asioita puuttuu: yhteinen metamalli, laajennettavuus
- Miten integroidaan folksonomiat ja ontologiat?

Tekoälyn eliminointi

- Tekoälyn osuutta alkuperäisessä visiossa on kritisoitu
- Tekoälyllä on "huono maine"
 - muoti-ilmiö 1980-luvulla, tämän jälkeen koitti "tekoälytalvi"
- Monet tekoälytekniikat ovat normaali osa tietojenkäsittelyä
 - (eivätkä siis mitään taikatemppuja...)
- Tekoälyn poistaminen Semanttisen Webin visiosta on mahdollista, mutta vain tilapäisesti

Uusi kehityssuunta: "Linked Data"

- "Tekoälyvapaa" ratkaisu, keskittyy lähinnä siihen miten eri tietolähteet voidaan linkittää toisiinsa
 - "Web of Data"
- Yhä enemmän dataa julkisesti tarjolla
 - "Linked Open Data"
 - data.gov.uk, DBpedia, jne.
- Olennaista on se että meillä on
 - yhteinen formaatti
 - yhteinen tapa identifioida käsitteet ja ilmentymät
 - selkeä käsitys siitä mitä esimerkiksi HTTP 303 tarkoittaa



"Open Your Data!"

Uudet haasteet *ja vaatimukset*

Haasteita

Ongelma	Tilanne	Oma mielipide
Ilmaisuvoima	Paljon valinnanvaraa	"Vähemmän on enemmän"
"Provenienssi"	Paljon valinnanvaraa	Tarvitaan selkeä standardi
Skaalautuvuus	Haasteellinen	Pahin kompastuskivi "myyntityössä"
Tietoturva ja tietosuoja	Haasteellinen	Helppo tietojen yhdistely voi johtaa tietosuojariskeihin
Ohjelmointimalli	Erittäin haasteellinen	Parempi integrointi ohjelmointikielten kanssa?
Käyttöliittymät	Olemme aivan eksyksissä...	

Skaalautuvuus

- RDF:n tietomalli perustuu suunnattuihin verkkoihin joiden kaaret on nimetty
 - jokaista kaarta vastaa "kolmikko" (alkusolmu, nimiö, loppusolmu)
 - kolmikoiden määrä on tapa mitata verkon kokoa
- Isoissa sovelluksissa tarvittava verkon koko voi asettaa merkittäviä haasteita tietokannan koolle
 - esim. 50 miljoonaa käyttäjää, 200 kolmikkoa per käyttäjä
= 10 miljardia kolmikkoa
- Kolmikoiden talletus relaatiotietokantaan johtaa usein kyselyihin joissa on monimutkaisia liitosoperaatioita
- RDF:n metamallin joustavuus tekee kolmikkotietokannoista vaikeasti optimoitavia

Tietosuoja ja tietoturva

- Helppo tietojen yhdistely sekä selkeä käsitteiden ja ilmentymien identifiointi tekevät helpommaksi muodostaa ison kokonaiskuvan
 - onko tämä uhka käyttäjän tietosuojalle?
- Tietoturvasääntöjen skaalautuvuus voi olla ongelma
 - (esim. Mika Mannermaan diplomityö, TKK/Aalto 2010)
- Miten määritellään tietoturvasäännöt?
 - käyttöliittymät ja käyttökokemus ovat haasteellisia

Mobiilipalvelut *ja Semanttinen Web*

Nettipalvelut: PC vai mobiililaite?

- Alkuperäinen visio:** Mobiililaitteet voisivat toimia käyttöliittymänä agenteille jotka suorittavat tehtäviä **käyttäjän puolesta**
- Nykysuuntaus:** Mobiililaitteet ovat enenevässä määrin PC:n kaltaisia (hyvä nettiselain, nopea verkko, ym.)
- Realiteetti:** Käyttökokemus ei kuitenkaan ole mobiililaitteessa sama kuin PC:ssä...

Miten Semanttisen Webin teknologiat ovat hyödyllisiä *tänään*?

- Palvelujen integrointi ja "tiedon jakaminen" ovat tärkeitä
- Semanttisen Webin teknologiat mahdollistavat
 - joustavan tiedon yhdistelyn ja integroinnin
 - joustavuuden tietomallien ja -skeemojen suhteen
- Käyttäjiä ei kiinnosta se miten vaikeita asiat ovat toteuttaa
 - heitä kiinnostaa helppo ja johdonmukainen käyttökokemus
 - traditionaaliset "datasiilot" johtavat keinotekoiseen rajoitukseen käyttökokemuksen suhteen – Semanttinen Web on tie ulos siiloista
- "Perceptual convergence"

Kiitos!

- Yhteydenotot: ora.lassila@nokia.com