




Semantic Web käytännön sovelluksissa

TkT Janne Saarela
Profium Oy

26.5.2004

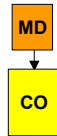


Sisällysluettelo

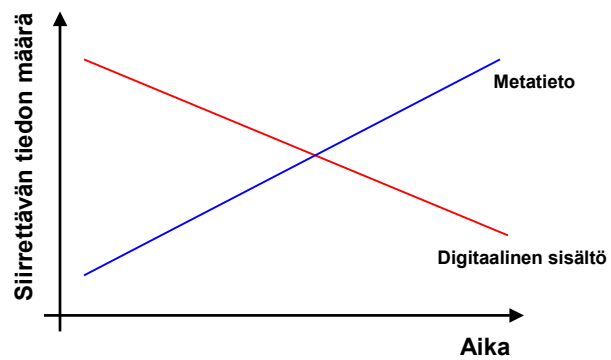
- Johdanto Semanttisen Webin maailmaan
- Mahdollisuudet
- Tämän päivän käyttökohteet
- Haasteet



Johdanto Semanttisen Webin maailmaan

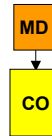


Digitaalinen sisältö ja metatieto



Metatieto

- Metatieto, mitä se on?
 - tietoa tiedosta
 - tietoa tietosisällöstä ja sen suhteesta muuhun sisältöön
 - saatekaaviot ja kirjaston arkistointikortit kuvailevat sisältöä, mutta ne on **tarkoitettu ihmisille -> ihminen osallistuu sisällönhallintaan**
 - tietokoneita varten tarvitaan formalisoidut sisältökuvaukset
-> **tietokone voi automatisoida sisällönhallintaa**
- Esimerkiksi:
 - kuvien automaattiseen hallintaan tarvitaan tieto kuvan sisällöstä **tietokoneen ymmärtämässä muodossa**
 - pelkkä avainsanojen käyttö ei vielä riitä (JPG-kuvaan liittyy sana 'Nokia')
 - Metatieto kuvaa merkityksen (semantiikan):
 - (kuva.jpg, 'yritys', 'http://www.nokia.com')



Semantic Web

- Semantic Web on tavoite tietoverkosta, jossa tietokone voi tehdä entistä enemmän käyttäjän puolesta asioita.
- Vaihtoehto 1: tee tietokoneesta älykkäämpi
 - Ja yritä saada ne ymmärtämään luonnollista kieltä, kuvia, graafeja, videoita, audiota jne.
- Vaihtoehto 2: tee sisällöstä tietokoneiden ymmärtämää
 - Luo sisältöä korkeammalla semantiikan tasolla kuin me tänä päivänä luomme
- Semantic Web tukeutuu lähestymistapaan 2

Mahdollisuudet

Visio 1: Merkitykseen perustuvat hakupalvelut

- Esim. kuinka hakea verkosta Janne Saarelan tekemät dokumentit?
- Esim. kuinka hakea verkosta Janne Saarelan ottamat valokuvat?
- Tämän päivän hakukoneet tarjoavat hakutuloksina linkkejä sivuihin, joissa esiintyvät sanat Janne ja Saarela. Hakukoneet eroavat toisistaan näiden tuloksien kattavuudessa ja tulosten järjestelyssä.

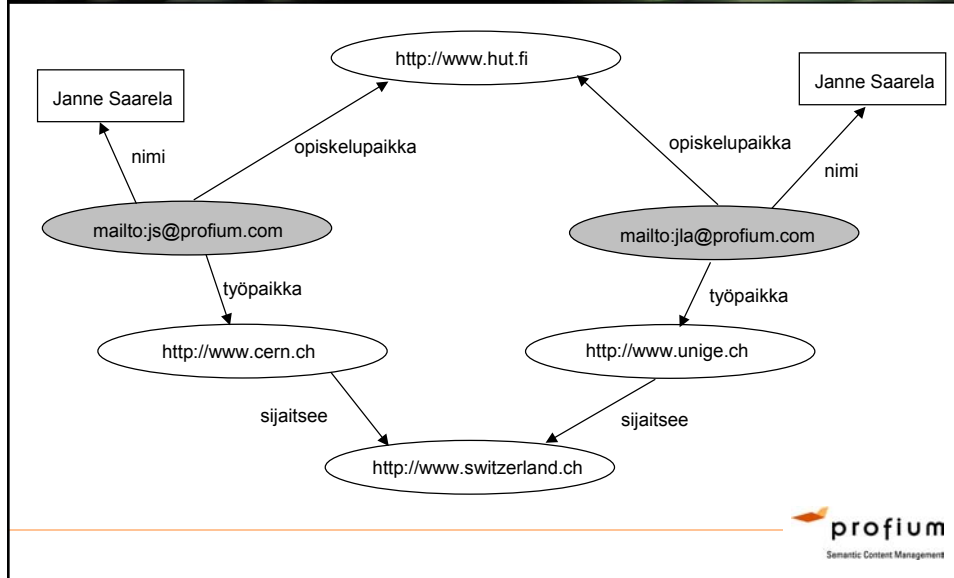
Visio 2: Sääntöpohjainen päättely

- Esimerkkisääntö: jos henkilö 1 on kirjoittanut henkilön 2 kanssa artikkelin yhdessä, he ovat ystäviä keskenään.
 - tietokoneen ymmärtämässä muodossa:
tekijä(X, H1), tekijä (X, H2) -> ystävä (H1, H2).
- Esim. hakupalvelu, jossa voi hakea Janne Saarelan ystäviä.
- Käyttäjä päättää mitä sääntöjä hän käyttää ja mihin hän luottaa.

Visio 3: asioiden vapaa assosiaatio

- Esimerkkihakue: mikä yhdistää Janne Saarelaa ja Janne Laaksoa?
 - tietokoneet edellyttävät että meillä on tapa yksikäsitteisesti tunnistaa Janne Saarela ja Janne Laakso.
 - henkilötunnus Suomessa toimisi hyvin, myös sähköpostiosoite on toimiva tunnistustapa
- Hakutulos:
 - Janne Saarela on työskennellyt CERNissä, CERN on Genevessä, Sveitsissä
 - Janne Laakso on työskennellyt Geneven yliopistossa, Geneven yliopisto on Genevessä, Sveitsissä.
 - Janne Saarela on opiskellut TKK:lla
 - Janne Laakso on opiskellut TKK:lla

Visio 3: asioiden vapaa assosiaatio (jatkuu)



Visio 4: todisteketjujen hyväksikäyttö

- Tietokoneiden suorittaessa päättelyä metatiedon ja sääntöjen avulla, tulee käyttäjän saada tieto miten päättelyn lopputulokseen on päästy.
- Käyttäjä tuskin luottaa "mustaan laatikkoon", joka kertoo vain lopputuloksen.
 - vrt. röntgenkuviin, joista
 - irroitetaan metatieto (esim. luun pituus ja leveys)
 - yhdistetään henkilön muihin metatietoihin (sukupuoli, pituus, paino, sairaushistoria)
 - sovelletaan päättelysääntöjä, joiden lopputulos on "jalka amputoitava"

Tämän päivän käyttökohteet

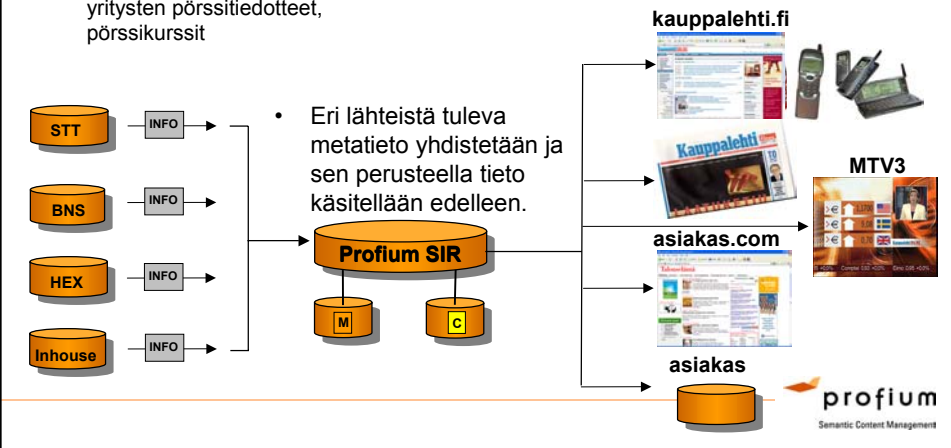
Esim. mediayritysten sisällönhallinta

- Esim 1. kuva-arkistoratkaisuissa kuvat luokitellaan niin että ne olisi helppo löytää myöhemmin ja että niitä voitaisiin käyttää myöhemmin kustannustehokkaasti.
 - tekijänoikeudet ovat myös tärkeää metatietoa
 - luokitteluissa usein käytetään valmiita metatietosanastoja (ontologioita) kuten esim. Tilastokeskuksen toimialaluokittelu
 - joskus oman toiminnan tehokas tukeminen vaatii oman sanaston/ontologian kehittämisen
- Esim 2. julkaisuprosesseissa metatieto kuvaa miten oikea tieto viedään oikeassa muodossa oikealle asiakkaalle.

Kauppalehti Online verkkopalvelu

- Kauppalehdellä on useita sisältölähteitä
 - omat online-uisartikkelit, yritysten pörssitiedotteet, pörssikurssit

- Sisältö julkaistaan eri muodossa eri käyttäjäryhmille.



Esim. museotoimialan kokoelmanhallinta

- Museot luokittelevat omia kokoelmiaan kokoelmanhallintaratkaisussa
 - metatieto auttaa käyttäjiä löytämään heitä kiinnostavan sisällön
- Museoiden verkottuminen sallii useita kokoelmanhallintaratkaisuja yhdistävät hakupalvelut
 - esim. Museoviraston Suomen Museot Online <http://www.suomenmuseotonline.fi>



Esim. yritysten portaaliratkaisut

- Suuryritykset ja julkinen sektori hakevat portaaliratkaisuilla tehokkuutta omaan toimintaansa.
- Käyttäjän tulisi löytää tieto riippumatta siitä missä se sijaitsee
 - ei siis tehdä neljää eri hakua eri järjestelmissä vaan vain 1 haku yhdessä paikassa
- Metatieto tuo erilaisia luokittelutapoja käyttävät järjestelmät yhden haun piiriin.
 - **tietokoneet voivat vihdoinkin käsitellä tietoa, jonka käsittelyyn niillä ei ole à priori –tietämystä!**

Haasteet

- Korkealaatuisen metatiedon tuottaminen
 - käyttäjälle tulisi saada hyvä mieli metatiedon tuottamisesta
 - joko metatieto hyödyttää häntä itseään suoraan tai välillisesti hänen asiakkaitaan/ystäviään
 - vrt. henkilökohtaisella digikameralla otettujen kuvien hallinta
- Metatiedon helppo tuottaminen
 - ohjelmistojen tulisi tehdä metatiedon tuotto helpoksi niin ettei sitä koettaisi suurena vaivana
- Avoimessa tietoverkossa metatiedon luotettavuus on kyseenalaista
 - yleiset hakukoneet eivät tänä päivänä uskalla vielä hyödyntää verkossa jo olevaa metatietoa

Yhteenveto

- Johdanto Semanttisen Webin maailmaan
- Mahdollisuudet
- Tämän päivän käyttökohteet
- Haasteet

Yhteystiedot

Janne Saarela
Profium Oy
Lars Sonckin kaari 12
02600 Espoo

Tel. +358 (0)9 855 98 000
Fax. +358 (0)9 855 98 002
Email. janne.saarela@profium.com
Internet. www.profium.com