



Kohti kansallista semanttisen webin sisältöinfrastruktuuria

Prof. Eero Hyvönen
TKK Viestintätekniikan laboratorio
Helsingin yliopisto, tietojenkäsittelytieteen laitos
Tietotekniikan tutkimuslaitos HIIT





- Miksi tarvitaan kansallisen semanttisen webin infrastruktuuri?
 - Ongelma
 - Teesit ratkaisumallille
- Suomalaiset semanttisen webin ontologiat -projekti
 - Ratkaisumallin kehityshanke
 - Tavoitteet
 - Ratkaisumallit
 - Tuloksia

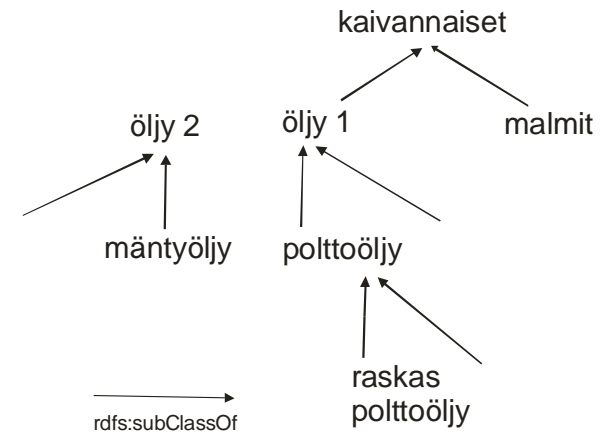
Ongelma



- Neljä faktaa
 - Semanttinen web = webin seuraava sukupolvi/kerros
 - Ontologiat = semanttisen webin ”silver bullet”
 - Suomalaiset ontologiat = ei ole
 - Tarttis terhä jotakin
- Ilman yhteisiä ontologoita
 - aitoa semanttisuutta ei webiin saada,
 - semanttinen web pirstaloituu yhteentoimimattomiksi saarekkeiksi,
 - sisältötyötä tehdään turhaan moneen kertaan ja
 - web-sovellusten kehittäminen tulee vaikeammaksi ja kalliimmaksi.

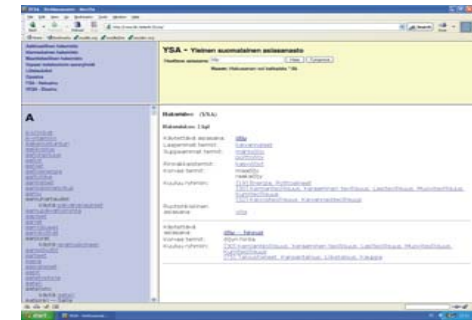


- Semanttisen web tarvitsee julkisen sisältöinfrastruktuurin
 - kuten liikenne tarvitsee tieverkoston
 - kuten energiahuolto tarvitsee sähköverkon ja standardit
 - kuten tietoliikenne tarvitsee GSM tai 3G-verkon
- Infrastruktuurin ytimenä ovat
 - laajat **referenssiontologiat**
 - näihin liittyvät keskeiset **ontologiapalvelut**
 - kansainväliset W3C:n **standardit**





- **Asiasanastoista on siirryttävä ontologioihin**
 - Ihmisen ymmärtämistä sanoista koneen ”ymmärtämiin” merkityksiin
- **Kansallinen ontologisointiprosessi tarpeen**
 - Eri alojen/organisaatioiden sisältöjen yhdistämiseksi
 - Ydinontologioita ylläpidetään julkisesti
 - » Laaja sitoutuminen välttämätöntä
 - Ydinontologiat julkaistaan open source
 - » Laaja hyödyntäminen ja yhteentoimivuus välttämätöntä
- **Liiketoimintamahdollisuudet lisäarvon tuottamisesta**
 - Sovelluskohtaiset ontologiat
 - Semanttisen webin sovellukset



Suomalaiset semanttisen webin ontologiat –projekti



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Tavoite: kansallisen semanttisen webin infrastruktuurin kehittäminen käyntiin
- Osahankkeet
 - Ontologioiden kehittäminen
 - Ontologiapalveluiden kehittäminen
 - Pilotointi semanttisissa portaalihankeissa
- Kaksi vaihetta
 - 1.9.2003-31.8.2005, käynnissä
 - 1.9.2005-31.5.2007, suunnitelma



UNIVERSITY OF HELSINKI



HELSINKI
INSTITUTE FOR
INFORMATION
TECHNOLOGY



SeCo
SEMANTIC COMPUTING

Organisaatio



HELSINKI UNIVERSITY OF TECHNOLOGY

Media Technology

- Tutkimus- ja kehitystyö 2003-2005
 - Pääosin Semanttisen laskennan tutkimusryhmässä (SeCo) TKK Viestintätekniikan laboratoriossa ja Helsingin yliopiston tietojenkäsittelytieteen laitoksella
 - Muina tutkimusosapuolina
 - » Tampereen yliopiston informaatiotutkimuksen laitos
 - » Helsingin yliopiston yleisen kielitieteen laitos
- Jatkosuunnitelmassa 2005-2007 myös
 - TKK Kartografian ja geoinformatiikan laboratorio
- Laaja yritysten ja julkisten organisaatioiden konsortio
 - 2003-2005, 14 organisaatiota
 - 2005-2007, 30 organisaatiota (suunnitelma)



UNIVERSITY OF HELSINKI



HELSINKI
INSTITUTE FOR
INFORMATION
TECHNOLOGY



SeCo
SEMANTIC COMPUTING

Projektin konsortio (18.4.2005)



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology



National Board
of Antiquities



Helsinki University Library
THE NATIONAL LIBRARY OF FINLAND

ALMA | MEDIA

KIASMA



TEKES

The Finnish Museum of Photography

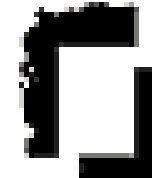


TietoEnator ^{TE}

Building the Information Society

m-cult

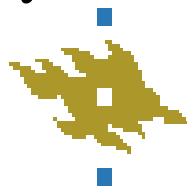
SanastokeskusTSK



Connexor Oy



HELSINKI
INSTITUTE FOR
INFORMATION
TECHNOLOGY



UNIVERSITY
OF HELSINKI

UNIVERSITY
OF TAMPERE

Finnish National
Gallery



UNIVERSITY OF HELSINKI



HELSINKI
INSTITUTE FOR
INFORMATION
TECHNOLOGY



SeCo
SEMANTIC COMPUTING

Tavoitteet



HELSINKI UNIVERSITY OF TECHNOLOGY

Media Technology

- Julkisten **ydinontologioiden kehittäminen** open source
 - Yleinen suomalainen ontologia YSO, perustana YSA
 - YSOon liittyvät alakohtaiset ontologiat
- Kansallisten **ontologiapalveluiden kehittäminen**
 - Toisistaan riippuvien ontologioiden rakentamisen, versioinnin ja julkaisemisen kehitysympäristö ONKI
 - Tiedon indeksointi web-palvelun avulla
 - Tiedon käsiteperustainen haku web-palvelun avulla
- **Pilotointi portaalihankkeissa**
 - KulttuuriSampo - Suomen kulttuuri semanttisessa webissä
 - » MuseoSuomen seuraava aste
 - Useita laajoja portaalihankkeita liittymässä projektiin



UNIVERSITY OF HELSINKI



HELSINKI
INSTITUTE FOR
INFORMATION
TECHNOLOGY



SeCo
SEMANTIC COMPUTING

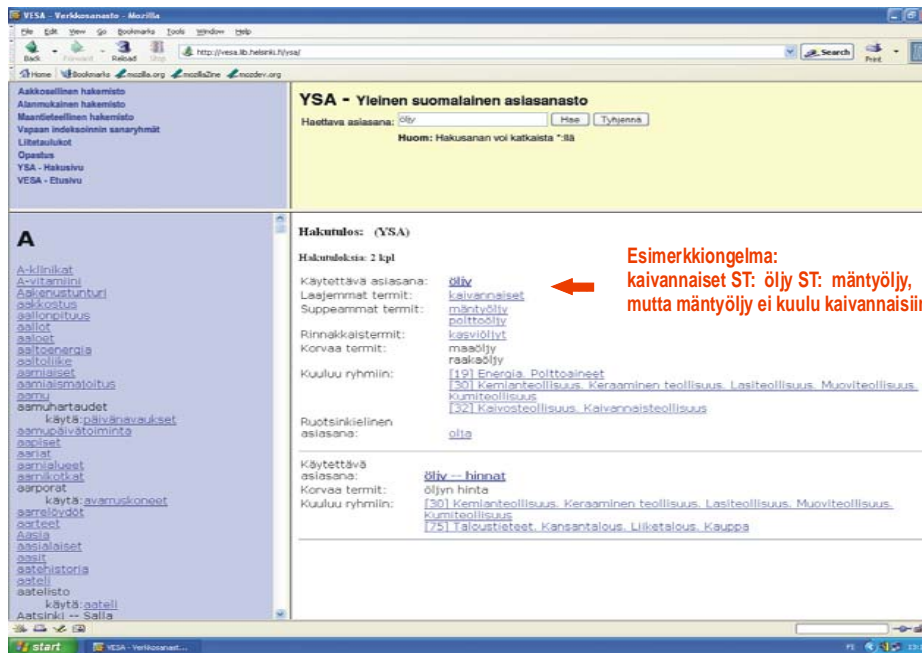
Julkisten ontologioiden kehittäminen



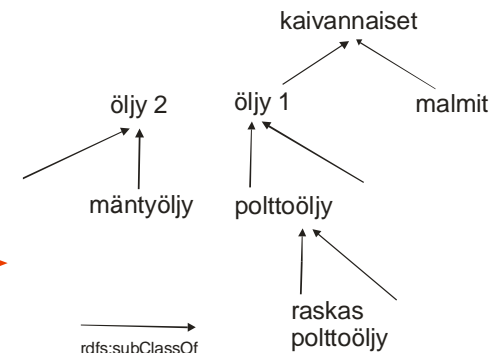
- Motto: Asiasanatesauruksista ontologioihin!
- Yleinen suomalainen ontologia YSO
 - Perustuu Yleiseen suomalaiseen asiasanastoon YSA

YSA

YSO



Esimerkkiongema:
kaivannaiset ST: öljy ST: mäntyöljy,
mutta mäntyöljy ei kuulu kaivannaisiin!



YSOon liittyviä muita ontologioita



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

Ontologia	Kuvaus
MAO	Museoviraston Museoalan asiasanastosta kehitetty ontologia.
YSO	Kansalliskirjaston Yleisen Suomalaisen Asiasanaston (YSA) ontologisoitu versio.
SAPO	Suomen Ajallinen Paikkaontologia
ICONCLASS	Kansainvälinen taiteen kuvailusysteemi, josta Valtion taidemuseo on tehnyt suomennoksen.
VALO	Valokuvataiteen museon kolmesta asiasanastosta muodostettu ontologinen kokonaisuus
TOIMO	Henkilöiden, yritysten, organisaatioiden yms. toimijoiden ontologia
PRO	Tapahtumien ja prosessien ontologia
TAITO	Taitojen kuvailun ontologia
MUSO	Musiikkialan ontologia
MATO	Maatalousalan ontologia



UNIVERSITY OF HELSINKI



HELSINKI
INSTITUTE FOR
INFORMATION
TECHNOLOGY



SeCo
SEMANTIC COMPUTING

Kansallinen ontologiapalvelin ONKI



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Ongelmana toisiinsa liittyvien ontologioiden
 - hajautettu kehitys ja versiointi,
 - julkaiseminen ja
 - hyödyntäminen webin avulla.
- Tavoitteet
 - semanttisen webin infrastruktuuripalveluiden tarjoaminen
 - ontologisointiprosessin tukeminen
- Ratkaisumallina ONKI-palvelimen palvelut
 - kehitys- ja julkaisuprosessit
 - palvelurajapinta web-palveluina (Web Service)



UNIVERSITY OF HELSINKI



HELSINKI
INSTITUTE FOR
INFORMATION
TECHNOLOGY



SeCo
SEMANTIC COMPUTING

ONKIn palvelut

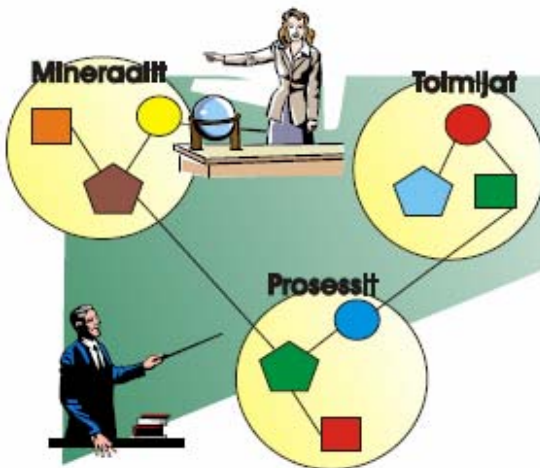


HELSINKI UNIVERSITY OF TECHNOLOGY

Media Technology

Ontologian kehittäjät

- Asiakas-palvelinmalli tukee hallittua hajautusta
- Käsitteiden turvallinen ristiinkäyttö antaa mahdollisuuden jakaa kehitys halutuille tahoille
- Muutostenhallinta pitää ajantasalla muiden ontologioiden muutoksista
- Versiohistoria tarjoaa näkymän mallien historiaan ja kunkin vaiheen muutoksiin



Tiedon hakija

- Selainnäkömä käsitteisiin ja niiden välisiin suhteisiin
- Käsitehaku



Nokia?
Palkka vai yritys?



Tiedon indeksoija

- Selain- ja Web Service-rajapinnat käsitteiden selailu-, haku- ja käyttöönottoihin, jotta automatisoinnista voidaan saada maksimaalinen hyöty



UNIVERSITY OF HELSINKI



HELSINKI
INSTITUTE FOR
INFORMATION
TECHNOLOGY



SeCo
SEMANTIC COMPUTING

Pilottisovellukset (SeCo-ryhmä)



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- MuseoSuomi
 - Suomen museot semanttisessa webissä
- KulttuuriSampo
 - Suomen kulttuuri semanttisessa webissä
- Lukuisia portaalihankkeita liittymässä projektiin (suunnitelma)
 - Suomi.fi, Opintoluotsi.fi, Tiedonportti.fi, Sosiaaliportti.fi, Kansalaisen terveystaali



UNIVERSITY OF HELSINKI



HELSINKI
INSTITUTE FOR
INFORMATION
TECHNOLOGY



SeCo
SEMANTIC COMPUTING

MuseoSuomi



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Ideat
 - Globaali web-näkymä kokoelmatietokantoihin
 - Semanttinen hakukone + samoilu
 - Museoiden yhteinen julkaisukanava
- <http://museosuomi.cs.helsinki.fi>
- Kansainvälisesti palkittu sovellus
 - Semantic Web Challenge Award 2004
 - Nordic Digital Excellence in Museums 2004
 - Pääministerin kunniamaininta





Suomen kulttuuri semanttisessa webissä

KulttuuriSampo

- **Ongelmana** monimuotoiset ja yhteensopimattomat kulttuurisisällöt
- Loppukäyttäjälle ei kyetä tarjoamaan kokonaisvaltaista sisältöpalvelua
- **Tavoitteena** on kehittää ontologioita ja niitä hyödyntäviä älykkäitä haku- ja samoilumenetelmiä
- **Ratkaisuna** sisällöntuottajille yhteinen julkaisukanava
- Loppukäyttäjälle saumaton näkymä eri kulttuurisisältöihin



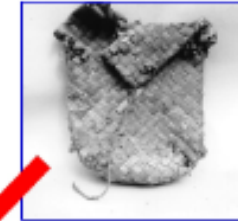
KulttuuriSampo: semanttista hakua ja samoilua



The screenshot shows a search result for 'Kullervon kireus'. The main image is a painting of a man in a white loincloth. To the right, there are several metadata sections:

- Tekijä:** Axel Waldemar Gallen-Kallela; ID: A II 765; Valmistusvuosi: 1899; Mitta: leveys 102,5 cm; Materiaali: kuitunetty nahkastaitee vähärillä; Tekniikka: öljy maalaus.
- Aineistolyyppi:** List of image types like 'maalaukset', 'kuvakortit', 'kuvakirjat', etc.
- Tapahtuma:** List of event types like 'näyttelyt', 'kokoelmien kuvaukset', etc.
- Toimija:** List of collection names like 'Kotimuseo', 'Kansallismuseo', etc.
- Toiminnan kohte:** List of thematic categories like 'Kotimuseo', 'Kansallismuseo', etc.

koosti, teehikonni



Axel Waldemar Gallen-Kallela



Runo 31 - osa 5

Kullervo lähtösän lausua kaksannen.
Kullerwogäky kuterkin kironaan kusten.

Runo 31 - osa 5

En tiedä, kuka painoi.
kalle työlle työtölmänsä.
Painoi kalle kalle kas'antapan?
Paini kalle kas'antapan?
Kullerw, Kullerwon paini,
tapan tuu sanooi vialki:
'Kalle Linnon mie sunni,
kas oon kullerwon kalle.
palljo konoo parempi,
retinän sunni.
Linnon mie' kille vialki.



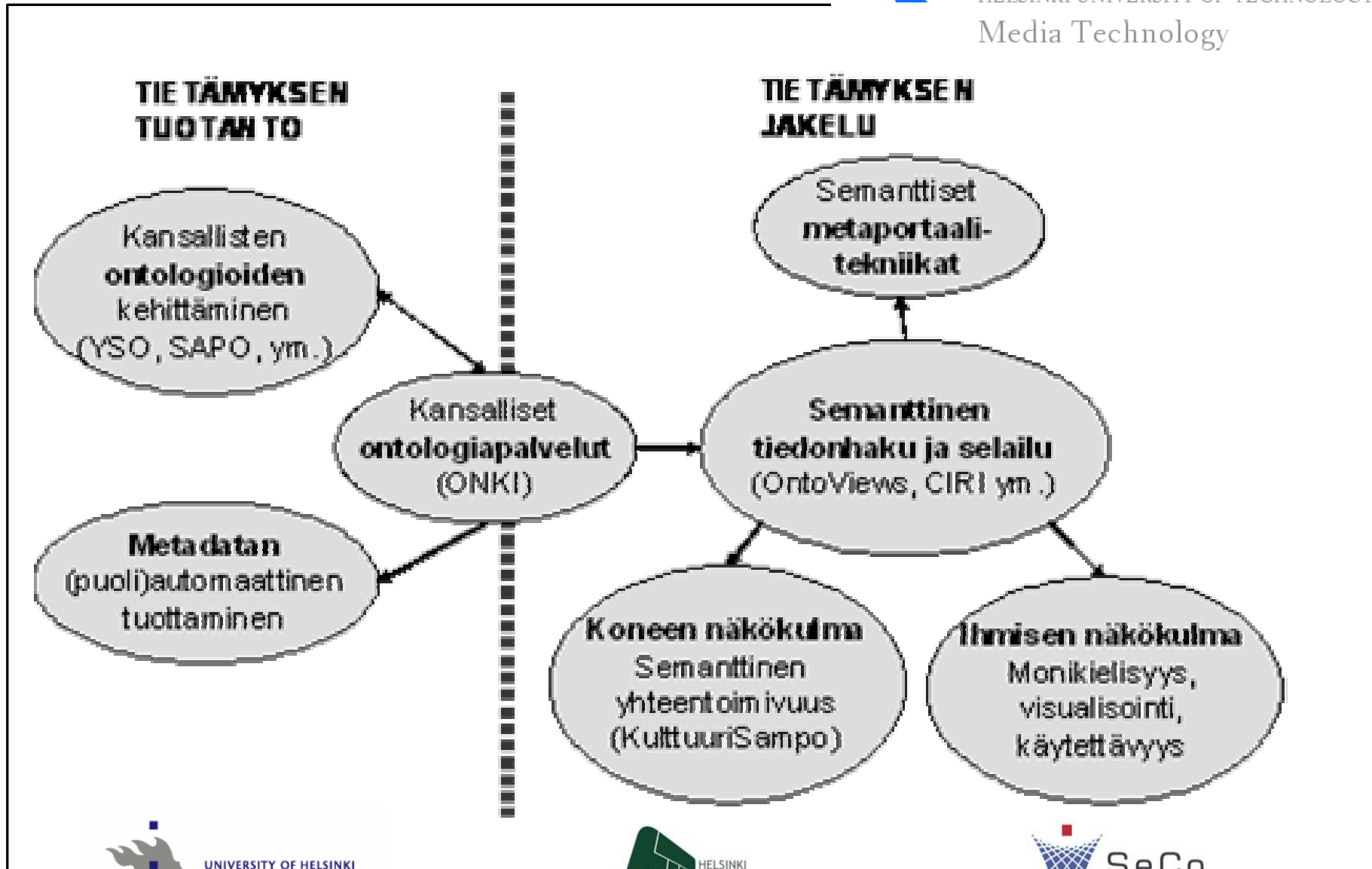
Tutkimushaasteita



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Ontologisen tietämyksen esittäminen ja päättely
- Hajautettu ontologiakehitys
- Semanttinen haku ja sisältöjen yhdistäminen
- Ontologioiden muutokset, versiointi ja aikasarjat
- Epätäsmälliset ontologiat
- Ontologioiden yhdistäminen

Tutkimuskohteita



Tietämyksen hallinnan ja semanttisen webin aiheoryhmä



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Rakentunut Suomalaiset semanttisen webin ontologiat konsortion ympärille
- Toimintaa
 - Älyä verkkoon – Web Intelligence
 - » Symposio Heurekassa 2004
 - Syksyllä 2005 uusi symposio?
 - Sähköpostilista
- Kaikki tervetulleita mukaan!
- Kaikki toimintaideat tervetulleita!
- Yhteydenotot
 - eero.hyvonen@tkk.fi