



Miten ja miksi asiasanastoista kehitetään ontologioita

Katri Seppälä

Semanttisen laskennan tutkimusryhmä (SeCo)
Teknillinen korkeakoulu, mediatekniikan laitos;
Helsingin yliopisto, tietojenkäsittelytieteen laitos;
Sanastokeskus TSK ry
<http://www.seco.tkk.fi/>

Mihin ontologioita tarvitaan?



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- ongelma: semanttisen webin tekniikat mahdollistavat älykkäiden palvelujen kehittämisen, mutta palvelujen tarvitsemaa käsitetietoa ei ole tarjolla tietokoneen ymmärtämässä muodossa
- aineistojen kuvailussa käytettävät asiasanastot on tarkoitettu ihmisen käyttöön
 - sisältävät monitulkintaisuutta
 - eivät aina anna riittävästi tietoja käsitteiden yksilöimiseksi
 - eivät aina anna riittävän monipuolisesti tietoa käsitteiden välisistä suhteista
- ontologiat soveltuvat sekä ihmisen että koneen käyttöön

Mihin ontologioita tarvitaan?



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- tavoite: ontologioiden saaminen julkiseen ja avoimeen käyttöön kehityskustannusten säästämiseksi, etuina mm.
 - koneymmärrettävä sisällönkuvailu
 - asiasanojen aiempaa tarkempi ja yhdenmukaisempi käyttö
 - » erikielisten vastineiden antaminen helpottuu
 - tarkemman kuvailun ansiosta tarkempia hakutuloksia
 - haku perustuu käsitteeseen pelkän merkkijonon sijaan
 - » helpottaa eri tavoin tuotettujen tietojen yhdistelyä ja yhteentoimivuutta
 - mahdollisuus ryhmitellä hakutuloksia tai muokata tiedonhakuperusteita käsitetiedon pohjalta

Miksi asiasanastoista kehitetään ontologioita?



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Yleinen suomalainen asiasanasto YSA ja erikoisalojen asiasanastot kuten Museoalan asiasanasto, Valokuvan asiasanasto, Muotoilun ja viestinnän asiasanasto ja Agriforest-asiasanasto ovat pitkään olleet laajassa käytössä aineistoja kuvailtaessa
 - YSA ja erikoisalojen asiasanastot täydentävät toisiaan
- asiasanastot sisältävät pohjatietoja käsitteistä, joten niitä muokkaamalla ja täydentämällä voidaan saada aikaan ontologioita, jotka sisältävät asiasanastojen käyttäjille tuttuja elementtejä ja näin ollen helpottavat ontologioiden käyttöönottoa

Miten asiasanastoista kehitetään ontologioita? Esimerkkinä YSO



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- perustuu Yleiseen suomalaiseen asiasanastoon (YSA)
- sisältää n. 20 000 ontologisoitua käsitettä (YSA:n vuoden 2005 lopun tilanne)
 - jokaisella käsitteellä on tunnisteena yksilöllinen URI (Uniform Resource Identifier)
- kielet: suomi
 - ruotsi ja englanti mukana ilman ontologista tarkistusta
- ontologian käsitteiden väliset suhteet
 - yläluokka-alaluokkasuhde
 - osa-kokonaisuussuhde
 - assosiatiivinen suhde

Miten asiasanastoista kehitetään ontologioita? Esimerkkinä YSO

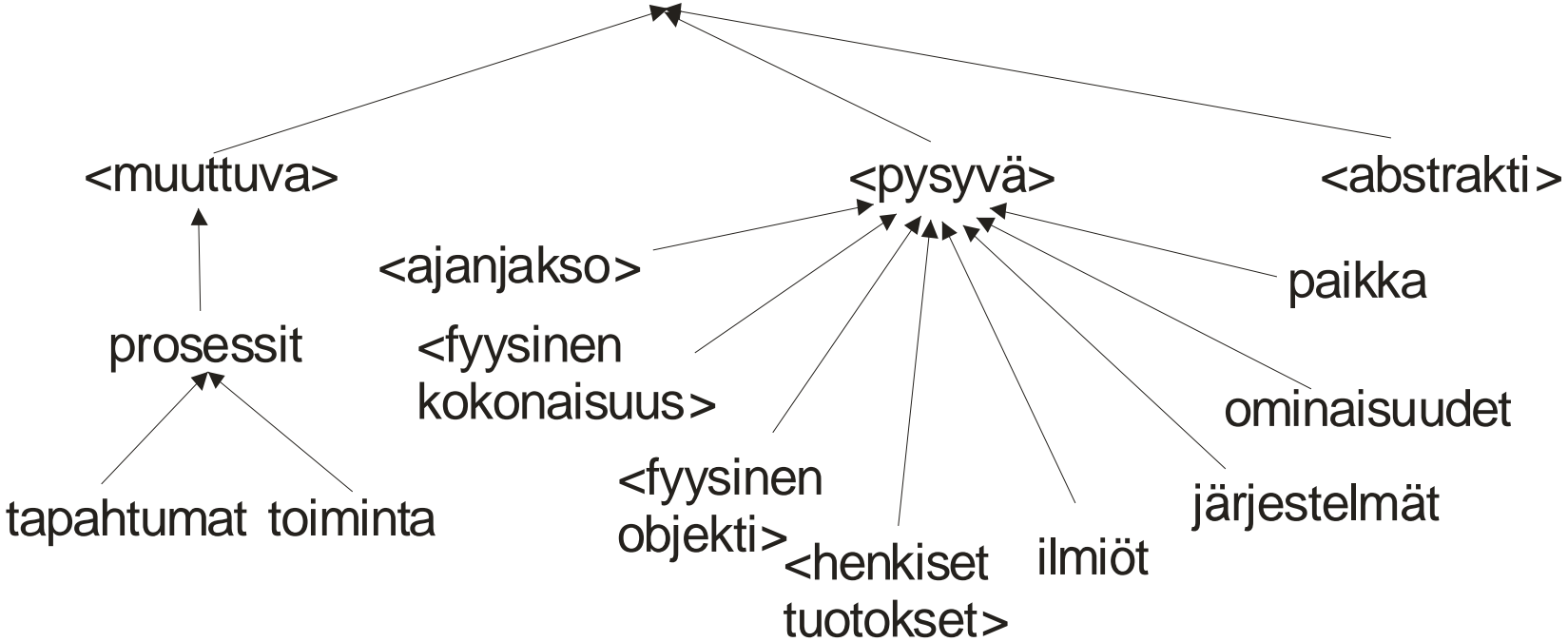


HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

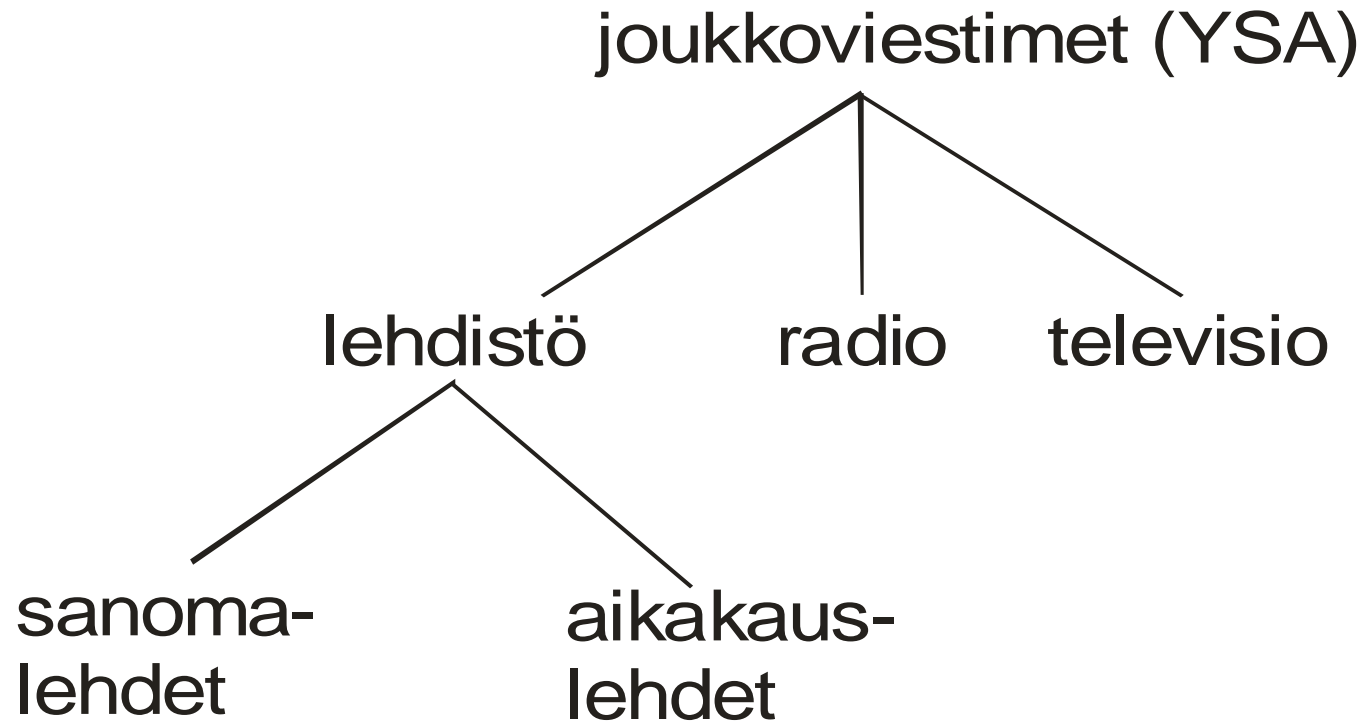
- YSO:n suomenkielistä käsitteistökokonaisuutta varten on rakennettu pääluokkamalli, jotta saataisiin aikaan koneelliseen päättelyyn tarvittava yhtenäinen hierarkia
 - käsitteet on sijoitettu pääluokkiin perustuvaan yhtenäiseen hierarkiaan
 - monimerkityksisten sanojen taustalla olevia käsitteitä on eritelty
 - » myös asiasanaston mukaiset tarkentamattomat käsitteet ovat ontologian ohessa koostekäsitteinä
 - käsitteistöön on lisätty rakenteen tueksi apukäsitteitä
 - yläluokka-alaluokka suhteiden ohkeen on otettu käyttöön osa-kokonaisuussuhteita
 - asiasanastosta periytyvä ja täydennetty hierarkia on tarkistettu karkealla tasolla



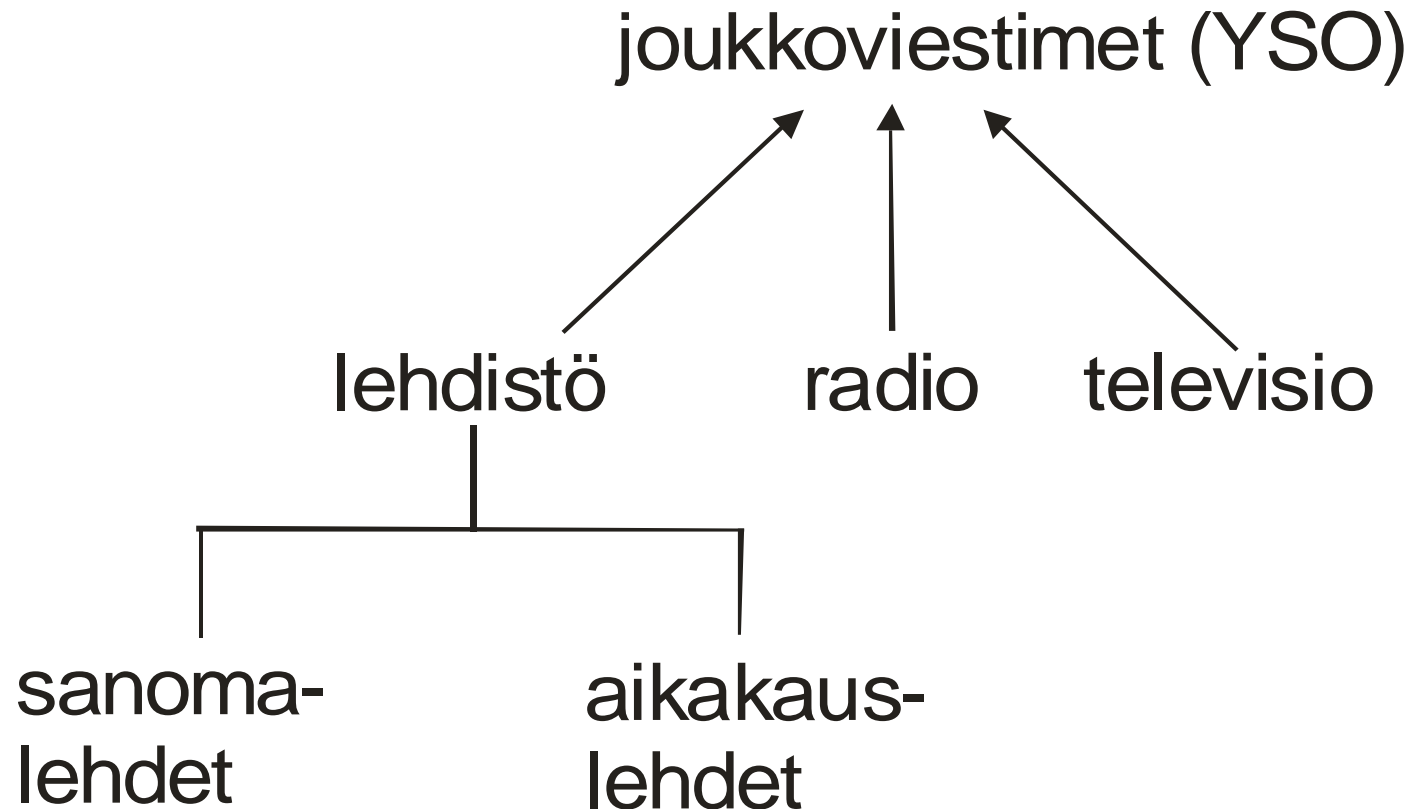
Miten asiasanastoista kehitetään ontologioita? Esimerkkinä YSO



Miten asiasanastoista kehitetään ontologioita? Esimerkkinä YSO



Miten asiasanastoista kehitetään ontologioita? Esimerkkinä YSO



Miten asiasanastoista kehitetään ontologioita? Esimerkkinä YSO



The screenshot shows the Protégé 3.3.1 interface for editing an ontology. The title bar indicates the project is 'yso-2008-06-12kauttaviivoilla' and the file is 'yso.pprj'. The main window is divided into several panes:

- Subclass Explorer:** Shows a hierarchy of classes. The selected class is 'yso_ajankausjulkaisut'.
- Class Editor:** Shows the 'yso_ajankausjulkaisut' class with its internal name 'yso:p1255'. It includes a table for 'rdfs:comment' and 'Asserted Conditions'.
- Property Tables:** Several tables are visible at the bottom, including 'yso-meta:prefLabel', 'yso-meta:altLabel', 'yso-meta:tempLabel', 'yso-meta:associativeR', 'yso-meta:partOf', 'om:definedConcept', 'yso-meta:semanticTag', 'yso-meta:ysoSource', 'yso-meta:ysoCommer', and 'dc:creator'.

Value	Lang
magazines	en
ajankauslehdet	fi
tidskrifter	sv

Value	Lang
lehdet	fi
tidsningar	sv

Value	Lang
-------	------

Value	Lang
-------	------

Value	Lang
yso_lehdistö	

Value	Lang
http://www.yso.fi/onto/ysa/Y448	

Value	Lang
yso_ajankauslehdet	

Value	Lang
-------	------

Value	Lang
-------	------

Value	Lang
-------	------



- Yleisellä suomalaisella asiasanastolla (YSA) ja erikoisalojen asiasanastoilla on huomattava määrä yhteisiä käsitteitä ja erikoisalojen asiasanastojen rinnalla käytetään usein YSAa
 - käsitteiden väliset suhteet eivät kuitenkaan aina ole yksiselitteisiä
- yhdistelmäontologiat, joissa YSOa on käytetty rakenteellisena mallina ja linkkinä erikoisalojen käsitteistöön
 - selkeyttävät eri ontologioiden yhteiset käsitteet
 - tuovat esiin monimerkityksiset termit ja niiden taustalla olevat eri käsitteet
 - helpottavat kokonaisuuden käyttöä ja kehitystä
 - » auttavat välttämään turhaa päällekkäisyyttä ja mahdollistavat entistä paremmin omaan käsitteistöön keskittymisen



- koneella tehtävä (asia)sanojen (label) merkkijonopohjainen vertailu
 - kerätään vertailtavien ontologioiden asiasanat ja muutetaan ne perusmuotoon
 - etsitään eri ontologioiden yhteiset asiasanat ja merkitään käsitteet yhteneviksi (owl:equivalent) yhteisten asiasanojen perusteella
 - » yhtenevien käsitteiden etsinnässä on mahdollista käyttää ensisijaisten asiasanojen lisäksi myös korvattuja asiasanoja/synonyymeja (altLabel) tai erikielisiä asiasanoja
 - yhdistelyssä käsitellään aina vain kahta ontologiaa kerrallaan, vaikka lopullisena tavoitteena olisi useamman ontologian yhdistäminen



- käsityönä tehtävä tarkistus
 - pitääkö ehdotettu ekvivalenssi paikkansa?
 - » onko sanaa käytetty eri ontologioissa samassa merkityksessä?
 - onko aineistosta jäänyt puuttumaan ekvivalenssimerkintöjä
 - » käytetäänkö eri ontologioissa eri synonyymeja? esim. ympäristöystävälliset tuotteet /ympäristöä säästävät tuotteet
 - mihin ei-ekvivalentit erikoisalan ontologian käsitteet sijoittuvat YSO:n hierarkiassa?
 - onko mahdollista tarkentaa erikoisalan ontologian sisäistä hierarkiaa?

Asiasanastoista yhdistelmä-ontologioita – Esimerkkinä TAO



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

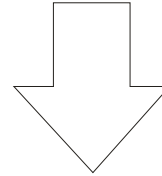
- Mikä on pergamentti?
 - Yleinen suomalainen asiasanasto: pergamentti
 - » Kuuluu ryhmiin: [82] Kustannustoiminta. Julkaisutoiminta. Kirjapainoala
 - » Ruotsinkielinen asiasana: pergament
 - Muotoilun ja viestinnän asiasanasto: pergamentti
 - » KT rumpunahka
 - » LT *nahka*
 - Museoalan asiasanasto: pergamentti
 - » LT *paperi*

Asiasanastoista yhdistelmä-ontologioita – Esimerkkinä TAO



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

pergamentti (YSA)
~pergamentti_kooste (YSO)



pergamentti
(paperi)
(YSO)

pergamentti
(nahka)
(YSO)

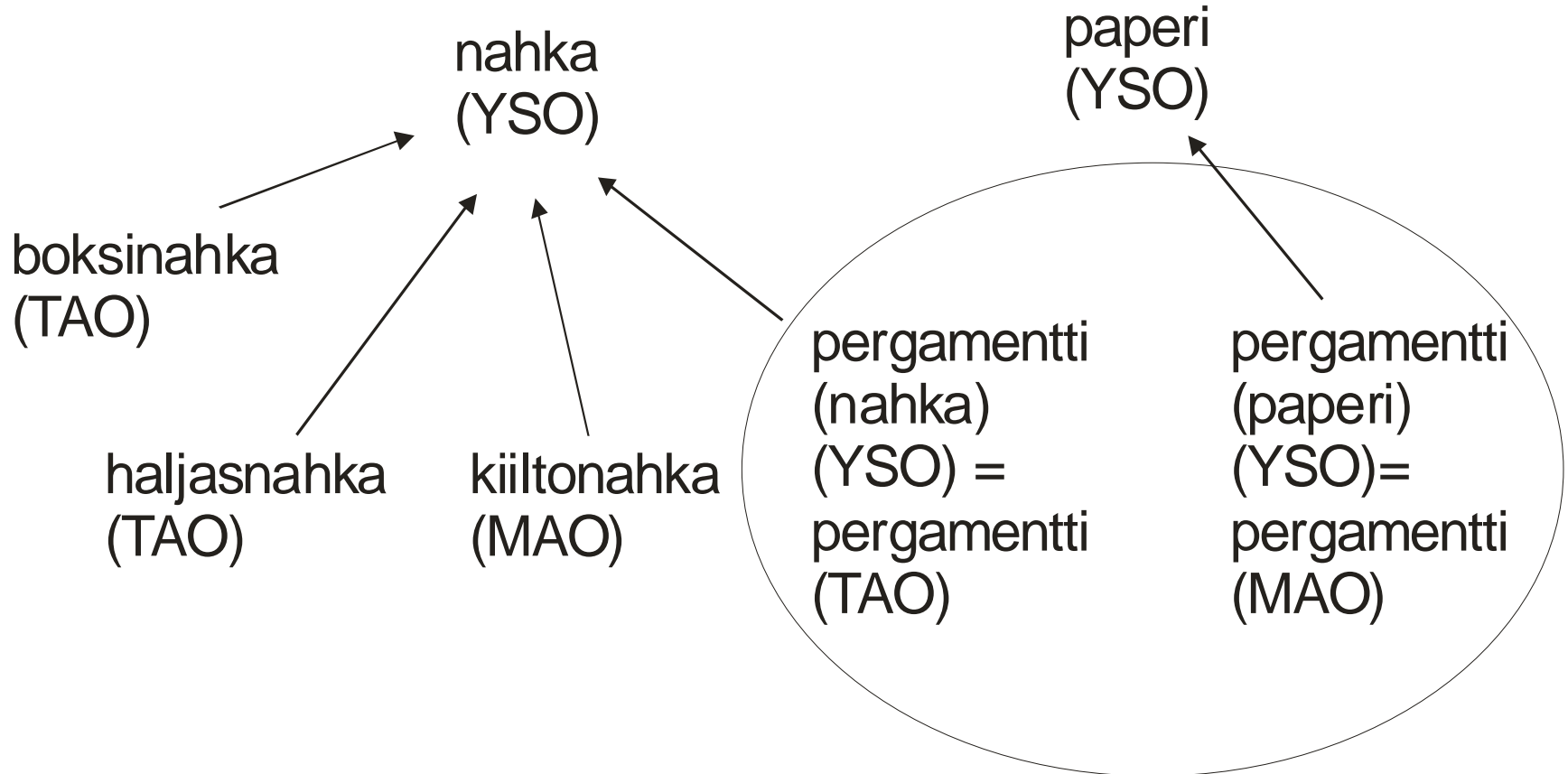
paperi
(MASA)

nahka
(MVA)

pergamentti
(MAO)

pergamentti
(TAO)

Asiasanastoista yhdistelmä-ontologioita – Esimerkkinä TAO





- YSO
 - Kansalliskirjasto: Juha Hakala, Tuula Haapamäki, Laila Heinemann ja Eeva Kärki
 - SeCo-tutkimusryhmä: mm. Anu Ylisalmi, Mirja Anttila, Eeva-Liisa Leppänen, Matias Frosterus, Eetu Mäkelä, Tuukka Ruotsalo, Tomi Kauppinen ja Osma Suominen
 - Helsingin kaupunginkirjasto: Erkki Lounasvuori
- TAO
 - Taideteollinen korkeakoulu: Ellen Karhulampi
- MAO
 - Espoon kaupunginmuseo, Suvi Kettula
 - Museovirasto