

# **Käyttöohje**

MAITO – metadatan hallintatyökalu

Helsinki 14.12.2005  
Ohjelmistotuotantoprojekti  
HELSINGIN YLIOPISTO  
Tietojenkäsittelytieteen laitos

## **Kurssi**

581260 Ohjelmistotuotantoprojekti (6 ov)

## **Projektiryhmä**

Väinö Ala-Härkönen  
Reima Halmetoja  
Antti Laitinen  
Kalle Pyykkönen  
Oskari Saarekas  
Tuomas Tanner  
Juuso Vanonen

## **Asiakas**

Olli Niinivaara

## **Johtoryhmä**

Juha Taina  
Joni Salmi

## **Kotisivu**

<http://www.cs.helsinki.fi/group/metadata/>

# Sisältö

Yleiskuvaus.....	1
Järjestelmän vaatimukset.....	1
MySQL:n asentaminen (Linux).....	1
Ohjelmiston asentaminen (Linux).....	1
Tietokanta-asetusten asettaminen.....	2
Ohjelmiston käyttäminen.....	2
Datan keruu, transformaatio, integraatio ja resurssiverkon luominen.....	2
Ohjelman käynnistäminen.....	2
Päävalikko.....	2
Datankeruu ja raakadatan transformointi.....	2
Järjestelmään syötettyjen datalähteiden listaaminen.....	2
Datalähteiden lisääminen.....	3
Datan keruu ja transformoiminen yhteiseen muotoon.....	3
Datalähteen poistaminen järjestelmästä.....	3
Transformoidun datan integrointi ja verkon muodostaminen.....	3
Resurssiverkon selaus ja tulostus.....	4
Ohjelman käynnistäminen.....	4
Haun tekeminen.....	4
Hakutuloksen vienti tiedostoon.....	5
Muuta.....	5
Liitteet.....	6



## Yleiskuvaus

MAITO-ohjelmiston avulla voidaan:

- harvestoida DC-XML muotoista metadataa OAI-PMH palvelimilta
- transformoida metadataa yhteiseen muotoon dcxml, oai-citeseer ja ohjelman oman pikaformaatin muotoista metadataa
- muodostaa transformoidusta metadatasta verkkoesitys
- selata verkkoa graafisen käyttöliittymän avulla
- tallentaa muodostunut verkko tai sen osa csv-listana tai Pajek-verkkona

Järjestelmä on testattu Helsingin yliopiston tietojenkäsittelytieteen laitoksen Linux-ympäristössä.

## Järjestelmän vaatimukset

MAITO-ohjelmiston ajamiseksi järjestelmään tulee olla asennettuna:

- Java 1.5 (tai uudempi)
- MySQL 5.0.16 (tai uudempi)

## MySQL:n asentaminen (Linux)

MySQL-Linux asennuspaketti (mysql-standard-5.0.16-linux-i686-glibc23.tar.gz) löytyy ohjelmiston asennuslevyltä ja kotisivuilta.

- Aseta asennusmedia cd-rom asemaan tai hae asennuspaketti kotisivuilta.
- Siirry hakemistoon, jonka alle haluat suorittaa asennuksen. (esim. oma kotihakemisto)
- Pura paketti, CD:ltä komennolla:
  - `tar -zxvf /mnt/cdrom/mysql/mysql-standard-5.0.16-linux-i686.tar.gz`
- Siirry asennushakemistoon:
  - `cd mysql-standard-5.0.16`
- Aja peruskantojen asennusskripti: `scripts/mysql_install_db`
- Käynnistä MySQL komennolla: `bin/mysql_safe`

Vaihtoehtoisesti MySQL:n voi asentaa Linuxiin RPM-paketista. Voit myös luonnollisesti käyttää uudempia MySQL-versioita valmistajan kotisivuilta. Toimivuus on testattu juuri tällä versiolla ja se oli uusin ohjelman kehitysaikoihin.

MySQL:n asentaminen Windowsille tai muihin ympäristöihin: Kts. asennuspaketin lataaminen ja asennusohjeet osoitteesta <http://www.mysql.com/>

## Ohjelmiston asentaminen (Linux)

- Aseta asennusmedia cd-rom asemaan tai hae asennuspaketti kotisivuilta.
- Siirry hakemistoon, jonka alle haluat suorittaa asennuksen.
- Pura asennuspaketti hakemistoon, esimerkiksi CD:ltä komennolla:
  - `unzip /mnt/cdrom/dist/maito-1.0.zip .`

## **Tietokanta-asetusten asettaminen**

Järjestelmän tietokanta-asetukset määritellään tiedostossa:  
config/dbconfig.properties

Oletuksena tiedosto vastaa MySQL:n oletusasennusta. Mikäli MySQL:n asennus on mukautettu tai käytetään toisella koneella olevaa MySQL kantaa editoidaan asetustiedostoa tekstieditorilla vastaamaan MySQL:n asennusta.

```
jdbc_url = <käytettävän MySQL-asennuksen URL ja portti>
jdbc_user = <MySQL käyttäjä>
jdbc_pass = <MySQL käyttäjän salasana>
dbname_rawdata = <Transformoidun datan tietokannan nimi>
dbname_resources = <Resurssiverkon tietokantojen nimien etuliite>
```

## **Ohjelmiston käyttäminen**

### ***Datan keruu, transformaatio, integraatio ja resurssiverkon luominen***

#### **Ohjelman käynnistäminen**

Datankeruuohjelma käynnistetään UNIX/Linux-ympäristöissä ohjelman työhakemistosta komennolla: ./datacollector.sh

Vastaavasti Windowsilla käynnistys hoituu työhakemistosta skriptillä  
datacollector.bat

#### **Päävalikko**

Datan keruu, transformaatio ja verkon muodostaminen käynnistetään datankeruuohjelman päävalikosta

**MAITO 1.0 - under GPL**

```
* Data collecting *           * Integration *****
1) list sources              5) list resource graphs
2) add a new source         6) list graph sources
3) remove a source          7) integrate a source
4) update a source
```

```
enter choice (1 - 7) or hit enter to quit :
>
```

#### **Datankeruu ja raakadatan transformointi**

#### **Järjestelmään syötettyjen datalähteiden listaaminen**

- valitse päävalikosta 1

## Datalähteiden lisääminen

- valitse päävalikosta 2
- syötä datalähteen tyyppi:
  - paikalliselle tiedostolle: `file`
  - OAI-PMH lähteelle: `oaipmh`
- syötä datalähteen formaatti
  - kaikelle harvestoitavalle datalle, sekä paikallisille DC-XML tiedostoille: `dcxml`
  - paikalliselle citeseer-tiedostolle: `oai_citeseer`
  - paikalliselle nimipikaformaattitiedostolle: `quick_format_name`
  - paikalliselle dokumenttipikaformaattitiedostolle: `quick_format_document`
- syötä datalähteen URL
  - paikalliselle tiedostolle: `file://<polku><tiedostonnimi>`
  - OAI-PMH lähteelle: `http://<repositorion URL>`

Mikäli tiedostolle ei määritellä polkua, oletetaan tiedoston löytyvän ohjelman asennushakemistosta.

*Huom! OAI-PMH protokollan yli harvestoitaessa datan tyyppi on aina dcxml. Oai-citeseer datatyyppiä käytetään ainoastaan transformoitaessa erillistä Citeseerin www-sivuilta ladattavissa olevaa laajennettua tietokantadumppia. Tietokantadumppi on pakattu tiedosto, joka sisältää ison määrän pieniä xml-tiedostoja. Dumppi voidaan lisätä haluttaessa yhtenä datalähteenä siten, että xml-tiedostot katenoidaan ensin yhteen tiedostoon (esimerkiksi Linuxissa `cat * > tiedostonnimi`).*

## Datan keruu ja transformoiminen yhteiseen muotoon

- Valitse päävalikosta 4
- valitse päivitettävät datalähteet tai \*, mikäli haluat päivittää kaikki järjestelmään lisätyt lähteet
- päivittäminen laukaisee OAI-PMH lähteiden harvestoinnin ja lähteiden transformoinnin yhteiseen muotoon

## Datalähteen poistaminen järjestelmästä

- Valitse päävalikosta 3
- valitse poistettava datalähteet tai \*, mikäli haluat poistaa kaikki järjestelmään lisätyt lähteet

## Transformoidun datan integrointi ja verkon muodostaminen

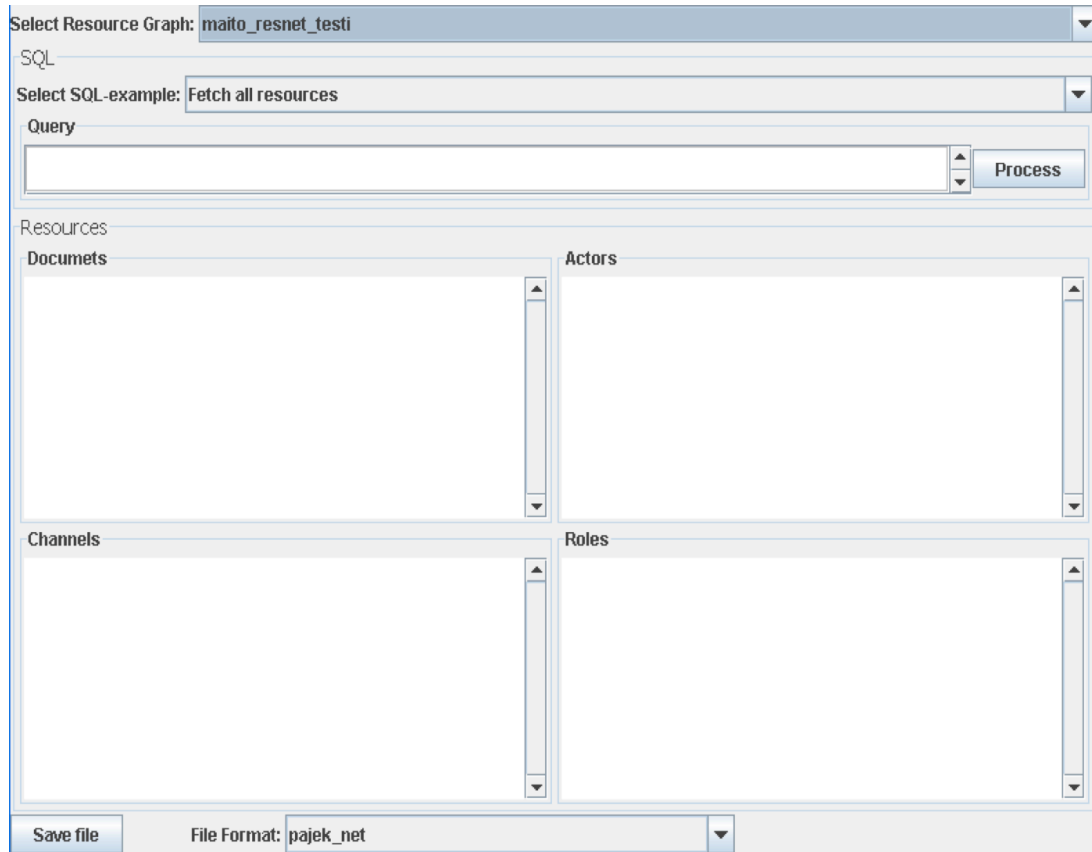
- valitse päävalikosta 7
- syötä kohdeverkon (tietokannan) nimi
- valitse kohdeverkkoon integroitavan datan lähde tai \*, jolloin kaikki järjestelmään lisätyt datalähteet integroidaan verkkoon

## Resurssiverkon selaus ja tulostus

### Ohjelman käynnistäminen

Ohjelma käynnistetään asennushakemistosta Linuxissa komennolla  
`./resourcebrowser.sh`

Windowsissa käynnistys puolestaan tapahtuu skriptillä `resourcebrowser.bat`



Kuval: Selauksen käyttöliittymä

### Haun tekeminen

- käsiteltävä verkko valitaan kohdasta: *Select Resource Graph*
- valmis haku voidaan valita kohdasta: *Select SQL-Example*
- SQL-lausetta voidaan editoida tai oma SQL-lause voidaan syöttää kenttään: *Query*
  - kts. liite tietokantakaavio
- SQL-haku käynnistetään painikkeesta: *Process*
- *Resource*-otsikon alapuolella on kentät kunkin tyyppisille resursseille, jotka suoritettu haku palauttaa.
- Haun tulos on jaettu useammalle eri sivulle. Kullakin sivulla näytetään tietty määrä resursseja. Sivun resurssit ovat jaettu tyyppin mukaan neljään eri kenttään: Documents, Actors, Channels ja Roles.



## **Hakutuloksen vienti tiedostoon**

- valitaan tiedoston tyyppi kohdasta: *File Format*
  - Pajek-verkolle: *pajek\_net*
  - CSV-listalle: *CSV-list*

## **Muuta**

Ohjelmassa ei ole toimintoa, jonka avulla resurssiverkkoja (tietokantoja) voitaisiin poistaa. Suosittelemme tietokantojen poistamiseen MySQL Query Browser-ohjelmaa, jonka voi ladata MySQL:n verkkosivuilta. Query Browserin avulla myös monet muut tietokantaoperaatiot tapahtuvat ohjelman sisäistä selaustoimintoa helpommin.

Jos kohtaat transformoinnissa tai integroinnissa outoja kaatumisen aiheuttavia virheitä, tarkista että olet lisännyt lähteen alunperin oikealla formaattimäärityllä. Jos yrität esimerkiksi käsitellä DCXML-dataa oai\_citeseer-tyyppisenä niin ohjelma saattaa reagoida arvaamattomasti.

# Liitteet

Resurssiverkon tietokantakaavio

