



Aalto-yliopisto
Perustieteiden
korkeakoulu

Ontologiamuokkaimen käyttö laskentaklusterin tehokkuusanalyysissä (aihe-esittely)

Santtu Klemettilä

25.2.2013

Ohjaaja: FT Markopekka Niinimäki

Valvoja: Prof. Ahti Salo

Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

Tausta: Ontologia

- Data on yleensä hajanaista ja heterogeenistä
 - Tietokoneella ei ole älyä ymmärtää sille määrittelemätöntä dataa
- Semanttinen web tarkoittaa verkkoa, jossa on tarjolla myös tiedon merkityksen kertova tieto
- Ontologia on kokonaisuus, jossa määrittyy sovellettavan aiheen käsitteet ja niiden väliset suhteet
 - Ei universaaleja ontologiastandardeja, vaan sovelluskohteille on määriteltävä tarkoituksenmukainen ontologia
- Tässä keskitytään RDF/XML-formaatissa oleviin OLAP-ontologioihin

Tausta: Laskentaklusteri

- Analyysin kohteena CERNin tutkimustyössä käytettävä laskentaklusteri
 - Accounting-data koostuu laskentatöihin liitettävistä tietueista (ajankohta, muistinkäyttö, dataliikenne ym.)
 - Energiankulutusdata on osaklusterikohtaista kahden tunnin aikajaksoihin jaettua dataa
- Kahdentyypisiä laskentatöitä: Monte Carlo –simulaatiot sekä analyysityöt
 - Simulaatioissa muistinkäyttö suurta
 - Analyysissä dataliikenne suurta

Tavoitteet

- Ontologiamuokkaimen arviointi dataintegraatiovälineenä
 - Muokkain toteutettu harjoittelussa CERNissä
 - Laskentaklusteriontologian määrittelemine ja sen avulla datan tuonti analysoitavaksi
- Laskentaklusterin tehokkuusanalyysi
 - Tutkitaan laskentatöiden jakautumista ja laskentatöiden vaikutusta klusterin energiankulutukseen
 - Työkaluna R-ohjelmisto, koska ontologiamuokkain tuottaa sille tarkoitettua komentosarjaa

Tietolähteet ja aineistot

- Analyysin pohjana laskentaklusteriin liitettävät accounting- ja energiankulutusdatat
 - Kaksi accounting-dataa, joista toista käytetään laajuutensa vuoksi klusterin käyttöluonneanalyysiin ja toista yhdessä energiadatan kanssa energiankulutusanalyysiin
- Teorian pohjana dataintegraatioon ja ontologioihin liittyvät julkaisut

Menetelmät ja työkalut

- Ontologiamuokkain toteutettu Javascriptillä ja Shell-scriptillä, jolla saadaan Javascriptillä tuotettu tieto talletettua tiedostoihin
- Ontologiamuokkaimella määritellään ontologia ja liitetään ontologian elementit datatiedostojen tietonimikkeisiin, minkä jälkeen ontologiamuokkaimesta saadaan R-ohjelmistossa suoritettava komentosarja
- Dataa analysoidaan R-ohjelmistolla regressio- ja varianssianalyysin keinoin
 - Haasteena eri datojen yhdistäminen (kaksituntisjaksoinen energiadata vs. sekalaispituinen laskentatyödata)

Aikataulu

- 6/2012: Aiheeseen ja aineistoon tutustuminen sekä ontologiamuokkaimen toteutus
- 8/2012: Analyysi ja työn kirjoittaminen
- 2/2013: Aiheen esittely
- 3/2013: Valmiin työn esittely