

Tartu Ülikool

Matemaatika-informaatikateaduskond

Arvutiteaduse instituut

Andmeladude automaatne moodustaja (AAM)

Projekti aruanne

Dan Bogdanov
Kadri Hendla
Liina Kamm

Tartu 2004

Sisukord

1 Eesmärgid.....	3
2 Projekti käik.....	3
3 Realiseeritud funktsionaalsus.....	4
3.1 Lähteandmebaaside analüüs.....	4
3.2 Andmelao loomine.....	4
3.3 Hierarhiate loomine.....	5
3.4 Andmekuupide loomine.....	5
3.5 Päringute tegemine.....	6
4 Rakenduse ülevaade	6
5 Tehnilised nõuded.....	6
6 Viited.....	7
7 Märkused.....	7

1 Eesmärgid

Rühma eesmärgiks seadsime andmeladude automaatse loomise süsteemi väljatöötamise. Esitasime rakendusele järgmised nõuded:

1. sisaldab võimalikult palju erinevaid funktsioone,
2. on lihtne kasutada,
3. on lihtne realiseerida.

Nendest punktidest lähtuvalt otsustasime süsteemi luua veebipõhisena keeles PHP ning kasutada MySQL andmebaase. See võimaldas kiiresti realiseerida mitmeid erinevaid funktsioone ning varustada kogu rakendus graafilise kasutajaliidesega.

2 Projekti käik

Visioonidokumendis [1] täpsustasime loodavale süsteemile esitatavad nõuded. Seejärel koostasime süsteemi protsesside analüüsi ning mudelid, mis on ära toodud arhitektuuri- ja disainidokumendis [2].

Paralleelselt alustasime kasutajaliidese, andmebaasi analüüsi ning andmelao tähtskeemi koostamise realiseerimisega. Hiljem lisasime dimensionide ja mõõdikute valiku ning hierarhiate loomise süsteemi. Kogu senise töö võttis kokku faktitabeli andmebaasi kandmine ning andmetega täitmine.

Järgmises etapis lisasime andmekuupide loomise vahendid koos agregeerimise võimalusega. Lõpuks lisasime lihtsa päringute süsteemi ning täiendasime kasutajaliidest.

Koostasime ka kasutusjuhendi [3], mis kirjeldab süsteemi paigaldamist ja kasutamist.

3 Realiseeritud funktsionaalsus

3.1 Lähteandmebaaside analüüs

AAM oskab etteantud MySQL andmebaasi põhjal koostada kõigi selles olevate tabelite analüüsi. Rakendus loeb sisse andmebaasi tabelite kirjelduse ning salvestab selle enda andmebaasi. Salvestatakse tabelite nimed, erinevate kirjete arv tabelis, atribuutide nimed, andmetüübid ja tähtsamad parameetrid.

AAM püüab ka rekonstrueerida tabelite vahel määratud seoseid. Selleks peab võtmeatribuut olema kujul *tabelinimi_atribuudinimi*. Kui AAM leiab atribuudi, mille nimes on alakriips, võrdleb ta kriipsule eelnevat osa tabelite nimekirjaga ning järgnevat osa leitud tabeli atribuutide nimekirjaga. Kui seos leitakse, siis märgitakse see ära andmebaasi salvestatud kirjelduses.

3.2 Andmelao loomine

Analüüsitud andmebaasi põhjal on võimalik luua andmeladusid. Etteantud andmebaasikirjelduse põhjal esitab AAM kasutajale tabelid, millest kasutaja valib vajalikud atribuudid. Seejärel saab kasutaja võimaluse dimensioneerida ja moodikute valikuks ning hierarhiate loomiseks. Hetkel kehtib piirang, mille järgi peab andmelaos olema vähemalt üks dimension ja üks moodik.

Igale atribuudile lubab rakendus luua vaid teatud tüüpi hierarhiaid. Arvulisi hierarhiaid saab luua täis- ja reaalarvuliste atribuutide peal. Kuupäeva hierarhiaid saab luua kuupäeva tüüpi atribuutidel. Käsitsi saab hierarhiaid teha kõigil eelpoolnimetatud atribuuditüüpidel ning lisaks veel sõne tüüpi tunnustel.

Kogutud andmete põhjal koostab AAM faktitabeli ja hierarhiatabelid ning täidab need lähtebaasist võetud andmetega.

Tähelepanu! Praegune realisatsioon on meetodite testimise eesmärgiga, mistõttu on lähtebaasist andmete ülekandmise päring veel ebaoptimaalne. Samuti on ühe päringu tööaeg piiratud veebiserveris PHP lehe töötlemiseks antud ajaga (tihti 30 sekundit). Seetõttu ei saa süsteemi hetkel kasutada mahukate andmestike töötlemiseks.

3.3 Hierarhiate loomine

Kõigi valitud hierarhiate jaoks luuakse eraldi tabelid, mille võtmeks on atribuudi unikaalsed väärtused ning ülejäänud väljadel hierarhiatasemete väärtused. Faktitabeli ja hierarhiatabelite vahel on välisvõtme seos: faktitabelisse ei saa lisada väärtust, mida hierarhiatabelis ei ole. Arvestades andmelao iseloomu ei ole see ka vajalik.

Arvulise hierarhia puhul saab kasutaja määrata vahemiku laiuse. Laius 10 tähendab, et arvtelg jaotatakse vahemikeks 10 kaupa. Väärtused 2, 7 ja 9 lähevad 0-10 alla, samas kui 11 ja 18 lähevad 10-20 alla. Hierarhiatabelis on kirjas vahemiku esimene väärtus (2, 7 ja 9 puhul siis 0). Arvulist hierarhiat saab koostada suvaliste vahemike (ka murdarvuliste) järgi. Atribuudi nimeks pannakse *skaalaNN*, kus NN on vahemiku suurus (komakoht asendatakse alakriipsuga). Hierarhiatasemete väärtused arvutatakse valemiga

$$H = \frac{v}{N} \cdot H_{max}$$

Kuupäevalise hierarhia puhul eraldatakse lätevärtusest vastav osa (kuupäev, kuu jne) ja muudetakse see eraldi väärtuseks. Nädalapäevade kodeering on vastavalt andmebaasiserveri konfiguratsioonile kas pühapäev = 0 või esmaspäev = 0. Näiteks: Kuupäeva 9. detsember 2004 puhul kirjutatakse hierarhiatabelisse atribuudid *kuup2ev* = 9, *kuu* = 12, *aasta* = 2004, *aastakymme* = 2000.

Käsitsi koostatud hierarhiate puhul peab kasutaja igale lähtebaasi atribuudi unikaalsele väärtusele määrama kategooria, mille alla see kuulub. Kategooria võib olla nii sõne kui arv. Erinevate kategooriate hulk salvestatakse hierarhiaatribuudi andmetüübi *ENUM* kirjeldusse. Hetkel toetab AAM vaid ühte käsitsi loodud hierarhiat atribuudi kohta. Selle nimeks andmebaasis on *klass1*.

3.4 Andmekuupide loomine

Andmekuupe saab luua andmeladude põhjal. Kasutaja saab valida kaks liht- ja ühe agregeeritava atribuudi. Agregeerimine ei ole kohustuslik. Toetatud funktsioonid on kogus, summa, keskmine, miinimum ja maksimum. Kuubi nimi koostatakse atribuutide ja agregeerimisfunktsiooni nimedest.

Andmed lähtelao faktitabelist kopeeritakse andmekuupi. Hetkel ei ole andmekuupide loomise käigus võimalik kasutada hierarhiaid. Sellest piirangust saab mööda minna päringumootoriga, mis loob seose andmekuubi ja andmelao hierarhiatabelite vahel.

3.5 Päringute tegemine

Ülesandepüstituse kohaselt ei sisalda AAM päringumootorit. Lisatud on vahendid SQL-keelsete päringute edastamiseks andmelaole või andmekuubile. Graafilises liideses annab rakendus päringu tegemisel ette SQL-lause malli, mida saab täiendada.

4 Rakenduse ülevaade

Tegemist on klient-server süsteemiga, mis koosneb veebirakendusest ja andmebaasiserverist. Veebirakendus sisaldab algoritme ja kasutajaliidest. Andmebaasiserverit kasutatakse süsteemsete andmete salvestamiseks ja andmebaaside töötlemiseks.

AAM kasutab andmebaaside kirjelduste salvestamiseks enda andmebaasi. Baasi kirjeldus on toodud arhitektuuridokumendis. Põhjalikum päringusüsteem võiks tabelite ning atribuutide kohta andmete saamisel samuti tugineda sellele andmestikule. Andmebaasis on kirjas lähteandmebaaside, andmeladude ja andmekuupide kirjeldused ning omavahelised seosed.

Tähelepanu! Hetkel salvestab rakendus andmebaasi kasutamiseks vajalikud autentimisandmed (kasutajanimi ja parool) lahtise tekstina. Seega palume süsteemi testida ainult välisvõrgust eraldatud arvutites ning testotstarbeliste kasutajaandmetega!

5 Tehnilised nõuded

Infosüsteemi arendasime Apache 2 veebiserverit, mis kasutab PHP versiooni 4 ja MySQL andmebaasi versiooni 4.1.

Tähelepanu! Süsteem ei tööta varasemate MySQL versioonidega, mis ei toeta alampäringuid. MySQL versioon 4.1 on ametlikult stabiilseks kuulutatud ning üleminek sellele on igati soovitatav.

6 Viited

- [1] Andmeladude Automaatne Moodustaja – Visioon
- [2] Andmeladude Automaatne Moodustaja – Arhitektuur ja disain
- [3] Andmeladude Automaatne Moodustaja – Kasutusjuhend

7 Märkused

AAM graafilise logo kujundas Oliver Oberg (oliver@bmonde.net). Logo kasutatakse autori teadmisel ja nõusolekul.