

# Projekta aktivitātes „Pētījumi par datu noliktavām” atskaite par 2011. gadu

## **Pētījumu virzieni: paveiktais un rezultāti**

Aktivitātes ietvaros tiek veikti pētījumi par trim virzieniem: datu noliktavu evolūcija, datu noliktavu personalizācija un procesu datu noliktavas

### Datu noliktavu evolūcijas pētījumi

Iepriekšējos pētījumos izstrādātā datu noliktavas evolūcijas ietvara arhitektūras sastāvdaļa ir arī OLAP atskaišu rīks. Rīka darbības pamatā ir vairāki metadatu līmeņi.

Atskaišu rīks tiek izstrādāts kā daļa no evolūciju atbalstošas un lietotāju orientētas datu noliktavas. Rīks atļauj automātiski vai pusautomātiski apstrādāt noteikta tipa izmaiņas datu noliktavā, tai pat laikā orientēts uz viegli saprotamu un viegli saprotamu datu analīzi lietotājiem.

Datu noliktavas evolūcijas ietvars ietver 2 vides – izstrādes un lietotāja vidi, kur izstrādes vide ir pamatā dažāda tipa metadatu pārvaldības atbalsts, kā arī datu noliktavas izmaiņu atbalsts. Lietotāja vidē ir atskaišu rīks ar iespēju korekti, neatkarīgi no datu noliktavas izmaiņām (versijām) definēt un skatīt atskaites.

Datu noliktavas evolūcijas ietvars atbalsta fiziskas, loģiskas un semantiskas izmaiņas datu noliktavas shēmai, kas rada jaunu datu noliktavas versiju. Fiziskās izmaiņas ir piemēram, jauna atribūta vai dimensijas pievienošana shēmai, loģisko izmaiņu piemērs ir hierarhijas līmeņa izmešana, semantiskās izmaiņas attiecas uz eksistējoša datu elementa nozīmes maiņu.

Loģiskie metadati tiek izmantoti, lai aprakstītu datu noliktavas shēmu, fiziskie metadati – lai aprakstītu, kā datu noliktavas shēma tiks glabāta relāciju datu bāzē, semantiskie metadati apraksta glabāto datu nozīmi biznesa valodā, atskaišu metadati apraksta atskaišu definīcijas uz datu noliktavas shēmu, OLAP preferenču metadati apraksta lietotāju iestatījumus attiecībā pret atskaišu struktūru un datiem.

Šajā pētījumu virzienā strādāts pie atskaišu rīka metadatu modeļa attīstīšanas, lai tas ietvertu arī datu noliktavas personalizācijas nolūkiem vajadzīgos metadatus.

Iepriekš jau izstrādātie un citos rakstos aprakstītie līmeņi tagad papildināti ar vēl vienu – iestatījumu (preferenču) metadatu līmeni.

### Datu noliktavas personalizācijas pētījumi

Turpināts darbs pie datu noliktavas atskaišu rīka attīstīšanas, iekļaujot personalizācijas iespējas, izstrādāta metode semantisko metadatu izmantošanai datu noliktavas iestatījumu definēšanai.

Atskaišu rīka metadatu modelis ir papildināts ar preferenču metamodeli, kas apraksta 2 veidu lietotāju preferences –shēmu specifiskas (definētas attiecībā pret OLAP shēmas

elementiem- piemēram, dimensiju, hierarhiju u.c. ) un atskaišu specifiskas (definētas attiecībā uz atskaišu datiem, kas izveidotas konkrētajā OLAP shēmā) . Lietotājs var definēt savu intereses pakāpi katrai OLAP preferencei.

Piedāvātajā metadatu modelī ir nedefinētas saistības starp dažādiem metadatu līmeņiem, preferenču metadatu līmenis ir saistīts gan ar loģiskiem metadatiem, gan atskaišu metadatiem (piemēram, preference var būt definēta attiecībā pret pieļaujamo agregācijas funkciju, kas ir loģiskie metadati, vai preferences definīcija var būt saistīta ar atskaites elementu no atskaišu metadatu līmeņa).

Atskaišu specifiskās preferences ietver ierobežojumus uz atskaišu datiem, katra no preferencēm var saturēt ierobežojumu kopu, kas saistītas ar loģiskiem operatoriem, kā arī var saturēt izteiksmes, salīdzināšanu un konstantes.

OLAP preferences savukārt tiek definētas izmantojot atskaišu rīka semantiskā līmeņa metadatu jēdzienus, vai precīzāk, komponentes lietotāju preferencēm, kas definētas uz atskaišu struktūru, ir OLAP shēmu elementi no loģiskā metadatu līmeņa, kas atbilst semantiskā metadatu līmeņa konceptiem. Tas pats attiecas uz preferencēm, kas ietver atskaišu datus. Tādējādi pastāv sakarība starp preferenču metadatiem un semantiskajiem metadatiem. Pētījuma ietvaros tiek piedāvāts uzskatīt semantiskos metadatus kā līdzekli, lai formulētu lietotāju preferences datu noliktavu atskaitēm, pielietojot iepriekš definētus datu noliktavu elementu aprakstus ar semantiskiem metadatiem.

Pētījuma ietvaros izstrādāts algoritms, kā izveidot preferences (iepriekš aprakstītā situācijā, kad datu noliktavā tiek izmantots atskaišu rīks ar metadatiem, kura modelim ir 5 minūtes metadatu līmeņi) un kā preferences pielietot, lai lietotājam piegādātu (kā rekomendācijas) viņa vēlmēm atbilstošākās atskaites, kas iepriekš ir definētas atskaišu rīkā.

Vēl viena pētījuma mērķis ir izstrādāt un implementēt jaunu metodi datu noliktavas personalizācijai, kuras pamatā ir datu noliktavas personalizācijas metamodelis, lai sniegtu lietotājam iespēju iegūt no datu noliktavas informāciju, kas maksimāli atbilst lietotāja vajadzībām.

Var ieviest personalizācijas iespējas atskaišu rīkā, lai lietotājs varētu uzdot iestatījumus atskaišu struktūrai, saturam un noformējumam. Pirms sākt veidot lietotāju iestatījumu metamodeli atskaišu rīkam, bija nepieciešams klasificēt lietotāju iestatījumus.

Lietotāju iestatījumi tika sadalīti divās grupās: iestatījumi atskaišu saturam, t.i. struktūrai un datiem (OlapPreference), un iestatījumi atskaišu noformējumam (VisualLayoutPreferences). Iestatījumu metamodelis ir izveidots tā, lai tas varētu papildināt jau eksistējošo iepriekšējos pētījumos izstrādāto LU atskaišu rīka metamodeli.

Iestatījumus var sadalīt arī pēc informācijas uzkrāšanas principa *tiešajos*, t.i., ja lietotājs pats manuāli piešķir intereses pakāpi iestatījumu elementiem, un *netiešajos*, t.i., ja intereses pakāpes iestatījumu elementiem tiek izrēķināti automātiski.

Rakstā tika piedāvātas 2 metodes, kā piedāvāt lietotājam potenciāli interesantas atskaites, balstoties uz netiešajiem iestatījumiem OLAP shēmas elementiem (t.i., shēmai, dimensijām, atribūtiem, faktu tabulām, mērījumiem, hierarhijām) un pieļaujamām

agregātfunkcijām. *Hot-start* metode tiek pielietota tad, ja lietotājam ir bagāta aktivitātes vēsture atskaišu rīkā. *Cold-start* metode tiek pielietota tad, ja (i) lietotājs pirmo reizi izmanto atskaišu rīku, vai ja (ii) lietotājs izmanto atskaišu rīku diezgan pasīvi (ierakstu skaits aktivitātes vēsturē ir mazāks par iepriekšdefinētu *sliksni*). Savukārt, *sliksnis* ir pozitīva konstantes vērtība, kas atbilst lietotāja aktivitātes ierakstu skaitam žurnālfailā un kas tiek uzskatīts par pietiekamu, lai pārietu no *cold-start* uz *hot-start* metodi.

*COLD-START* metodes pamatā ir atskaišu struktūras aprakstīšana ar vektoriem, kuru koordinātes raksturo atskaišu OLAP shēmas elementu instanču esamību katrā atskaitē. Tālāk notiek vektoru līdzības izrēķināšana atskaišu pāriem (Vector/Cosine Similarity) un beigās notiek Top-*N* atskaišu atlasīšana ar visaugstākajām vektoru līdzības vērtībām, sakārtotām dilstošā secībā.

*HOT-START* metodes pamatā ir īpašā veidā svaru piekārtošana visām datu noliktavas shēmām un to elementiem, ko izmanto sekojošos algoritmos - (a) tiek aprēķināta un piefiksēta lietotāja profilā intereses pakāpe shēmas elementiem, kas ir sastopami lietotāja aktivitātes vēsturē; (b) katrai atskaitē tiek noteikti tajā izmantotie shēmas elementi; (c) katrai atskaitē tiek izrēķināta līdzība starp lietotāja intereses pakāpēm shēmas elementiem atskaitē un lietotāja profilā. Beigās notiek Top-*N* atskaišu atlasīšana ar visaugstākajām lietotāja profila un atskaišu intereses pakāpes līdzības vērtības, sakārtotām dilstošā secībā.

#### Aprobācijas vides izveide personalizācijas metožu pārbaudei.

Ir izveidota jauna aprobācijas vide, kur pielietot atskaišu rīku, kurā tiek realizētas pētījumos piedāvātās metodes, t.i. datuve datu noliktavas atskaitēm par studentu vērtējumiem, galvenā faktu tabula ir Studentu vērtējumi Moodle e-kursos. Dimensijas ir Laiks, E-kurss, E-kursa kategorija, Pasniedzējs, Students, Programma, Struktūrvienība. Analīzes mērķis ir gūt priekšstatu par to, kas un cik bieži lieto studentu vērtējumu grāmatas rīku.

#### Procesu datu noliktavu pētījumi

Pārskata periodā veikts pētījums, kā rezultātā izstrādāta datu noliktavas datu avotu lietotāju modeļu integrācijas metode, kas izmanto vispārīgu lietotāju modeli. Integrētais lietotāja modelis paredzēts informācijas uzkrāšanai par lietotāju un viņa darbībām datu avotos, kā rezultātā integrēto lietotāja informāciju iespējams izmantot par pamatu eksistējošas datu noliktavas jaunu datu modeļu izstrādei jaunām datu noliktavas daļām.

Turpināts darbs pie biznesa procesu pārvaldības sistēmas arhitektūras izstrādes, kas iekļāta, kā informācijas sistēmu adaptācijas platformas sastāvdaļa, iestrādājot arhitektūrā komponentes, kas nodrošina biznesa likumu definēšanas, implementācijas un kontroles atbalstu biznesa procesu izpildes kontroles vajadzībām.

Dažādu biznesa likumu implementācijas vajadzībām vienas programmatūras ietvaros kā risinājumu pielietojot programmatūras kā servisa pieeju (SaaS – *Software as a Service*) un to realizējot kā daudzu nomnieku vienas instances risinājumu, papildinot ar divu līmeņu adaptācijas pieeju, iepriekšējā pētījumā piedāvātai biznesa procesu pārvaldības sistēmas arhitektūrai izstrādāts modelis datu glabāšanai vairākām adaptētām vienas IS instancēm, kas realizē viena biznesa procesa dažādas variācijas.

Biznesa procesu izpildes kontrolei kā risinājums tiek piedāvāta datu noliktava, kuras modelis jāveido atbilstoši veikspējas indikatoru definīcijām. Veikta veikspējas indikatoru īpašību analīze un klasifikācija, izmantojot literatūrā pieejamu biežāk lietoto indikatoru datubāzi. Pētījuma ietvaros izstrādāts formāls indikatoru pieraksts, kā arī definēts 5 soļu veikspējas indikatoru dzīves cikls un indikatorus aprakstošie atribūti katra dzīves cikla posma kontekstā.

Veikts pētījums par reālā laika datu noliktavu realizācijas iespējām un nepieciešamību biznesa procesu pārvaldības datu noliktavām.

### **Dalība starptautiskās konferencēs un publikācijās**

No 6. oktobra līdz 8. oktobrim Rīgā norisinājās starptautiskā konference „10th International Conference on Perspectives in Business Informatics Research” (BIR 2011).

Konferences ietvaros tika organizēti arī semināri (workshops) un doktorantu konsorcijs (doctoral consortium). Lai veicinātu domu apmaiņu par datu noliktavu pētījumu aktivitātes virzienam tuviem informācijas integrācijas un personalizācijas jautājumiem, datu noliktavas grupas dalībnieki iesaistījās starptautiskās zinātniskās konferences BIR (*Business Informatics Research*)’2011 asociētās zinātniskās darbnīcas UOII (*User Oriented Information Intrgration*) organizēšanā un programmu komitejas darba koordinēšanā. Kopā tika akceptēti 10 raksti publicēšanai un prezentēšanai zinātniskajā darbnīcā UOII2011. No projekta datu noliktavu grupas UOII zinātniskajai darbnīcai iesniegti 3 raksti.

(UOII 2011) organizēšanā piedalījās projekta datu noliktavu aktivitātes dalībnieki Laila Niedrite (viena no priekšsēdētājiem), Darja Solodovņikova un Natālija Kozmina, sadarbībā ar otru semināra priekšsēdētāju LU DF profesoru Gunti Arnicānu. Laila Niedrite bija arī viena no konferences publikāciju krājumu redaktorēm.

BIR 2011 konferences un UOII 2011 semināra ietvaros datu noliktavu grupas dalībnieki prezentēja zinātniskos rezultātus, kā arī raksti tika iespiesti ar konferenci saistītos 3 dažādos izdevumos:

Janis Grabis, Marite Kirikova (Eds.), *Perspectives in Business Informatics Research*, 10th International Conference, BIR 2011, Riga, Latvia, October 2011, Proceedings, Springer, ISBN 978-3-642-24510-7.

- Aivars Niedritis, Laila Niedrite and Natālija Kozmina, *Performance Measurement Framework with Formal Indicator Denitions*,
- Janis Zuters, *Near Real-Time Data Warehousing with Multi-Stage Trickle and Flip*.

J.Grunspenķis (Ed.), *Scientific Journal of Riga Technical University*, Computer Science, Ser. 5, Vol. 46, Special issue for 10th International Conference on Perspectives in Business Informatics Research, Riga, RTU, ISSN 1407-7493

- Natālija Kozmina and Darja Solodovņikova, *On Implicitly Discovered OLAP Schema-Specific Preferences in Reporting Tool*,

Laila Niedrite, Renate Strazdina, Benkt Wangler (Eds.), *Perspectives in Business Informatics Research, Local Proceedings*, 10th International Conference, BIR 2011 Associated Workshops and Doctoral Consortium, Riga Technical University, 2011

- Aivars Niedritis, *Delivery of Consistent and Integrated User's Data within a Multi-Tenant Adaptive SaaS Application*,
- Aivars Niedritis, Laila Niedrite and Natalija Kozmina, *Integration Architecture of User Models*,
- Darja Solodovnikova and Natalija Kozmina, *Determining Preferences from Semantic Metadata in OLAP Reporting Tool*,

No UOII 2011 seminārā prezentētajiem rakstiem viens datu noliktavu grupas raksts N. Kozmina, D. Solodovņikova, “*Determining Preferences from Semantic Metadata in OLAP Reporting Tool*” izvirzīts pēckonferences rakstu izlasei - krājumam Springer izdevniecībā, sagatavota raksta paplašinātā versija un iesniegta atkārtotai izvērtēšanai.

Laila Niedrite kopā ar datu noliktavu grupas dalībnieču Darjas Solodovņikovas un Natālijas Kozminas atbalstu strādāja pie BIR 2011 semināru pēckonferences paplašināto rakstu izlases krājuma sastādīšanas. Rakstu krājums tiks izdots Springer izdevniecībā un Laila Niedrite ir viena no rakstu krājuma redaktoriem (kopā ar vēl diviem redaktoriem - Renate Strazdina, Benkt Wangler).

### **Rezultātu izmantošana akadēmiskajā darbā**

Literatūras studiju un pētījumu rezultāti izmantoti mācību procesā datorzinātņu bakalaura programmas studentiem paredzētajā kursā „Specseminārs I” 2011.g. rudens semestrī (6 semināri), sagatavoti materiāli un nolasīti studentu specsemināru laikā par personalizāciju datu noliktavās, par lietotāju profilu veidošanas tehnikām, par lietotāju iestatījumiem datu noliktavu jomā, kā arī par procesu datu noliktavu pētījumiem.

### **LU 69. zinātniskā konference**

LU 69. zinātniskajai konferencē nolasīts referāts L.Niedrite, N.Kozmina, D.Solodovņikova „Datu noliktavu pētījumi Latvijas Universitātē”, prezentēts 2011. gada 11. februārī, Datorzinātņu un informācijas tehnoloģiju apvienotajā sekcijā (uzstājās N.Kozmina).

### **Projekta seminārs**

Viedo sensoru un kvantu skaitļošanas seminārā, ko organizēja A.Ambaiņa un L.Seļavo jauno zinātnieku grupas, 27.05.2011. tika nolasīts referāts Natālija Kozmina, „Personalizācijas un evolūcijas atbalsts datu noliktavas atskaišu rīkā”.

### **Grupās dalībnieku doktorantūras studijas**

Natālija Kozmina 2011. gadā pabeidza doktorantūras 1. studiju gadu un reģistrējās 2. studiju gadam, regulāri piedalījās datorzinātņu doktorantu seminārā un arī 1 reizi referēja par savu paveikto pētniecībā, saistībā ar paveikto projekta ietvaros.

Darjas Solodovņikovas promocijas darbs (vadītāja Dr.sc.comp. Laila Niedrite), „Uz datu noliktavas shēmas evolūciju orientēts vaicājumu definēšanas un attēlošanas rīks” tika aizstāvēts 2011. gada 9. februārī.

Aivars Niedritis pabeidza rakstīt sava promocijas darba melnrakstu, un 2011. gada septembrī nokārtoja promocijas eksāmenu, līdz ar to pabeidza LU datorzinātņu doktora studiju programmu. Turpinās darbs pie promocijas darba gala variantā.