

Projekta aktivitātes „Pētījumi par datu noliktavām” atskaite

Pētījumu virzieni: paveiktais un rezultāti

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

Datu noliktavas personalizācijas pētījumi

Projekta laikā veiktas literatūras studijas par personalizāciju datubāzu kontekstā - lietotāju iestatījumu organizēšana pēc *iestatījumu tīkla* principa; datu noliktavu kontekstā - OLAP personalizācijas dimensiju apraksts: personalizācijas faktori un darbības, dinamiskā personalizācija, lietotāju iestatījumiem un personalizāciju informācijas izguves (*Information Retrieval*) kontekstā - lietotāju profilu veidošana un lietotāju iestatījumu novērtēšana. Tika izpētīti raksti, kas attiecās uz rekomendācijas sistēmām un to īpašībām, par rekomendācijām atsevišķām cilvēku grupām, rekomendācijas sistēmu lietotāju novērtēšanu, rekomendāciju sistēmu projektēšanas metodēm

Balstoties uz literatūras studijām, notika darbs pie pārskata raksta veidošanas par datu noliktavu personalizācijas problēmām. Pētījumu rezultāti apkopoti zinātniskā rakstā N. Kozmina, L. Niedrīte „Research Directions of OLAP Personalization”. Rakstā tika aplūkotas piecas pieejas OLAP personalizācijas ieviešanai: iestatījumu konstruktori, dinamiskā personalizācija, vizuālais OLAP, rekomendācijas ar lietotāju sesiju analīzi un rekomendācijas ar lietotāju profilu analīzi. Tika piedāvāts šo pieeju novērtējums, lai noskaidrotu a) personalizācijas pielietojamību OLAP shēmas elementiem, agregātfunkcijām, OLAP operācijām, b) ierobežojumus (*hard/soft constraints*), kas ir pielietotas katrā pieejā, c) metodes lietotāju iestatījumu un lietotāju informācijas iegūšanai. Šī raksta mērķis ir sistematizēt jau piedāvātas idejas OLAP personalizācijas jomā, lai atrastu iespējas tālākai ideju papildināšanai vai jaunu pieeju veidošanai OLAP personalizācijas jomā.

Raksts N. Kozmina, L. Niedrīte „Research Directions of OLAP Personalization” tika iesniegts 19. konferencei par Informāciju Sistēmu Attīstību – Information Systems Development (ISD 2010, Prague, Czech Republic). Raksts tika pieņemts publicēšanai pēckonferences rakstu krājumā Springer izdevniecībā. Raksts prezentēts konferencē.

Tālāk tiek turpināta iestatījumu ieviešanas tēma datu noliktavās un detalizētāk izpētīta viena no pieejām – rekomendācijas ar lietotāju profilu analīzi. Darba gaitā izstrādāts lietotāju profilu metamodelis, kas iekļauj sevī lietotāja, laika, atrašanas vietas, lietotāja tiesību, iestatījumu, aparatūras, atskaites un rekomendācijas profilus. Profilu metamodeļa veidošana balstās uz doktorantes Darjas Solodovņikovas promocijas darba ietvaros izstrādātā atskaišu rīka (turpmāk - EVO) īpašību pētīšanu, kā arī uz zinātniskās literatūras pētīšanu, kas attiecās uz lietotāju modeļiem un lietotāju kontekstu. Tika uzsākts pētījums, kas ir saistīts ar eksistējošo rekomendāciju sistēmu īpašībām – izmantotajiem kritērijiem rekomendāciju atlasei, rekomendāciju vizuālo izkārtojumu, u.c. Šī pētījuma mērķis ir piedāvāt algoritmu rekomendāciju ģenerēšanai, kritērijus rekomendāciju atlasei, kā arī,

iespējamo LU atskaišu pārvaldības rīka EVO prototipa jaunu versiju, kurā ir integrētas datu noliktavas atskaišu rekomendācijas.

Pētījumu rezultātā tika izstrādāta lietotāju aprakstošo profilu klasifikācija, kas iekļauj sevī sešus profilus – mijiedarbības, lietotāja, laika, atrašanās vietas, iestatījumu un rekomendāciju profilus un modelis, kas apraksta lietotāju profilu savstarpējās sakarības, kā arī diagramma, kas raksturo profilu datu veidus (kopā 5 veidi – konteksta, statistiskie, analīzes, aktivitātes un tiešā ievada dati). Lai konstruētu atribūtu kopas mijiedarbības, lietotāja, laika, atrašanās vietas profiliem, tika izmantota metode, kas iekļauj sevī dažādu informācijas avotu pētīšanu, t.i., datu noliktavas literatūru, Common Warehouse Metamodel (CWM) standartu, zinātniskie un tehniskie rakstus, tīmekļa servisu, LU datu noliktavu u.c. (kopā 13 dažādi avoti). Rezultātā tika veikts projektējums mijiedarbības, lietotāja, laika, atrašanās vietas profilu klašu diagrammām. Tika piedāvāti vairāki scenāriji, kas apraksta iespējamus lietotāja iestatījumu tipus. Tika izveidots metamodelis, kas apraksta lietotāja atskaites vizuālā noformējuma iestatījumus un iestatījumus OLAP shēmas elementiem. Tiek attīstītas idejas par personalizācijas pieejas, kas balstās uz lietotāja iestatījumiem, ieviešanu LU atskaišu pārvaldības rīkā. Tika piedāvāta un formāli pierakstīta kā UML aktivitāšu diagramma rekomendāciju veidošanas algoritma ideja.

Darba gaitā tiek izpētītas eksistējošo rekomendāciju sistēmu rekomendāciju atlases kritērijus un izmantotās metodes. Rekomendāciju sistēmas var iedalīt piecās lielās grupās: satura balstītas (Content-based), uz sadarbības filtrēšanas (Collaborative filtering), uz zināšanu jeb noteikumu balstītas (Knowledge Engineering or Rule-based), demogrāfiskās iedalījuma balstītas (Demographic systems) un hibrīdās (Hybrid) jeb kombinētās. Katrai metodei tika piefiksētas savas priekšrocības un trūkumi. Šī pētījuma mērķis ir noskaidrot, kurus kritērijus un metodes varētu pielietot atskaišu rīka rekomendāciju atlasei.

Tika izstrādāts raksts par lietotāju aprakstošo profilu izmantošanu personalizācijas ieviešanai datu noliktavās – N. Kozmina, L. Niedrite „*OLAP Personalization with User-describing Profiles*” un tika iesniegts starptautiskai konferencei 9th International Conference on Perspectives in Business Informatics Research – BIR2010 (Rostock, Germany). Raksts prezentēts konferencē un publicēts:

N.Kozmina, L.Niedrite. OLAP Personalization with User-Describing Profiles. // Perspectives in Business Informatics Research, Lecture Notes in Business Information Processing, 2010, Volume 64, Springer, p. 188-202

Sākot no 2010. gada septembra tika uzsākts aprobācijas vides, kur notiks OLAP personalizācijas ieviešana, sagatavošanas darbs. Tiek veikts darbs, kas ir vērsts uz jaunas studentu vērtējumu datuves izveidošanu. Šī darba gaitā tapa dokumentācija par Moodle rīka „Vērtējumu grāmata” datu modeli. Tika izstrādāta dokumentācija, kurā piedāvāti iespējami faktu tabulu un dimensiju varianti, to implementācijas scenāriji un īpatnības, izvēlēts konkrēts implementācijas scenārijs un uzsākta datuves izstrāde aprobācijas videi. Datuves izstrādes procesā tika uzprojektētas un uzbūvētas nepieciešamās faktu tabulas, uzrakstītas un notestētas procedūras datu pārvešanai no Moodle vides uz datu noliktavu, datu apstrādei un ielādei faktu tabulā, datuves datu pārvešanai no izstrādes vides darba vidē. Pašlaik tika uzsākts pēdējais aprobācijas vides sagatavošanas etaps – notiek loģisko un fizisko atskaišu datu sasaiste un atskaišu veidošana atskaišu rīkā.

Datu noliktavu evolūcijas pētījumi

Datu noliktavām ir tendence attīstīties mainīgo datu avotu un lietotāju darījumprasību dēļ. Visas šīs izmaiņas ir atbilstoši jāapstrādā, tādēļ datu noliktavas izstrādes procesu var uzskatīt par nebeidzamo. Tika piedāvāts uz evolūciju orientētu un uz lietotāju vērstu datu noliktavas metamodeli, kas, no vienas puses, atļauj pārvaldīt datu noliktavu evolūciju automātiski vai pusautomātiski, un, no otras puses, tas nodrošina lietotājiem saprotamas, vieglas un pārredzamas datu analīzes iespējas. Piedāvātā pieeja atbalsta datu noliktavas shēmu versijas un datu semantikas versijas.

Pētījuma rezultāti apkopoti rakstā Solodovnikova D., Niedrite L. 'Evolution-Oriented User-Centric Data Warehouse'. *Proceedings of the 19th International Conference on Information Systems Development by Springer, Prague, Czech Republic, 2010* (tiks publicēts pēckonferences rakstu krājumā).

Projekta ietvaros tika turpināts darbs pie Darjas Solodovņikovas promocijas darba pabeigšanas, promocijas darbs „Uz datu noliktavas shēmas evolūciju orientēts vaicājumu definēšanas un attēlošanas rīks” tika aizstāvēts 2011. gada 9. februārī.

Promocijas darba galvenais ieguldījums ir pieeja daudzversiju datu noliktavas atbalstam, kas iekļauj gan fizisko shēmas versiju reprezentāciju datu bāzē, gan iespēju modelēt datu noliktavas loģisko shēmu vairākām versijām, kā arī algoritms atskaišu konstruēšanai un izpildīšanai uz vairākām datu noliktavas shēmas versijām. Viens no galvenajiem promocijas darba rezultātiem ir piedāvātais datu noliktavas metamodelis, kas veidots balstoties uz CWM metadatu standartu, atbilstošās UML pakotnes paplašinot ar shēmas versiju aprakstīšanai vajadzīgiem metadatiem. Metadati sastāv no fiziskā, loģiskā un semantiskā līmeņa un atskaišu specifikācijas.

Darjas Solodovņikovas promocijas darba ietvaros izstrādāts atskaišu rīka prototips, veikta rīka aprobācija LU e-studiju datuvē, realizējot e-studiju datuves izmaiņu apstrādi rīka metadatos, versiju definēšanai izmantojot promocijas darbā piedāvāto rīka metamodeli.

Procesu datu noliktavu pētījumi

Veikti pētījumi biznesa procesa pārvaldības sistēmas arhitektūras izveidei, kurā iekļautas datu noliktavai nepieciešamās komponentes un kas nodrošinātu iespēju elastīgas un izmaksu efektīvas IS izstrādei tīmeklī. Tika pētītas sekojošas pieejas: 1) „SPL – software product line”, kur pamatā iepriekš definēta arhitektūra un zināmas pamatfunkcijas. Visi atkalizmantojamie komponenti tiek iepriekš jau iekļauti, lai iegūtu atšķirīgu instanci programmatūrai, tradicionāli tiek veiktas izmaiņas programmatūras kodā. 2) ”Software as a service (SaaS)”: jauns veids, kā nodrošināt piekļuvi programmatūrai, ko piedāvā kā pakalpojumu dažāda lieluma organizācijām, lai nodrošinātu šīm organizācijām līdzīgus biznesa procesus; pakalpojuma sniedzējs izvieto SaaS programmatūru uz saviem serveriem, uztur gan programmatūru, gan infrastruktūru. Eksistē dažādi realizācijas veidi, par piemērotāko tika atzīts konfigurējams daudzu nomnieku (multi-tenant) vienas instances risinājums uz pakalpojuma piegādātāja servera.

Biznesa procesu pārvaldības sistēmas arhitektūras izstrāde iekļauta, kā informācijas sistēmu adaptācijas platformas sastāvdaļa, iestrādājot arhitektūrā komponentes, kas nodrošina biznesa likumu definēšanas, implementācijas un kontroles atbalstu biznesa

procesu izpildes kontroles vajadzībām, veiktā pētījuma rezultāts, ir informācijas sistēmu adaptācijas arhitektūras koncepcija.

Tīmekļa informācijas sistēmas (TIS) arhitektūra, kas nodrošinātu tās izmantošanu vairākās vienas biznesa jomas organizācijās ar līdzīgiem biznesa procesiem, adaptējot to katras atsevišķas organizācijas vajadzībām, sastāv no sekojošiem komponentiem: 1) Konteksta monitora, kas tajā brīdī, kad lietotājs pirmo reizi pieslēdzas sistēmai, nosaka sistēmas lietošanas konteksta īpašības, 2) Adaptācijas komponenta, kas nodrošina tai divu līmeņu adaptāciju: a) “rupjo adaptāciju” – kas veic sistēmas lietotājam pieejamo funkciju grupu noteikšanu atbilstoši sistēmas lietošanas kontekstam (organizācijai), kā rezultātā tiek izveidota adaptēta TIS organizācijas instance; b) “detalizētu adaptāciju”, ko nosaka savukārt lietotāja profils un kam ir divi soļi: “detalizēta adaptācija – navigācijai”, un “detalizēta adaptācija – saturam”: atbilstoši lietotāja profilam tiek izveidots lietotāja TIS gala variants, iekļaujot tajā arī lietotājam atļautos datus. 3) Darbību monitora, kas tiek izmantots, lai uzkrātu datus par izmantotajām TIS funkcijām, lai šo informāciju izmantotu TIS darbības operatīvai analīzei un vēlākai šo datu izmantošanai tālākai TIS attīstīšanas stratēģijas izstrādei. 4) TIS datu slāņa, kas satur profilus, kas tiek izmantoti abu līmeņu adaptācijā un TIS atbilstošos biznesa datus. TIS datu slānī tiek glabāta arī informācija par TIS funkciju darbību vēsturi, ko savāc darbību monitors.

Darba rezultāti par informācijas sistēmas arhitektūras izveidi, kurā iekļauta procesu monitorēšanas komponente ar datu noliktavu un izmantotas pieejas, kas nodrošinātu elastīgas un izmaksu efektīvas IS izstrādi, izmantoti raksta sagatavošanai A. Niedrītis, L.Niedrīte „The Adaptation of a Web Information System: a Perspective of Organizations”, kas pieņemts 19. starptautiskajā ISD konferencē (International Conference on Information Systems Development). Raksts prezentēts konferencē un tiks publicēts Springer izdevniecībā pēckonferences rakstu krājumā.

Biznesa procesu pārvaldības ietvaram kā vienu no alternatīvām iespējām, ko ņemt par pamatu ietvara definēšanai, tika izskatīti ISTehnoloģijas realizācijas principi un metamodelis. ISTehnoloģija sastāv no metamodela un lietojumprogrammām, lai definētu platformneatkarīgus organizācijas biznesa modeļus un realizētu šos modeļus specifiskās platformās. ISTehnoloģijas pamatnostādnes tika izvērtētas un papildinātas ar jaunākām atziņām no Informācijas Sistēmu izstrādes pētījumiem – par modeļu bāzētu arhitektūru (MDA) un par SPL pieeju (Software product line). Atbilstoši jaunajām nostādnēm aprakstīts ISTehnoloģijas metamodelis, kura metamodelis sastāv no vairākiem moduļiem: darba vietas (Workplace), organizācijas (Organizational), objektu (Object), procesu (Process), atskaišu (Report) un monitorēšanas (Monitoring) moduļiem. Darba rezultāti par ISTehnoloģijas pamatā esošiem principiem un metamodēli izmantoti raksta sagatavošanai J.Iljins, M.Treimanis „From Organization Business Model to Information System: One approach and Lessons Learned”. 19. starptautiskai ISD konferencē. Raksts pieņemts publicēšanai springer izdevniecībā pēckonferences rakstu krājumā.

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

Tika veikti pētījumi saistībā ar galveno veikspējas indikatoru (KPI) definēšanas, implementācijas un kontroles atbalstu procesu izpildes kontroles vajadzībām. Tika pētītas pieejas mērķu modelēšanai, mērķu un biznesa likumu definīciju saistībai ar galvenajiem veikspējas indikatoriem – lai varētu veikt biznesa procesu pārvaldību, šim nolūkam definējot piemērotu ietvaru.

Tika izveidots vispārējs „Plāns mašīnmācīšanās metožu izmantošanas lietderības pārbaudei procesu vadībā, izmantojot datu noliktavas”, t.sk. mērījumu analīzes modeļa precizēšanai. Ietvaros tika veikta dažādu datizraces metožu izpēte, kas varētu tikt izmantotas risinājuma izstrādē no laika apstrādes viedokļa. Tika apzinātas galvenās prasības un iespējamie optimizācijas mehānismi biznesa procesu datu apstrādes „tieši laikā” nodrošināšanai datu noliktavā.

Dalība starptautiskās konferencēs

2010. gada 24. augusts - 28. augusts, konferencē par informāciju sistēmu attīstību – Information Systems Development (ISD 2010, Prague, Czech Republic) piedalījās vairāki grupas dalībnieki ar prezentācijām:

Natālija Kozmina uzstājās ar prezentāciju par rakstu N. Kozmina, L. Niedrīte „Research Directions of OLAP Personalization”.

Darja Solodovņikova uzstājās ar prezentāciju par rakstu Solodovņikova D., Niedrīte L. ‘Evolution-Oriented User-Centric Data Warehouse’.

Aivars Niedrītis uzstājās ar prezentāciju par rakstu A. Niedrītis, L. Niedrīte „The Adaptation of a Web Information System: a Perspective of Organizations”.

Laila Niedrīte piedalījās konferencē kā visu 3 rakstu līdzautore.

2010. gada 28. septembris - 2. oktobris: Natālija Kozmina piedalījās 9. starptautiskajā konferencē BIR 2010 („Perspectives in Business Informatics Research”), kas notika Rostokā, Vācijā, un uzstājās ar prezentāciju par rakstu N. Kozmina, L. Niedrīte „*OLAP Personalization with User-describing Profiles*”.

Rezultātu izmantošana akadēmiskajā darbā

Literatūras studiju un pētījumu rezultāti izmantoti mācību procesā datorzinātņu bakalaura programmas studentiem paredzētajosursos „Specseminārs I” un „Specseminārs II”. 2009.g. rudens semestrī (decembrī – 4 semināri), 2010.g. pavasara semestrī (3 semināri) un 2010.g. rudens semestrī (4 semināri) sagatavoti materiāli un nolasīti studentu specsemināru laikā par vaicājumu personalizāciju datu noliktavās un lietotāju profilu veidošanas tehnikām, par lietotāju iestatījumiem datu noliktavu jomā un vaicājumu personalizāciju datubāzēs, par personalizāciju informācijas izguves jomā. Studentu patstāvīgam darbam tika atlasīti 8 raksti par rekomendāciju sistēmām.

LU 68. zinātniskā konference

Tika nolasīts referāts N. Kozmina, L. Niedrīte, D. Solodovņikova. “Datu noliktavu personalizācija - pētījumu virzieni” Referāts LU 68. zinātniskās konferences sekcijā “Datorzinātne un informācijas tehnoloģijas”, 19.02.2010.

Projekta seminārs

2010. gada jūnijā Datu noliktavu pētnieku grupa organizēja projekta 4. semināru, kas notika divās daļās - 01.06.2010 un 17.06.2010 un tika prezentētas sekojošas tēmas par pētījumiem projektā:

Darja Solodovņikova "Datu noliktavu evolūcijas problēmas" (01.06.2010)

Laila Niedrīte "Procesu datu noliktavas" (01.06.2010)

Laila Niedrīte "Metodes datu noliktavu konceptuālo modeļu izveidei: mērķu bāzēta metode procesu datu noliktavas gadījumā" (17.06.2010)

Natālija Kozmina "Datu noliktavu personalizācija" (17.06.2010)

Grupās dalībnieku doktorantūras studijas

Natālija Kozmina 2010. gada augustā iestājās LU Datorzinātņu doktorantūrā, kopā ar vadītāju Dr.sc.comp. Lailu Niedrīti sagatavoja pieteikumu promocijas darba tēmai – „Uz metadatiem balstīta datu noliktavu personalizācija”.

2010. gada decembrī Natālija Kozmina datorzinātņu doktorantūras seminārā uzstājās ar prezentāciju par sava promocijas darba mērķi, uzdevumiem, iestrādātajiem un rezultātiem, kas iegūti pētījumos, strādājot projektā.

2010. gada 26. maijā Aivars Niedrītis (vadītājs prof. Māris Treimanis) datorzinātņu doktorantūras seminārā uzstājās ar prezentāciju par promocijas darba izstrādes gaitu un rezultātiem, kas ietver arī pētījumus, kas izmantoti rakstā A.Niedritis, L.Niedrīte „The Adaptation of a Web Information System: a Perspective of Organizations”, ISD, 2010.

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