



Eiropas Sociālā fonda projekts

“Datorzinātnes pielietojumi un tās saiknes ar kvantu fiziku”

Nr.2009/0216/1DP/1.1.1.2.0/09/APIA/VIAA/044

Personalizācijas un evolūcijas atbalsts datu noliktavas atskaišu rīkā

Natālija Kozmina

DATU NOLIKTAVU PĒTNIECĪBAS GRUPAS TĒMAS

- Procesu datu noliktavas
 - Autori: L.Niedrīte, A.Niedrītis
- Datu noliktavu evolūcija
 - Autori: D.Solodovņikova, L.Niedrīte
- Datu noliktavu personalizācija
 - Autori: N.Kozmina, L.Niedrīte



DATU NOLIKTAVU EVOLŪCIJA

○ Izmaiņas

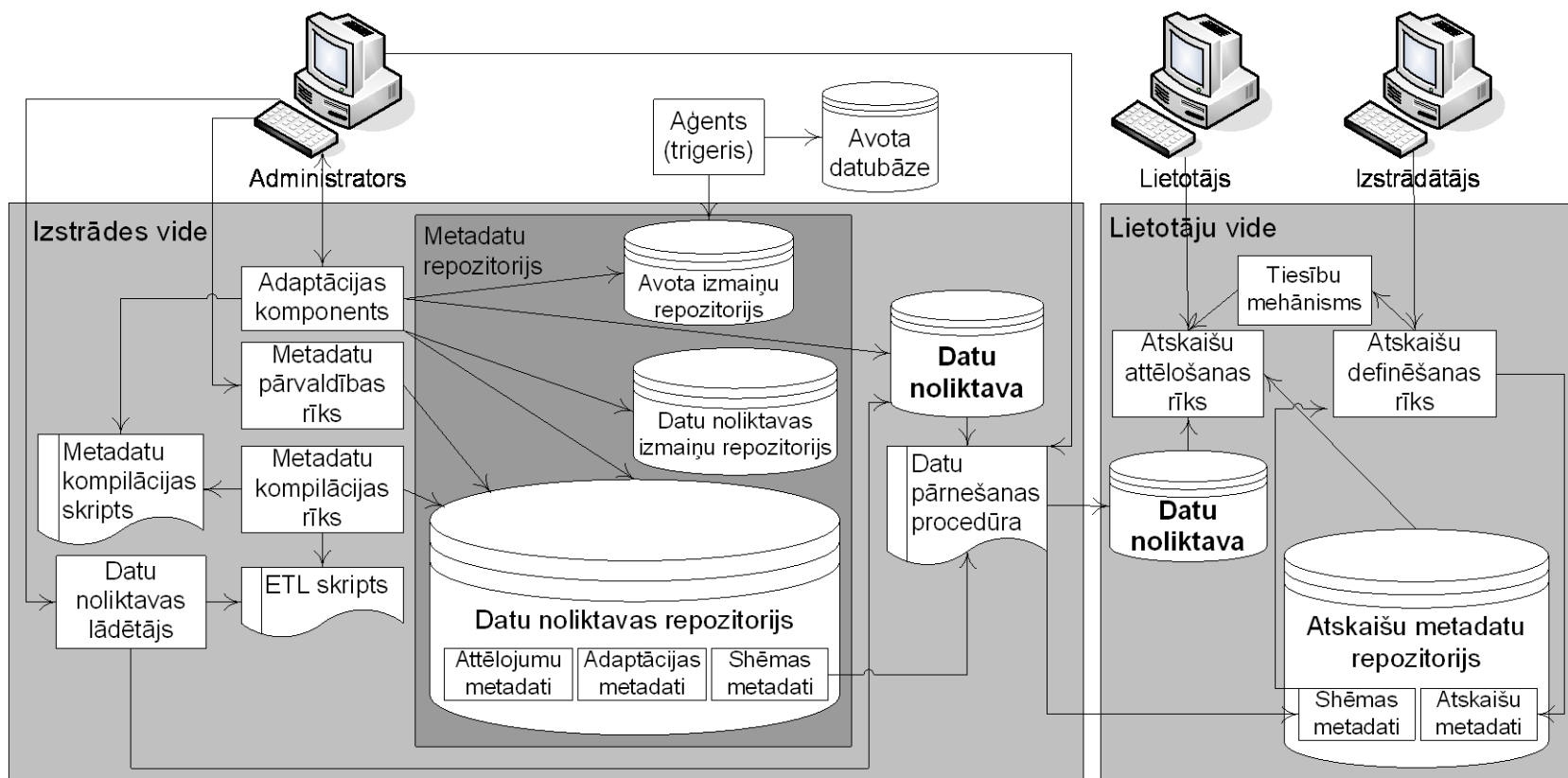
- Datu avotos
- Darījumprasībās

○ Sekas

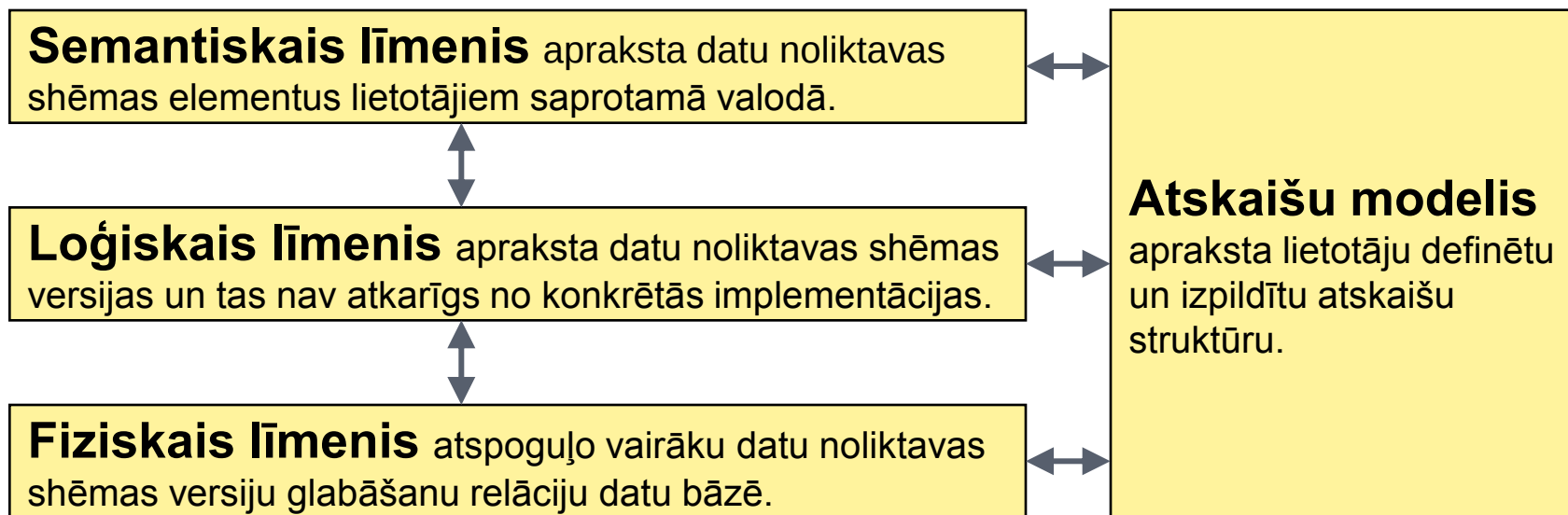
- Izmaiņas datu noliktavas shēmās
- Kļūdainie datu iegūšanas, pārveidošanas un ielādes (ETL) procesi
- Kļūdainas atskaites uz datu noliktavas shēmu



DATU NOLIKTAVAS EVOLŪCIJAS ARHITEKTŪRA



DAUDZVERSIJU DATU NOLIKTAVAS MODELIS UN METADATI



- Darjas Solodovņikovas promocijas darbs tika veiksmīgi aizstāvēts 2011. gada 9. februārī.



ATSKAIŠU ATTĒLOŠANAS RĪKS (1/2)

Shēmu versijas - Mozilla Firefox

Shēmu versijas

Fails ▼ Labot ▼ Skatīt ▼ Noformēt ▼ Darbarīki ▼ Palīdzība

Pasniedzēju atskaite

Studentu klikšķu skaits komponentos

Versijas atskaitei

Šo atskaiti ir iespējams attēlot saskaņā ar sekojošajām versijām:

Nosaukums	Spēkā no	Spēkā līdz	Ir pēdējā	
E-studiju datuve - 1. versija	2004-11-14	2005-01-27		Izvēlēties

Šo atskaiti ir iespējams attēlot atsevišķi katrai no sekojošajām versijām: [Izvēlēties](#)

Nosaukums	Spēkā no	Spēkā līdz	Ir pēdējā	
E-studiju datuve - 2. versija	2005-01-28		Ir pēdējā	
E-studiju datuve - 1. versija	2004-11-14	2005-01-27		

Atcelt

Done

ATSKAIŠU ATTĒLOŠANAS RĪKS (2/2)

Atskaišu rīks - Mozilla Firefox

Atskaišu rīks

Fails ▾ Labot ▾ Skatīt ▾ Noformēt ▾ Darbarīki ▾ Palīdzība

Pasniedzēju atskaite

Studentu klikšķu skaits komponentos

Atskaite ir attēlota saskaņā ar sekojošo versiju:

Nosaukums	Spēkā no	Spēkā līdz	Ir pēdējā
E-studiju datuve - 1. versija	2004-11-14	2005-01-27	

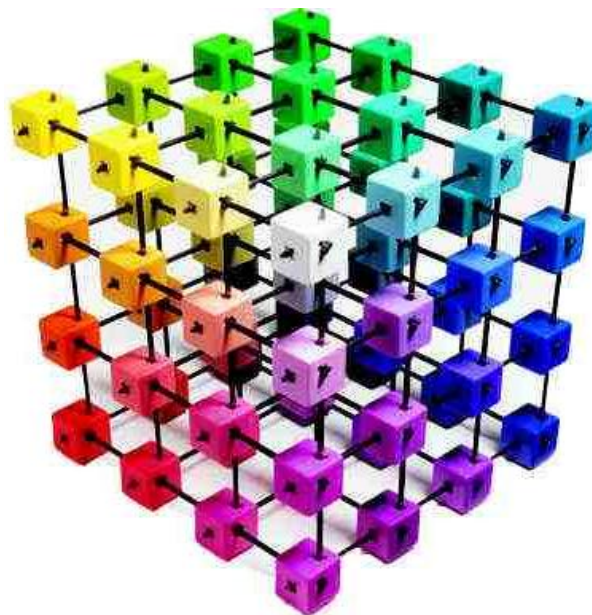
Studentu klikšķu skaits komponentos
Parametru vērtības: Kurša kods: DatZ4029

Gads	Mēnesis	Komponentu grupa	Mācību materiāli			Pārējās darbības	Sadarbības rīki		Saskarne	Saziņas rīki	
		Komponentu apakšgrupa	Darbs ar mācību materiāliem	Mācību materiālu rīki	Vārdnīca un priekšmetu rādītājs	Pārējās darbības	Mājas lapas	Prezentācijas	Saskarne	Diskusijas	Kalendārs
2004	Decembris		3569	6	1	1383		2042	3623	6118	753
	Novembris		1635	1	2	1075		229	951	266	188
2005	Aprīlis		1623	2		13	58	64	313	82	112
	Augusts									9	6
	Decembris		2721	4		2	123	1566	1902	1730	314
	Februāris		2066	1		2	28	46	329	284	223
	Janvāris		1997					7	90	962	417
	Jūnijs		217				4	6	40	6	39
	Maijs		306			6	19	31	336	48	60
	Marts		1857			3	29	82	542	89	132
	Novembris		4756		1		93	348	1418	437	262
	Oktobris		4297				87	305	747	465	326
	Septembris		1541				69	236	492	303	246

Lappuse << 1 / 2 >> 1 Rādīt

Done

DATU NOLIKTAVU PERSONALIZĀCIJA



PERSONALIZĀCIJAS PIEMĒRS NO GOOGLE

o iGoogle

The screenshot shows the iGoogle homepage with a tropical background featuring palm trees and a blue sky. The main search bar is at the top, with the iGoogle logo on the left. Below the search bar are buttons for "Google Search" and "I'm Feeling Lucky". To the right of the search bar are links for "Advanced search" and "Language tools". At the bottom right of the header, there are links for "Change theme from Key Discayne..." and "Add gadgets".

Home

- Gadget shares
- Updates
- Friends
- Chat**
- Search contacts...
- Natalie Kozmina
- Set status here

DELFI

- [+ 'Moody's': Griekijas maksātnešpēja negatīvi ietekmēs citas Eirozonas valstis](#)
- [+ Video: 'VW' tirgotājs ar ķermeņa svaru demonstrē 'Golf' durvju izturību](#)
- [+ Valdība piešķir 5.87 miljonus latu ceļu remontiem, kompensējot ziemas darbus](#)

www.lu.lv: jaunākās ziņas

- [+ Noslēgusies Jauno fiziku skolas 1. sezona](#)
- [+ LU Cietvielu fizikas institūtā noslēgušās IV Latvijas Saules kausa sacensības](#)
- [+ Apmeklētājus iepriecina pārmaiņas LU Bibliotēkas Centrālajā bibliotēkā](#)

Motivational Quotes of the Day

 *Hope is necessary in every condition.*
Samuel Johnson (1709 - 1784)

Windows Cloud Hosting
Public or Private Clouds Configured in Minutes!
  **Ads by Google**

Sticky Note

Color Junction

apollo.lv - Visi raksti, Laika ziņas, horoskopi, labākie val

- [+ DJ Ella nodibina savu ierakstu leiblu un ar Tito Torres izdod albumu](#)
- [+ No izlases kandidātu pulka atskaita Dillon un Reki](#)
- [+ Kļiedē aizspriedumus un neziņu par dzemdes kakla vēzi](#)

Weather

Rīga

 **19°C** Current: Clear
Wind: S at 23 km/h
Humidity: 49%

Tue	Wed	Thu	Fri
			
22° 6°	14° 6°	17° 5°	20° 9°

PERSONALIZĀCIJAS PIEMĒRS NO DATUBĀZU JOMAS

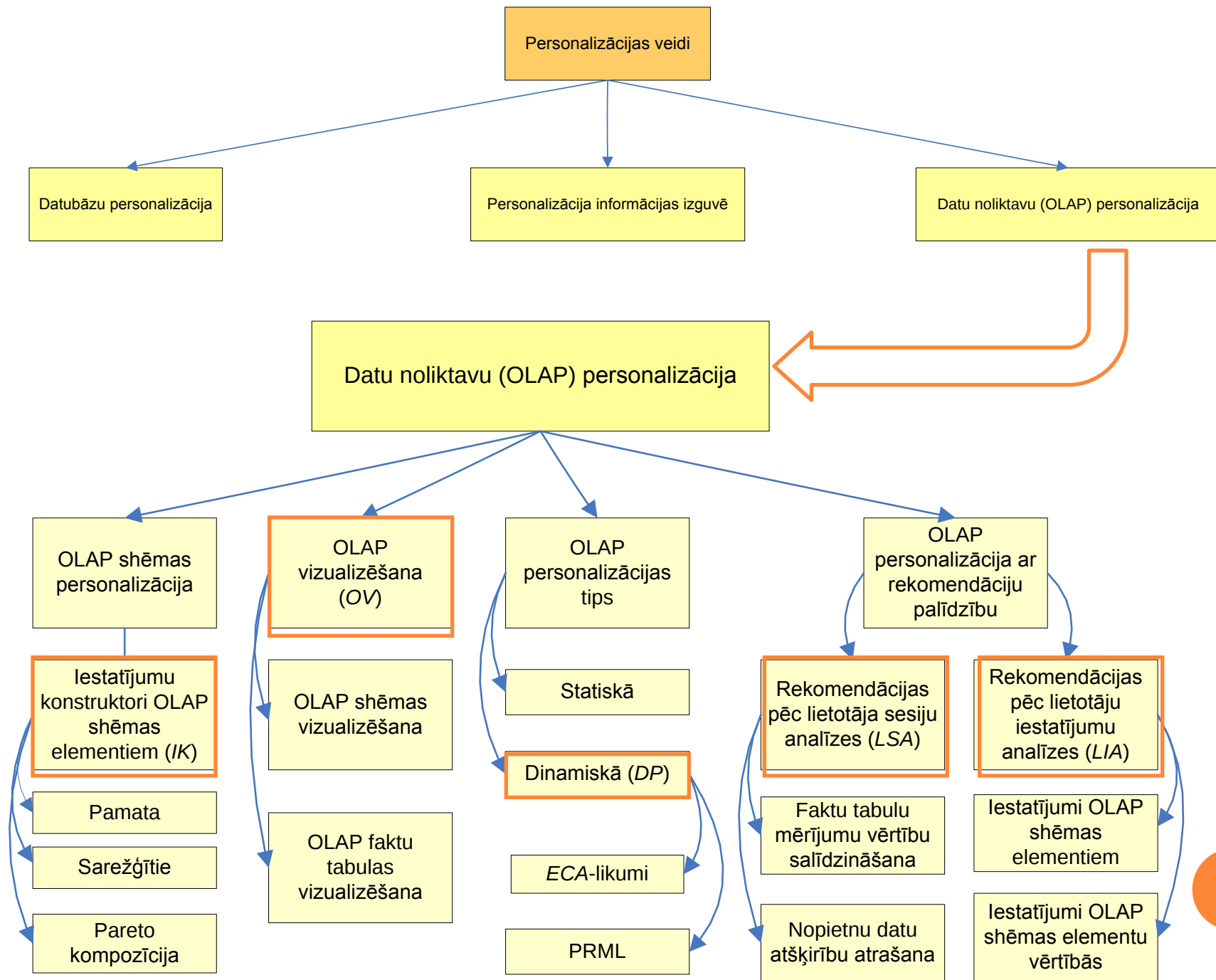
- Ar personalizāciju
 - Lietotāja iestatījumi

I1	Annai visvairāk gribētos aizbraukt ceļojumā uz Eiropas valsti (it īpaši uz Franciju), nekā uz Āfriku
I2	Vislabākais ceļojuma ilgums, pēc Annas domām, ir 10 dienas, bet viņai šķiet pieņemams arī ceļojums uz 9 vai 8 dienām.
I3	Annai patīk ceļot kopā ar lielu (30 - 40 cilv.) grupu, un jo vairāk cilvēku būs grupā, jo labāk.

- Lietotāja profils

<i>Iestatījums</i>	<i>Intereses pakāpe</i>
Kontinents = „Eiropa”	1
Kontinents = „Āfrika”	0,5
Valsts = „Francija”	1
Valsts = „Vācija”	0,6
Valsts = „Lielbritānija”	0,6
Dienas = 10	1
Dienas = 9	0,6
Dienas = 8	0,3
Grupa = 30	0,5
Grupa = 40	1



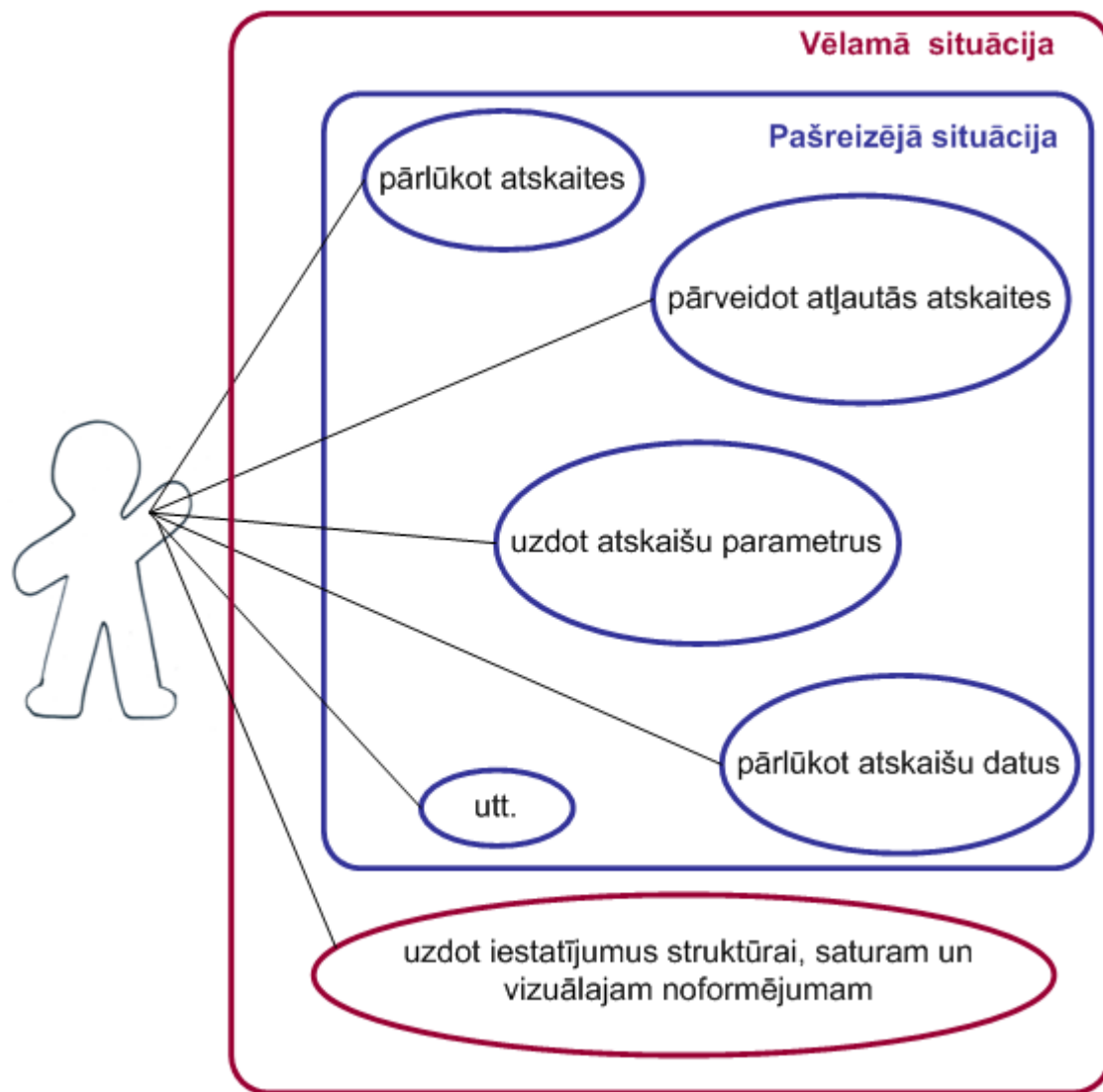


PĒTĪJUMA MĒRĶIS

- Izstrādāt un implementēt jaunu metodi datu noliktavas personalizācijai, kuras pamatā ir datu noliktavas personalizācijas metamodeli, lai sniegt lietotājam iespēju iegūt no datu noliktavas informāciju, kas maksimāli atbilst lietotāja vajadzībām.
- Rezultāti nopublicēti:
 - Pārskats par OLAP personalizācijas virzieniem: N. Kozmina & L. Niedrite: “Research Directions in OLAP Personalization” (ISD 2010, Prāga, Čehija).
 - Metode, kas piedāvā vispusīgu aprakstu lietotāja un datu noliktavas mijiedarbībai: N. Kozmina & L. Niedrite: “OLAP Personalization with User-describing Profiles” (BIR 2010, Rostoka, Vācija).



ATSKAIŠU RĪKA PERSONALIZĀCIJAS PAMATOJUMS



*Kā piedāvāt
lietotājam
datus,
kurus viņš/viņa
meklē?*

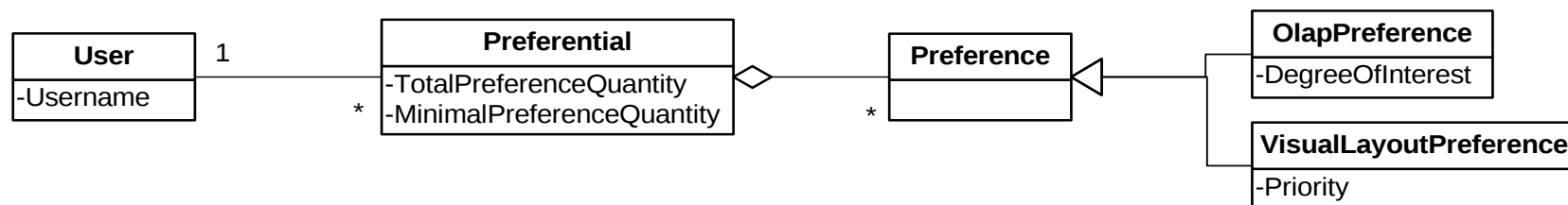


*leviest
personalizāciju
atskaišu
pārvaldības
rīkā!*



LIETOTĀJA IESTATĪJUMI

- Divu veidu lietotāju iestatījumi, kas var tikt ieviesti atskaišu rīkā:
 - Iestatījumi atskaišu saturam
 - Iestatījumi atskaišu noformējumam

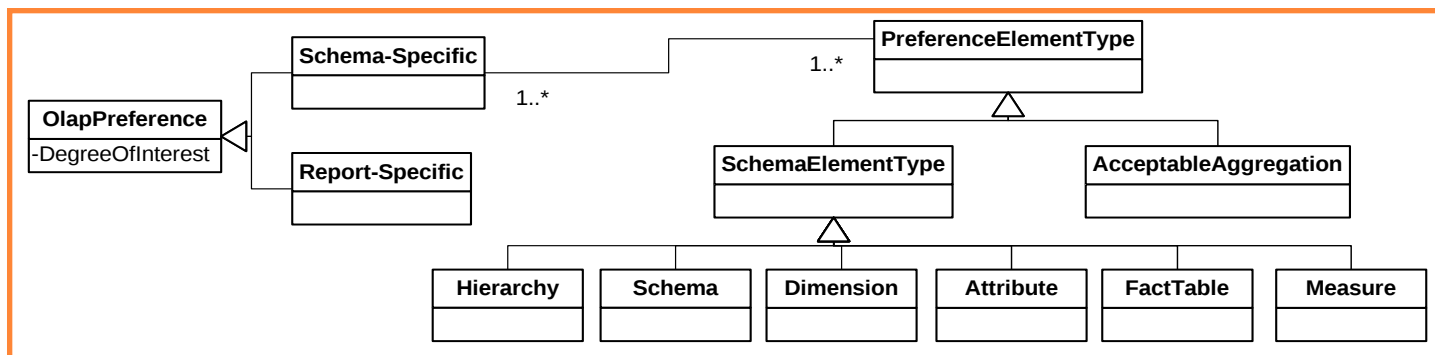
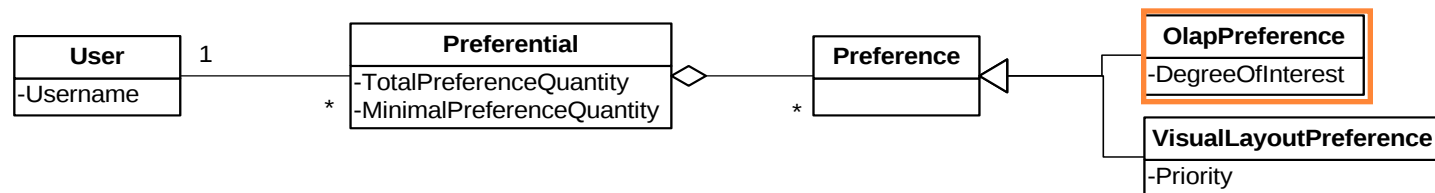


- Lietotāju iestatījumi pēc informācijas uzkrāšanas principa:
 - Tiešie (intereses pakāpi definē pats lietotājs)
 - Netiešie (intereses pakāpe tiek izrēķināta automātiski bez lietotāja dalības)



LIETOTĀJA NETIEŠIE OLAP SHĒMAS-SPECIFISKIE IESTATĪJUMI

- *Hot-start* un *Cold-start*: 2 metodes, kā piedāvāt lietotājam potenciāli interesantas atskaites, balstoties uz netiešajiem iestatījumi OLAP shēmas elementiem un pieļaujamām agregātfunkcijām (pāreju no cold-start uz hot-start metodi nosaka sliekšnis)
- Rezultāti tika iesniegti BIR2011 konferencei: N. Kozmina & D. Solodovnikova: “On Implicitly Discovered OLAP Schema-Specific Preferences in Reporting Tool”



COLD-START METODE

- 1. *solis* – Atskaišu struktūras aprakstīšana ar vektoriem, kuru koordinātes raksturo atskaišu OLAP shēmas elementu instanču esamību katrā atskaitē
- 2. *solis* – Vektoru līdzības izrēķināšana atskaišu pāriem (Vector/Cosine Similarity)
- 3. *solis* – Top- N atskaišu atlasīšana ar visaugstākajām vektoru līdzības vērtībām, sakārtotām dilstošā secībā

	Attributes				Dimensions			Fact Tables		Measures			Acceptable Aggregation		Hierarchies			OLAP Schemas	
$\vec{r}_1$	1	1	1	0	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1
$\vec{r}_2$	1	0	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1
	year	semester	faculty	program	time	program		registrations		student	PhD student		AVG	COUNT	time	faculty		students	

$$sim = \frac{\vec{r}_1 \cdot \vec{r}_2}{|\vec{r}_1| * |\vec{r}_2|}$$



HOT-START METODE

- 1. *solis* – Īpašā veidā svaru piekārtošana visām datu noliktavas shēmām un to elementiem.
- 2. *solis* –
 - (a) Ar algoritmu, kas balstās uz svariem, tiek aprēķināta un piefiksēta lietotāja profilā intereses pakāpe shēmas elementiem, kas ir sastopami lietotāja aktivitātes vēsturē;
 - (b) Katrai atskaitei tiek noteikti tajā izmantotie shēmas elementi;
 - (c) Katrai atskaitei tiek izrēķināta līdzība starp lietotāja intereses pakāpēm shēmas elementiem atskaitē un lietotāja profilā.
- 3. *solis* – Top- N atskaišu atlasīšana ar visaugstākajām lietotāja profila un atskaišu intereses pakāpes līdzības vērtības, sakārtotām dilstošā secībā.

$$sim = \frac{\sum_{i=1}^n DOI(E_i)}{\sum_{j=1}^m DOI(G_j)}$$

$E_1, \dots, E_n$ shēmas
elementi atskaitēs,
 $G_1, \dots, G_m$ visi
shēmas elementi
lietotāja profilā



TĀLĀKIE DARBI

- Novērtēt piedāvātās pieejas lietojamību dažādām lietotāju grupām
- Aplūkot iespējas kā paplašināt piedāvāto pieeju, ar metodēm
 - Tiešo OLAP un vizuālo iestatījumu definēšanai un apstrādei,
 - Netiešo iestatījumu uz atskaites datiem apstrādei,
 - Lietotāja demogrāfiskās informācijas vākšana un izmantošanai,
 - Sadarbības filtrēšanas principu izmantošanai.



The left side of the slide features a series of vertical stripes in shades of brown, tan, and grey. Overlaid on these stripes are several orange circles of varying sizes, arranged in a cluster that tapers towards the bottom.

PALDIES PAR UZMANĪBU!