

DATU NOLIKTAVU PERSONALIZĀCIJA

Natālija Kozmina

Seminārs 17.06.2010.



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

PERSONALIZĀCIJA

- *Personalizācija* ir process, kura gaitā lietotājam tiek piedāvāta noteiktā informācija, kas atbilst lietotāja vajadzībām*.
- Personalizācija var tikt nodrošināta, pielāgojot datus un to vizualizēšanu, saskaņā ar lietotāju iestatījumiem.

* - BNET Business Dictionary.

<http://dictionary.bnet.com/definition/personalization.html>



PERSONALIZĀCIJAS PIEMĒRS NO GOOGLE

o iGoogle

The screenshot displays the iGoogle homepage with a personalized dashboard. At the top, the iGoogle logo is on the left, followed by a search bar and buttons for "Google Search" and "I'm Feeling Lucky". Below the search bar, there are links for "Advanced Search" and "Language Tools". A notification bar indicates "Show this page every time I start to browse the web. Make iGoogle my homepage" and a link to "Change theme from Beaches of Eu... | Add stuff »".

The dashboard is organized into several sections:

- Home:** A sidebar on the left with "Updates Friends", "Chat" (with a search bar and "Natalie Kozmina Set status here"), and "Make iGoogle my homepage".
- www.lu.lv: jaunākās ziņas:** A news widget with three items:
 - Sadarbībā ar kompāniju „JTI” uzsākts projekts LU senioru integrācijai
 - “Zinātnes kafejnīca” ar saviem pētījumiem iepazīstinās jaunie zinātnieki
 - Ar uzvaru un zaudējumu ienemam 2 vietu
- Weather:** A weather widget for Riga showing 9°C, current conditions (Clear, Wind: N at 3 km/h, Humidity: 87%), and a 4-day forecast:

Day	Icon	Temp
Wed		15° 7°
Thu		18° 9°
Fri		23° 13°
Sat		19° 9°
- Motivational Quotes of the Day:** A quote by J. K. Rowling: "Indifference and neglect often do much more damage than outright dislike. J. K. Rowling, Harry Potter and the Order of the Phoenix, 2003". Below it is a link "Are you european or what" and "Ads by Google".
- apollo.lv - Visi raksti, Laika ziņas, horoskopi, labākie valūtas kursi:** A news widget with three items:
 - Hicfelds: šī ir vēsturiska uzvara
 - Tiesa atzīst Kačinska vainu konkurenta apmelošanā
 - Samoilovs/Sorokins kvalificējas Prāgas posma pamatturnīram
- ColorJunction:** A widget for color inspiration.
- Sticky Note:** A widget for notes.

PERSONALIZĀCIJAS PIEMĒRS NO GOOGLE (2)

- Google personalizētā meklēšana

Īpašība	Anonīmais lietotājs	Google lietotājs
Kas tiek pierakstīts?	Tas, uz ko lietotājs klikšķina meklēšanas rezultātu sarakstā	Tas, uz ko lietotājs klikšķina meklēšanas rezultātu sarakstā un mājaslapas, kuras lietotājs apmeklē (ja Google Toolbar izsekošanas iespēja ir ieslēgta)
Cik ilgi dati glabājas?	180 dienas	Vienmēr, vai kamēr lietotājs nenolems to izdzēst
Vai var pārlūkot meklēšanas vēsturi?	Nē	Jā
Vai var atslēgt šo iespēju?	Jā	Jā



PERSONALIZĀCIJAS PIEMĒRS NO DATUBĀZU JOMAS

- Bez personalizācijas

```
select ID, apraksts from celojumi
```

ID	Valsts	Kontinents	Grupa	Dienas	Cena	Apraksts
1	Tunisija	Āfrika	20	8	400	...
2	Vācija	Eiropa	30	10	350	...
3	Francija	Eiropa	15	5	450	...
4	Lielbritānija	Eiropa	40	6	500	...
5	Kuba	Z-Amerika	12	12	1500	...
...



PERSONALIZĀCIJAS PIEMĒRS NO DATUBĀZU JOMAS (2)

- 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.



PERSONALIZĀCIJAS PIEMĒRS NO DATUBĀZU JOMAS (3)

- 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



PERSONALIZĀCIJAS PIEMĒRS NO DATUBĀZU JOMAS (4)

○ Ar personalizāciju

```
Q1 = select ID, apraksts  
from celojumi  
where kontinents in („Eiropa”,  
„Āfrika”)
```

```
Q2 = select ID, apraksts  
from celojumi  
where dienas between 8 and 10
```

```
Q3 = select ID, apraksts  
from celojumi  
where grupa = 30 or grupa = 40
```

```
select ID, apraksts  
from Q1 union all Q2 union all Q3  
having count(*) >= 2
```

Lai izveidotu personalizētu vaicājumu, vajag:

- Piefiksēt lietotāja iestatījumus,
- Piekārtot katram iestatījumam intereses pakāpi,
- Atlasīt vismaz L iestatījumus,
- Sakārtot iestatījumus pēc intereses pakāpes.



MOTIVĀCIJA DATU NOLIKTAVU PERSONALIZĀCIJAS PĒTĪŠANAI

Pastāv daži iemesli, kāpēc ir jāveic pētījumus OLAP*
personalizācijas jomā:

- Pirmkārt, iestatījumi dod lietotājiem iespēju koncentrēties uz tieši viņu interesējošiem datiem.
- Otrkārt, OLAP sesijas laikā, lietotājs var pats nezināt, ko tieši viņš grib atrast.
- Treškārt, ir svarīgi piedāvāt lietotājam iespēju veidot iestatījumus agregācijas datiem, jo bieži vien apkopota informācija ir daudz vērtīgākā un pieprasītākā, nekā detalizētā.

* - OnLine Analytical Processing (tiešsaistes analītiskā apstrāde)



DARBA MĒRĶIS

- Izstrādāt konceptuālo modeli datu noliktavas personalizācijas pieejai, lai nākotnē ieviestu to aprobācijas vidē.



DARBA UZDEVUMI

- Iepazīties ar personalizācijas aspektiem datu noliktavu jomā, izpētot zinātnisko literatūru,
- Piedāvāt datu noliktavas eksistējošo personalizācijas pieeju klasifikāciju un analīzi,
- Izstrādāt konceptuālo modeli datu noliktavas personalizācijas pieejai,
- Piedāvāt lietotāju rekomendāciju ģenerēšanas algoritmu, kas pamatojas uz lietotāju iestatījumiem.



IZMANTOTAS METODEDES

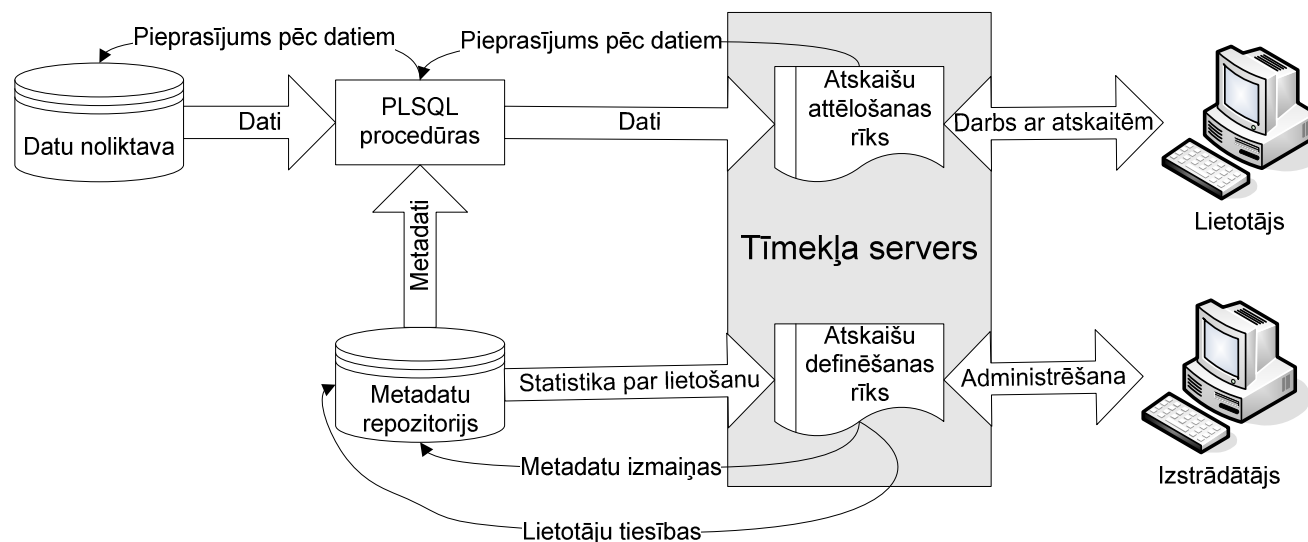
- Grāmatu, standartu, zinātnisko un tehnisko rakstu izpēte,
- Rīku personalizācijas iespēju analīze, izmantojot rīku Oracle Warehouse Builder,
- Lietotāju profilu projektēšana, izmantojot UML klašu un aktivitāšu diagrammas.



APROBĀCIJAS VIDE PERSONALIZĀCIJAS IEVIEŠANAI

LU tika izstrādāts bezlicenču rīks darbam ar datu noliktavas atskaitēm. Izstrādāto rīku sauc par *atskaišu pārvaldības rīku*, un tas sastāv no diviem komponentiem:

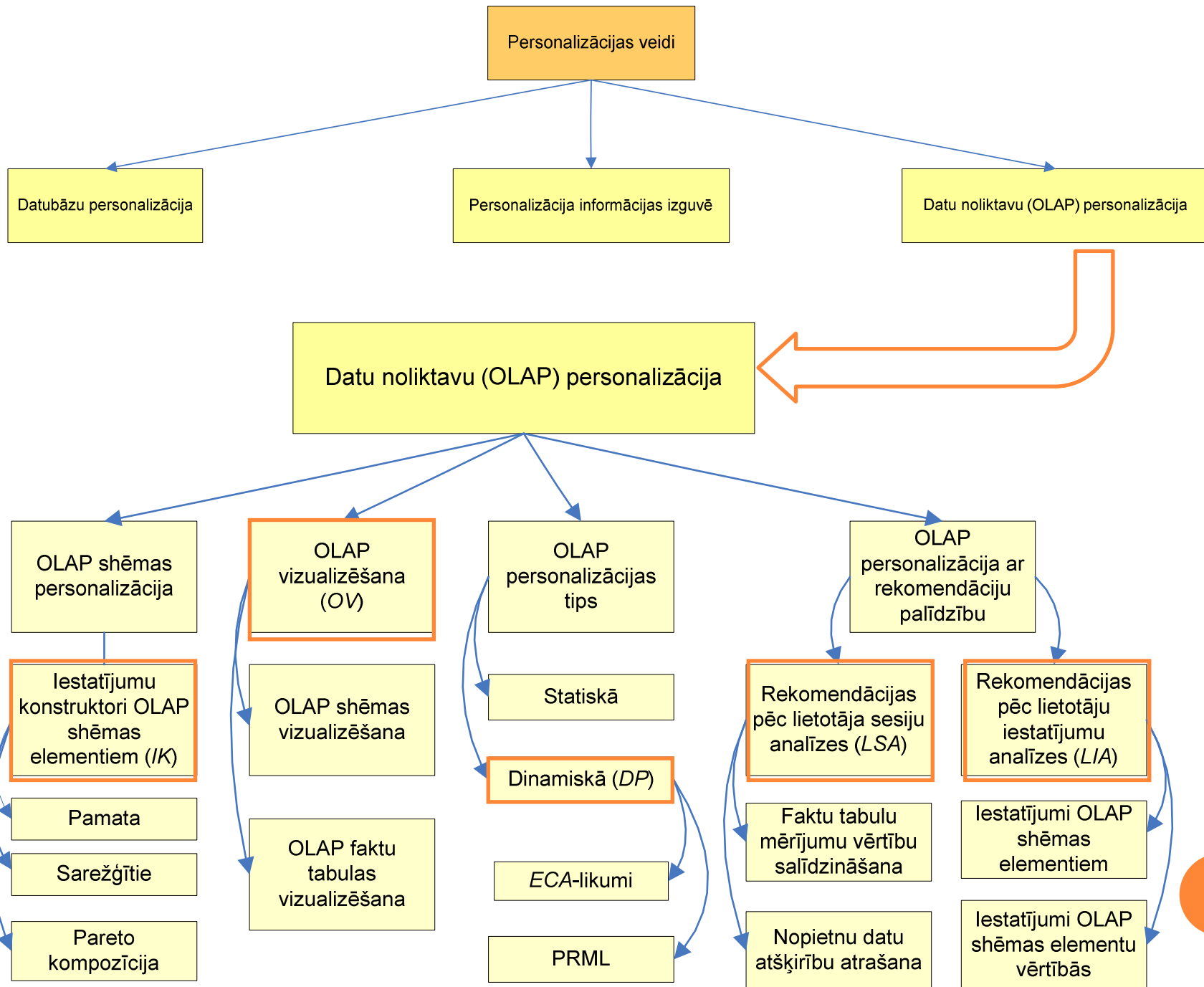
- *atskaišu attēlošanas rīks*
- *atskaišu definēšanas rīks*



ATSKAIŠU RĪKA PERSONALIZĀCIJAS PAMATOJUMS

- Datu noliktavas atskaišu rīka lietotājus var loģiski sadalīt uz vairākām grupām (piemērām, studenti, pasniedzēji, LU darbinieki, utt.).
- Katrai lietotāju grupai vai atsevišķam lietotājam:
 - Atšķiras vajadzības pēc atskaitēm,
 - Interesē atšķirīgais izvēlēto atskaišu saturs (dati),
 - Liekas piemērots dažāds atskaišu noformējums.





IESTATĪJUMU KONSTRUKTORI OLAP SHĒMAS ELEMENTIEM (IK)

- Pamata iestatījums = konstruktors un atribūts, mērījums, hierarhija vai agregācija
- Sarežģītais iestatījums = pamata iestatījumu kombinācijas
- Abus iestatījumu veidus var aprakstīt, lietojot formālo gramatiku:

$\langle \text{expr} \rangle ::= \langle \text{baseConstr} \rangle \mid \langle \text{expr} \rangle \otimes \langle \text{baseConstr} \rangle$
 $\langle \text{baseConstr} \rangle ::= \text{POS} \mid \text{NEG} \mid \text{BETWEEN} \mid \text{LOWEST} \mid$
 $\text{HIGHEST} \mid \text{CONTAIN} \mid \text{NEAR} \mid \text{COARSEST} \mid \text{FINEST}$

- Piemērs #1: pamata iestatījumi atribūtiem
 - POS(Month, 'Jan-10')
 - NEG(Month, 'Jan-10')
- Piemērs #2: pamata iestatījumi mērījumiem
 - BETWEEN(AvgIncome, 0, 1000)
 - LOWEST(AvgIncome), HIGHEST(AvgIncome)
- ...



OLAP VIZUALIZĒŠANA (OV)

- Lietotāja saskarne sarežģītu daudzdimensiju datu pārlūkošanai OLAP vidē
- Lietotāji var
 - virzīties pa dimensijām, mērījumiem un hierarhijām,
 - izvēlēties mērījumus un agregātfunkcijas,
 - uzlikt filtrus uz dimensiju atribūtu vērtībām.
- Dekompozīcijas koks

Orders

DIMENSIONS

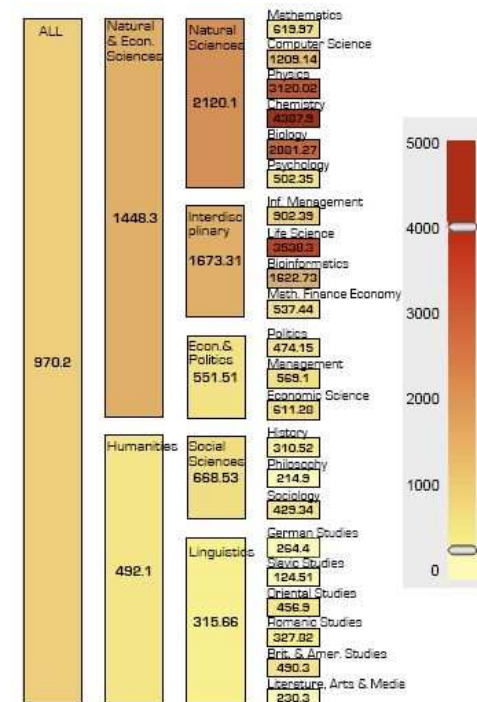
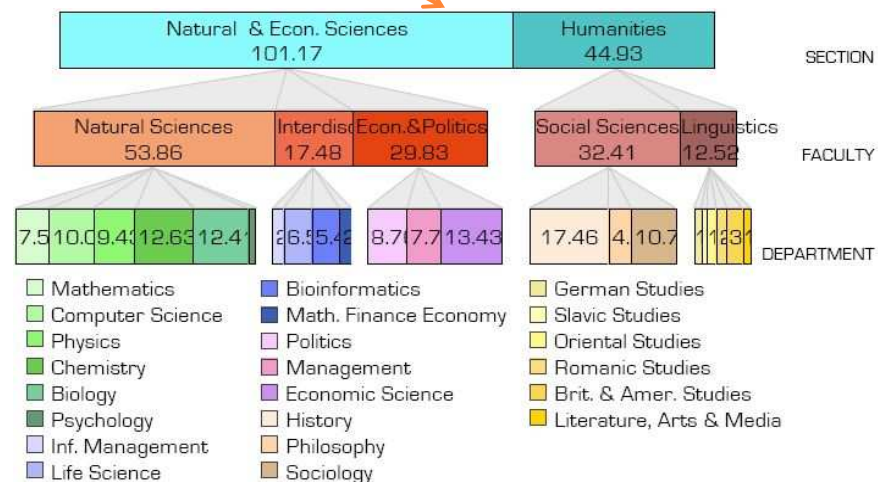
- Cost Class
- Funding
- Project
- Period
 - Calendar Year
 - 2001
 - 2002 (checked)
 - 2003
 - Semi-annual
 - show (checked)
 - hide (unchecked)
 - Quarter
 - Week
 - Academic Year
- Purchaser

MEASURES

- amount (checked)

FUNCTIONS

- Total (SUM) (checked)
- Average (AVG) (unchecked)
- Occurrences (COUNT) (unchecked)
- Maximum (MAX) (unchecked)
- Minimum (MIN) (unchecked)



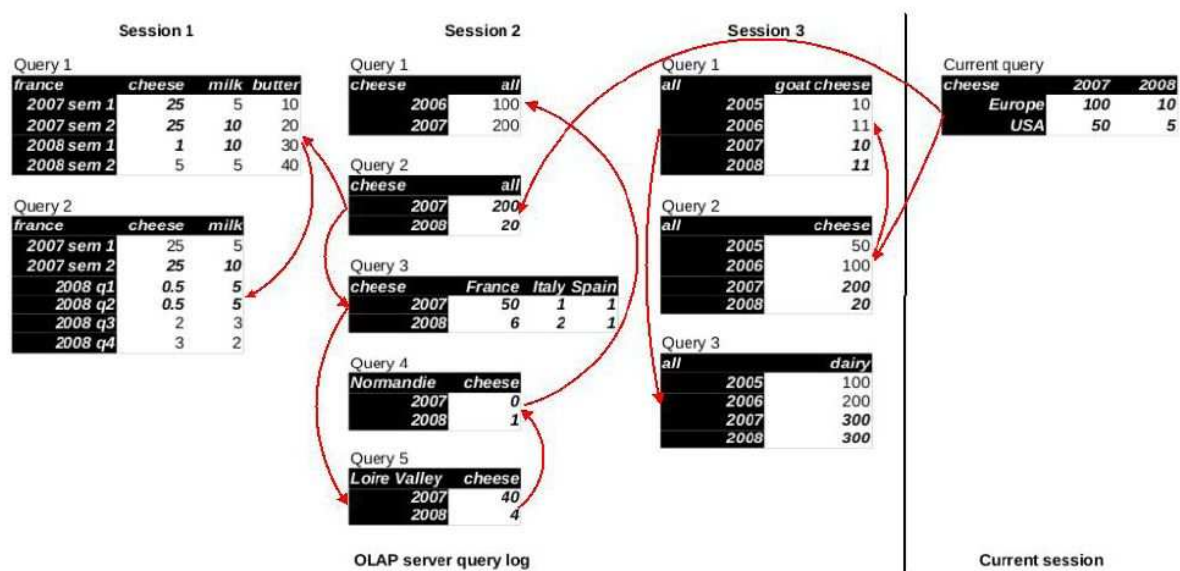
DINĀMISKĀ PERSONALIZĀCIJA (*DP*)

- OLAP personalizācija ir *dināmiskā* tad, ja OLAP shēma tiek uzģenerēta izpildes laikā un lietotāja uzvedība un konteksts tiek ņemti vērā.
- ECA-likuma pamata struktūra:
When *event* do
If *condition* then *action* endIf
endWhen
- Divi darbības veidi (*action*):
 - *hide*
 - *set*
- Piemērs: *Set*-darbība var kalpot, lai aprēķinātu lietotāja intereses pakāpi noteiktajā dimensiju atribūtā. *Hide*-darbība tiek izpildīta tad, kad lietotāja intereses pakāpe noteiktajā atribūtā ir zemākā par iepriekšdefinēto “sliksni”.



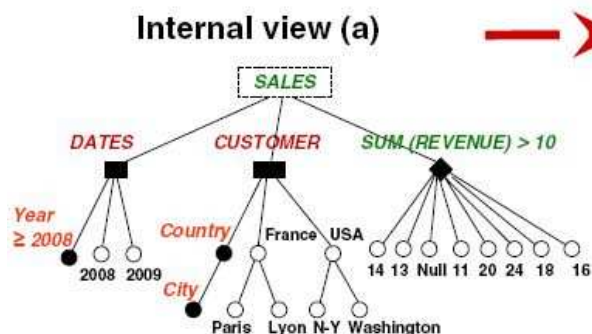
REKOMENDĀCIJAS PĒC LIETOTĀJA SESIJU ANALĪZES (LSA)

- Lietotāju iepriekšējo sesiju struktūras
 - OLAP servera žurnādatne
- Faktu tabulu mērījumi tiek salīdzināti un nopietnas datu atšķirības tiek atrastas.
- Tiek analizētas visas lietotāja sesijas, kurās figurē tādas pašas datu atšķirības, kā pašreizējā sesijā.



REKOMENDĀCIJAS PĒC LIETOTĀJA IESTATĪJUMU ANALĪZES (LIA)

- Konteksta-bāzēta metodi lietotāju rekomendāciju sniegšanai.
- Analīzes kontekstā ietilpst 2 elementu kopas: *OLAP shēmas elementi* – faktu tabulas, mērījumi, dimensijas, atribūti, utt. – un to *vērtības*.
- Ierobežojumi uz mērījumiem un atribūtiem ir piefiksēti *lietotāja profilā* un skaitliski novērtēti (reālais skaitlis intervālā [0; 1]).
- Ir iespējams izteikt lietotāja ieinteresētību agregātdatos, kā arī noteiktās hierarhijās un hierarhiju līmeņos.



Dimensional Table

SALES SUM (REVENUE)		CUSTOMER HGEO				
		CITY	London	Milan	N-Y	Paris
DATES HMONTH	YEAR					
	2006		(205)	(108)	(380)	(180)
	2007		(185)	(40)	(410)	(280)
	2008		(240)	(77)	(82)	(310)
	2009		(168)	(135)	(110)	(415)

Recommendation: [Sales/Sum\(Revenue\); Dates/Month, Customer/City; Sum\(Revenue\)>100'](#)

Explanations:

- Product/Description='Toshiba U300' --> DATES.level = Month
- Product/Description='Toshiba U300' --> Sum(Revenue)>100

DATU NOLIKTAVU PERSONALIZĀCIJAS PIEEJU SALĪDZINĀJUMA ANALĪZE

- Dažādu personalizācijas pieeju pielietošana OLAP objektiem

<i>Pers. tips/ Pers. Objekts</i>	<i>Dimensija</i>	<i>Dimensijas atribūts</i>	<i>Hierarhija</i>	<i>Līmenis</i>	<i>Faktu tabula</i>	<i>Mērījums</i>	<i>Agreg. funkcija</i>	<i>Select operācija</i>	<i>Drilldown operācija</i>	<i>Rollup operācija</i>	<i>Rotate operācija</i>
IK	-	A	-	A	-	A	-	D	D	A	-
DP	A	A	A	A	A	A	A	A [RT09], D [GPM+09]	A	A	A [RT09], - [GPM+09]
OV	A	A	A	A	A	A	A	A	A	A	A
LSA	A	A	A	A	A	A	D	A	A	A	-
LIA	A	A	A	A	A	A	A	A	D	D	-

Līdzīgi tiek novērtēti:

- OLAP personalizācijas tipi un ierobežojumi,
- Iestatījumu iegūšanas un lietotāju datu uzkrāšanas metodes.
- Raksts “Research Directions of OLAP Personalization” tika pieņemts konferencei ISD2010.



PAMATOJUMS IZVĒLĒTAI OLAP PERSONALIZĀCIJAS PĪEEJAI

- *LIA* pieeja tika veiksmīgi īstenota atskaišu personalizēšanai, balstoties uz viena konkrēta lietotāja atskaišu iestatījumiem.
- Lai dotu iespēju lietotājam izmantot arī citu lietotāju iestatījumus, tika piedāvāta paplašināta metode, kuras pamatā arī ir rekomendācijas (sadarbības metodes princips).
- Rekomendācijas ir lietotāju iestatījumu komplekti atskaitēm.



LIETOTĀJU PROFILU KLASIFIKĀCIJA PĒC ZAHMANA IETVARA KONCEPCIJAS

- *Lietotāja profils* ir informācijas kopums, kas apraksta lietotāju
- **Kāds** ir sistēmas lietotājs?
- Pamata dati par lietotāju (tiesības, personas dati, sistēmas ģenerētie dati, piemēram, reģistrācijas datums, utt.).
- *User – Lietotāja profils*
- **Kas** ir tas, ko sistēmas lietotājs sagaida kā rezultātu?
- Tiek aprakstīti dati par sistēmas lietotāja (vaicājumu) iestatījumiem.
- *Preferential* – Iestatījumu profils



LIETOTĀJU PROFILU KLASIFIKĀCIJA PĒC ZAHMANA IETVARA KONCEPCIJAS (2)

- **Kur** sistēmas lietotājs atrodas?
- Dati par lietotāja fizisko atrašanās vietu, kā arī ģeogrāfiskās īpašības, atvasinātas no IP-adreses.
- *Spatial – Atrašanās vietas profils*
- **Kad** lietotājs izmanto sistēmu?
- Sistēmas lietotāja aktivitāšu laika īpašības.
- *Temporal – Laika profils*

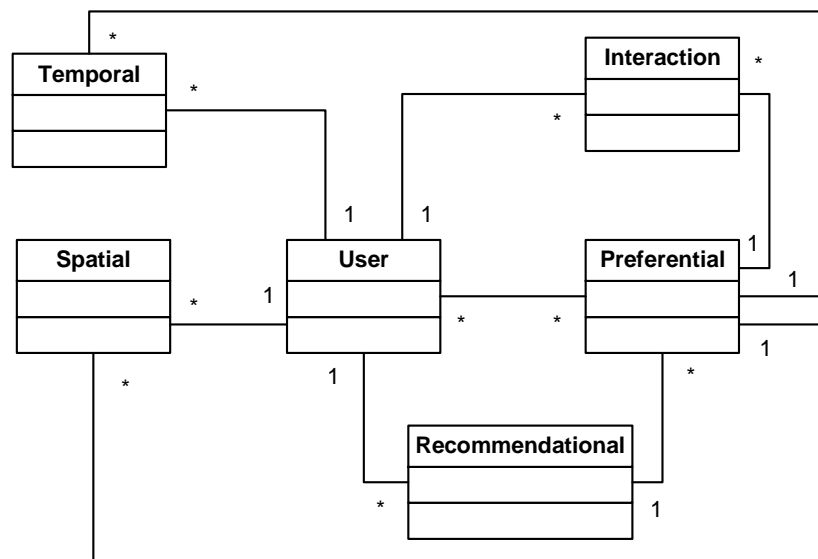


LIETOTĀJU PROFILU KLASIFIKĀCIJA PĒC ZAHMANA IETVARA KONCEPCIJAS (3)

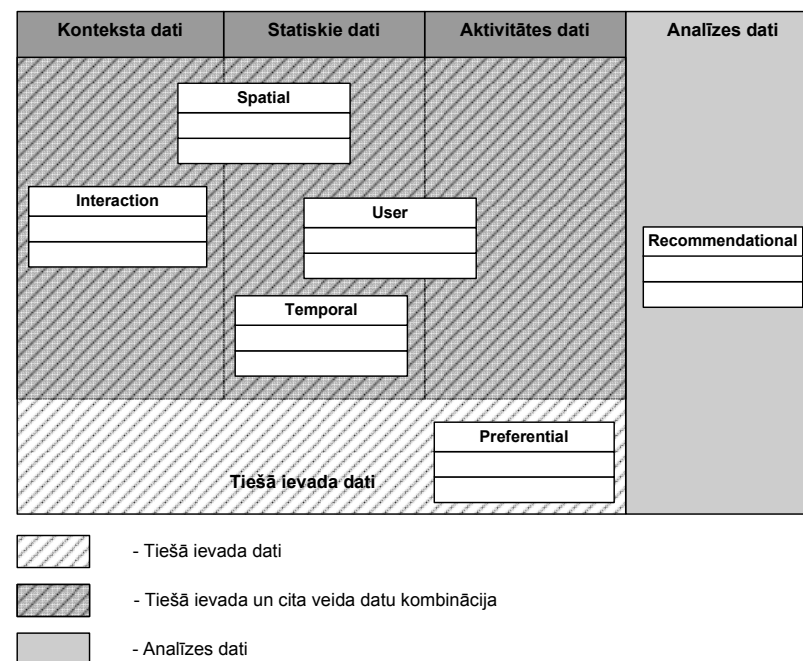
- **Kā** notiek lietotāja un sistēmas mijiedarbība?
- Lietotāja ierīces (t.i. dators, klēpjdaturs, mobilais tālrunis, utt.) īpašības, kuru viņš/viņa izmanto, lai pieslēgties sistēmai, kā arī programmatūras (piemēram, tīmekļa pārlūka) īpašības.
- *Interaction* – Mijiedarbības profils
- **Kāpēc** lietotāju interesē šī sistēma?
- Dati par sistēmas lietotāja aktivitātēm un iestatījumiem tiek savākti un analizēti. Piemēram, ja lietotājs pārlūko noteiktā tipa atskaites, rekomendācijas, kas satur saites uz līdzīgām atskaitēm, var tikt uzģenerētas.
- *Recommendational* – Rekomendāciju profils



LIETOTĀJU PROFILU SAKARĪBAS



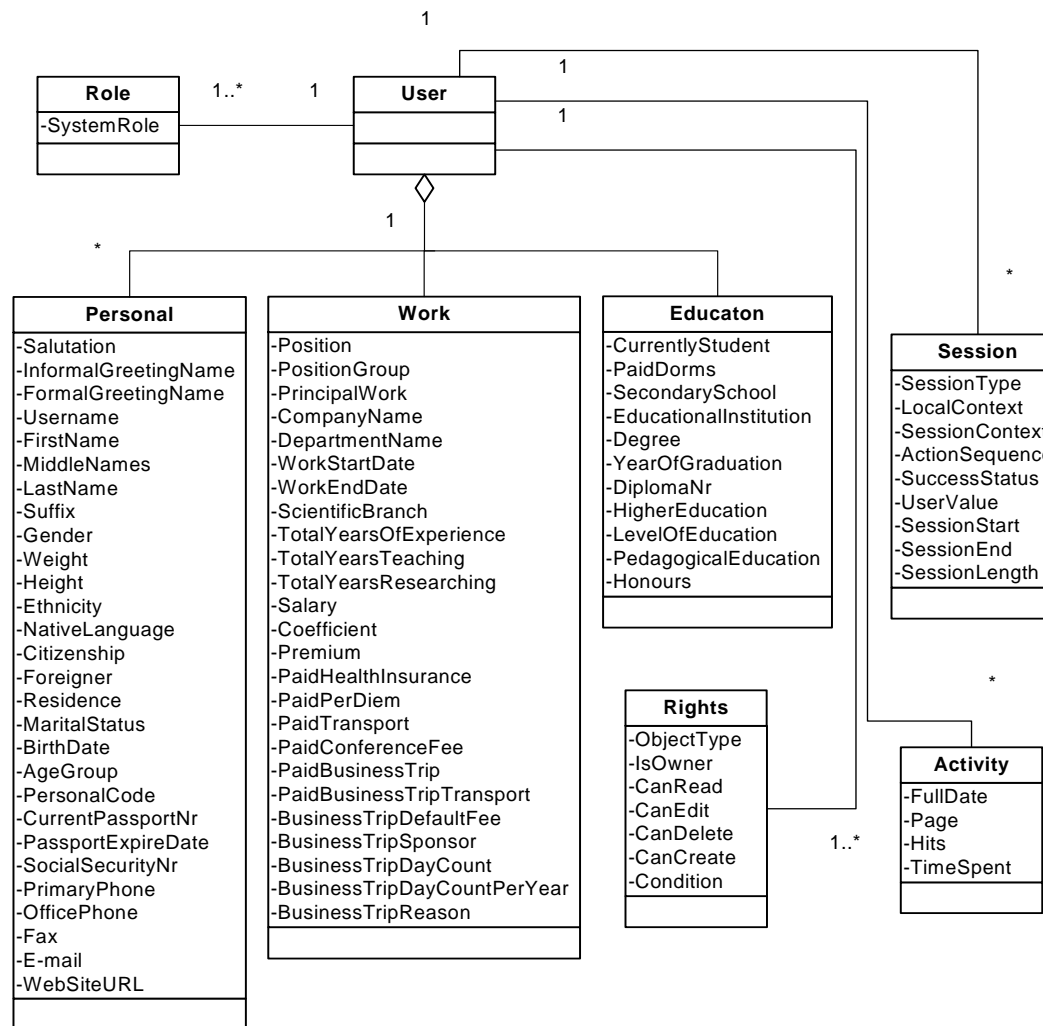
LIETOTĀJU PROFILU DATU AVOTI



- Raksts “OLAP Personalization with User-describing Profiles” tika iesniegts konferencei BIR2010. Pašlaik notiek rakstu izvērtēšana un recenzēšana.



LIETOTĀJU APRAKSTOŠĀ PROFILA* PIEMĒRS

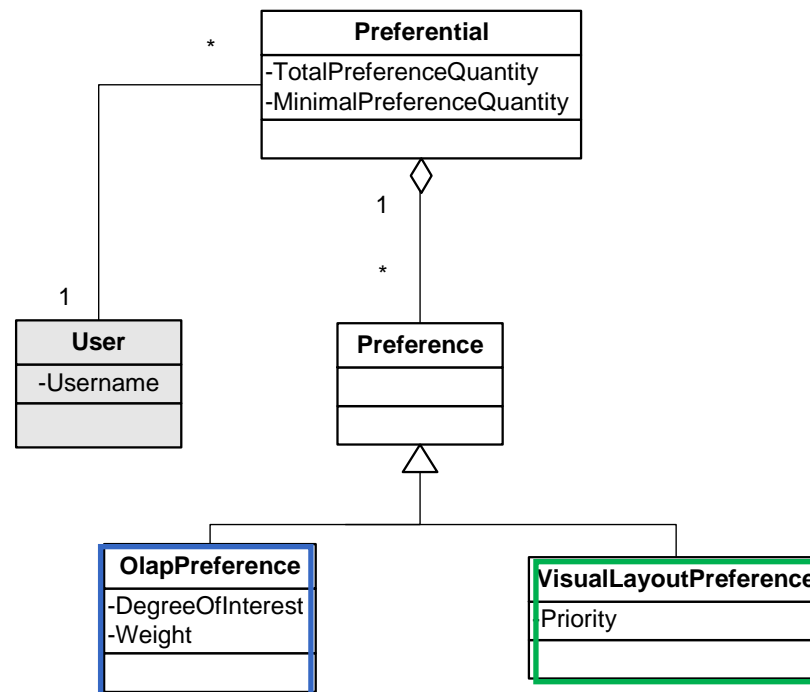


* - lietotāja, mijiedarbības, atrašanās vietas, laika profili



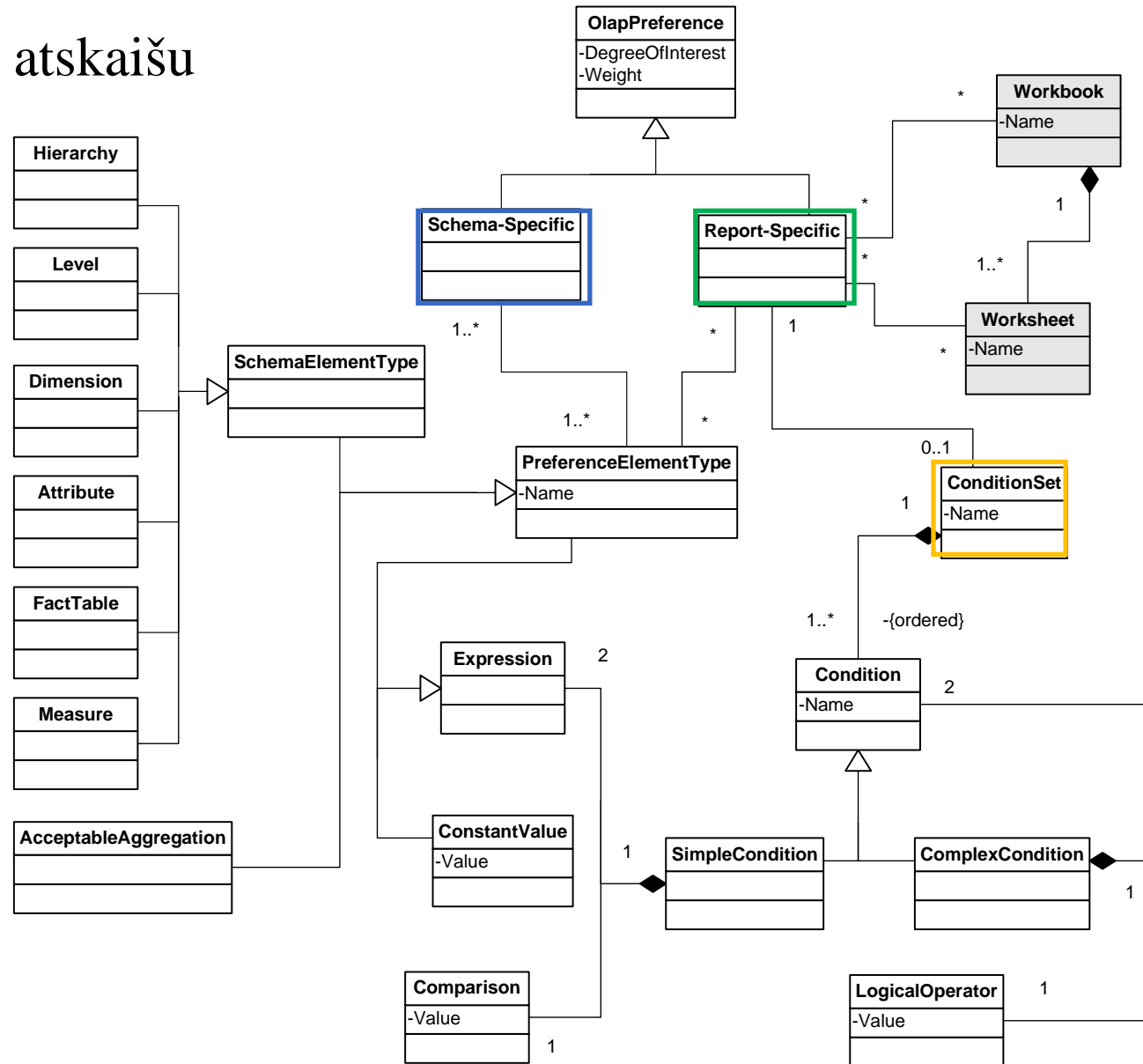
LIETOTĀJA IESTATĪJUMU PROFILA METAMODELIS

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

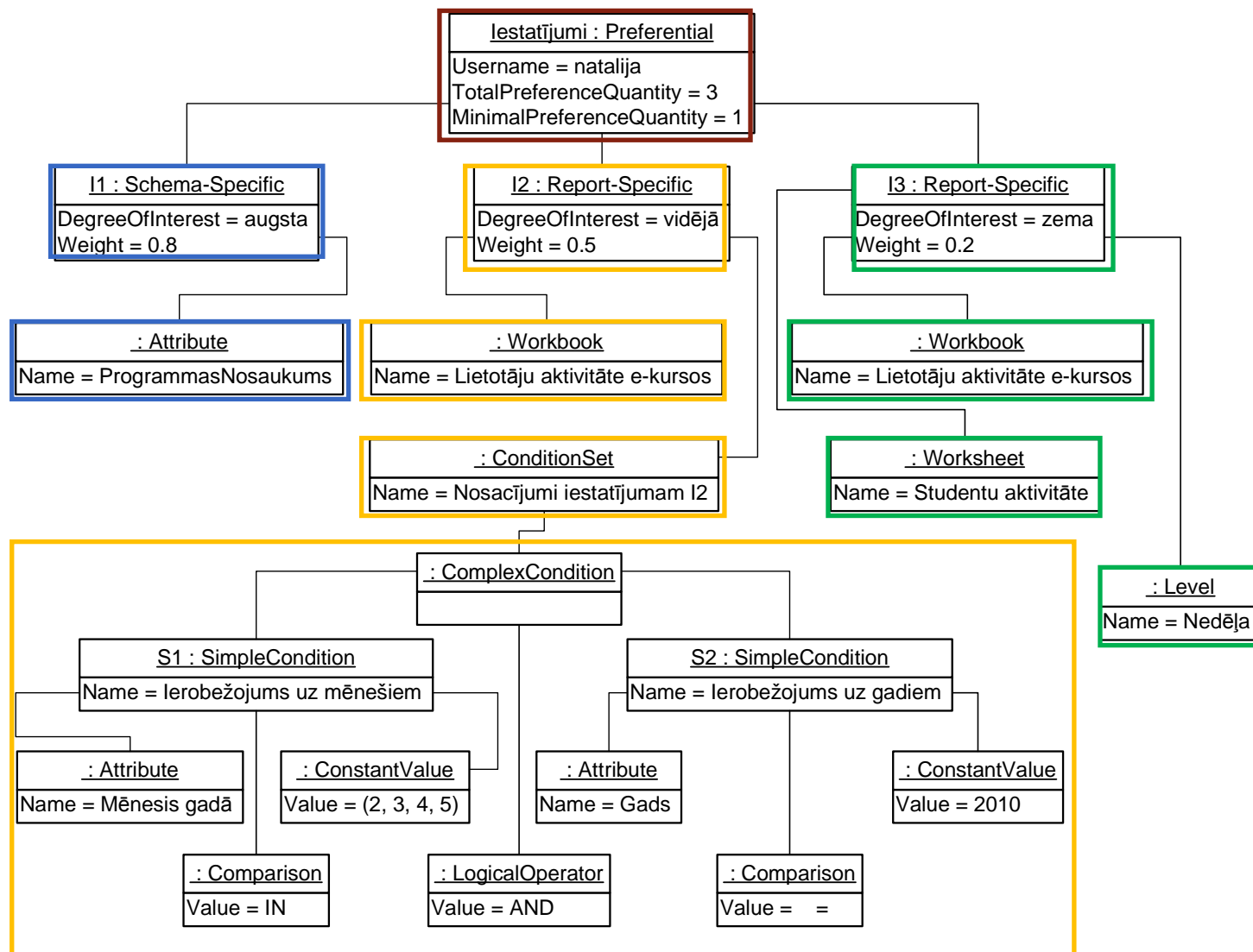


LIETOTĀJA IESTATĪJUMU PROFILA METAMODELIS (2)

- Iestatījumi atskaišu saturam

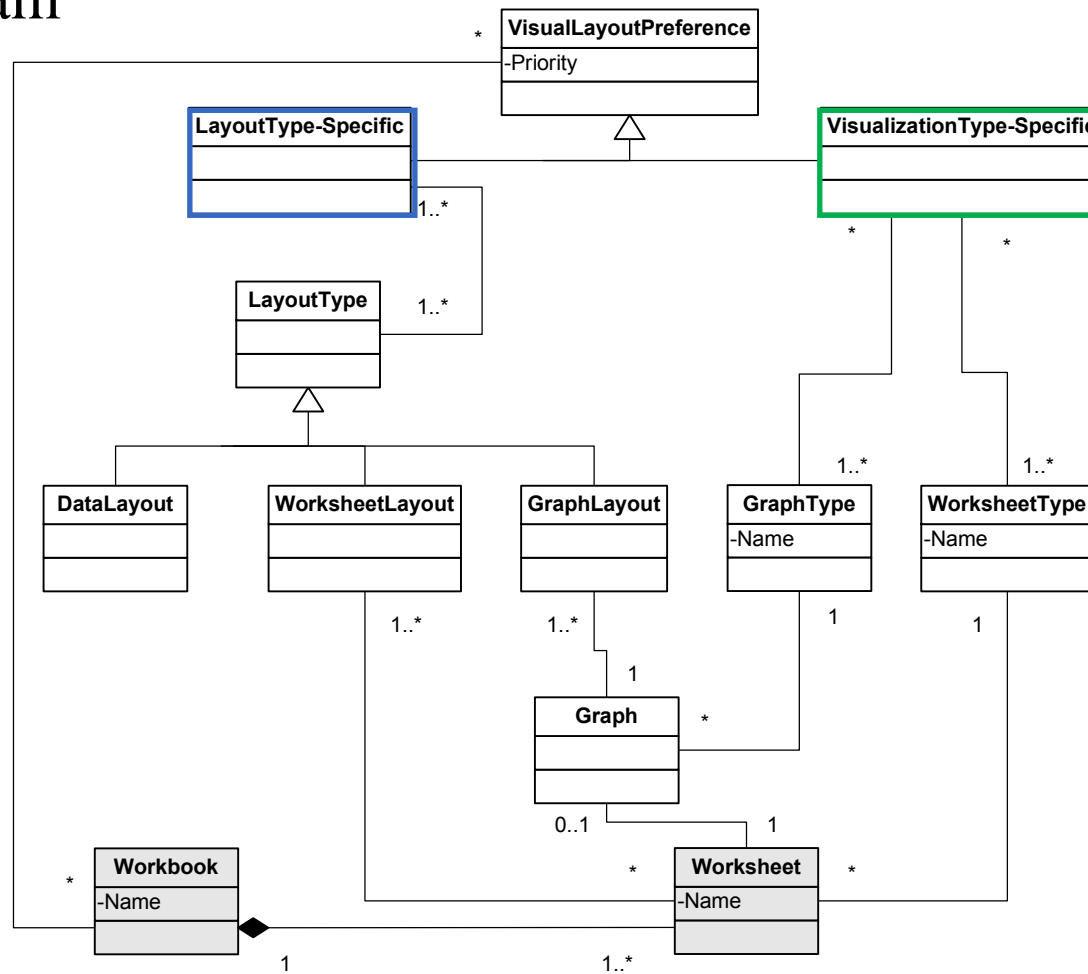


IESTATĪJUMI ATSKAIŠU SATURAM – PIEMĒRS

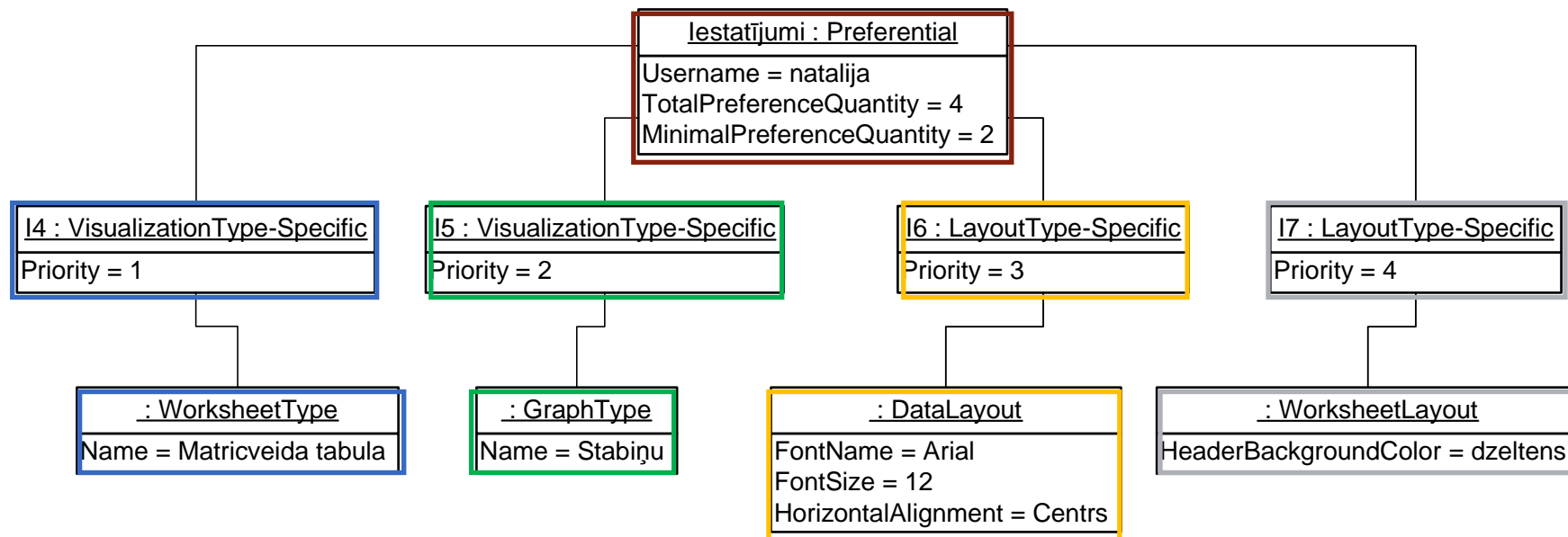


LIETOTĀJA IESTATĪJUMU PROFILA METAMODELIS (3)

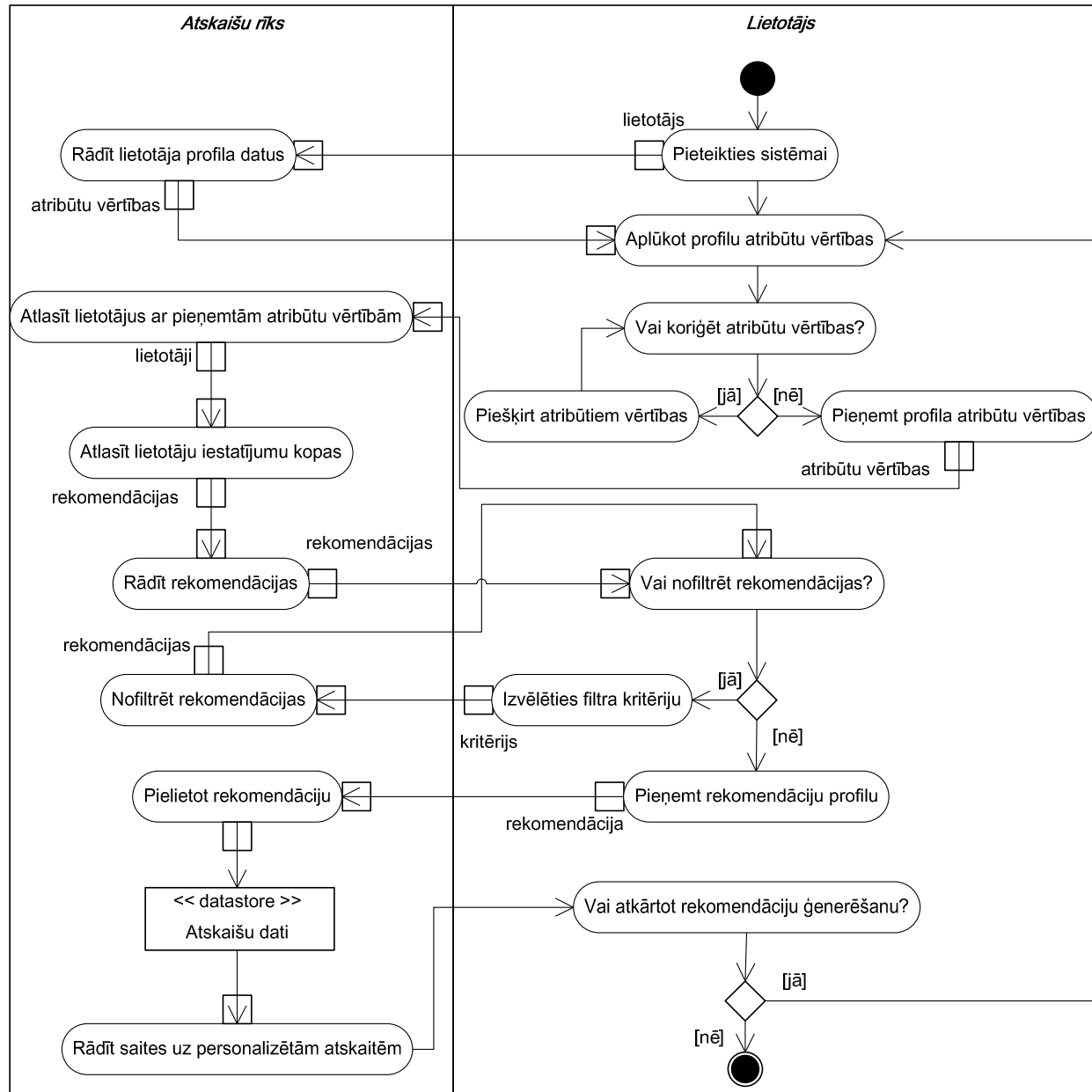
- Iestatījumi atskaišu noformējumam



IESTATĪJUMI ATSKAIŠU NOFORMĒJUMAM – PIEMĒRS



REKOMENDĀCIJU PROFILA VEIDOŠANA



REZULTĀTI

- Lai nonāktu pie risinājuma, bija jāveic pamatīgas literatūras studijas, kas attiecās uz personalizāciju datu noliktavu jomā.
- Tika izveidota eksistējošo datu noliktavas personalizācijas pieeju klasifikācija un analīze.
- Tika piedāvāta jauna metode, kas paplašina *LIA* pieejas ideju un izmanto Zahmana ietvara koncepciju.
- Darbā tika izveidoti vairāki profili, kas apraksta lietotāju, un aprakstīta rekomendāciju ģenerēšanas algoritma ideja.
- 2 raksti tika iesniegti starptautiskajās konferencēs (ISD2010 un BIR2010), 1 no tiem ir jau pieņemts.



TĀLĀKAIS DARBS

- Viens no tālākā darba iespējamiem virzieniem ir piedāvātās personalizācijas pieejas ieviešana LU atskaišu pārvaldības rīkā.
- Pirms tam būtu vērtīgi izpētīt eksistējošo rekomendāciju sistēmu rekomendāciju ģenerēšanas algoritmus un rekomendāciju atlasē kritērijus.



PALDIES PAR UZMANĪBU!

