



RĪGAS TEHNISKĀ
UNIVERSITĀTE

Henrihs Gorskis

ONTOLOĢIJĀ BĀZĒTAS INTELEKTUĀLAS DATU IZGŪŠANAS METODOLOĢIJAS IZSTRĀDE

Promocijas darba kopsavilkums



RĪGAS TEHNISKĀ UNIVERSITĀTE
Datorzinātnes un informācijas tehnoloģijas fakultāte
Informācijas tehnoloģijas institūts

Henrihs GORSKIS

Doktora studiju programmas „Informācijas tehnoloģija” students

**ONTOLOĢIJĀ BĀZĒTAS
INTELEKTUĀLAS DATU IZGŪŠANAS
METODOLOĢIJAS IZSTRĀDE**

Promocijas darba kopsavilkums

Zinātniskie vadītāji:
profesors *Dr. habil. sc. comp.*

Arkādijs Borisovs
profesore *Dr. sc. ing.*
Ludmila Aleksejeva

Zinātniskā konsultante
docente *Dr. sc. ing.*
Inese Poļaka

Rīga 2018

Gorskis, H. **Ontoloģijā bāzētas intelektuālas datu izgūšanas metodoloģijas izstrāde**. Promocijas darba kopsavilkums. – Rīga: RTU Izdevniecība, 2018. 40 lpp.

Iespiests saskaņā ar 2018. gada 24. janvāra ITI padomes sēdes lēmumu, protokols Nr. 12100-2/1.

Promocijas darbs izstrādāts ar RTU Datorzinātnes un informācijas tehnoloģijas fakultātes projekta “Zinātniskās darbības attīstība augstskolā” atbalstu.

ISBN 978-9934-22-097-5 (print)

ISBN 978-9934-22-098-2 (pdf)

PROMOCIJAS DARBS
IZVIRZĪTS INŽENIERZINĀTŅU DOKTORA GRĀDA
IEGŪŠANAI RĪGAS TEHNISKAJĀ UNIVERSITĀTĒ

Promocijas darbs inženierzinātņu doktora grāda iegūšanai tiek publiski aizstāvēts 2018. gada 1. oktobrī plkst. 14.30 Rīgas Tehniskās universitātes Datorzinātnes un informācijas tehnoloģijas fakultātē, Sētas ielā 1, 3. korpusā, 202. auditorijā.

OFICIĀLIE RECENZENTI

Profesors *Dr. habil. sc. ing.* Jānis Grundspenķis,
Rīgas Tehniskā universitāte, Latvija

Profesors *Dr. sc. ing.* Pēteris Grabusts,
Rēzeknes Tehnoloģiju akadēmija, Latvija

Profesors *Dr. sc.* Aleksandrs Božņuks,
Dienvidu Federālās universitātes Taganrogas Tehnoloģiju institūts, Krievija

APSTIPRINĀJUMS

Apstiprinu, ka esmu izstrādājis šo promocijas darbu, kas iesniegts izskatīšanai Rīgas Tehniskajā universitātē inženierzinātņu doktora grāda iegūšanai. Promocijas darbs zinātniskā grāda iegūšanai nav iesniegts nevienā citā universitātē.

Henrihs Gorskis
paraksts

Datums

Promocijas darbs uzrakstīts latviešu valodā, tajā ir ievads, četras nodaļas, rezultātu analīze un secinājumi, 13 tabulu, 35 attēli, kopā – 181 lappuse. Literatūras sarakstā ir 121 vienība.

SATURS

VISPĀRĒJS DARBA RAKSTUROJUMS	5
Problēmsfēra.....	5
Aktualitāte	5
Problēmas nostādne.....	5
Pētījuma mērķis un uzdevumi	6
Pētījuma objekts un priekšmets.....	7
Pētījuma hipotēzes.....	7
Pētījuma metodes	7
Zinātniskā novitāte	7
Praktiskais nozīmīgums.....	8
Aprobācija	9
Promocijas darba struktūra un saturs	11
DARBA NODAĻU SATURA APRAKSTS	12
1. Ar ontoloģijā bāzētu datu izgūšanu saistītas tehnoloģijas.....	12
Ontoloģijas pamata tehnoloģijas	12
Ontoloģijas informācijas tehnoloģijā	13
Ontoloģijā bāzētas datu izgūšanas jēdziens.....	14
2. Ontoloģijā bāzētas datu izgūšanas metodes	15
Ontoloģijā bāzētas datu izgūšanas sistēmas	15
Datubāzes un ontoloģijas sasaistīšanas pieejas	16
Ontoloģijas izveidošanas metodes.....	17
3. Ontoloģijā bāzētas datu izgūšanas metodoloģijas izstrāde.....	18
Metodoloģijas prasības sistēmai.....	18
Metodoloģijas pamati un posmi	19
Ontoloģijas un datubāzes sasaistes realizācija	20
Zināšanu bāzes izveidošanas posms.....	20
Sistēmas izveidošanas posms	21
4. Intelektuālās respondentu datu izgūšanas sistēmas izstrāde.....	22
Zināšanu bāzes sagatavošana	22
Ontoloģijas funkciju realizācija.....	23
Prototipa saskarne.....	24
Datu izgūšana un apstrāde	25
Sistēmas struktūra.....	26
Izstrādātās sistēmas piemērotības noteikšana.....	26
REZULTĀTI UN SECINĀJUMI	28
IZMANTOTĀS LITERATŪRAS SARAKSTS.....	32

VISPĀRĒJS DARBA RAKSTUROJUMS

Problēmsfēra

Pateicoties jaunākajām informācijas tehnoloģijām, ir iespējams glabāt un apstrādāt lielu informācijas daudzumu. Gandrīz katrā nozarē ir nepieciešamība veikt pierakstus par notikumiem un objektiem, kā arī vākt un analizēt cita veida informāciju. Informācijas glabāšana ir fundamentāla tālākai analīzei un informācijas apstrādei. Gan lielākām, gan mazākām organizācijām un uzņēmumiem ar laiku rodas sarežģītumi ar datu pārskatāmību [106]. Jo vairāk datu ir datubāzē, jo sarežģītāk ir izgūt nepieciešamo informāciju. Arī projekta dzīves gaitā datubāzes iekšējā uzbūve mainās. Šādu parādību īpaši bieži ir iespējams novērot zinātniskos pētījumos, kad izpētes sākumā vēl nav zināms, kā nākotnē projekts attīstīsies vai mainīsies un kādā veidā dati var tikt analizēti. Pat gadījumos, kad uzprojektētā datu struktūra ir adekvāta informācijai, tik un tā var rasties problēma ar liela apjoma heterogēnu datu glabāšanu. Jebkurā gadījumā rodas nepieciešamība apkopot datus tā, lai atvieglotu darbu ar tiem. Viena no tehnoloģijām, kas radusies tieši šādas nepieciešamības dēļ, ir ontoloģijas. Šis darbs apskata ontoloģiju lietošanu datu izgūšanas vienkāršošanai.

Aktualitāte

Darbā ir izveidota metodoloģija, kas ļauj izstrādāt ontoloģijā bāzētu datu izgūšanas sistēmu. Lai gan šobrīd eksistē citas pieejas un sistēmas, kas sola ieviest līdzīgu funkcionalitāti, esošās sistēmas piedāvā pārveidot klasisko datubāzi par semantiskiem tīkliem vai citām struktūrām, kas lielā mērā atšķiras no datubāzēs lietotajām struktūrām. Esošās sistēmas veic spriešanu par visiem datiem vienlaikus, kas lielā mērā palēnina darbību liela datu apjoma gadījumā. Dažas no esošajām sistēmām lieto tikai daļu no tā, ko spēj piedāvāt ontoloģija (skat. 2. nodaļu). Šīs sistēmas piedāvā tikai klašu pievienošanu un šo klašu hierarhijas izveidošanu. Parasti šāda veida sistēmas izlaiž sarežģīto klašu (klases ar ierobežojumiem) veidošanu, kurām piemīt īpašības un ierobežojumi. Lietojot šādas sistēmas, rodas nepilnība, jo tieši sarežģītās klasēs tiek glabātas noklusētas zināšanas, kas ļauj individuus klasificēt pēc īpašībām.

Problēmas nostādne

Pastāv situācija, kad eksistējošā informācijas apstrādes sistēmā tiek izmantota atsevišķa datu glabāšanas sistēma (datubāze), kas satur sarežģīti strukturētu informāciju, kas sastāv no liela skaita heterogēnu ierakstu. Dati ir atrodami vairākās tabulās bez skaidrām saitēm starp tām, kā arī bez zināmiem ierobežojumiem. Tabulas ir lielas un sarežģītas. Tajās ir daudz kolonnu. Nav zināmi noteikumi, pēc kuriem kolonnas tiek aizpildītas vai atstātas tukšas. Datu izgūšana no šādas sistēmas ir sarežģīts un laikietilpīgs process, ko spēj veikt tikai konkrētās datubāzes eksperts. Pareizo un

vajadzīgo datu izgūšana prasa padziļinātas zināšanas par datubāzes struktūru, kas nav vienkārši iemācamas visiem datu lietotājiem.

Tāpēc eksistē uzdevums aprakstīt eksperta zināšanas kopā ar zināšanām no citiem avotiem koplietojamā zināšanu modelī. Šāda modeļa izveidošanas mērķis ir, modeli integrējot, iegūt sistēmu, kurā būtu skaidri definēti termini datubāzes satura aprakstam un atspoguļošanai. Šiem terminiem ir jābūt balstītiem uz domēnā lietojamiem terminiem. Ir svarīgi, ka šāds modelis nav datubāzes modelis, bet gan domēna modelis ar domēna terminoloģiju, kurā ir ieliktas saites ar datubāzi un datu izgūšanas noteikumiem, lai to varētu izmantot domēna eksperts. Ontoloģija dod iespēju papildināt esošu datubāzes shematisko informāciju ar domēna aprakstiem un klasēm.

Darba aprobācijā izpētīts medicīniska pētījuma datubāzes fragments. Aprobācijas mērķis – izstrādāto metodoloģiju izmantot ontoloģijās bāzētas datu izgūšanas sistēmas realizācijai. Pēc metodoloģijas izstrādāta sistēma ļautu pieslēgties un izgūt datus no šīs datubāzes atvieglotā veidā, lietojot pētījumā nodarbinātā personāla ikdienas terminoloģiju. Tas jāizdara, ontoloģijā ievietojot nepieciešamās zināšanas par domēnu un datubāzi.

Pētījuma mērķis un uzdevumi

Promocijas darba **mērķis** – izstrādāt metodoloģiju ontoloģijā bāzētas datu izgūšanas realizācijai, kas vērsta uz lietojamības vienkāršošanu un ir paredzēta eksistējošas relāciju datubāzes papildināšanai. Izvirzītā mērķa sasniegšanai nepieciešams izpildīt šādus **uzdevumus**:

- izpētīt ontoloģijas un ar ontoloģijām un datu izgūšanu saistītas tehnoloģijas, gūstot ieskatu šo tehnoloģiju iespējās;
- izpētīt un salīdzināt esošās uz datu izgūšanu vērstās intelektuālās tehnoloģijas un programmlīdzekļus un to uzbūves pieejas, lai identificētu iespējamās nepilnības un izvirzītu prasības metodoloģijai;
- izstrādāt metodes un pieejas, kas ļauj realizēt visas datu izgūšanai nepieciešamās darbības ontoloģijā bāzētā veidā;
- izstrādāt metodoloģiju, kas apraksta metodes, pieejas un pamatnostādnes ontoloģijā bāzētas datu izgūšanas realizācijai;
- izveidot ontoloģijā bāzētās datu izgūšanas sistēmas prototipu atbilstoši izstrādātajai metodoloģijai, lai aprobētu prototipu un metodoloģiju.

Intelektuālas datu izgūšanas sistēmas izstrādes metodoloģija tiek veidota pieeju un metožu kopu aprakstam, ar kurām ir iespējams realizēt ontoloģijā bāzētu datu izgūšanas sistēmu kā papildinājumu eksistējošai relāciju datubāzei.

Pētījuma objekts un priekšmets

Pētījuma objekts ir ontoloģijā bāzētas intelektuālas datu izgūšanas sistēmas izveidošanas metodoloģija.

Pētījuma priekšmets ir ontoloģijas, ontoloģiju izveidošanas metodes, ontoloģiju pielietošana datu izgūšanas uzdevumā, zināšanu aprakstīšanas iespējas ar ontoloģiju.

Pētījuma hipotēzes

Promocijas darba izstrādes gaitā, pamatojoties uz ontoloģiju izveides tehnoloģiju un datu izgūšanas pieeju izpēti un analīzi, pārbaudei izvirzītas trīs hipotēzes.

1. Ontoloģijas iespējas ir pietiekamas ontoloģijas un datubāzes sasaistei, datu apakškopas definēšanai, datu validācijai un papildu zināšanu atklāšanai par klasēm un indivīdiem, nelietojot papildu tehnoloģijas.
2. Lai izveidotu meklējamo datu izgūšanai nepieciešamo *SQL* vaicājumu no ontoloģijā pieejamās informācijas, pietiek aprakstīt datubāzes objektus ar ontoloģijas konceptiem.
3. Pēc metodoloģijas izstrādāta sistēma sniedz interaktīvu semantisko slāni tādā mērā, lai lietotājam nebūtu jāsaskaras ar datubāzes terminiem un tehnoloģijām, veidojot *ad hoc* datu pieprasījumu.

Pētījuma metodes

Promocijas darbā izmantotas manuālās ontoloģijas izveidošanas metodes un vairākas ontoloģijas apmācības pieejas (*ontology learning*) – automātiskas ontoloģijas izveidošanas metodes, izmantojot mašīnāpmācības algoritmus [37, 41, 44], apvienošanas (*merging*) un kartēšanas (*mapping*) metodes, ar datubāzes apstrādi saistītas metodes [37], datu analīzes metodes [37], datubāzes “veselības” pārbaudes metodes [37], ontoloģijas spriešanas realizāciju [39, 42]. Promocijas darbā izmantotais ontoloģijas izveidošanas process izstrādātās ontoloģijas definēšanai un papildināšanai (skat. 4. nodaļu) ir aprakstīts dokumentā “Ontoloģijas 101”, kas ir fundamentāls materiāls par ontoloģijām [87].

Zinātniskā novitāte

Promocijas darbā izstrādāta metodoloģija, kas apraksta ontoloģijā bāzētas datu izgūšanas sistēmas izstrādes procesu. Tajā tiek aprakstīts datu izgūšanas sistēmas izstrādes process, kas implementē datubāzes abstrakciju un vienkāršo vajadzīgo datu aprakstu. Metodoloģijas izstrādes procesā ir izstrādātas vairākas jaunas pieejas, kas ir nepieciešamas šādas sistēmas realizācijai.

1. Ir izstrādāta jauna ontoloģijas un datubāzes sasaistes pieeja. Šajā pieejā ontoloģijā tiek izmantoti datubāzes objektus aprakstoši koncepti, kas kalpo par savienošiem punktiem starp ontoloģijas domēna konceptiem un informāciju no datubāzes.

2. Ir izstrādāta jauna ontoloģijas izveidošanas pieeja. Šīs pieejas mērķis ir izveidot ontoloģiju, kas izmanto savienojošus elementus. Šīs ontoloģijas izveidošanas pieejas lietošanas rezultātā tiek iegūta ontoloģija, kas izmanto datubāzes konceptus kā savienojošos punktus. Izmantojot šo pieeju, tiek izgūta ar datubāzi savienota ontoloģija, kas var kalpot par zināšanu bāzi izstrādājama datu izgūšanas sistēmai.
3. Ir izstrādāta ar datubāzi sasaistītas ontoloģijas novērtēšanas pieeja. Izmantojot šo novērtēšanas pieeju, var noteikt, cik lielā mērā ontoloģijas domēna koncepti ir saistīti ar datubāzes elementiem, kā arī atklāt domēna konceptus bez atspoguļojuma datubāzē.
4. Ir izstrādāta pieeja datubāzes konceptu un domēna konceptu atšķiršanai. Lai izstrādātā sistēma spētu ņemt vērā konceptu dažādību, atšķirtu konceptu veidus un izmantotu pareizos konceptus *SQL* vaicājuma izveidei, nepieciešams identificēt datubāzes konceptus. Šim nolūkam metodoloģijā tiek piedāvāta pieeja, kas izmanto konceptu identifikatoru prefiksa daļas, lai norādītu konceptu veidus.
5. Ir izstrādāta *SQL* vaicājuma izveidošanas pieeja, kurā izmantota ar vaicājuma aprakstu saistītu konceptu meklēšana, lai atklātu vaicājumā iesaistāmus datubāzes konceptus. Izmantojot ontoloģijā pieejamo informāciju un spriešanas funkcijas, tiek izveidoti vaicājumi datubāzes indivīdu izgūšanai.
6. Ir izstrādāta lietotāja saskarne, kas ļauj lietotājam izvēlēties domēna ontoloģijas konceptus, lai aprakstītu meklējamos datus. Lietotājs, izvēloties viņam pazīstamus terminus, veido ontoloģijā interpretējamu konceptuālu aprakstu. Ontoloģijā apraksts tiek papildināts un sistēmā pārveidots par *SQL* vaicājumu.

Praktiskais nozīmīgums

Izstrādātā sistēma palīdz iegūt datus no kompleksas datubāzes, piedāvājot izvēlēties domēnu terminoloģiju sarežģītu vaicājumu vietā. Darbā izstrādātais medicīnas datu izgūšanas sistēmas IRDIS (intelektuālā respondentu datu izgūšanas sistēma) prototips ļauj piekļūt datiem medicīniskā pētījuma medicīniskajam personālam, kuram nav tehnisku zināšanu par datubāzes darbību. Sistēma ļauj viņiem izgūt nepieciešamo informāciju pilnīgā un ērtā veidā. Izmantojot sistēmas saskarni, lietotājiem ir iespēja aprakstīt meklējamo informāciju saprotamā medicīniskā un ar pētījumu saistītā terminoloģijā, kurā ir neizteiktas definīcijas un datu apraksti. Sistēma veic pārveidojumu, kurā tā aizstāj šos terminus un aprakstīto struktūru ar datu iegūšanas vaicājumu. Iegūtie dati tiek strukturēti, lai atbilstu lietotāja prasībām. Tas ļauj personālam iegūt viņiem vajadzīgo informāciju ērti un tieši no datubāzes. Pirms šī rīka ieviešanas katrreiz, kad bija nepieciešams izgūt iepriekš neparedzētu datu apakškopu, medicīnas personālam bija jāvērsas pie datubāzes speciālistiem. Tas prasīja papildu

laiku, eksperta pūles, kā arī ieviesa kļūdas, kas varēja rasties šādas starpniecības rezultātā. Izstrādātais rīks atvieglo eksperta darbu, kā arī ļauj lietotājiem ātri, ērti un iteratīvi strādāt tiešā veidā ar datubāzes saturu.

Pamatojoties uz darbā pārbaudītām un izstrādātām pieejām, tika realizēta datu izgūšanas programmatūra medicīnas pētījumam. Izstrādātā sistēma tika ieviesta medicīnas pētījumā, kurā to lieto, lai no datubāzes izgūtu vajadzīgos datus.

Tika izstrādāta sistēma, kas, izmantojot ar pētījuma darbību saistītus ontoloģijas konceptus, ļauj pieslēgties datubāzei un izgūt informāciju, kas ir nepieciešama medicīnas darbiniekam. Darbinieks atrod un atlasa viņam vajadzīgos konceptus. Koncepti atbilst tam, ko lietotājs izvirzīja kā vēlamu. Pēc atlasītiem konceptiem sistēma piekļūst datubāzei un iegūst datus.

Aprobācija

Promocijas darba izstrādāšanas gaitā iegūtie rezultāti un sasniegumi ir prezentēti, uzstājoties ar ziņojumu 11 starptautiskās zinātniskās konferencēs.

1. Riga Technical University 58th International Scientific Conference 13.10.2017. (*Gorskis H. Database Object Concepts for Ontology Based Data Access*).
2. 11th International Scientific and Practical Conference. 15–17 June, 2017, Rezekne Academy of Technologies, Rezekne, Latvia (*Gorskis H., Aleksejeva L., Polaka I. Ontology-Based System Development for Medical Database Access*).
3. Riga Technical University 57th International Scientific Conference 18.10.2016. (*Gorskis H., Using ontology reasoning in data queries*).
4. 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, Vien, Austria, 29-30 August, 2016 (*Gorskis H., Aleksejeva L., Polaka I. Database analysis for ontology learning*).
5. 22rd International Conference on Soft Computing Mendel 2016, 8–10 June, 2016, Brno, Czech Republic (*Gorskis H., Borisov A., Aleksejeva L. Genetic algorithm based random selection-rule creation for ontology building*).
6. Research and Technology – Step into the Future TTI Research and Academic Conference, 22 April, 2016, Riga, Latvia (*Gorskis H. Location Based Reminder System with Reusable Ontology*).
7. Riga Technical University 56th International Scientific Conference 16.10.2015. (*Gorskis H., Borisovs A. Automated Ontology Building using Classification Trees*).
8. 10th International Scientific and Practical Conference, 18–20 June, 2015, Rezekne Higher Education Institution, Rezekne, Latvia (*Gorskis H., Borisovs A. Storing an OWL 2 Ontology in a Relational Database Structure*).
9. 6th International Conference on Knowledge Engineering and Ontology Development, KEOD 2014, 21–24 October, Rome, Italy (*Gorskis H., Borisov A. Location based Reminder System with Reusable Ontology*).

10. Riga Technical University 55th International Scientific Conference 17.10.2014.
(Gorskis H., Borisov A. *An ontological basis for a reminder system*).
11. Riga Technical University 54th International Scientific Conference 16.10.2013.
(Gorskis H., Borisovs A. *Reusable components in knowledge-based systems*).

Rezultāti publicēti zinātniskajos rakstos.

1. Gorskis, H. Aleksejeva, L., Poļaka, I. Database Concepts in a Domain Ontology. *Information Technology and Management Science*. 2017, vol. 20, pp. 69–73. ISSN 2255-9086. Indexed in: EBSCO, Google Scholar, Ulrich's International Periodicals Directory.
2. Gorskis, H., Aleksejeva, L., Poļaka, I. Ontology-Based System Development for Medical Database Access. In: *Environment. Technology. Resources : Proceedings of the 11th International Scientific and Practical Conference. Vol.2, Latvia, Rezekne, 15–17 June, 2017*. Rezekne: Rezekne Academy of Technologies, 2017, pp. 24–29. ISSN 1691-5402. Indexed in: **Scopus**.
3. Kiršners, A., Paršutins, S., Gorskis, H. Entropy-Based Classifier Enhancement to Handle Imbalanced Class Problem. *Procedia Computer Science*. 2017, vol. 104, pp. 586–591. ISSN 1877-0509. Indexed in: **Scopus**.
4. Gorskis, H. Improved database schema development for OWL2. *Information Technology and Management Science*. 2016, vol. 19, pp. 85–91. ISSN 2255-9094. Indexed in: EBSCO, Google Scholar, Ulrich's International Periodicals Directory.
5. Gorskis, H., Aleksejeva, L., Poļaka, I. Database Analysis for Ontology Learning. *Procedia Computer Science*. 2016, vol. 102, pp. 113–120. ISSN 1877-0509. Indexed in: **Scopus** and **ISI Web of Knowledge**.
6. Gorskis, H., Borisovs, A., Aleksejeva, L. Genetic Algorithm Based Random Selection-Rule Creation for Ontology Building. In: *Recent Advances in Soft Computing: Proceedings of the 22nd International Conference on Soft Computing (MENDEL 2016). Czech Republic, Brno, 8–10 June, 2016*. Cham: Springer International Publishing AG, 2017, pp. 34–45. (Advances in Intelligent Systems and Computing. Vol. 576). ISBN 978-3-319-58087-6. Indexed in: **Scopus**, **ISI Web of Knowledge**.
7. Gorskis, H., Borisovs, A. Ontology Building Using Classification Rules and Discovered Concepts. *Information Technology and Management Science*. 2015, vol. 18, pp. 37–41. ISSN 2255-9086. e-ISSN 2255-9094. Available from: doi:10.1515/itms-2015-0006. Indexed in: Baidu Scholar, Cabell's Directory, Celdes, CNKI Scholar, CNPIEC, EBSCO, Google Scholar, Ulrich's International Periodicals Directory, VINITI and others.

8. Gorskis, H., Borisovs, A. Storing an OWL 2 Ontology in a Relational Database Structure. In: *Environment. Technology. Resources: Proceedings of the 10th International Scientific and Practical Conference, Latvia, Rezekne, 18–20 June, 2015*. Rezekne: Rezekne Higher Education Institution, 2015, pp. 71–75. ISBN 978-9984-44-173-3. Indexed in: **Scopus**.
9. Gorskis, H., Borisovs, A. Location based Reminder System with Reusable Ontology. In: *KEOD 2014 : Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Italy, Rome, 21–24 October, 2014*. [S.l.]: SciTePress, 2014, pp. 161–167. ISBN 978-989-758-049-9. Indexed in: **Scopus**, Engineering Village, dblp, Inspec, Thomson Reuters.
10. Gorskis, H., Borisovs, A. Knowledge-based System Design Based on Generic Method Conception. *Information Technology and Management Science*. 2014, vol. 17, pp. 55–60. ISSN 2255-9086. Indexed in: Cabell's Directory, Celdes, CNKI Scholar, CNPIEC, EBSCO, Google Scholar, Ulrich's International Periodicals Directory, VINITI and others.
11. Gorskis, H., Borisovs, A. Reusable Components in Knowledge-based Configuration Design Systems. *Information Technology and Management Science*. 2013, vol. 16, pp. 66–72. ISSN 22559086. Indexed in: EBSCO, ProQuest, Cabell's Directory, Celdes, CNKI Scholar, CNPIEC, Google Scholar and others.
12. Gorskis, H., Čižovs, J. Ontology Building Using Data Mining Techniques. *Information Technology and Management Science*. 2012, vol. 15, pp. 183–188. ISSN 2255-9086. Indexed in: EBSCO, ProQuest, Versita, VINITI.

Promocijas darba pētījuma rezultāti ir izstrādāti saistībā ar šādiem projektiem:

- dalība projektā “Viedtālrunis slimības noteikšanai no izelpojamā gaisa”. Pasūtītājs: European Commission, Research Programme, DG Research & Innovation – HORIZON 2020 (no 10.2015.);
- Latvijas Universitātes projekts 2014/0035/2DP/2.1.1.1.0/14/APIA/VIAA/102 “Riska stratifikācijas metodes izstrāde kuņģa vēža un pirmsvēža stāvokļiem, izmantojot biomarkierus” (09.2014.–09.2015.).

Darbs izstrādāts ar RTU Datorzinātnes un informācijas tehnoloģijas fakultātes projektu “Zinātniskās darbības attīstība augstskolā” “34-12000-DOK.DITF/16” (12.2016.–08.2017.) un “34-12000-DOK.DITF” (01.2016.–12.2016.) atbalstu.

Promocijas darba struktūra un saturs

Pirmajā nodaļā apskatītas ontoloģijas pamattehnoloģijas, kas tiešā veidā ietekmē ontoloģijas darbību un funkcionalitāti. Šī nodaļa sniedz ieskatu par to, kas ir

ontoloģijas, kādas citas tehnoloģijas tās ietekmē, kāda ir to vēsture un kādas funkcijas tās piedāvā zināšanu un informācijas aprakstam.

Otrā nodaļa apskata ontoloģijā bāzētās datu izgūšanas pašreizējo stāvokli. Tajā apkopots, kādas darbības tiek veiktas un kādas metodes ir pieejamas, lai realizētu ontoloģijā bāzēto datu izgūšanu ar līdzšinējām pieejām. Turpat tiek apskatītas un salīdzinātas arī eksistējošas lietotnes, kas realizē ontoloģijā bāzētu datu izgūšanu.

Trešā nodaļa veltīta ontoloģijā bāzētās datu izgūšanas sistēmas realizācijas metodoloģijas aprakstam. Tajā tiek aprakstītas izmantoto eksistējošo metožu modifikācijas, kā arī ontoloģijas izveidošanas un īstenošanas iespējas, kas līdz šim nepastāvēja. Nodaļā aprakstīta visas metodoloģijas struktūra un darbības soļi.

Ceturtajā nodaļā izklāstīta intelektuālās respondentu datu izgūšanas sistēmas (IRDIS) prototipa realizācija, kura ir izveidota pēc izstrādātās metodoloģijas.

Rezultātu un secinājumu daļā ir aprakstīti ieskatī, kas iegūti darba izpildes gaitā, kā arī sasniegto rezultātu uzskaitījums un interpretācija.

DARBA NODAĻU SATURA APRAKSTS

1. AR ONTOLOĢIJĀ BĀZĒTU DATU IZGŪŠANU SAISTĪTAS TEHNOLOĢIJAS

Nodaļā aprakstītas tehnoloģijas, kas kalpo par ontoloģijas pamatu, pašas ontoloģijas un ontoloģijas lietojums datu izgūšanas uzdevumā. Šo tehnoloģiju apskats dod ieskatu ontoloģiju potenciālā un izmantošanas iespējās. Ontoloģijas pamatojas uz semantiskām tehnoloģijām, kas eksistēja pirms tām. Semantiskās tehnoloģijas ir tehnoloģijas, kas apstrādā informāciju, kas ir sniegta brīvā un nestrukturētā veidā, izmantojot jēdzienus. Semantisko tehnoloģiju pamats ir semantiskie tīkli, kas apraksta informāciju, lietojot mezglus un saites.

Ontoloģijas pamata tehnoloģijas

Semantiskās tehnoloģijas ir daudzveidīgas, un tām ir liels skaits realizāciju. Ir dažādi semantisko tīklu veidi, kas veic aprakstu, lietojot mezglus. Viens no semantisko tīklu paveidiem ir konceptu kartes (*CM, concept map*), kas tiek lietotas daļēji formālu semantisko zināšanu reprezentācijai [50]. Resursu aprakstīšanas ietvars (*RDF, Resource Description Framework*) ir tehnoloģija, kas izveidota, lai standartizētu resursu aprakstu [92]. Parasti, kad ir nepieciešams aprakstīt informāciju, tiek veidotas datu struktūras. Šī struktūra ir sava veida modelis, kas ietekmē vācamo informāciju. Gadījumos, kad visa iespējamā informācija nav iepriekš zināma vai ir ļoti daudzveidīga, ir sarežģīti veidot stingri strukturētu modeli. Viena no iespējamām pieejām

nestrukturētu datu izmantošanas problēmas risināšanai ir semantisko tīklu lietošana kopā ar *RDF*. Semantiskie tīkli ļauj aprakstīt informāciju brīvā veidā. Jebkuru ienākošo informāciju ir iespējams apskatīt kā centrālo mezglu, kas reprezentē informācijas vienību (objektu), un saišu uzskaitījumu, kas savienotu šo mezglu ar datiem vai citiem objektiem. Šāda veida semantiskos tīklus ir iespējams aprakstīt, nosaucot saistītos objektus (vai objektu un datu vienību) un saiti, kas tos savieno. Resursu aprakstīšanas ietvarā tas tiek saukts par izteicienu, kas sastāv no trijnieka. Trijnieks ir veidojums no subjekta, predikāta un objekta, kur:

- izteiciena subjekts ir elements, par kuru tiek veikts apgalvojums;
- predikāts ir savienojums, saite vai īpašība, kas norāda uz attiecības vai īpašības veidu;
- objekts ir trijnieka daļa, kas pasīvi piedalās šajā apgalvojumā un tiek savienota ar subjektu izmantojot predikātu.

Ar vienkāršu *RDF* nav pietiekami sarežģītas domēna informācijas konceptuālam aprakstam, jo visa informācija ir viendabīga un *RDF* nav pietiekami spējīgs aprakstīt visas informācijas aspektus. Tajā pašā laikā *RDF* spēj atrisināt informācijas daudzveidības problēmu. Stingras struktūras trūkums ļauj aprakstīt jebkāda veida informāciju. Tomēr tādā gadījumā šāda veida tīkls var izauzt nesamērīgi liels un iegūt eksponenciāli lielu saišu skaitu. Tīkls, kas veidojas no resursiem, atsaucoties citam uz citu, ar lielu mezglu skaitu kļūst cilvēkam nepārskatāms. Turklāt visi resursi ir vienādi. Tie ir vienkāršas īpašību kolekcijas. Līdz ar to rodas nepieciešamība pēc klasēm vai veidiem, kā tos iedalīt, lai vienkāršotu vajadzīgo resursu atlasī. Lai papildinātu *RDF* ar lielākām spējām aprakstīt informāciju, tika izveidota *RDF* shēma [93].

RDF shēma ir *RDF* paplašinājums [56], kura mērķis ir piešķirt *RDF* iespēju aprakstīt datu meta līmeni. *RDF* shēma ievieša tādus jēdzienus kā klase un īpašības saite. Ontoloģija lieto *RDF* shēmā izveidotu klašu līmeni un papildina to ar noteiktu spriešanas veidu [1] un dažiem citiem klašu līmeņa elementiem. Ir svarīgi saprast, ka ontoloģija nav aizvietoājums ieviestajam klašu līmenim, kas ir definēts *RDF* shēmā. Tas ir noteikta veida paplašinājums. Klašu līmeņa ieviešana *RDF* datos ļauj atšķirt dažāda veida īpašības. Tas dod iespēju aprakstīt klases un klašu īpašības. Tomēr šajā pieejā nav stingru mehānismu klasifikācijai ar *RDF* shēmu vien, un nav vienotas pieejas tam, kas ir klase un kādi resursi var piederēt vai nepiederēt klasēm.

Ontoloģijas informācijas tehnoloģijā

Ontoloģija ir zināšanu aprakstīšanas veids, kas ļauj definēt klases un klašu savstarpējas attiecības, pēc kurām ir iespējams veikt klasifikāciju un īpašību mantošanu. Klasifikācija tiek veikta, apskatot informācijas vienības īpašību kopu. Informācijas vienība ontoloģijā tiek saukta par individu. Katram individuam piemīt īpašību kopa, pēc kuras to ir iespējams klasificēt, kā arī secināt par noklusētām īpašībām

pēc klašu piederībām. Informācijas tehnoloģijās par ontoloģiju tiek saukta īpašā veidā strukturēta zināšanu bāze, kurā tiek definēti visi koncepti, kas ir uzskatāmi par svarīgiem kādā domēnā. Koncepti apraksta klases, datu tipus un dažāda veida īpašību saites. Vairāki autori piedāvā savas formālas ontoloģijas definīcijas [31, 87, 108]. Pamatojoties uz šīm definīcijām, darbā ontoloģija tiek formāli definēta kā šādu elementu kopa (skat. izteiksmi (1)):

$$O = \{C, R^D, R^I, D, H^C, H^{R^D}, H^{R^I}, H^D, I_C, I_R, A\}, \quad (1)$$

kur	C	–	klašu kopa;
	D	–	datu tipu kopa;
	R^D	–	datu īpašību kopa;
	R^I	–	objektu īpašību kopa;
	H^C	–	klašu hierarhija;
	H^D	–	datu tipu hierarhija;
	H^{R^D}	–	datu īpašību hierarhija;
	H^{R^I}	–	objektu īpašību hierarhija;
	I_C	–	klašu instances – indivīdi;
	I_R	–	saišu instances;
	A	–	indivīdu aksiomas.

Ontoloģija satur vairākus klašu C , datu tipu D , datu īpašību R^D , objektu īpašību R^I un indivīdu I_C veidus. Šie elementu konceptu veidi ir izmantojami ontoloģijā domēna definīcijas veidošanai. Katram konceptu veidam ir sava hierarhija (H^C , H^D , utt.), pēc kura tiek noteiktas konceptu attiecības. Ar aksiomu palīdzību tiek veikti visi apgalvojumi par domēna konceptiem, kas apraksta modeļa saturu.

Ontoloģija atšķiras no citām semantiskām tehnoloģijām un klašu modeļiem ar to, ka tā apraksta konceptu eksistences kritērijus (līdzīgi filozofiskai ontoloģijai). Šai unikālajai pieejai ir ietekme uz klašu definīciju un spriešanas mehānismiem. Bieži filozofiskā daļa tiek ignorēta; tas rada domstarpības un pārpratumus par to, kā ontoloģijas tiek veidotas un lietotas. Dažreiz ontoloģijas tiek jauktas ar vienkāršu klašu modeļiem. Dažreiz ontoloģija tiek kļūdaini pielīdzināta objektorientētajai klašu modelēšanas pieejai. Ontoloģija nav paredzēta objektu šablonu definēšanai, tajā ir iespējams norādīt uz klases īpašībām, kas piemīt visiem tās indivīdiem, bet tas attiecas tikai uz eksistenciāli svarīgām īpašībām nevis uz visām īpašībām. Ontoloģijā klase tiek saistīta ar īpašību tikai tad, ja šī īpašība ir eksistenciāli svarīga.

Ontoloģijā bāzētas datu izgūšanas jēdziens

Ontoloģijā bāzētā datu izgūšana ir iepriekš minēto tehnoloģiju un pieeju apvienojums ar mērķi izgūt datus no ontoloģijas vai cita datu avota ar ontoloģijas

palīdzību. Ontoloģijā bāzētās datu izgūšanai jārada iespējas izmantot zināšanas un spriešanu, kas piemīt ontoloģijai, lai aprakstītu vēlamos datus. Pēc izveidotā apraksta tiek izmantotas ontoloģijā glabātas zināšanas un ar spriešanas procedūrām veikta datu atlase. Šī procesa rezultātā no datu avota tiek izgūti dati.

Ontoloģijā bāzētās datu izgūšanas izcelsme ir ontoloģijas vaicājumu izmantošana ontoloģijas informācijas izgūšanai [72]. Ar *SPARQL* vaicājumu valodu ir iespējams izgūt mezglu un saišu informācijas no *RDF* un ontoloģiju aprakstiem [105]. Ontoloģijā bāzēta datu izgūšana balstās uz šīs vaicājumu valodas lietošanas datu avotam apvienojumā ar ontoloģijas aprakstu. Ar to tiek panākts, ka, izmantojot ontoloģiju, tiek izgūta informācija ne tikai par ontoloģiju, bet arī par datiem no datu avota. Šādas funkcionalitātes mērķis ir izgūt ne tikai vēlamu informāciju, bet arī papildinātus, klasificētus datus no datu avota.

2. ONTOLOĢIJĀ BĀZĒTAS DATU IZGŪŠANAS METODES

Nodaļā dziļāk apskatīts pats ontoloģijā bāzētās datu izgūšanas process un pieejas tā realizācijai. Tiek pētīts, no kādiem elementiem ontoloģijā bāzēts datu izgūšanas process sastāv, kādas tehnoloģijas ir saistītas ar šo procesu un kādi principi tiek lietoti to realizācijās – apskatītas vairākas sistēmas, kas izmanto ontoloģijā bāzētu datu izgūšanu dažādos veidos, kā arī – kā sistēmas realizē šo procesu.

Ontoloģijā bāzētās datu izgūšanas sistēmas

Lai saprastu, kādā veidā ir izstrādātas līdzšinējās sistēmas un pēc kādām metodoloģijām tās ir veidotas, nepieciešams apskatīt to funkcionalitāti un izmantotās tehnoloģijas. Darbā apskatītas sistēmas, kas lieto ontoloģijas, lai realizētu intelektuālo datu izgūšanu. Lielākajā daļā līdzšinējo sistēmu dažādos mēros tiek lietota *SPARQL* vaicājuma valoda. Tā var būt lietota tiešā veidā, grafiskā veidā, kā pamats citam risinājumam vai kā piekļuves punkts, kuram ir jāslēdzas klāt ar citu rīku, lai izgūtu vēlamos datus. Apskatā analizēti Latvijas Universitātē (LU) piedāvātie risinājumi [6, 7, 8, 9, 95], Bozen-Bolzano Brīvās universitātes projekts *Ontop* [18], Eiropas komisijas atbalstītais lielo datu projekts *Optique* [67], kā arī daži atsevišķi raksti par dažādiem pētījumiem ontoloģijā bāzētā datu izgūšanā. Apskata rezultātā šie rīki tiek salīdzināti pēc šādiem kritērijiem: ieviešanas vienkāršība, vaicājuma tehnoloģijas, piemērotība semantiskā slāņa funkciju izpildei, modeļa neatkarība un izpildes dinamika. Salīdzinājums ir parādīts 1. tabulā.

Ontop un *Optique* veido virtuālus *RDF* aprakstus no datubāzes, izmantojot sasaistes dokumentu, kurā datubāzes vaicājumi tiek pielīdzināti ontoloģijas konceptiem. Datu izgūšana tiek realizēta ar *SPARQL* vaicājumiem. *Optique* gadījumā ir pieejams vizuālais rīks, kas palīdz izveidot *SPARQL* vaicājumu. Izpilde notiek

dinamiskā veidā, modelis eksistē atsevišķi no datiem, bet semantiskais slānis netiek pilnībā realizēts vaicājuma sarežģītības dēļ. LU izstrādātā zvaigznes ontoloģija apvienojumā ar ātru vaicājuma valodu lieto datubāzes aprakstu ontoloģijas formā. Izmantojot īpašu vaicājuma valodu, kas ir līdzīga *SQL* ar dažiem papildinājumiem un latviskotu sintaksi, tiek piedāvāts rīks, kas no datubāzes atlasa atmiņā glabātus datus. Gan rīki, kas izmanto ontoloģiju, lai glabātu datus, gan rīki, kas izmanto datubāzi, kurā ir ontoloģijas un datu apvienojums, nav spējīgi strādāt neatkarīgā veidā. Datu izgūšana tiek realizēta ar *SPARQL* vai *SQL* palīdzību. Lietotājs, kuram ir jāizmanto sistēma, var nezināt *SPARQL*, un viņš nebūs spējīgs izveidot piemērotu vaicājumu. Līdz ar to metodoloģijā par nosacījumu tiek pieņemts, ka uzdevuma risināšanai ir jābalstās uz ontoloģijas iespējām. Lai gan *SPARQL* vaicājums spēj izgūt ontoloģijas elementus, pats vaicājums nav koncepts un nav tiešā veidā saistīts ar ontoloģiju.

1. tabula

Rīku un pieeju salīdzinājums

Rīks vai pieeja	Ieviešanas vienkāršība	Vaicājumā iesaistītās tehnoloģijas	Semantiskais slānis	Datu un modeļa saistība	Dinamiskais izpildījums
<i>Ontop</i>	Vienkārši	<i>SPARQL</i>	Slikta abstrakcija	Neatkarīgi	Jā
	+	-	-	+	+
<i>Optique</i>	Vidēji	Vizuālais rīks, <i>SPARQL</i>	Slikta abstrakcija	Neatkarīgi	Jā
	±	±	-	+	+
<i>Star ontology un QL</i>	Vidēji	Ātra vaicājuma valoda	Daļēja abstrakcija	Daļēji apvienoti	Jā
	±	±	±	±	+
<i>RDF glabātuve</i>	Sarežģīti	<i>SPARQL</i>	Nav abstrakcijas	Apvienoti	Nē
	-	-	-	-	-
Ontoloģijas DB	Sarežģīti	<i>SQL</i>	Nav abstrakcijas	Apvienoti	Nē
	-	-	-	-	-
Metodoloģijas sistēma	Vienkārši	Koncepti	Laba abstrakcija	Neatkarīgi	Jā
	+	+	+	+	+

Pēc metodoloģijas izveidotai sistēmai jāminimizē šīs nepilnības un jāpiedāvā risinājums, kas ir pieejams lietotājiem saprotamā veidā un realizē gan atsevišķu modeli datiem, gan ļauj realizēt dinamisku datu izgūšanu.

Datubāzes un ontoloģijas sasaistīšanas pieejas

Eksistē vairākas pieejas, kas realizē datu sasaisti starp ontoloģiju un datubāzi, un šīs pieejas tiek dažādi realizētas. Visbiežāk izmantotās sasaistīšanas pieejas iedalās šādi.

- Atsevišķa savienojuma dokumenta izmantošana [2, 18, 96, 103]. Šajā pieejā tiek izveidots atsevišķs dokuments, kurā ir aprakstīts, kā ontoloģijas koncepti ir saistīti ar datu avotu. Šādu dokumentu veido, izmantojot tehnoloģijas, kas nav tieši saistītas ar ontoloģijām. Parasti šāds dokuments sastāv no saistīto konceptu uzskaitījuma apvienojumā ar datubāzes vaicājumu.
- Ontoloģijas anotācija [17, 23]. Šajā pieejā ontoloģijas dokuments tiek paplašināts ar anotācijas elementiem, kas apraksta, kā iegūt klašu individuus. Šīs pieejas priekšrocība ir tajā, ka ontoloģijas izstrādātājs uzreiz var redzēt vaicājumu pie elementa. Pieejas trūkums – izstrādātājam nepieciešams apskatīt katru elementu atsevišķi, lai atrastu visas ar datubāzi saistītās anotācijas.
- Datubāzes ontoloģijas pieeja [22, 94]. Šajā pieejā visa ontoloģija izveidota pēc datubāzes shēmas [115, 116]. Šāda veida ontoloģija bieži ir automatizētas ontoloģijas izveidošanas rezultāts [34, 36, 43]. Ontoloģija pārtop par sava veida datubāzes aprakstu, kas piedāvā ontoloģijas semantikas iespējas [89]. Šīs pieejas priekšrocība ir tajā, ka pati ontoloģija ir ciešā veidā saistīta ar datubāzi [82] un ontoloģijas izstrādātājam nav nepieciešams rūpēties par datubāzes savienojumu kā no ontoloģijas atsevišķu elementu. Pieejas trūkums – ontoloģijas pārlieks fokuss uz datubāzi, vājš domēna apraksts.
- Tikai ontoloģija [102]. Tā vietā, lai sasaistītu datubāzi un ontoloģiju, tiek lietota tikai ontoloģija, kas papildu modelim satur pilnīgi visus datus no datubāzes. Šādā veidā nav nepieciešams implementēt savienojumu ar datubāzi, un visas darbības tiek veiktas tikai ontoloģijas ietvaros. Šāda pieeja ir iespējama tikai neliela datu apjoma gadījumā, jo ontoloģijas apstrādes tehnoloģijas nav optimizētas lielu datu apjomu pārstrādei.
- Ontoloģijas datubāze [3, 4, 28, 57]. Līdzīgi iepriekšējai pieejai ontoloģija un dati tiek apvienoti kopā, atšķirība ir tā, ka ontoloģijas pārvalde tiek implementēta datubāzē [120].

Ontoloģijas izveidošanas metodes

Ontoloģijas uzdevums ir aprakstīt dažādas augstākā līmeņa, domēna, uzdevuma, lietotnes vai datu zināšanas. Ontoloģijām gūstot popularitāti, tiek veidotas ontoloģijas, kas jau apraksta konkrētu problēmas sfēru, bet lielākajā daļā gadījumu nepieciešams izveidot jaunus zināšanu aprakstus, jo piemērotas ontoloģijas vēl nav izveidotas. Ontoloģijas var būt dažādas, tāpēc eksistē dažādi veidi, kā izveidot ontoloģiju. Vienam un tam pašam domēnam ir iespējams izveidot dažādas, bieži – nesavienojamas, ontoloģijas. Ne visām ontoloģijām var garantēt, ka tās būs noderīgas izvirkzītajam uzdevumam. Ontoloģijas izveidošana pārsvarā ir manuāls process. Šis process ir interaktīvs, jo nepieciešams validēt, vai izveidotā ontoloģija atbilst prasībām [101]. Nav viena pareiza veida, kā izveidot ontoloģijas, vienmēr pastāv alternatīvas, kas ir vai nav pieņemamas atkarībā no situācijas [87]. Eksistē arī vairāki mēģinājumi daļēji automatizēt ontoloģijas izveidošanu, lietojot eksistējošus formālus materiālus [41, 44, 45, 66].

3. ONTOLOĢIJĀ BĀZĒTAS DATU IZGŪŠANAS METODOLOĢIJAS IZSTRĀDE

Balstoties uz visu ar ontoloģijām un ontoloģijās bāzēto datu izgūšanas (OBDI) saistīto pieeju un risinājumu izpēti, ir izstrādāta jauna ontoloģijā bāzētas datu izgūšanas sistēmas izveidošanas metodoloģija. Šādas metodoloģijas izveidošana tiek pamatota ar to, ka visām līdzšinējām OBDI sistēmām ir trūkumi. Tās sistēmas, kas realizē ontoloģijas iespējas, nav ērtas lietošanā, jo tās pārāk stingri fokusētas uz resursu aprakstiem un vaicājumu valodām. Tās sistēmas, kas ir ērtas lietošanā, neizmanto visas ontoloģijas iespējas vai attālinās no tām, izveidojot sistēmas, kas ir tikai attālināti saistītas ar ontoloģijām, bet neizmanto tās pilnā mērā [6].

Izstrādātā metodoloģija ir paredzēta:

- OBDI sistēmu izveidošanai, kas ir pilnībā balstītas uz ontoloģiju;
- visu ar datu izgūšanu saistīto funkciju izpildīšanai, lietojot galvenokārt ontoloģiju;
- lietošanā ērtas sistēmas izveidošanai.

Šajā nodaļā aprakstīta izstrādātā metodoloģija, kuras uzdevums ir apkopot pieejas un metodes, kas ļauj realizēt ontoloģijā balstītas datu izgūšanas sistēmas.

Metodoloģijas prasības sistēmai

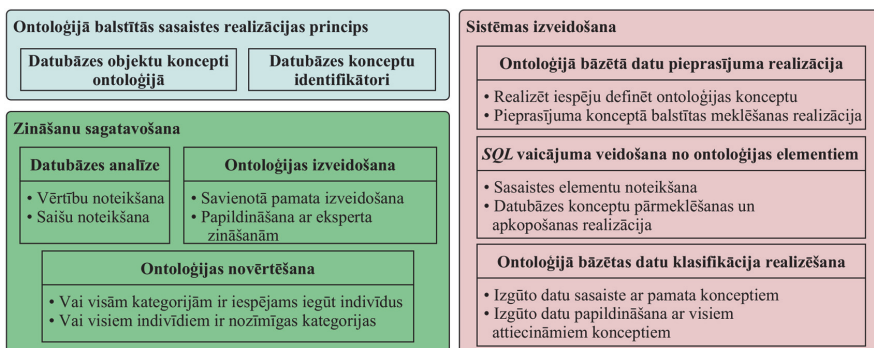
Metodoloģijas uzdevums ir piedāvāt procesu un metožu kopu, kas ļauj realizēt ontoloģijā balstītu datu izgūšanas sistēmu, kas papildina eksistējošu relāciju datubāzi ar lietotājiem ērtu funkcionalitāti datu izgūšanai. Otrajā nodaļā apskatītajām sistēmām ir konstatēti trūkumi un nepilnības (skat. 1. tabulu). Pēc izstrādātās metodoloģijas radītai sistēmai tiek izvirzītas prasības, kurām ir jānovērš identificētie trūkumi. Sistēmas prasības tiek izvirzītas šādiem sistēmas aspektiem:

- ieviešanas prasības;
- vaicājumā iesaistītās tehnoloģijas;
- semantiskā slāņa nodrošinājums;
- datu un modeļa sasaiste;
- vaicājuma izpildes dinamika.

Ieviešanas prasībām ir jābūt pēc iespējas minimālām. Tas ir saistīts ar noteikumu pēc iespējas mazāk skart eksistējošo sistēmu un datubāzi. Saskaņā ar izvirzīto darba mērķi izstrādājamai sistēmai jābūt papildinājumam visiem pašreiz lietojamiem risinājumiem. Tai ir jāspēj pieslēgties un izmantot eksistējošu relāciju datubāzi bez nepieciešamības veikt tajā struktūras vai satura izmaiņas. Tāpat tiek izvirzīts, ka izstrādājamai sistēmai nav jārealizē aizstājums eksistējošai datubāzei, tā nerealizē datu glabāšanas alternatīvu.

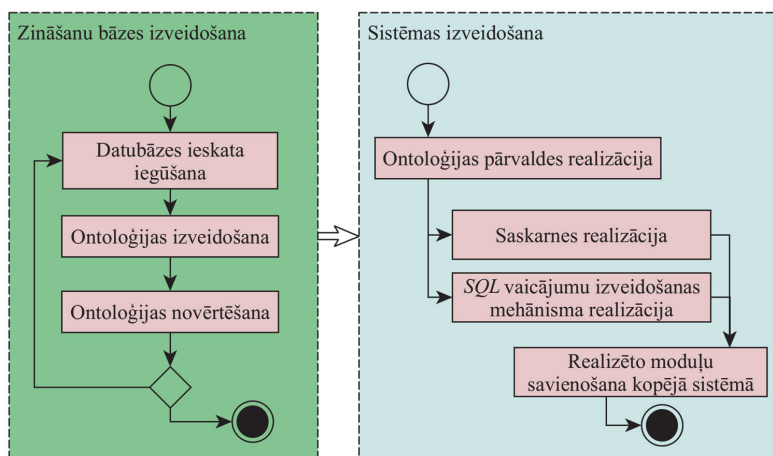
Metodoloģijas pamati un posmi

Izstrādātās metodoloģijas pieejas un posmi parādīti 1. attēlā. Ontoloģijā bāzētās datu izgūšanas sistēmas izveidošanas metodoloģija sastāv no trim lielām daļām. Metodoloģijas pamatdaļa ir ontoloģijā balstītās sasaistes realizācijas princips. Piedāvātais veids, kā sasaistīt ontoloģijas konceptus un datubāzes objektus, ietekmē veidu, kurā nepieciešams realizēt visas pārējās darbības. Tālāk metodoloģija iedalās divās lielās ar sistēmas izstrādi saistītās daļās. Pirmajā – zināšanu sagatavošanas – daļā tiek aprakstītas darbības un pieejas, kas ir saistītas ar ontoloģijas izveidošanu. Šī ir ontoloģija, kas būs izstrādātās sistēmas centrālā zināšanu bāze.



1. att. Metodoloģijas pieejas un posmi.

Otrā – sistēmas izveidošanas – daļa apraksta darbības, paņēmienus un pieejas, kas ir jālieto un jārealizē sistēmas darbības laikā. Nodaļā aprakstīti visi metodoloģijas posmi, pieejas un principi. Šie posmi tiek iedalīti noteiktās darbībās, kas ir jāveic katrā posmā. Darbības ir parādītas 2. attēlā.



2. att. Darbību process atbilstoši metodoloģijai.

Pieejas ietver ontoloģijas un datubāzes sasaistes veidu un sasaistes elementu identifikācijas veidu. Metodoloģija sastāv no diviem posmiem:

- zināšanu bāzes izveidošana,
- sistēmas izveidošana.

Ontoloģijas un datubāzes sasaistes realizācija

Izstrādātās metodoloģijas pamatā ir datubāzes un ontoloģijas sasaistes pieeja, kas ietelmē zināšanu bāzes un sistēmas izveides posmus. Savienojuma realizācijai izstrādātajā pieejā tiek lietoti trīs ontoloģijas konceptu veidi. Tie ir tabulu (vai skatu), kolonnu un relācijas koncepti.

- Tabulas koncepts ir klase. Tas norāda uz datubāzes tabulu vai skatu, jo datubāzē skats tiek apstrādāts tieši tādā pašā veidā kā tabula.
- Kolonna ir datu īpašība. Datu īpašības ir tiešā veidā attiecināmas uz tabulu kolonnām. Tas ir saistīts ar to, ka ontoloģijā datu īpašības apraksta indivīdam piemītošos datus, bet tabulu kolonna satur ieraksta datus.
- Relācijas koncepts ir objektu īpašība. Objektu īpašība norāda uz attiecībām starp klašu indivīdiem. Ir iespējams aprakstīt datubāzes tabulu relācijas, izmantojot objektu īpašības.

Papildus šiem trim lielajiem datubāzes konceptu veidiem rodas vēl papildu koncepti, kas arī ir saistīti ar datubāzi. Datubāzes tabulas atspoguļojums ontoloģijā ir klases koncepts, tāpēc šādas klases indivīdam ir jābūt datubāzes ierakstam. Tas nozīmē, ka datubāzes ieraksta atspoguļojums ontoloģijā ir indivīds. Datu izgūšanas uzdevumā datiem un modelim ir jābūt atdalītiem, tāpēc indivīdi, kas netiek lietoti klases definēšanai, netiek pievienoti ontoloģijai.

Lai realizētu izstrādāto ontoloģijas un datubāzes sasaistes pieeju, ir nepieciešams atšķirt un identificēt datubāzes un domēna konceptus. Tas ir nepieciešams tajos gadījumos, kad jādefinē domēna koncepts ar tādu pašu nosaukumu, kas jau ir izmantots datubāzes koncepta identificēšanā. Šim nolūkam tika izstrādāta pieeja, kas izmanto dažādus prefiksus domēna un datubāzes identifikatoros. Lietojot dažādus prefiksus datubāzes konceptiem un domēna konceptiem, ir iespējams tos atšķirt, kā arī lietot vairākus objektus ar vienādu pamatnosaukumu.

Zināšanu bāzes izveidošanas posms

Zināšanu bāzes izveidošana ir metodoloģijas posms, kurā tiek izveidota ar datubāzi sasaistīta ontoloģija, kas kalpos par sistēmas zināšanu bāzi. Atbilstoši metodoloģijai ontoloģijā ir jābūt elementiem, kas atsaucas uz datubāzes elementiem. Lai izveidotu piemērotu ontoloģiju, ir nepieciešams apskatīt datubāzi un iegūt nepieciešamos ieskatus par elementiem, kuriem ir jābūt ontoloģijā. Datubāzes analīzes laikā tiek noteikti svarīgi elementi (tabulas un kolonnas), meklēti uzskaitījumliterāļi (īpašas vērtības kolonnās)

un saites starp tabulām. Šī informācija var būt pieejama vai nu no datubāzes apraksta, vai atrasta analīzes laikā. Tabulu, kolonnu un vērtību pārņemšana uz sasaistītu ontoloģiju pārsvarā ir manuāls darbs ekspertam. Analīze palīdz noteikt norādes uz elementiem, ko ir nepieciešams apskatīt ekspertam lēmumu pieņemšanai. Uzskaitījuma literāli tiek atrasti ar šādu algoritma palīdzību:

```
1. Select count (F) from T where not F is null and not F = ''
2. If count < min then not enum
3. Select count(distinct F) where not F is null and not F = ''
4. if distinct count < sqrt(count) then is enum
```

kur F – apskatāmā kolonna un T – apskatāmā tabula. Atrodot visas unikālās (netukšās) kolonnas vērtības un salīdzinot tās ar netukšu ierakstu skaitu, tiek atklāti potenciāli uzskaitījumliterāli.

Tabulu saites tiek atklātas, pārbaudot, vai visas kolonnas vērtības ir atrodamas kādas citas tabulas atslēgu kolonnā.

Pēc analīzē iegūto datubāzes objektu un vērtību apkopojuma tiek izveidots sasaistīts ontoloģijas pamats. Šis pamats tiek papildināts ar domēna konceptiem, kas ir iegūti no problēmvides domēna apskates.

Ontoloģijas izveidošana beidzas ar ontoloģijas novērtēšanu. Izveidotā ontoloģija tiek novērtēta, pārbaudot, vai visi sasaistes koncepti ir izmantoti domēna konceptu definīcijās.

Sistēmas izveidošanas posms

Sistēmas izveidošanas daļā metodoloģijā apskatīti tie paņēmieni, principi un darbības, ko nepieciešams lietot, lai realizētu OBDI sistēmu.

Izstrādātā metodoloģija ir saistīta ar pilnīgi jaunu pieeju datu vaicājuma izveidošanai, tāpēc nepieciešams izstrādāt piemērotu saskarnes veidu. Atbilstoši metodoloģijai izstrādājamai sistēmai ir jāizpilda visus ar datu izgūšanu saistītus uzdevumus, lietojot ontoloģiju kā pamatu un ontoloģijā iebūvētās spriešanas spējas.

Lai realizētu saskarni, kas nelieto vaicājumu valodu, kas ir jāapstrādā atsevišķi no ontoloģijas, tiek izvirzīta pieeja, kurā vaicājuma vietā tiek lietots vaicājuma koncepts. Vaicājuma koncepts ir īpašs klases koncepts, kas reprezentē vēlamo indivīdu jeb informācijas kopu.

Izveidotais koncepts apraksta indivīdu kopu, kam piemīt lietotāja norādītās īpašības. Izpildot ontoloģijā iebūvēto spriešanas procesu, ir iespējams noteikt visus ar vaicājuma konceptu saistītus datubāzes konceptus un veidot datu vaicājumu.

SQL vaicājuma izveidošana balstās vaicājuma koncepta analīzē. Ontoloģijā tiek atrasts vaicājuma koncepts. Vaicājuma koncepta hierarhija tiek pārmeklēta, lai atrastu ar datubāzi saistītus konceptus. Tie var būt gan datubāzes tabulu klašu koncepti, gan datubāzes kolonnu datu īpašības, kas ir izmantotas sarežģītu konceptu definīcijās. Tās

var arī būt objektu īpašības, kas norāda uz tabulu relācijām. Visi atrastie datubāzes koncepti ietekmē izveidojamo *SQL* vaicājumu.

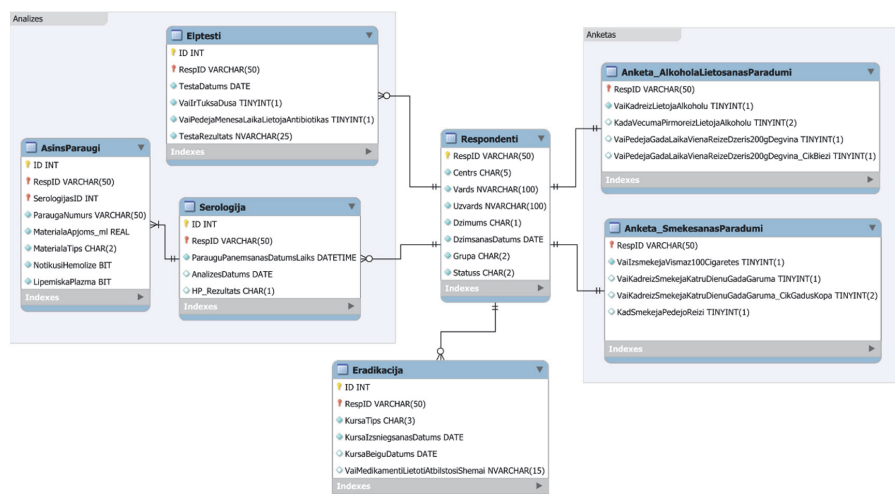
Visas aprakstītās daļas un procesus jāapvieno kopējā sistēmā. Šim nolūkam katrai sistēmas daļai tiek realizēts modulis. Visi moduļi ir spējīgi darboties patstāvīgi, atsaucoties uz kopējo ontoloģiju un izgūstot nepieciešamās zināšanas.

4. INTELEKTUĀLĀS RESPONDENTU DATU IZGŪŠANAS SISTĒMAS IZSTRĀDE

Pamatojoties uz izstrādāto metodoloģiju (skat. 3.nodaļu), tika izveidota intelektuāla medicīnas pētījuma datu izgūšanas sistēma IRDIS [38]. Šīs sistēmas mērķis ir piedāvāt medicīnas personālam (datu lietotājiem) iespēju iegūt datus par izpētāmām personām (turpmāk tekstā – respondenti). Līdz ar to sistēma kalpo par starpnieku starp datu lietotāju un eksistējošu respondentu datubāzi. Šīs sistēmas intelektuālās spējas (spriešana, zināšanu pārvalde) nodrošina ontoloģijas izmantošana. Nodaļā aprakstīta izstrādes gaita un izstrādātā sistēmas prototipa realizācija, kā arī aprobējamās sistēmas izmantošanas rezultāti.

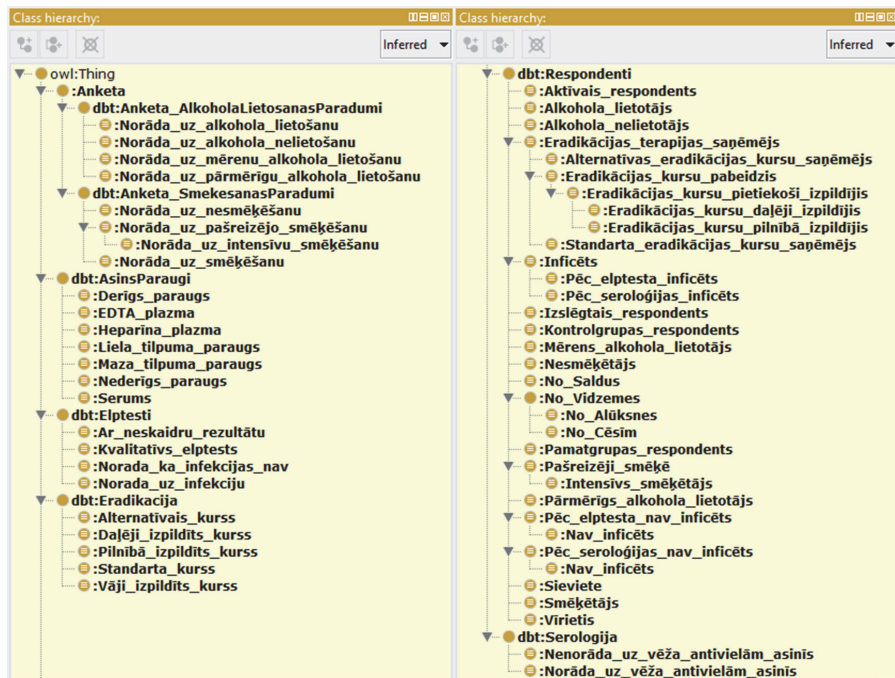
Zināšanu bāzes sagatavošana

Darbā ir apskatīta datubāze, kas glabā medicīnas datus par respondentiem. Šajā datubāzē ir 60 tabulu. Dažas no tām satur metadatus un tehnisku informāciju. Lietotājam pieeja tehniskajai informācijai nav nepieciešama. No šīm tabulām aprobācijas prototipam ir noteiktas un izvēlētas septiņas svarīgākās (skat. 3. attēlu), kas pārņemtas aprakstīšanai ontoloģijā.



3. att. Datubāzes fragmenta struktūra.

Lēmumu par izmantojamām tabulām pieņem datubāzes eksperts. Pamatojoties uz izvēlētajām tabulām, ir veikta datubāzes analīze, kuras rezultātā ir atklātas visas izmantojamās īpašās vērtības un relācijas starp tabulām. No šīs informācijas ir izveidots ontoloģijas pamats. Pēc pamata papildināšanas ir iegūta sistēmas ontoloģija, kas kalpo par izstrādātā rīka zināšanu bāzi. Izveidotās ontoloģijas klašu līmenis ir redzams 4. attēlā.



4. att. IRDIS ontoloģija.

Ontoloģijas funkciju realizācija

Izveidotās sistēmas centrā ir ontoloģija un tās spriešanas mehānismi. Lai gan ir iespējams realizēt ontoloģijas pārvaldes mehānismus, izmantojot eksistējošas ontoloģijas programmatūras bibliotēkas [58], izveidotās sistēmas ontoloģijas implementācija tika izveidota no jauna, lai vienkāršotu *SQL* vaicājuma izveidošanai nepieciešamo konceptu pārmeklējumu. Ontoloģijas pārvaldes funkcijas, kā arī pats prototips, ir izveidotas *Java* valodā. Ontoloģijas pārvaldes modulis sastāv no 53 *Java* klasēm, kas nodrošina visas ontoloģijas pamatfunkcijas, un vēl divām klasēm, kas nodrošina *OWL/XML* dokumenta ielasīšanu.

Visi ontoloģijas resursi ir iedalāmi vienā no pieciem elementu veidiem un tiek glabāti un apstrādāti attiecīgā elementa veida modulī. Tie ir klašu, objektu īpašību, datu īpašību, anotāciju īpašību un datu tipu moduļi. Klašu modulis apstrādā visas zināšanas, kas ir saistītas ar klašu konceptiem, to definīcijas un klašu hierarhiju. Datu īpašību

modulis dara to pašu, bet satur tikai datu īpašību konceptus. Objektu īpašību modulis pārvalda objektu īpašību konceptus un hierarhiju. Datu tipu modulis pārvalda ontoloģijā definētos datu tipus, kā arī standarta datu tipu konceptus un šo konceptu hierarhiju. Anotācijas īpašību modulis pārvalda gan ontoloģijā reģistrētās, gan standarta anotācijas īpašības un to hierarhiju.

Katrs ontoloģijas elementu modulis pārvalda unikālus ontoloģijas konceptu elementus. Visi ontoloģijā apstrādājami elementi (skat. 2. tabulu) ir iedalīti elementu veidos pēc funkcijas un kategorijas.

2. tabula

Elementu veidi

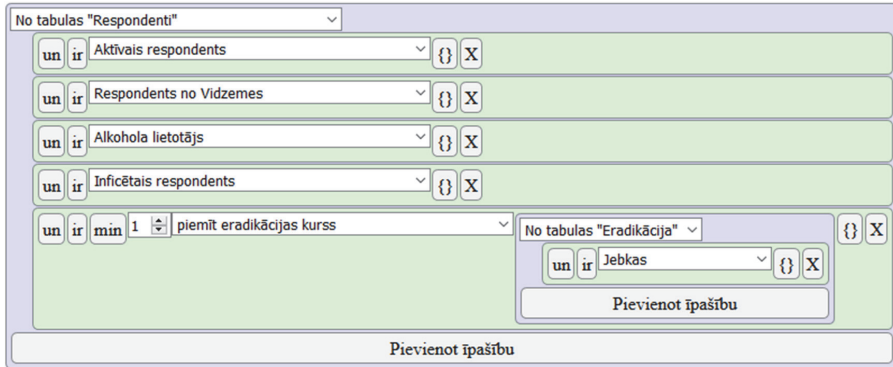
		Klase	Datu tips	Objektu īpašība	Datu īpašība	Anotāciju īpašība	
T-kastes elementa funkcija	Elements ar nosaukumu	Klase ar nosaukumu	Datu tips ar nosaukumu	Objektu īpašība ar nosaukumu	Datu īpašība ar nosaukumu	Anotāciju īpašība ar nosaukumu	
	Komplekta elements	Klases komplekts	Datu tipa komplekts	Objektu īpašības komplekts	Datu īpašības komplekts	Anotācijas īpašības komplekts	
	Elements kas satur vērtību	Indivīdu saturoša klase	Vērtību saturošs datu tips				
	Grupās elementi	Konjūncijas elements	KLases konjūncija	Datu tipu konjūncija			
		Disjūncijas elements	Klases disjūncija	Datu tipu disjūncija			
		Citi grupas elementi			Objektu īpašību ķēde		
	Citi elementi	Kardinalitātes klase		Datu tipa ierobežojums	Inversa objektu īpašība		
Klase ar atslēgu			Valodas datu tips				

Ontoloģijas spriešanas mehānismi tiek daļēji realizēti elementu hierarhijā (pievienojot jaunu elementu hierarhijai, ir iespējams izdarīt vairākus secinājumus) un daļēji spriešanas modulī, kurā eksistējošās zināšanas iteratīvi tiek papildinātas, meklējot noteiktas situācijas, kas atbilst spriešanas modulī aprakstītajiem likumiem.

Prototipa saskarne

Promocijas darba izstrādātās sistēmas saskarne ir izveidota kā vizuāls rīks, kurā definējams koncepts ir klašu kopa. Lietotājs, no saraksta izvēloties saistītās klases, papildina izveidojamo klašu kopu un veido klašu konjūnciju. Saskarnes implementācija ir redzama 5. attēlā. Tajā ir redzams, ka meklējams koncepts jeb vaicājuma koncepts tiek definēts kā citu klašu grupa (konjūncija vai disjūncija). Katrs panelis (zilā krāsā) reprezentē atsevišķu konceptu. Vislielākais panelis, kas ietver pilnīgi visu, ir definētā vaicājuma klase. Šajā gadījumā vaicājuma koncepts ir balstīts uz ontoloģijas klasi – no tabulas “Respondenti”, kas ir etiķete datubāzes konceptam, kas norāda uz datubāzes tabulu “Respondenti”. Šāds pamata koncepts ir nepieciešams,

lai ierobežotu lietotājiem izvēlnēs parādītos konceptus. Pēc tam seko klases, kas veido vaicājuma koncepta konjunkciju.

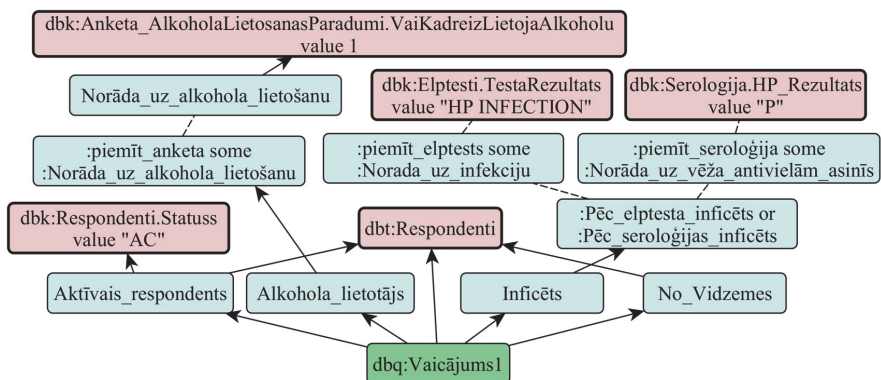


5. att. Vaicājuma izveides saskarnes logs.

Saskarnes logā redzams, ka tā ir konjunkcija, jo visās rindās sākumā ir vārds “un”. Ja uzklikšķinātu uz šī atslēgvārda, konjunkcija pārveidotos par disjunktiju, un visas rindas sāktos ar “vai”. Katrā rindā ir iespējams norādīt klases konceptu, kas ir šīs grupas (konjunkcijas vai disjunktijas) elements.

Datu izgūšana un apstrāde

Saskarnē izveidotais vaicājuma elements tiek pievienots ontoloģijai. Pēc spriešanas funkciju izpildes tas ir papildināts ar visām izsecināmām attiecībām ar citiem elementiem. Pārmeklējot visus ontoloģijā noteiktos virselementus (koncepti, kas atrodas hierarhijā augstāk), tiek atklāti visi sasaistes punkti ar datubāzi. Šī procesa piemērs ir parādīts 6. attēlā.



6. att. Vaicājuma koncepta virskoncepti.

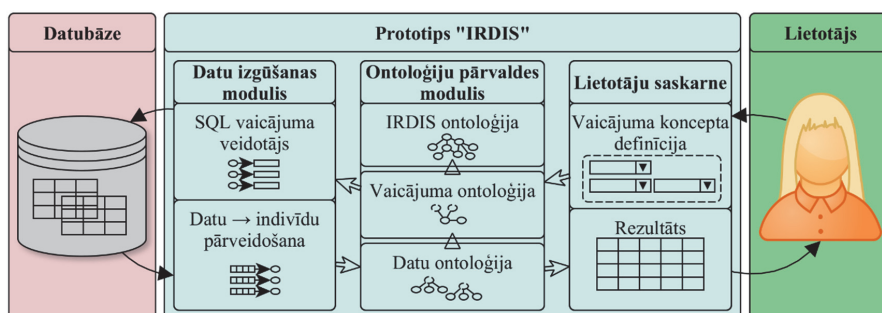
Definētais vaicājuma koncepts vienmēr ir saistīts ar vairākām eksistējošām klasēm. Pārmeklējot vaicājuma koncepta virskonceptus, tiek meklētas tās klases, kas ir

definētas kā datubāzes sasaistes punkti vai savās definīcijās izmanto sasaistītos elementus.

Apkopojot atklātos sasaistes punktus, *SQL* ģenerators izveido un izpilda vaicājumu datubāzei. Vaicājuma rezultāts tiek pievienots ontoloģijai. Katrs datubāzes ieraksts reprezentē indivīdu. Apstrādājot katru iegūto ierakstu, tiek izveidots jauns indivīds, kuram piemīt programmatiski uzģenerēts nosaukums. Šim indivīdam tiek pievienotas datu īpašības, kas sakrīt ar kolonnas nosaukumu. Ierakstu apstrādes procesā parādās indivīdi. Izpildot ontoloģijas spriešanas procesu, indivīdiem tiek noteiktas visas izsecināmās klašu piederības. Atlasot visus indivīdus, kas ir klasificēti kā piederoši vaicājuma klasei, tie un visas tām piemītošas īpašības tiek atgriezti kā vaicājuma rezultāts.

Sistēmas struktūra

Ar metodoloģijas palīdzību izstrādātās daļas kopā veido sistēmu (skat. 7. attēlu). Sistēmas centrā ir ontoloģijas pārvaldes modulis, kas atbild par zināšanu glabāšanu un lietošanu.



7. att. Sistēmas struktūra.

Visas pārējās sistēmas daļas izmanto centrālo ontoloģiju, lai realizētu savas funkcijas. Lietotāja saskarne izmanto ontoloģiju, lai piedāvātu lietotājam konceptus vaicājuma aprakstam. Datu izgūšanas modulis, kas veido *SQL* vaicājumus un ielādē datus ontoloģijā, lieto ontoloģiju, lai veidotu vaicājuma konceptam atbilstošu *SQL* vaicājumu. Pats ontoloģijas modulis veido vairāku slāņu ontoloģiju. Katrs slānis papildina iepriekšējo ontoloģijas slāni. Sākumā ontoloģijas modulis satur tikai sistēmas (IRDIS) ontoloģiju. Pēc vaicājuma koncepta izveides tiek izveidots īslaicīgs vaicājuma ontoloģijas slānis, kas satur ontoloģiju, kas atsaucas uz pamata ontoloģiju, bet papildus satur vaicājuma konceptu. Dati tiek ielādēti datu ontoloģijā, kas satur tikai datu indivīdus, bet atsaucas uz iepriekš izveidotām ontoloģijām. Kad lietotājs pārtrauc darbu, izveidotās ontoloģijas tiek atmetas, saglabājot tikai nemainīgu pamata ontoloģiju.

Izstrādātās sistēmas piemērotības noteikšana

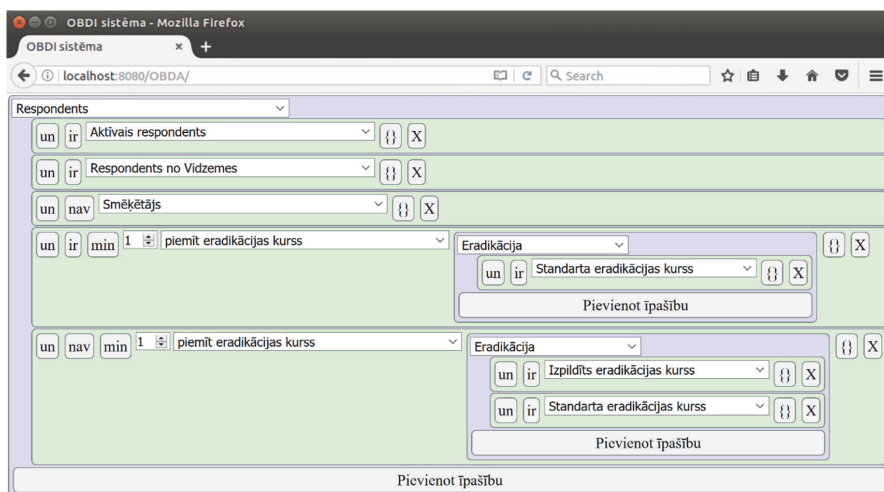
Izstrādātās intelektuālās respondentu datu izgūšanas sistēmas IRDIS funkcionalitātes izpētei un piemērotības noteikšanai paredzētas divas eksperimentu sērijas.

Pirmās eksperimentu sērijas uzdevums – noteikt izstrādātā prototipa darbības pareizību. Šajā eksperimentu sērijā tiek pārbaudīts, vai izstrādātajā ontoloģijā aprakstītās zināšanas ir pietiekamas un piemērotas datu validācijai un atļasei. Šajā eksperimentu sērijā tiek izveidoti vairāki konceptu apraksti, kuru mērķis ir pārbaudīt izsecinātās un izgūtās datu kopas. Eksperimentu izpilde parādīja, ka izstrādātā ontoloģijas funkcionalitāte un sistēma kopumā ir spējīga veidot aprakstītajiem konceptiem piemērotus *SQL* vaicājumus. Eksperimenta gaitā tika veiksmīgi izveidoti vaicājumi konceptu komplimentiem, kā arī atrastas tukšās klases.

Otrās eksperimentu sērijas nolūks – pārbaudīt, vai izveidotā sistēma ir piemērots semantiskais slānis datu izgūšanas uzdevumam. Šajā sērijā lietotāji izveido konceptu aprakstus pēc iepriekš izvirzīta uzdevuma. Vaicājuma atbilstība tiek noteikta, salīdzinot sistēmas izsniegto rezultātu ar rezultātu pēc eksperta vaicājuma. Lietošanas eksperimenti sastāv no trim daļām.

1. Lietotājs izveido eksperimenta uzdevuma nostādni. Brīvā apraksta veidā tiek izveidots uzdevums datu apakškopas izgūšanai. Tajā ir aprakstīts, kādu informāciju ir nepieciešams izgūt no datubāzes.
2. Lietotājs izveido meklējamo datu vaicājuma konceptu atbilstoši uzdevuma nostādnei. Lietojot izstrādātā prototipa saskarni, lietotājs izveido meklējamo datu konceptuālu aprakstu. Sistēma izpilda datu izgūšanu.
3. Datubāzes eksperts izveido *SQL* vaicājumu un izgūst datus salīdzināšanai ar sistēmas rezultātu.

Sarežģīta datu apraksta koncepta definīcija parādīta 8. attēlā. Ekrānuzņēmums ilustrē darbā izstrādāto lietotāja saskarni, kas ļauj definēt konceptus, izvēloties terminus no ontoloģijas.



8. att. Vaicājuma koncepta definīcija.

Piemērā parādītais koncepta apraksts definē visus aktīvos respondentus no Vidzemes, kuri nav smēķētāji un kuriem ir nepabeigts eradikācijas kurss. Šis konceptu apraksts sniedz vienādu rezultātu ar *SQL* vaicājumu, kuru salīdzinājuma nolūkiem izveidoja eksperts. Atbilstošais *SQL* vaicājums:

```
select * from Respondenti, Anketa_SmekšanasParadumi, Eradikacija
where
Respondenti.Statuss = 'AC' and
(Respondenti.Centrs = 'LV101' or Respondenti.Centrs = 'LV102') and
Anketa_SmekšanasParadumi.RespID = Respondenti.RespID and
Anketa_SmekšanasParadumi.VaiIzsmekejaVismaz100Cigaretes = 0 and
Eradikacija.RespID = Respondenti.RespID and
Eradikacija.KursaTips = 'STD' and
Eradikacija.VaiMedikamentiLietotiAtbilstosiShemai is null;
```

Eksperimentu sērija parādīja, ka izstrādātā sistēma ļauj realizēt datu izgūšanu, pārveidojot konceptuālu aprakstu par *SQL* vaicājumu un izgūstot tos pašus datus, ko izgūst eksperts. Eksperimentos izgūtais ierakstu skaits ir parādīts 3. tabulā. Ceturtā eksperimentā spriešanās izgūtais rezultāts atšķiras no eksperta rezultāta, jo lietotājs nepilnīgi aprakstīja meklējamo konceptu. Pēc koncepta pilnveidošanas rezultāti bija vienādi.

3. tabula

Eksperimentos izgūto ierakstu skaits

	Eksperiments			
	1	2	3	4
Sistēma	3521	702	673	0 / 434
Eksperts	3521	702	673	434

Nodaļā apskatīta izveidotās metodoloģijas izmantošana, izveidojot intelektuālu respondentu datu izgūšanas sistēmu. Izstrādātā pieeja datubāzes analīzei ir lietota, lai iegūtu ieskatu datos, analīzes rezultātu izmantojot par informācijas avotu ontoloģijas izstrādei. Apskatītās un izpildītās darbības ļauj iegūt informāciju par datubāzes saturu, kas dod iespēju izveidot datus bāzētus konceptus. Izveidotā sistēma ir apbēta un pārbaudīta ar vairākiem eksperimentiem, kas parādīja tās atbilstību izvīrztajam uzdevumam.

REZULTĀTI UN SECINĀJUMI

Darbā izpētītas ontoloģijas izmantošanas iespējas piekļūt datubāzei un izgūt no tās datus, izpētītas ar to saistītās tehnoloģijas un eksistējošie risinājumi. Šīs analīzes rezultātā ir izstrādāta metodoloģija, kas apraksta pieejas un metodes ontoloģijā bāzētu intelektuālu datu izgūšanas sistēmu izstrādei. Sistēma vērsta uz uzlabotu lietojamību, lai nodrošinātu semantisko slāni datu izgūšanas uzdevumam. Metodoloģija piedāvā

pieeju ontoloģijas izveidošanai, kas ir sasaistīta ar eksistējošu relāciju datubāzi. Lai izveidotu sistēmai nepieciešamo sasaistīto ontoloģiju, metodoloģija apraksta datubāzes analīzes pieejas. Šīs pieejas izmantošanas rezultātā tiek iegūta ontoloģija, kas satur datu savienojamības elementus. Realizējot metodoloģijā aprakstītās ontoloģijas spriešanas iespējas par ontoloģijā iekļautiem datu elementiem, ir iespējams izmantot zināšanas par sasaistītās datubāzes struktūru. Šādā veidā zināšanas par domēnu un zināšanas par datiem ir savstarpēji papildinošas. Zināšanas par datiem un datubāzes struktūru palīdz fokusēt domēna zināšanas un aprakstīt tikai to, kas ir izgūstams no datubāzes. Domēna zināšanas papildina datubāzes datus ar piederībām konceptuālām domēnu klasēm. Šāda pieeja atvieglo ontoloģijā bāzētās datu izgūšanas uzdevuma tehnisko daļu. Lai intelektuālā datu izgūšana būtu pieejamāka gala lietotājam, metodoloģijā piedāvāta pieeja lietotāja saskarnes izveidei. Viena no ontoloģijas priekšrocībām ir tās spēja veicināt komunikāciju un savstarpējo sapratni. Tomēr, ja šīs iespējas tiek paslēptas aiz sarežģītām tehnoloģijām, kas tās mistificē, lietojamība un cilvēka izpratne tiek pazaudēta. Lai ontoloģijā bāzēta datu izgūšana būtu lietojama arī tiem, kas nav ontoloģiju eksperti, metodoloģijā tiek piedāvāta pieeja semantiskā slāņa nodrošināšanai. Semantiskais slānis tiek nodrošināts, realizējot lietotāja saskarni, kurā ir piedāvāti domēna koncepti. Izmantojot domēna konceptus, lietotājs definē jaunu konceptu, kam piemīt noteiktas īpašības. Lietotājs ir spējīgs aprakstīt vēlamu datu klasi, lietojot viņam pazīstamu terminoloģiju bez zināšanām par datu glabāšanā iesaistītajām tehnoloģijām un datu struktūrām.

Darba izstrādes un pētījuma gaitā sasniegtie rezultāti

- Izpētītas ar ontoloģijām un ontoloģijā bāzētu datu izgūšanu saistītas tehnoloģijas. Tas deva ieskatu ontoloģijas iespējās, tās spriešanas funkcijās un lietojumā, ļāva izveidot trūkstošās pieejas ontoloģijā bāzētai sasaistei, datubāzes konceptu identificēšanai un ontoloģijā bāzētai *SQL* vaicājumu izveidošanai. Pamatojoties uz *OWL 2* valodas analīzi, ir izstrādāta *Java* programmatūras bibliotēka, kas realizē ontoloģijas pārvaldi un spriešanas funkcijas.
- Izpētīti un salīdzināti esošie intelektuālās datu izgūšanas programmlīdzekļi un to uzbūves pieejas. Atklāti trūkumi datubāzes un ontoloģijas sasaistē un sistēmu lietojumā, kas tiešā veidā ietekmē metodoloģijas pamatus. Šīs analīzes rezultāts ir eksistējošu rīku salīdzinājums un tajos identificēto trūkumu apkopojums. Uz analīzes pamata ir izstrādātas metodoloģijas prasības datu izgūšanas sistēmas izveidei.

- Izstrādātas metodes un pieejas, kas ļauj realizēt intelektuālu datu izgūšanu ar ontoloģijas palīdzību. Izstrādāta ontoloģijā bāzēta datubāzes un ontoloģijas sasaistes pieeja, veids, kā izveidot *SQL* vaicājumus no sasaistes konceptiem un ontoloģijā izpildāma datubāzes ierakstu klasifikācijas pieeja.
- Pamatojoties uz starprezultātiem, sasniegts promocijas darba galvenais rezultāts – izstrādāta ontoloģijā bāzētas datu izgūšanas sistēmu realizācijas metodoloģija. Metodoloģija apkopo datubāzes analīzes metodiku, ontoloģijas izveidošanas metodiku, sistēmas realizācijas metodiku, kā arī sasaistes un konceptu identificēšanas pieejas.
- Metodoloģijas aprobācijas nolūkam ir izstrādāts intelektuālas respondentu datu izgūšanas sistēmas (IRDIS) prototips, kas realizēts, vadoties pēc izstrādātās metodoloģijas. Prototipa aprobācijā tas uzrādīja visas metodoloģijā aprakstītas īpašības, izgūstot datus no saistītās relāciju datubāzes vienkāršotā veidā.

Promocijas darba izstrādes un pētījuma gaitā, pamatojoties uz iegūtajiem rezultātiem un sasniegumiem, ir izdarīti šādi galvenie secinājumi.

- Izstrādātais prototips parādīja, ka ontoloģijas iespējas bez papildu tehnoloģijām ļauj realizēt datubāzes sasaisti, datu apakškopas definēšanu, datu validāciju un papildu zināšanu atklāšanu par klasēm un indivīdiem. Izstrādātajā prototipā nav lietotas ontoloģiju papildinošas tehnoloģijas (sasaistes dokuments vai *SPARQL* vaicājumu interpretators). Visas funkcijas ir realizētas tikai ar ontoloģiju, kas apstiprina pirmo hipotēzi.
- Izstrādātajā sistēmas prototipā *SQL* vaicājumi nepieciešamo datu izgūšanai tiek veidoti, lietojot ontoloģijā ieliktais zināšanas par datubāzes tabulām, kolonnām, relācijām un vērtībām. Izstrādātajā prototipā ir iespējams lietot ontoloģijas spriešanas funkcijas un konceptu hierarhiju pārmeklēšanu, lai izgūtu pietiekamu informāciju *SQL* vaicājuma izveidošanai. Tas apstiprina otro hipotēzi.
- IRDIS prototipa lietotājiem, veidojot *ad hoc* datu pieprasījumu, nebija jāsaskaras ar datubāzes terminiem vai tehnoloģijām. Prototipā piedāvātā saskarne dod lietotājam iespēju definēt vēlamo datu klasi, lietojot tikai domēna terminoloģiju. Lietotājam nav jāzina par datubāzes tehnoloģijām, datubāzes struktūru vai lietotām īpašām vērtībām. Lietotājam arī nav nepieciešams zināt semantiskās tehnoloģijas un nav jāspēj veidot *SPARQL* vaicājumus. Līdz ar to apstiprinās trešā hipotēze.

- Izmantojot vaicājuma konceptu un ontoloģijas spriešanu, ir iespējams izpildīt datu validēšanu, nevēršoties pie datubāzes. Ontoloģijā aprakstītais modelis apvienojumā ar ontoloģijas spriešanas iespējām ir spējīgs noteikt neizpildāmus datu vaicājumus modeļa līmenī bez nepieciešamības vērsties pie datubāzes. Kļūdainu vaicājumu filtrēšana sistēmas līmenī ļauj minimizēt datubāzes noslodzi.

Izstrādātā metodoloģija apraksta ontoloģijā bāzētas datu izgūšanas sistēmas realizāciju, tās nolūks ir papildināt eksistējošas relāciju datubāzes ar semantisko informāciju, intelektu (ontoloģijas zināšanu veidā) un nodrošināt semantisko slāni starp lietotājiem un datubāzes tehnoloģijām. Nākotnes pētījumi var būt vizīti uz ontoloģijas iespēju pielāgošanu tieši darbam ar datubāzēm. Tas ļautu racionalizēt zināšanu pārvaldi un spriešanas procesus saistībā ar relāciju datubāžu pieejām datu glabāšanai, novēršot ontoloģijai sarežģītus realizācijas momentus.

IZMANTOTĀS LITERATŪRAS SARAKSTS

1. Akama, K., Nantajeewarawat, E. Solving Query-answering Problems with If-and-Only-If Formulas. In: *Proceedings of the 6th Intern. Conf. on Knowledge Engineering and Ontology Development – KEOD 2014, (IC3K 2014), Italy, Rome, 21–24 October 2014*. Vol. 1. Rome: 2014, pp. 333–344.
2. An, Y., Topaloglou, T. Maintaining Semantic Mappings between Database Schemas and Ontologies. In: *Semantic Web, Ontologies and Databases: VLDB Workshop, SWDB-ODSIS 2007, 24 September, 2007, Vienna, Austria*. Springer Intern. Publ., 2008, pp. 138–152. (LNCS 5005).
3. Astrova, I., Kalja, A., Jaeger, E., et al. Storing OWL Ontologies in SQL 3 Object-relational Databases. In: *AIC'08 Proceedings of the 8th Conf. on Applied Informatics and Communications, Rhodes, Greece, 20–22 August, 2008*. Wisconsin, USA: 2008, pp. 99–103.
4. Astrova, I., Nahum, K., Ahto, K. Storing OWL Ontologies in SQL Relational Databases. *Int. J. of Electrical, Computer and Systems Engineering*. 2007, vol. 1, No. 4, pp. 242–247.
5. Baader, F., Borgwardt, S., Lippmann, M. Temporal Query Entailment in the Description Logic SHQ. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*. 2015, vol. 33, pp. 71–93.
6. Barzdins J., Grasmanis M., Rencis E., et al. Ad-Hoc Querying of Semistar Data Ontologies Using Controlled Natural Language. In: *Databases and Information Systems IX: Selected Papers from the 12th Intern. Baltic Conf., DB&IS 2016*. Amsterdam: IOS Press, 2016, pp. 3–16.
7. Barzdins, J., Grasmanis, M., Rencis, E., et al. Self-service Ad-hoc Querying Using Controlled Natural Language. In: *Databases and Information Systems: 12th Int. Baltic Conf., DB&IS 2016, Latvia, Riga, 4–6 July, 2016. Proceedings*. Springer Int. Publ., 2016, pp. 18–34.
8. Barzdins, J., Grasmanis, M., Rencis, E., et al. Towards a More Effective Hospital: Helping Health Professionals to Learn from their Own Practice by Developing an Easy to use Clinical Processes Querying Language. *Procedia Computer Science*. 2016, vol. 100, pp. 498–506.
9. Barzdins, J., Rencis, E., Sostaks, A. Fast Ad Hoc Queries Based on Data Ontologies. In: *Databases and Information Systems VIII: Selected Papers from the 11th Intern. Baltic Conf., DB&IS 2014*. Amsterdam: IOS Press, 2014, pp. 43–56.
10. Bechhofer, S. *OWL Reasoning Examples* [online]. UK, University of Manchester, 2003 [Accessed December 2017]. Available from: <http://owl.man.ac.uk/2003/why/latest/>.
11. Belghiat, A., Bourahla, M. Automatic Generation of OWL Ontologies from UML Class Diagrams Based on MetaModelling and Graph Grammars. *Int. J. of Computer and Information Engineering*, 2012, vol. 6, No. 8, pp. 967–972.
12. Benazzouz, T., Echchtabi, A., Charkaoui, A. Using Ontology as a Decision Support System for Manage Risks in Medicines Supply Chain: Case of Public Hospitals in Morocco. In: *Proceedings of 7th Int. Conf. on Industrial Engineering and Operations Management, Rabat, Morocco, 11–13 April, 2017*. IOEM Society, 2017, pp. 281–292.

13. Benharzallah, S., Kazar, O., Caplat, G. Intelligent Query Processing for Semantic Mediation of Information Systems. *Egyptian Informatics Journal*. 2011, vol. 12, No. 3, pp. 151–163.
14. Berners-Lee, T. *Semantic Web on XML*. The World Wide Web Consortium, 2000. Available from: <https://www.w3.org/2000/Talks/1206-xml2k-tbl/all.htm>
15. Berners-Lee, T., Fielding, R., Masinter, L. *Uniform Resource Identifier (URI): Generic Syntax*. The Internet Society, 2005. Available from: <https://www.ietf.org/rfc/rfc3986.txt>.
16. Borisovs, A., Kuļešova, G., Zmanovska, T. *Introduction to Ontology Engineering*. Riga: RTU Press, 2014. 196 p.
17. Bumans, G. Mapping between Relational Databases and OWL Ontologies: An Example. *Computer Science and Information Technologies: Scientific Papers University of Latvia*. 2010, vol. 756, pp. 99–117.
18. Calvanese, D., Cogrel, B., Komla-Ebri, S., et al. Ontop: Answering SPARQL Queries over Relational Databases. *Semantic Web*. 2017, vol. 8, no. 3, pp. 471–487.
19. Calvanese, D., Liuzzo, P., Mosca, A., et al. Ontology-based Data Integration in EPNet: Production and Distribution of Food during the Roman Empire. *Engineering Applications of Artificial Intelligence*. 2016, vol. 51, p. 212–229.
20. Cameron, J. D., Ramaprasad, A., Syn, T. An ontology of and roadmap for mHealth research. *Int. J. of Medical Informatics*. 2017, vol. 100, pp. 16–25.
21. Cerans K., Barzdins J., Sostaks A., u.c., Extended UML Class Diagram Constructs for Visual SPARQL Queries in ViziQuer/web. In: *Proceedings of the 3rd Int. Workshop on Visualization and Interaction for Ontologies and Linked Data (VOILA '17)*, Vienna, Austria. *CEUR Workshop Proceedings*, vol. 1947, pp. 87–98.
22. Čerāns, K., Būmans, G. Database to Ontology Mapping Patterns in RDB2OWL Lite. In: *Databases and Information Systems: 12th Int. Baltic Conf., DB&IS 2016, Riga, Latvia, 4–6 July 2016*. Switzerland: Springer Int. Publ., 2016, pp. 366–380.
23. Čerāns, K., Būmans, G. RDB2OWL: A Language and Tool for Database to Ontology Mapping. In: *Databases and Information Systems VI: Selected Papers from the 9th Int. Baltic Conf., DB&IS 2010, Riga, Latvia, 5–7 July, 2010*. Amsterdam: IOS Press, 2011, pp. 139–152.
24. Chuprina, S., Postanogov, I., Nasraoui, O. Ontology Based Data Access Methods to Teach Students to Transform Traditional Information Systems and Simplify Decision Making Process. *Procedia Computer Science*. 2016, vol. 80, pp. 1801–1811.
25. Collins, A.M., Quillian, M.R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*. 1969, vol. 8, No. 2, pp. 240–247.
26. Corcho, O., Gómez-Pérez, A. A Roadmap to Ontology Specification Languages. In: *Knowledge Engineering and Knowledge Management Methods, Models, and Tools: 12th Int. Conf. Proceedings, EKAW 2000, Juan-les-Pins, France, 2–6 October, 2000*. Berlin Heidelberg: Springer, 2000, pp. 80–96.
27. Cormen, T.H., Leiserson, C.E., Rivest, R.L., et al. *Introduction to Algorithms*. 3rd ed. MIT Press, 2009.

28. Cudré-Mauroux, P., Enchev, I., Fundatureanu, S., et al. NoSQL Databases for RDF: An Empirical Evaluation. In: *The Semantic Web–ISWC, Int. Semantic Web Conf. ISWC-13, Australia, Sydney, 21–25 October 2013*. Part 2. Berlin Heidelberg: Springer–Verlag, 2013, pp. 310–325.
29. Daramola, O., Adigun, M., Ayo, C. Building an Ontology-Based Framework for Tourism Recommendation Services. In: *Information and Communication Technologies in Tourism 2009: Proceedings of the Int. Conf. in Amsterdam, The Netherlands, 2009*. Wien: Springer, 2009, pp. 135–147.
30. del Mar Roldán García M., García-Nieto J., Aldana-Montes J.F., An ontology-based Data Integration Approach for Web Analytics in e-commerce. *Expert Systems with Applications*. 2016, vol. 63, pp. 20–34.
31. Dieng-Kuntz, R., Minier, D., Růžička, M., et al. Building and Using a Medical Ontology for Knowledge Management and Cooperative Work in a Health Care Network. *Computers in Biology and Medicine*. 2006, vol. 36, No. 7–8, pp. 871–892.
32. El Idrissi, B., Bařna, S., Bařna, K. Ontology Learning from Relational Database: How to Label the Relationships Between Concepts? In: *Beyond Databases, Architectures and Structures: 11th Int. Conf., BDAS 2015, Ustroř, Poland, 26–29, May 2015, Proceedings*. Springer, 2015, pp. 235–244.
33. Fragkou, P., Galiotou, E., Matsakas, M. Enriching the e-GIF Ontology for an Improved Application of Linking Data Technologies to Greek Open Government Data. *Procedia – Social and Behavioral Sciences*. 2014, vol. 147, pp. 167–174.
34. Gali, A., Chen, C. X., Claypool, K. T., et al. From Ontology to Relational Databases. In: S. Wang et al. (eds.) *ER Workshops 2004*. LNCS, vol. 3289. Heidelberg: Springer, 2004, pp. 278–289.
35. Gorskis, H. Aleksejeva, L., Pořaka, I. Database Concepts in a Domain Ontology. *Information Technology and Management Science*. 2017, vol. 20, pp. 69–73.
36. Gorskis, H. Improved Database Schema Development for OWL2. *Information Technology and Management Science*. 2016, vol. 19, pp. 85–91.
37. Gorskis, H., Aleksejeva, L., Pořaka, I. Database Analysis for Ontology Learning. *Procedia Computer Science*. 2016, vol. 102, pp. 113–120.
38. Gorskis, H., Aleksejeva, L., Pořaka, I. Ontology-Based System Development for Medical Database Access. In: *Environment. Technology. Resources : Proceedings of the 11th Int. Scientific and Practical Conf. Vol. 2, Latvia, Rezekne, 15–17 June, 2017*. Rezekne: Rezekne Academy of Technologies, 2017, pp. 24–29.
39. Gorskis, H., Borisovs, A. Knowledge-based System Design Based on Generic Method Conception. *Information Technology and Management Science*. 2014, vol. 17, pp. 55–60.
40. Gorskis, H., Borisovs, A. Location based Reminder System with Reusable Ontology. In: *KEOD 2014: Proceedings of the Int. Conf. on Knowledge Engineering and Ontology Development, Italy, Rome, 21–24 October, 2014*. [S.l.]: SciTePress, 2014, pp. 161–167.
41. Gorskis, H., Borisovs, A. Ontology Building Using Classification Rules and Discovered Concepts. *Information Technology and Management Science*. 2015, vol. 18, pp. 37–41.

42. Gorskis, H., Borisovs, A. Reusable Components in Knowledge-based Configuration Design Systems. *Information Technology and Management Science*. 2013, vol. 16, pp. 66–72.
43. Gorskis, H., Borisovs, A. Storing an OWL 2 Ontology in a Relational Database Structure. In: *Environment. Technology. Resources: Proceedings of the 10th Int. Scientific and Practical Conf., Latvia, Rezekne, 18–20 June, 2015*. Rezekne: Rezekne Higher Education Institution, 2015, pp. 71–75.
44. Gorskis, H., Borisovs, A., Aleksejeva, L. Genetic Algorithm Based Random Selection-Rule Creation for Ontology Building. In: *Recent Advances in Soft Computing: Proceedings of the 22nd Int. Conf. Soft Computing (MENDEL 2016). Czech Republic, Brno, 8–10 June, 2016*. Cham: Springer Int. Publ. AG, 2017, pp. 34–45.
45. Gorskis, H., Čižovs, J. Ontology Building Using Data Mining Techniques. *Information Technology and Management Science*. 2012, vol. 15, pp. 183–188.
46. Grabusts, P., Borisov, A., Aleksejeva, L. Decision Tree Creation Methodology Using Propositionalized Attributes. *Information Technology and Management Science*. 2016, vol. 19, pp. 34–38.
47. Grabusts, P., Borisov, A., Aleksejeva, L. Ontology-Based Classification System Development Methodology. *Information Technology and Management Science*. 2015, vol. 18, pp. 129–134.
48. Gruber, T.R. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*. 1993, vol. 5, pp. 199–220.
49. Grundspenkis J., Towards the Formal Method for Evaluation of Concept Map Complexity from the Systems Viewpoint. In: *Databases and Information Systems IX: Selected Papers from the 12th Int. Baltic Conf., DB&IS 2016, Riga, Latvia, 4–6 July 2016*. Amsterdam: IOS Press, 2016, pp. 341–354.
50. Grundspenkis, J. Initial Steps Towards the Development of Formal Method for Evaluation of Concept Map Complexity from the Systems Viewpoint. In: *Databases and Information Systems: 12th Int. Baltic Conf., DB&IS 2016, Riga, Latvia, 4–6 July 2016*. Springer, 2016, pp. 366–380.
51. Grundspenkis, J. Knowledge Creation Supported by Intelligent Knowledge Assessment System. In: *Proceedings of the 12th World Multi-Conf. on Systemics, Cybernetics and Informatics, Orlando, Florida, USA, 29 June–2 July, 2007*. Vol. VII. IIS, 2008, pp. 135–140.
52. Gundla, N.K., Chen, Z. Creating NoSQL Biological Databases with Ontologies for Query Relaxation. *Procedia Computer Science*. 2016, vol. 91, pp. 460–469.
53. Hastings, J., Magka, D., Batchelor, C., et al. Structure-based classification and ontology in chemistry. *Journal of Cheminformatics*. 2012, vol. 4, No. 8. Available from: doi: 10.1186/1758-2946-4-8
54. Hazber, M.A.G., Li, R., Gu, X., et al. Integration Mapping Rules: Transforming Relational Database to Semantic Web Ontology. *Applied Mathematics & Information Sciences*. 2016, vol. 10, No. 3, pp. 881–901.
55. Henss, J., Kleb, J., Grimm, S., et al. A Database Backend for OWL. In: *Proceedings of the 5th International Workshop on OWL: Experiences and Directions (OWLED 2009), Chantilly, Virginia, USA, 23–24 October 2009*. Vol. 529. Washington: CEUR-WS, 2009. Available from: <http://ceur-ws.org/Vol-529/>

56. Hitzler, P., Krötzsch, M., Parsia, B., et al. *OWL 2 Web Ontology Language Primer* [Online]. W3C, 2009 [Accessed 15 March, 2015]. Available from: <http://www.w3.org/TR/owl2-primer/>
57. Ho, L.T.T., Tran, C.P.T., Hoang, Q. An Approach of Transforming Ontologies into Relational Databases. In: *Intelligent Information and Database Systems: 7th Asian Conf., ACIIDS 2015, Bali, Indonesia, 23–25 March, 2015*. Springer, 2015, pp. 149–158.
58. Horridge, M., Bechhofer, S. The OWL API: A Java API for OWL ontologies. *Semantic Web*. 2011, vol. 2 (1), pp. 11–21.
59. Kalibatiene, D., Vasilecas, O. Survey on Ontology Languages. In: *Perspectives in Business Informatics Research: Proceedings of the 10th Int. Conf., BIR 2011, Riga, Latvia, 6–8 October, 2011*. Berlin: Springer, 2011, pp. 124–141.
60. Kang, D., Kim, M. Propositionalized Attribute Taxonomies from Data for Data-driven Construction of Concise Classifiers. *Expert Systems with Applications*. 2011, vol. 38, No. 10, pp. 12739–12746.
61. Kapłański, P., Seganti, A., Cieśliński, K., et al. Automated Reasoning Based User Interface. *Expert Systems with Applications*. 2017, vol. 71, pp. 125–137.
62. Kassahun, Y., Perrone, R., De Momi, E., et al. Automatic Classification of Epilepsy Types Using Ontology-based and Genetics-based Machine Learning. *Artificial Intelligence in Medicine*. 2014, vol. 61, No. 2, pp. 79–88.
63. Katib, A., Slavov, V., Rao, P. RIQ: Fast processing of SPARQL queries on RDF quadruples. *Web Semantics: Science, Services and Agents on the World Wide Web*. 2016, vol. 37–38, pp. 90–111.
64. Kaulins, A., Borisovs, A. Building Ontology from Relational Database. *Information Technology and Management Science*. 2014, vol. 17, pp. 45–49.
65. Kaulins, A., Borisovs, A. Learning Ontology from Object-Relational Database. *Information Technology and Management Science*. 2015, vol. 18, pp. 78–83.
66. Kharbat, F., Ghalayini, H. New Algorithm for Building Ontology from Existing Rules: A Case Study. In: *2009 Int. Conf. on Information Management and Engineering (ICIME '09), Kuala-Lumpur, Malaysia, 3–5 April, 2009*. IEEE, 2009, pp. 12–16. Available from: DOI: 10.1109/ICIME.2009.16.
67. Kharlamov, E., Hovland, D., Skjæveland, M., et al. Ontology Based Data Access in Statoil. *Web Semantics: Science, Services and Agents on the World Wide Web*. 2017, vol. 44, IN PRINT.
68. Kirshners, A., Parshutin, S., Borisov, A. Combining clustering and a decision tree classifier in a forecasting task. *Automatic Control and Computer Science*. 2010, vol. 44, pp. 124–132.
69. Kiršners, A., Paršutins, S., Gorskis, H. Entropy-Based Classifier Enhancement to Handle Imbalanced Class Problem. *Procedia Computer Science*. 2017, vol. 104, pp. 586–591.
70. Knorr, M., Alferes, J.J., Hitzler, P. Local closed world reasoning with description logics under the well-founded semantics. *Artificial Intelligence*. 2011, vol. 175, No. 9, pp. 1528–1554.
71. Konstantinou, N., Spanos, D., Mitrou, N., Ontology and Database Mapping: A Survey of Current Implementations and Future Directions. *Journal of Web Engineering*. 2008, vol. 7, No. 1, pp. 1–24.

72. Kontchakov, R., Lutz, C., Toman, D., et al. The Combined Approach to Ontology-Based Data Access. In: *Proceedings of the 22nd Int. Joint Conf. on Artificial Intelligence IJCAI, Barcelona, Catalonia, Spain, 16–22 July, 2011*. Menlo Park, California: AAAI Press, 2011, pp. 2656–2661.
73. Kostylev, E.V., Reutter, J.L. Complexity of Answering Counting Aggregate Queries over DL-Lite. *Web Semantics: Science, Services and Agents on the World Wide Web*. 2015, vol. 33, pp. 94–111.
74. Kumar, A., Yip, Y.L., Smith, B., et al. Bridging the Gap between Medical and Bioinformatics: An Ontological Case Study in Colon Carcinoma. *Computers in Biology and Medicine*. 2006, vol. 36, No. 7–8, pp. 694–711.
75. Kuo, Y., Lonie, A., Sonenberg, L., et al. Domain Ontology Driven Data Mining: A Medical Case Study. In: *Proceedings of the 2007 Int. workshop on Domain driven data mining (DDDM '07), San Jose, CA, USA, 12–15 August, 2007*. New York: ACM, 2007, pp. 11–17.
76. Kwon, O., Choi, S., Park, G., NAMA: a context-aware multi-agent based web service approach to proactive need identification for personalized reminder systems. *Expert Systems with Applications*. 2005, vol. 29 (1), pp. 17–32.
77. Lambrini, S., Achilles, K. Biomedical Engineering through Ontologies. In: *Proceedings of the 6th Int. Conf. on Knowledge Engineering and Ontology Development – KEOD 2014, (IC3K 2014), Italy, Rome, 21–24 October 2014*. Vol. 1. Rome: 2014, pp. 240–247.
78. Leinweber, A., Freiberg, M., Spence, P., et al. Enterprise Ontologies: Open Issues and the State of Research – A Systematic Literature Review. In: *Proceedings of the 6th Int. Conf. on Knowledge Engineering and Ontology Development – KEOD 2014, (IC3K 2014), Italy, Rome, 21–24 October 2014*. Vol. 1. Rome: 2014, pp. 280–287.
79. Lembo, D., Lenzerini, M., Rosati, R., et al., Inconsistency-tolerant query answering in ontology-based data access. *Web Semantics: Science, Services and Agents on the World Wide Web*. 2015, vol. 33, pp. 3–29.
80. Li, M., Du, X.Y., Wang, S. Learning Ontology from Relational Database. In: *Proceedings of 4th Int. Conf. on Machine Learning and Cybernetics, ICMLC 2005, Guangzhou, China, 18-21 August, 2005*. Vol. 6, pp. 3410–3415.
81. Martínez-Cruz, C., Noguera, J.M., Vila, M.A. Flexible Queries on Relational Databases using Fuzzy Logic and Ontologies. *Information Sciences*. 2016, vol. 366, pp. 150–164.
82. Mecca, G., Rull, G., Santoro, D., et al. Ontology-based mappings. *Data & Knowledge Engineering*. 2015, vol. 98, pp. 8–29.
83. Motik, B., Horrocks, I., Sattler, U. Bridging the Gap Between OWL and Relational Databases. *Web Semantics: Science, Services and Agents on the World Wide Web*. 2009, vol. 7, No. 2, pp. 74–89.
84. Munir, K., Anjum, M. S. The Use of Ontologies for Effective Knowledge Modelling and Information Retrieval. *Applied Computing and Informatics*. 2017. Available from: DOI: 10.1016/j.aci.2017.07.003
85. Nicola, A., Missikoff, M., Navigli, R. A Software Engineering Approach to Ontology Building. *Information Systems*. 2009, vol. 34, No. 2, pp. 258–275.

86. Nikolaou, C., Koubarakis, M. Querying Incomplete Information in RDF with SPARQL. *Artificial Intelligence*. 2016, vol. 237, pp. 138–171.
87. Noy, N. F., McGuinness, D. L. *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Tech. Report KSL-01-05 and Stanford Medical Informatics Tech. Report SMI-2001-0880, March 2001.
88. *Optique Project: Scalable End-user Access to Big Data* [online]. Norway, 2014. [Accessed December 2017]. Available from: <http://optique-project.eu/>.
89. Ouyang, D. T., Cui, X. J., Ye, Y. X. Mapping Integrity Constraint Ontology to Relational Databases. *The Journal of China Universities of Posts and Telecommunications*. 2010, vol. 17 (6), pp. 113–121.
90. OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax. 2nd ed. [Online]. W3C, 2009 [Accessed 15 March, 2015]. Available from: <https://www.w3.org/TR/owl2-syntax/>.
91. Paganelli, F., Giuli, D. An Ontology-based Context Model for Home Health Monitoring and Alerting in Chronic Patient Care Networks. In: *Advanced Information Networking and Applications Workshops, 2007, AINAW'07. Proceedings of 21st Int. Conf., Canada, Ontario, 21–23 May, 2007*. Vol. 2. IEEE, 2007. pp. 838–845.
92. *RDF Primer* [Online]. W3C, 2004 [Accessed 15 March, 2015]. Available from: <https://www.w3.org/TR/rdf-primer/>.
93. *RDF Schema 1.1* [Online]. W3C, 2004-2014 [Accessed 15 March, 2015]. Available from: <https://www.w3.org/TR/rdf-schema/>.
94. Redmond, T. An Open Source Database Backend for the OWL API and Protege 4. In: *Proceedings of the 7th Int. Workshop on OWL: Experiences and Directions (OWLED 2010), San Francisco, California, USA, 21–22 June, 2010*. Vol. 614. Washington: CEUR-WS, 2010. Available from: <http://ceur-ws.org/Vol-614/>
95. Research Programme SOPHIS [online]. Riga, University of Latvia, 2014 [Accessed 29 December 2017]. Available from: <http://sophis.edi.lv/>.
96. Rodríguez-Muro, M., Rezk, M. Efficient SPARQL-to-SQL with R2RML mappings. *Web Semantics: Science, Services and Agents on the World Wide Web*. 2015, vol. 33, pp. 141–169.
97. Rothenfluh, T. E., Gennari, J. H., Eriksson, H., et al. Reusable Ontologies, Knowledge-Acquisition Tools, and Performance Systems: Protégé-II Solutions to Sisyphus-2. *Int. J. of Human-Computer Studies*. 1994, vol. 44 (3–4), pp. 303–332.
98. Russell, S., Norvig, P. *Artificial Intelligence: A Modern Approach*. 3rd ed. Pearson, 2009.
99. Saïs, F., Thomopoulos, R. Ontology-aware Prediction from Rules: A Reconciliation-based Approach. *Knowledge-Based Systems*. 2014, vol. 67, pp. 117–130.
100. Santoso, H. A., Haw, S. C., Abdul-Mehdi, Z. T. Ontology Extraction from Relational Database: Concept Hierarchy as Background Knowledge. *Knowledge-Based Systems*. 2011, vol. 24 (3), pp. 457–464.
101. Sigov, A., Baranyuk, V., Nechaev, V., et al. Approach for Forming the Bionic Ontology. *Procedia Computer Science*. 2017, vol. 103, pp. 495–498.

102. Sir, M., Bradac, Z., Fiedler, P. Ontology versus Database. *IFAC-PapersOnLine*, 2015, vol. 48, No. 4, pp. 220–225.
103. Skjæveland, M.G., Giese, M., Hovland, D., et al. Engineering Ontology-based Access to Real-world Data Sources. *Web Semantics: Science, Services and Agents on the World Wide Web*. 2015, vol. 33, pp. 112–140.
104. Smirnov, A., Levashova, T., Shilov, N. Knowledge Fusion in Context-Aware Decision Support Systems. In: *KEOD 2014: Proceedings of the Int. Conf. on Knowledge Engineering and Ontology Development, Italy, Rome, 21–24 October, 2014*. [S.l.]: SciTePress, 2014, pp. 186–194.
105. *SPARQL Query Language for RDF* [Online]. W3C, 2004-2014 [Accessed 15 March, 2015]. Available from: <https://www.w3.org/TR/rdf-sparql-query/>.
106. Stair, R., Reynolds, G. *Fundamentals of Information Systems*. 6th ed. Cengage Learning, Inc., 2011.
107. Ta, C. D. C., Thi, T. P. Improving the Algorithm for Mapping of OWL to Relational Database Schema. In: *Machine Learning and Data Mining in Pattern Recognition: 11th Int. Conf., MLDM 2015, Germany, Hamburg, 20–21 July, 2015*. Springer, 2015, pp. 130–139.
108. Tarus, J. K., Niu, Z., Yousif, A. A Hybrid Knowledge-based Recommender System for e-learning based on Ontology and Sequential Pattern Mining. *Future Generation Computer Systems*. 2017, vol. 72, pp. 37–48.
109. Thomopoulos, R., Destercke, S., Charnomordic, B., et al. An Iterative Approach to Build Relevant Ontology-aware Data-driven Models. *Information Sciences*. 2013, vol. 221, pp. 452–472.
110. Uschold, M., Gruninger, M. Ontologies: Principles, Methods and Applications. *Knowledge Engineering Review*. 1996, vol. 11, No. 2, pp. 93–136.
111. Vega-Gorgojo, G., Slaughter, L., Giese, M., et al. Visual Query Interfaces for Semantic Datasets: An Evaluation Study. *Web Semantics: Science, Services and Agents on the World Wide Web*. 2016, vol. 39, pp. 81–96.
112. Vicknair, C., Macias, M., Zhao, Z., et al. A Comparison of a Graph Database and a Relational Database: A Data Provenance Perspective. In: *Proceedings of the 48th Annual Southeast Regional Conference, Oxford, MS, USA, 15–17 April, 2010*. New York: ACM, 2010. Available from: DOI: 10.1145/1900008.1900067
113. *VizQuer/web Tool for RDF Data Analysis Queries* [Online]. Riga: Institute of Mathematics and Computer Science, University of Latvia, 2004 [Accessed 15 March, 2015]. Available from <http://viziquer.lumii.lv/>.
114. Voigt, M., Mitschick, A., Schulz, J. Yet Another Triple Store Benchmark? Practical Experiences with Real-World Data. In: *Proceedings of the 2nd Int. Workshop on Semantic Digital Archives (SDA 2012), Cyprus, Paphos, September 2012*, pp. 85–94.
115. Vysniauskas, E., Nemuraite, L. Mapping of OWL Ontology Concepts to RDB Schemas. In: *Information Technologies' 2009: Proceedings of the 15th Int. Conf. on Information and Software Technologies, Lithuania, Kaunas, 23–24 April, 2009*. Kaunas: Kaunas University of Technology, 2009, pp. 317–327.
116. Vysniauskas, E., Nemuraite, L. Transforming Ontology Representation from OWL to Relational Database. *Information Technology and Control*. 2015, vol. 35 (3), pp. 333–343.

117. Wache, H., Voegele, T., Visser, U., et al. Ontology-based Integration of Information – A Survey of Existing Approaches. In: *Proceedings of the IJCAI-01 workshop on Ontologies and Information Sharing, USA, Seattle, 4–6 August, 2001*, pp. 108–117.
118. Wimmer, H., Rada, R. Good versus Bad Knowledge: Ontology Guided Evolutionary Algorithms. *Expert Systems with Applications*. 2015, vol. 42, No. 21, pp. 8039–8051.
119. Zarembo, I. Automatic Transformation of Relational Database Schema into OWL Ontologies. In: *Environment. Technology. Resources: Proceedings of the 10th Int. Scientific and Practical Conf., Latvia, Rezekne, 18–20 June, 2015*. Vol. 3. Rezekne: Rezekne Higher Education Institution, 2015, pp. 217–222.
120. Zhang, F., Ma, Z. M., Li, W. Storing OWL Ontologies in Object-oriented Databases. *Knowledge-Based Systems*. 2015, vol. 76, pp. 240–255.
121. Zhang, L., Li, J. Automatic Generation of Ontology Based on Database. *Journal of Computational Information Systems*. 2011, vol. 7 (4), pp. 1148–1154.



Henrihs Gorskis ir dzimis 1987. gadā Rīgā. 2011. gadā ieguvis inženierzinātņu akadēmisko bakalaura grādu informācijas tehnoloģijā Rīgas Tehniskajā universitātē (RTU). 2013. gadā ar izcilību absolvējis maģistratūru un ieguvis inženierzinātņu akadēmisko maģistra grādu informācijas tehnoloģijā RTU.

H. Gorskis ir RTU Datorzinātnes un informācijas tehnoloģijas fakultātes Informācijas tehnoloģijas institūta Modelēšanas un imitācijas katedras pētnieks. Pētījuma virziens – ontoloģijas, datizrace, lēmumu pieņemšanas atbalsts un klasifikācijas algoritmi.