LATVIJAS UNIVERSITĀTE

DATORIKAS FAKULTĀTE

STRUKTURĒTU DATU GRAFU VIZUALIZĀCIJA VIRTUĀLĀ REALITĀTĒ KĀ RĪKS LAI ASISTĒTU NOZIEGUMU DIGITĀLĀ IZMEKLĒŠANĀ

MAĢISTRA DARBS

Autors: **Ronalds Rodiks**
Stud. Apl. Nr.: rr07009

Darba vadītājs: Dr. Dat. Leo Seļāvo

RĪGA 2022

ANOTĀCIJA

Maģistra darbā „Strukturētu datu grafu vizualizācija virtuālā realitātē kā rīks lai asistētu noziegumu digitālā izmeklēšanā” tiek apskatīti pētījumi par datu apstrādes metodēm digitālā izmeklēšanā un datu attēlošanas paņēmieniem. Darba mērķis ir, balstoties uz pētījumu analīzi, izstrādāt reālus un funkcionējošus, divus savstarpēji atkarīgus, rīku prototipus, kas veiktu kriminālprocesa laikā izgūto datu apstrādi, izveidojot starp tām saistības, vienā un veiktu skaidru un saprotamu apstrādāto datu attēlošanu otrā.

Atslēgvārdi: Digitālā izmeklēšana, Datu apstrāde, Datu attēlošana.

ABSTRACT

The master's thesis "Visualization of structured data graphs in virtual reality as a tool to assist in digital forensics" deals with research on data processing methods in digital forensics and data portrayal techniques. The aim of the thesis is to develop two interdependent, real and functional prototypes that would process the data obtained during criminal proceedings by creating links between the data in one tool and to make a clear and understandable visual representation of the processed data in the other.

Keywords: Digital forensics, Data processing, Data visualization.

AUTOREFERĀTS

Darba tapšanas laikā tika apskatīti vairāki savstarpēji atšķirīgi risinājumi problēmjautājumiem, kas saistīti ar datu apstrādi digitālā izmeklēšanā. Tika apskatīti dažādie datu attēlošanas paņēmieni. Lai optimizētu risinājumu, tika ievākta informācija no izmeklētājiem par tiem interesējošo izmeklēšanas laikā, apskatāmo datņu raksturīpašībām.

Analizējot pētījumu, kas saistīti ar datu apstrādi, piedāvātos risinājumus, tika secināts, ka tajos piedāvātie rīki ir galēji produkti, kas paredz padziļinātu iesaisti no izmeklētājiem, lai radītu sekmīgus rezultātus. Balstoties uz izmeklētāju atgriezenisko saiti, tika izstrādāts risinājums, kas prasītu pēc iespējas mazāku iesaisti no izmeklētājiem, atvieglojot viņu darbu, un būtu kā papildus līdzeklis izmeklēšanas procesā.

Balstoties uz iegūtās informācijas, tika izstrādāti divi savstarpēji atkarīgi rīki datu apstrādei un attēlošanai. Izstrādes procesa laikā, risinājums tika testēts uz vairāku kriminālprocesu laikā izgūtām spogūļkopijām. Testēšanas un atgriezeniskās saites rezultātā tapa rīku prototipu gala versija, kas tiek prezentēta šajā maģistra darbā.

Gala risinājums tika prezentēts 23 izmeklētājiem. 18 izmeklētājiem šis risinājums bija viegli uztverams un uzskatīts par lietderīgu. 15 izmeklētājiem šajā risinājumā trūka papildus iesaistes un iedarbības ar vidē attēlojamiem datiem. 5 izmeklētājiem risinājums nelikās saprotams.

SATURS

1. IEVADS

Mūsdienās digitālā izmeklēšana ir daudznozaru darbs, kas aptver vairākas jomas, tostarp tiesību zinātnes, datorzinātnes, finanšu zinātnes, tīklu veidošanu, datizraci un krimināltiesības. Izmeklētāji un analītiķi arvien biežāk saskaras ar dažādiem izaicinājumiem un problēmām saistībā ar digitālo pierādījumu apstrādes efektivitāti un tā lielo apstrādājamo apjomu.

1.1. Darba mērķis

Problēmjautājumi, kuri šī darba ietvaros tiek aplūkoti, tiek risināti ar mērķi, lai samazinātu lielo izmeklētāju un analītiķu darbu, kas tiek ieguldīts, lai izveidotu pierādījumu ķēdi kriminālllietās.

Pirmais šī darba mērķis ir, balstoties uz pētījumu par datu rekonstrukciju analīzi, izstrādāt risinājuma prototipu, kas spētu automātiski rekonstruēt un kategorizēt datora notikumus un to plūsmu, atsaucoties uz pašām datnēm, kas tajos ir iesaistīti un, vienlaikus, nodrošināt izmeklētājiem un analītiķiem iespēju analizēt konstruētos un sasaistītos datus faktiski neielūkojoties pašās datnēs, kas šīs notikumu ķēdes ir uzbūvējuši.

Otrais šī darba mērķis ir, balstoties uz darbā analizētām datu attēlošanas pieejām, atrast risinājumu – kā, izmantojot datu attēlošanu, var padarīt sākotnēji nesakārtotus, neskaidrus un abstraktus datus par strukturētu informāciju, kas būtu intuitīvi un viegli uztverama, un, izveidot šāda rīka prototipu.

1.2. Darba Struktūra

Darba otrajā nodaļā ir izklāstīts īss pārskats par digitālās izmeklēšanas vēsturi. Trešajā nodaļā tiek aplūkoti problēmjautājumi un iepriekš veikto pētījumu risinājumu priekšlikumi. Darba ceturtajā nodaļā tiek salīdzināti daži no pieejamiem rīkiem un to lietojumu metodēm. Piektajā nodaļā tiek apskatītas datu attēlošanas metodes un to analīze. Sestajā nodaļā tiek aprakstīts un piedāvāts risinājums balstoties uz esošo rīku metodoloģiju un pieejām. Septītajā nodaļā tiek veikts izstrādātā rīka apraksts.

2. DIGITĀLĀ IZMEKLĒŠANA

Mūsdienās digitālā izmeklēšana ir svarīgs rīks, lai atrisinātu noziegumus, kas izdarīti izmantojot datorus vai viedierīces, piemēram, banku krāpšana, pikšķerēšana u.c., kā arī tādu noziegumu atrisināšanai, kur apsūdzēto datorā vai viedierīcē var atrasties pierādījumi piemēram, naudas atmazgāšana, bērnu izmantošana u.c. Programmatūras, kas tiek izmantotas kā rīki izmeklēšanā ir kļuvušas par svarīgu informācijas nodrošināšanas rīku, jo tie spēj rekonstruēt atstātos pierādījumus, piemēram, kiberuzbrukumu gadījumos u.c., taču, tā kā mūsdienās vienā ierīcē vienā minūtē operētājsistēma un dažādas programmatūras ražo, kā arī reģistrē simtiem procesu [1], digitālā izmeklēšanas jomā ir radies jauns šķērslis, kuru nepieciešams risināt, proti, kā efektīvi un ātri šos datus apstrādāt.

Augstāk minētā tendence digitālās izmeklēšanas jomā, digitālo notikumu identificēšanā un, vispārīgāk, to rekonstrukcija, un datu plūsmas izsekošana ir kļuvusi par lielu slogu ar ko saskaras izmeklētāji un analītiķi tikai salīdzinoši nesen. Šīs nodaļas apakšpunktos tiks izklāstīts īss pārskats par digitālās izmeklēšanas vēsturi ar mērķi pamatot, kādēļ problēmjautājumi saistībā ar digitālo izmeklēšanu ir radušās tieši pēdējo desmit gadu laikā, kā arī nedaudz tiks minētas būtiskākās tehnoloģiskās iezīmes vēstures posmos.

2.1. Ieskats digitālās izmeklēšanas vēsturē

Digitālai izmeklēšanai kā nozarei ir aptuveni, vairāk kā 50 gadu. Vēsturiski, digitālā izmeklēšana ir atvasinājusies no 70-tajos gados veiktajiem datu atgūšanas paņēmieniem. Savukārt, 80-to gadu beigās datu atgūšanai pievienojās citas metodes, kas bija datu atformatēšana, dzēsto datu atjaunošana un cieto disku diagnostika. [1]

Autori M. Polits (*Mark Pollitt*), K. Čovs (*Kam-Pui Chow*) un S. Šenojs (*Sujeet Shenoi*) savā darbā, [2] kā raksturīgākās iezīmes šim divdesmit gadu sākumposmam izdalīja – 1) aparatūras, programmatūras un lietojumprogrammu daudzveidību; 2) datu formātu izplatību, daudzi no kuriem nebija dokumentēti; 3) lielu atkarību no centralizētām ar datortehniku aprīkotām iestādēm, jo reti kad lietotāju vai apsūdzēto mājās atradās liela datu glabātuve, kurai bija nepieciešams veikt analīzi; 4) formāla procesa vadlīniju, rīku un apmācību trūkums.

Šajā laikā, pašu digitālo izmeklēšanu, galvenokārt, veica datoru speciālisti, kuri strādāja kopā ar tiesībsargājošām valsts institūcijām, pieaicinot augstāk minētos speciālistus katrā atsevišķā izmeklēšanas procesā, kurā tie bija nepieciešami. [2] Turklāt, nepieciešamība veikt digitālo izmeklēšanu bija salīdzinoši maza. Piemēram, saskaņā ar autora S. L. Garfinkela (*Simson L.*

Garfinkel) izklāstīto [2], pierādījumus, kuri tika atrasti uz laika sadales sistēmām, bieži vien varēja atgūt neizmantojot specializētus rīkus. Pie tam, šajā laikā datu glabāšanai izmantoja pēc izmēra mazus datu nesējus, līdz ar to noziedznieki, lielākoties, izdrukāja visu informāciju tādejādi, ļoti maz izmeklēšanas procesos bija nepieciešamība pēc digitālo mediju analīzes. Savukārt, paša datoru uzlaušana neskaitījās kā noziegums līdz 1984. gadam, jeb līdz brīdim, kad ASV tika izdots pirmais likumdošanas projekts, kas sevī iekļāva nelikumīgas darbības saistībā ar datoriem. [2]

Pēc autora S. L. Garfinkela domām, gadi no 1999 līdz 2007 bija sava veida digitālās izmeklēšanas "*zelta laikmets*". [1] Savā darbā viņš aprakstīja, ka šajā laikā digitālā izmeklēšana kļuva par logu, caur kuru var atskatīties pagātnē - ar to domājot, ka ar digitālās izmeklēšanas rīkiem un paņēmieniem varēja atjaunot datus, kuri, domājams, bija sen izdzēsti, kā arī kā logu noziedznieka prātā - ar to domājot, e-pastos un terzētavās saņemto ziņojumu atjaunošanu. Savukārt, tīklu un datora atmiņas izmeklēšana padarīja iespējamu novērot nozieguma vēsturi soli pa solim, pat ja pēc nozieguma ir pagājis ilgs laiks. Šajā laikā digitālā izmeklēšana kļuva tik plaši izplatīta un uzticama, ka par to veidoja TV raidījumus (*CSI*) [1].

Autori M. Polits un K. Čovs savā darbā [2], kā raksturīgākās iezīmes šim posmam norādīja – 1) plašo *Microsoft Windows*, it sevišķi *Windows XP* izmantošanu; 2) pastāvēja salīdzinoši maz digitālo izmeklēšanu interesējošo datņu - galvenokārt *Microsoft Office* dokumenti, *JPEG* digitālās fotogrāfijas un *AVI* un *WMV* video; 3) datu analīze lielākoties aprobežojās ar vienu datorsistēmu, kas piederēja izmeklēšanā esošam subjektam; 4) atmiņas ierīces lielākoties bija aprīkotas ar standarta interfeisiem (*IDE/ATA*), kas pievienotas izmantojot noņemamus kabeļus un savienotājus, kas nostiprināti ar skrūvēm, respektīvi – to atrašanās vieta un pieslēguma savienojumi bija standartizēti; 5) tirgū vairāki uzņēmumi piedāvāja rīkus, kas bija pietiekami labi, lai atgūtu dzēstos datus bez specializētam zināšanām.

2.2. Digitālā izmeklēšana mūsdienās

Savā darbā [1] S. Garfunkels apgalvo, ka mūsdienās liela daļa no progresa, kas tika veikta pēdējo divdesmit gadu laikā - ātri kļūst nenozīmīga un ka digitālā izmeklēšana ir saskārusies ar krīzi. S. Garfunkels to pamato, ka iepriekš praktizētās metodes ir mazāk efektīvas vai pat ir nepielietojamas datoru industrijas straujo tehnoloģisko sasniegumu un secīgo pārmaiņu dēļ. [1] Pieaugošais lielo datu apjoms ierīču diskos nozīmē, ka izmeklētājiem bieži nepietiek laiks, lai izveidotu ierīces datu nesēja spoguļkopijas vai arī apstrādāt to datus.

Operētājsistēmu un datņu formātu daudzveidība ir ievērojami palielinājusies, un to apstrādei ir nepieciešami sarežģīti rīki, kuru izstrādes izmaksas ir ievērojami pieaugušas. Pirms tam kriminālprocesos iesaistītā digitālā izmeklēšana bieži vien aprobežojās ar vienas ierīces analīzi, taču mūsdienās lielākoties ir nepieciešams to veikt vairākām ierīcēm, starp kurām, papildus ir jāveic atrasto pierādījumu korelācija. Arvien biežāk dati tiek šifrēti, kas nozīme, ka pat tad, ja tos var atgūt, labākajā gadījumā, ir nepieciešams ilgs laiks, pirms tos var apstrādāt, vai arī tos vispār nevar apstrādāt. Mākoņpakalpojumu izmantošana nozīmē, ka bieži vien datu struktūra un atrašanās vieta ir sadalīta, kas nozīmē, ka iegūstot ierīci, pilnīgu datu apjomu pat nevar iegūt.

Ņemot vērā, ka ir izskaidrots kādi bija apstākļi digitālā izmeklēšanas nozarē 50-to gadu mijā, un kādas ir problēmas, ar ko digitālā izmeklēšana saskaras mūsdienās vispārēji, tālāk tiks padziļināti izskaitīti datu apstrādes problēmjautājumi, kas ir saistīti ar pašu datu izgūšanu un to attēlošanu.

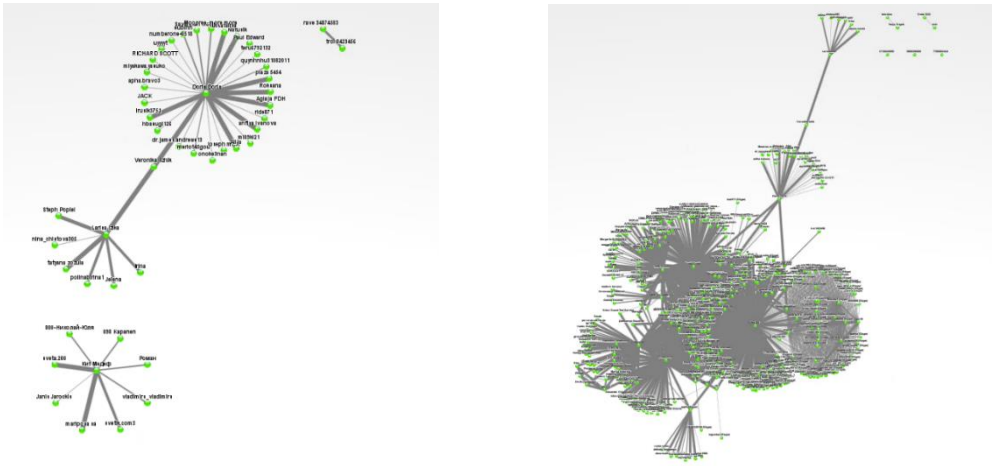
3. DATU APSTRĀDES PROBLĒMJAUTĀJUMI

Datu plūsmu un datoru procesu rekonstrukcija no datņu informācijas ir sarežģīts process, lai to būtu vieglāk atrisināt savā darbā [3] par datora notikumu rekonstrukciju autori J. Čabots (*Yoan Chabot*), A. Bertauksa (*Aurélie Bertaux*), K. Nikola (*Christophe Nicolle*) un T. Kečadi (*Tahar Kechadi*) izvirzīja galvenos problēmjautājumus, kuru atrisinājumus ir nepieciešams iestrādāt risinājumā, lai tas būtu vērtīgs ieguldījums digitālā izmeklēšanā. Pirmais problēmjautājums ir lielais datu apjoms. Otrais problēmjautājums ir minēto datu neviendabīgums. Pēdējais problēmjautājums ir likumiskās prasības, kuras jāievēro apstrādājot datus. Problēmjautājumi un to risinājumu priekšlikumi tiek aplūkoti turpmākās šīs sadaļas apakšnodaļās.

3.1. Datu apjoma problēma

Pārskatot vairākus pieejamos zinātniskos darbus saistībā ar datu izgūšanu, kas nepieciešama digitālai izmeklēšanai, tika konstatēts, ka tajos ir minēti dažādi informācijas iegūšanas rīki, no kuriem izplatītākie ir *Plaso toolbox, Cyberfonsic TimeLab, Vound Intella , FTK Imager* un *Cellebrite*. No iepriekš minētiem rīkiem, *Plaso toolbox* un *Cyberforensic TimeLab* izveido datu laika joslu no cieto disku spoguļkopijām, *Vound Intella* un *Cellebrite,* izveido datu laika joslu un klasificē datus, pamatojoties uz to paplašinājumiem, un *FTK Imager* no izveidotās spoguļkopijas eksportē cietā diska saturu kā sākotnējo failu sistēmu.

Atspoguļojot dažus gigabaitus datu, izmantojot iepriekš minētos rīkus, piemēram, *Windows* operētājsistēmām, tiek identificēti tūkstošiem procesu no dažādiem avotiem, sākot ar *Apache* žurnāla (*.log*) datņu datiem, operētājsistēmas žurnāla, pārlūkprogrammu vēstures žurnāliem utt. kā rezultātā no izgūtās informācijas tiek izveidota milzīga datu bāze ar izolētiem datiem. Šāda veida risinājumi nerekonstruē to skaidrā un intuitīvā veidā, jo dati tiek sadalīti un ierobežoti ar attiecīgo klasifikāciju datu bāzē. Piemēram, zemāk attēlā Nr.1 parādīti *Vound Intella* izveidotie laika joslu grafiki pie samērā maziem datu apjomiem.



Attēls Nr.1 **Laika josla no 335 857 datnēm (2GB) un 1 154 935 datnēm (5GB)**

Visu sākotnējā paragrāfā minēto rīku (izņemot, *FTK Imager*) apstrādes metodoloģija ir datu glabāšana datu bāzēs, kas sastāv no tabulām, kurās atrodas vispārīga tehniskā informācija, un informācija par konkrēto datni. Šī datu apstrādes metodoloģija tiek dēvēta par notikumu korelāciju digitālā izmeklēšanā (*Event correlation for forensics*) jeb *ECF*. [4] Autori K.Džeims (*Christopher James*) un H.Dž. Patersons (*Hargreaves Jonathan Patterson*) savā darbā [5] raksturoja, ka tāda tipa datu bāzes struktūras ir domātas, lai attēlotu datus kanoniskā formā, respektīvi - neatkarīgu no avota izcelsmes, un, vienlaicīgi, informācijai piešķirot divus līmeņus - vispārējs informācijas līmenis un specializētais jeb tehniskās informācijas līmenis. Augstākminētie autori papildus norādīja, ka, lai gan šī konceptuālā nodalīšana ļauj ieviest iepriekš minēto kanonisko notikumu sadalīšanu, kas, savukārt, atvieglo informācijas apstrādi, vienlaikus saglabājot datu jeb šajā gadījumā - datoru procesu un datu specifiku - tajā pašā laikā, šo datu bāzes izmantošana neļauj gala lietotājam, šajā gadījumā, izmeklētājiem un analītiķiem, pilnībā izmantot šo iespēju, jo tā skaidri neatspoguļo datu semantiku, tādejādi ierobežojot izpratni par šiem datiem. [5]

Savā darbā J.Čabots, A. Beatriksa, K. Nikola, T. Kečadi piedāvā alternatīvu metodoloģiju *ECF*, kas ir kvalitatīvo notikumu digitālā izmeklēšana (*Forensics of Rich Events*) jeb *FORE* [3], kas, kā iepriekš minētie autori skaidroja, piedāvā korelācijas izveidošanas rīku balstoties uz apstākļiem, kas identificē cēloņskarības starp datiem. Piemēram, dati A izraisa notikumu B, ja dati A kā notikums noticis pirms B. Tomēr, iepriekš minētie autori norāda, ka neskatoties uz šāda rīka lietderību, tas pielieto nosacījumu sistēmu, kura, attiecīgi, pieprasa no lietotāja manuālus iepriekšdefinētus nosacījumus. Tā kā izmeklētāji un analītiķi nevar paredzēt visus nepieciešamos

notikumus, piemēram, tie nevar paredzēt informāciju, kas tiem nav zināma, nezināmu nosacījumu ievadīšana būtu, visdrīzāk, neiespējams uzdevums.

Turklāt, pat ja visi faktori būtu zināmi, nosacījumu ievade joprojām būtu ļoti ilgstošs uzdevums. Tādejādi, iepriekšminēto autoru darbos, kas bija koncertēti uz datu laika joslu rekonstrukciju un datu attēlošanu tieši digitālās izmeklēšanas jomā, bija vienprātība par to, ka notikumu un datu plūsmu rekonstrukcijai ir jābūt izstrādātai tādā veidā, ka automatizēti atlases nosacījumi būtu spējīgi pielāgoties jebkurai situācijai, vai vismaz tuvu šim mērķim, it sevišķi, situācijās, kad izmeklētājiem un analītiķiem nav skaidras izpratnes par to, ko no šiem datiem vajadzētu attēlot, t.i. nav iepriekš noteiktas izpratnes par to, kas izmeklētājiem un analītiķiem ir jāatrod datu sistēmās. Līdz ar to, rīks, kurš varētu risināt iepriekš minēto, ir jāizstrādā tā, lai tas nepaļautos uz lietotāja definētiem nosacījumiem.

Ņemot vērā minēto, J.Čabots, A. Beatriksa, K. Nikola, T. Kečadi secina, ka ir nepieciešams, lai digitālo notikumu un datu plūsmu rekonstrukcijas, kā arī to analīzei ir jānotiek automātiski. [6] Turklāt automātisko rīku koncepcijā, ir nepieciešams, ka dati, kas nolasīti no ierīcēm, ir saprotami un nolasāmi līdz ar to, tiem jābūt strukturētiem, formāliem un ar skaidri definētu semantiku. [6] Tādejādi ir nepieciešams izvērtēt pieejamo datu struktūras līmeņus kopā ar mehānismiem, kas veicina automatizācijas lietošanas ērtumu. Tajā pašā laikā ir nepieciešams izcelt svarīgāku informāciju, kurai lietotāja saskarnē ir jābūt viegli redzamai un pieejamai izmeklētājiem un analītiķiem. Turklāt, digitāliem notikumiem un datu plūsmai ir jābūt viegli lasāmai, filtrējamai un apkopojamai ar uzreiz redzamām korelācijām starp datiem un notikumiem t.i. gala rezultātam jāproducē redzamu informāciju no kuras izmeklētāji un analītiķi varētu izdarīt secinājumus. Vienlaikus, izstrādātajam modelim jānodrošina viegla piekļuve uzrādīto datu avotam t.i. izcelsmes datnēm, no kurām ir iegūta informācija, jābūt viegli pieejamai caur skata modeļa tipa lietotāja interfeisu, kurā informāciju var nolasīt saprotamā un intuitīvā veidā. [6]

3.2. Datu neviendabīguma problēma

Saistībā ar datu neviendabīgumu autori K.Džeims un H.Dž. Patersons savā darbā [5] norādīja, ka parasti izmeklēšanas laikā tiek lietoti dažādi datu avoti, piemēram, pārlūkvēsture, operētājsistēmu notikumu žurnāli, datņu sistēmas, dažādas lietojumprogrammas, kas satur dokumentu žurnālus, e-pasti, saglabātās datnes (tajā skaitā mediju datnes, *MS Office* vai *Open libre* vai citas līdzīgi producētas datnes) utml. Lai nepalaistu garām informāciju un gūt precīzu priekšstatu par izmeklētājiem un analītiķiem interesējošiem notikumiem vai datu plūsmu,

augstākminētie autori uzskata, ka rekonstrukciju metodēm ir jābūt spējīgām iegūt informāciju no visiem avotiem un attiecīgi tos apstrādāt.

Dažādās metodes un rīki, kas tika minēti iepriekš, jau spēj tikt galā ar lielāko daļu neviendabīgiem informācijas avotiem. Autori J.Čabots, A. Beatriksa, K. Nikola, T. Kečadi savā darbā [3] norādīja, ka šo neviendabīgo avotu apstrādei ir nepieciešams automatizēts risinājums, kas varētu izgūt informāciju no tiem, būtu specifisks katram avotam un vienlaikus spētu izveidot pilnīgu informācijas modeli, kas attēlotu minētos datus kā daļu no visiem avotiem, kas atrodas datu plūsmā vai notikumā.

Lai atrisinātu neviendabīguma problēmu, iepriekš minētie autori savā darbā [3] nodefinēja nosacījumus kuriem ir jāatbilst izstrādātajam risinājumam. Informāciju veidojošiem datiem ir jābūt pilnīgiem, lai tie precīzi varētu atspoguļot datu plūsmu vai notikumus t.i. modelim ir jābūt pietiekami attīstītai vārdnīcai, lai tas var korekti attēlot datus ar to semantiskām sasaistēm, to īpašības un savstarpējās attiecības. Ir jāievieš tādi mehānismi kā datu dalītājus (*data parsers*), kas spēj apstrādāt neviendabīgus avotus.

3.3. Datu likumiskās prasības

Lai dati tiktu uzskatīti par pierādījumiem, tiem jābūt iegūtiem atbilstoši likumiskām prasībām. Pēdējais problēmjautājums, kuru jāatrisina gala rezultātā, ir juridiskās prasības, kas jāievēro jebkurā digitālās izmeklēšanas laikā. Rezultātiem, kas iegūti, izmantojot jebkuru metodoloģiju, ir jāatbilst noteiktām prasībām. K.Džeims un H.Dž. Patersons savā darbā [5]apkopoja šīs prasības un tās uzskaitīja - rezultātu ticamība, datu integritāte un to reproducējamība. Minētie autori norāda, ka šie kritēriji ir svarīgi, jo nodrošina, ka risinājuma iznākums ir pieņemams tiesā. Autori K.Džeims un H.Dž. Patersons savā darbā [5] norādīja, ka, lai izpildītu likumiskās prasības, datu rekonstrukcijas metodoloģijai ir jāatbilst noteiktiem punktiem, kas uzskaitīti zemāk.

Risinājumā informācija ir jāintegrē tā, lai to varētu izsekot. Piemēram, pierādījumu ticamības līmenis ir tieši saistīts ar informācijas precizitāti un metodi, kas tika lietota, lai to iegūtu, kā arī konkrētā izmantotā avota precizitāti un ticamību, vai šajā gadījumā pierādījumu. Tādejādi, detalizējot izmeklēšanas procesu, ar avota modelēšanas metodoloģiju, no datiem iegūtā informācija var būtiski palielināt pierādījumu ticamību. Savukārt, datu reproducējamība ļauj neatkarīgām iestādēm atkārtot datu iegūšanas procesu ar tādiem pašiem rezultātiem, un izdarīt attiecīgos secinājumus, vienlaikus izprotot izmeklētāju dedukcijas, kā arī, vienlaikus, izvērtēt pašu

pierādījumu kvalitāti, lai tiesneši varētu viegli izsekot un izvērtēt pierādījumus tiesā. Ņemot vērā iepriekš minēto, autori secina, ka ir jābūt modelim, kurš ir balstīts uz formālu teoriju. [5]

4. DATU REKONSTRUKCIJU PIEEJU PIEMĒRI

Šajā nodaļā ir izklāstīts par pētnieku izveidotiem modeļiem iepriekšminēto digitālās izmeklēšanas problēmjautājumu atrisināšanai. Lai gan tālāk minētās pieejas ir tikai daļa no vairākām, konkrētie modeļi ir izvēlēti, pamatojoties uz to atšķirīgām idejām par to, kā var panākt datu un likumsakarību klasifikāciju. Nodaļas beigās visas minētās metodes tika apkopotas un ir veikti secinājumi, uz kuriem daļēji balstoties tiek piedāvāts risinājums.

4.1. Pētījumu par datu rekonstrukciju atlase

Pētījumi tika apvienoti pēc to risinājuma metodoloģijas atsevišķās šīs nodaļas apakšnodaļās. Minētie pētījumi tika atlasīti posmā no 2021. gada oktobra līdz 2022. gada martam. Pētījumu atlasei salikumos tika izmantoti atslēgvārdi: digitālā izmeklēšana (*digital forensics*), izmeklēšana (*forensics*), datu analīze (*data analysis*), digitālo notikumu rekonstruēšana (*digital event reconstruction*). Platformas, kurās tika veikta atlase ir *Google Scholar, ResearchGate* un *Nature*.

4.2. Uz Beiesa tīkliem balstīta pieeja

Viena no populārākajām metodikām, kuru izmanto digitālo noziegumu izmeklēšanā ir Beiesa (*Bayesian*) tīklu modeļi. [7] Autori M. Kvāns, K. Čojs (*K. Chow*) [7] u.c. savā darbā [8] skaidroja, ka Beiesa tīkli tiek izmantoti, lai kvantificētu pierādījumu stiprās puses, tādā veidā uzlabojot pierādījumu, šajā gadījumā, atrasto un izvadīto rezultātu ticamību, to izsekojamību un palīdz pamatot to argumentāciju kā pierādījuma pievienošanu pie krimināllietas vai kriminālprocesa.

Praksē, Beiesa tīklu pielietojuma metodoloģijas lietderīgumu autori M. Kvāns, K. Čojs u.c. sīkāk raksturoja, [8] ka Beiesa teorēmas kopā ar grafu teoriju pielietojums nodrošina veidu kā raksturot cēloņsakarības starp mainīgajiem, kas digitālās izmeklēšanas kontekstā ir pierādījumu kopa, kas izdalīta konkrētās atrastās datnes izmeklējamā elektroniskajā ierīcē. [8] Iepriekš minētie autori skaidroja, ka veidojot Beiesa tīklu, cēloņu struktūra un noteiktās varbūtību vērtības izriet no vairākiem eksperimentiem un ekspertu viedokļiem, respektīvi - paši izmeklētāji izvēlās katra pierādījuma, jeb atsevišķo datņu nozīmīgumu, pamatojoties uz izmeklēto krimināllietu kopumu.

[7] Ņemot vērā iepriekš minēto, tika secināts, ka sarežģījumi, kuri rodas konstruējot šādu Beiesa tīklu ir veidot skaidru priekšstatu par pierādījumu svarīgumu, balstoties uz izmeklētāju viedokļiem t.i., ja izmeklētāju novērtējums par konkrētā pierādījuma svarīgumu ir nepilnīgs un/vai ir nekonsekvents, tad arī analīze, kas tiks veikta modelī būs nepilnīga, un, rezultātā, izvade būs neprecīza. Tādejādi, pielietojot Beiesa tīkla modeli, izmeklētājiem ir jāizprot izrietošo secinājumu pareizība, kuru izvada modelis.

Autori M. Kvāns, K. Čojs u.c. bija veikuši Beiesa modeļa analīzi balstoties uz reālu krimināllietu. [8] Šajā analīzē viņi bija izmantojuši objektīvus varbūtības mainīgos, kuri tika piešķirti noteiktā skalā konkrētiem pierādījumiem. Šīs skalas tika iegūtas apkopojot pieredzējušu tiesībaizsardzību iestāžu darbinieku un analītiķu argumentus, par to cik konkrētais pierādījums ir piederīgs krimināllietai. Pēc analīzes rezultātiem M. Kvāns, K. Čojs u.c. secināja, ka noteikto pierādījumu būtiskums, kurus bija manuāli atlasījuši tiesībaizsardzību iestāžu darbinieki un analītiķi, salīdzinājumā ar rezultātiem, kurus atlasīja Beijiesa tīkla modelis sakrita 92.7% [8].

4.2.1. Beiesa tīklu pieejas analīze

Trīs gadus vēlāk (2011. gadā) pēc iepriekš minētā pētījuma, daži no augstāk minēto analīžu veikšanas autoriem - M. Kvāns, R. Overils (*R. Overill*) u.c. savā darbā [7] analizēja Beiesa tīklu modeļu izmantojuma digitālo noziegumu izmeklēšanas rezultātā iegūto datu derīgumu, ticamību un kvalitāti. Sava darba [8] ievadā viņi norādīja, ka, lai gan, ir bijuši veikti daudzi pētījumi par Beiesa tīklu izmantošanu kā rīku, lai uzlabotu digitālās kriminālistikas izmeklēšanu (t.sk. viņu iepriekš minētais pētījums), taču līdz viņu veiktam darbam netika izvērtēta Beiesa tīklu izvades datu kvalitāte. Savā darbā autori šo novērtējumu veica izmantojot zināmo ievaddatu sajaukumu, kurā iepriekš novērtēts šo datu nozīmīgums balstoties uz to lietderību kriminālprocesam, salīdzinājumā ar pierādījumu aprēķināmo datu nozīmīgumu, kas tika iegūti izvades rezultāta.

Balstoties uz 4.2. apakšnodaļā minēto, autori M. Kvāns, R. Overils u.c. savā darbā [8] izveidoja analīzes metodoloģiju, kas kalpoja kā līdzeklis, lai izvērtētu Beijesa tīkla modeļu pielietojumu kā rīku pierādījumu analīzē, pamatojoties uz kuru tie veica secinājumus par šo modeļu efektivitāti. Analīzei autori M. Kvāns, R. Overils u.c. savā darbā izmantoja viņu piedāvāto asuma analīzes tehnikas (*sensitivity analysis techniques*) [8], lai novērtētu Beiesa tīkla modeļu pielietojumu pareizību, attiecinot to uz vairākiem reāliem izmeklēšanas procesiem, kurā Beiesa tīkls tika izveidots izmantojot informāciju, kas gūta no sodāmības ziņojuma, kurā tiek aprakstīti

digitālie pierādījumi. Autori šo analīzes metodiku izstrādāja ar domu, lai pārbaudītu tīkla, jutīgumu attiecībā uz nelielām un lielām izmaiņām atsevišķu pierādījumu esamības iespējamībā.

M. Kvāns, R. Overils u.c. norādīja [8], ka Beiesa tīkls ir atkarīgs no izvades rezultāta stabilumu pret izmaiņām ievaddatu iespējamības vērtībām, papildus precizējot, ka Beiesa tīkls ir stabils, ja izvades asuma līmeņi nav ietekmēti no mazām izmaiņām iespējamības vērtībās. Šajā ziņā, asuma analīze ir svarīga, jo praksē ir grūti precīzi novērtēt pieņēmumus un saistības balstoties uz kurām pamatā tiek būvēts Beiesa tīkla modelis un tā konstruēšana.

Asuma analīze pēta šīs tīkla īpašības, pētot tā izvades variācijas, kas tiek radītas no varbūtisko vērību izmaiņām. M. Kvāns, R. Overils u.c. raksturoja, ka viena no izplatītām pieejām asuma novērtēšanā ir iteratīvi mainīt katru iespējamības vērtību visās iespējamās kombinācijās un novērtēt to ietekmi uz rezultātu izvadi. Gadījumā, ja tiek novērots, ka lielas izmaiņas iespējamības vērtībās rada nenozīmīgu ietekmi uz izvades rezultātu, tad tiek novērtēts, ka konkrētie pierādījumi ir pietiekami ietekmīgi un tām nav ietekmes uz modeli. Savukārt, ja tiek veiktas nelielas izmaiņas, un secīgi izvadītie rezultāti būtiski izmainās, tad ir nepieciešams pārskatīt tīkla struktūru un iepriekšējās varbūtības vērtības. [8]

Tā kā varbūtību vērtību piešķiršana konkrētiem pierādījumiem un to pamatojumi Beiesa tīklā, kas izveidots digitālās kriminālistikas izmeklēšanai, lielākoties, ir diskrēti, tad parametru asuma analīzes funkcijas, asuma vērtība un virsotņu tuvums tiek izmantots, lai noteiktu katra pierādījuma robežu asumu un to stabilitāti gan pie mazām, gan pie lielām iespējamības vērtībām. Tādejādi, parametru jūtīguma analīze novērtē izvades rezultātu pamatojoties uz sniegto pierādījumu variācijām.

4.2.2. Beiesa tīklu pielietojuma analīzes rezultāti

Analīzes rezultātā, M. Kvāns, R. Overils u.c. secināja [8], ka rezultāti, kuri tiek iegūti no Beiesa tīklu modeļiem ir statiski ticami un stabili attiecībā uz gadījumiem, kuros pierādījumu varbūtībās ir pieļaujamas salīdzinoši mazas svārstības. Taču, papildus M. Kvāns, R. Overils u.c. norādīja, [8] ka izmeklētājiem un analītiķiem ir kritiski jāpārskata visi pierādījumi, kad tiem tiek piešķirta nozīmīguma vērtības, jo izmaiņas pie šī mainīgā visvairāk ietekmē Beiesa tīkla izvadi.

Tādejādi autori secināja, ka lai gan Beiesa modelis ir efektīvs veids kā veikt pierādījumu analīzi, tas paļaujas uz ievadi no izmeklētājiem un analītiķiem, un atkarībā no tā, ļoti lielā mērā ir atkarīgs tās izvades rezultāts. [8]

4.3. Uz ontoloģijām balstīta pieeja

Autori J.Čabots, A. Beatriksa, K. Nikola, un T. Kečadi savā darbā [6] kā pamatu savam risinājumam piedāvā risinājumu, kas sevī ietver strukturētu datu attēlošanu kopā ar to standartizētu formālu informācijas attēlojumu un vienlaikus iekļauj semantisko informāciju. Savā darbā autori [6] piedāvā metodikā izmantot ontoloģijas, kas tiek implementētas atvasinājumā no *OWL2* ontoloģijas valodas. Autori M. Smits (*M. Smith*), I. Horoks (*I. Horrocks*) un M. Kročzs (*M. Krotzsch*) *OWL2* raksturo kā ontoloģiju valodu, kas ir semantiska tīkla valoda ar formāli definētu nozīmi. *OWL2* ontoloģijas nodrošina atribūtu tipus kā klases, to īpašības un mainīgo vērtības glabāšanu semantiska tīkla dokumentos. [9]

4.3.1. ORD2I ontoloģija

Autori J.Čabots, A. Beatriksa, K. Nikola, un T. Kečadi *OWL2* ontoloģijas atvasinājuma izvēli pamato ar to, ka šāda valoda ontoloģiju modeļos ļauj attēlot informāciju konkrētā domēnā strukturējot tās pēc entītijām (*entities*), to savstarpējām attiecībām un loģiskiem ierobežojumiem, kas, attiecīgi, ļauj korektāk attēlot informāciju, kas iegūta kriminālprocesa izmeklēšanas laikā. Šī piedāvātā pieeja, ir balstīta uz trīs slāņu ontoloģiju, kuru autori nosauca par *ORD2I* jeb Digitālo incidentu un izmeklēšanu reprezentēšanas ontoloģiju (*Ontology for the Representation of Digital Incidents and Investigations*). Augstākinētie autori savā pētījumā [6] norādīja, ka šī pieeja varētu reprezentēt visus digitālos notikumus. Tā tika raksturota kā ontoloģija, kurā ir saistīts operatoru kopums, kas analizē iegūto laika joslu, tādejādi nodrošinot izmeklējamā objekta atkārtojamību.

Savukārt, par *OWL2* autori norādīja, [3] ka tas tiek lietots kā pamats, jo šī ontoloģijas valoda ļauj ierobežot klases izteicienu, lai samazinātu rīka izpildes laiku, jo rīka darbības tiek veiktas ar lielu apjomu datiem, un tajā pašā laikā rīkam jāspēj veikt secinājumus un analīzi. Papildus, autori nonāca pie secinājuma, ka šai valodai ir spēcīgs teorētisks pamats, jo tā balstīta uz aprakstu loģiku, kas formāli definē semantiskos konceptus un to savstarpējās attiecības. Autori norādīja, [6] ka šīs ontoloģijas komponentes un loģika, kas ar tām ir saistīta, ir matemātiski definēta, kas ļauj pārbaudīt to zināšanu nemainīgumu un sakarību.

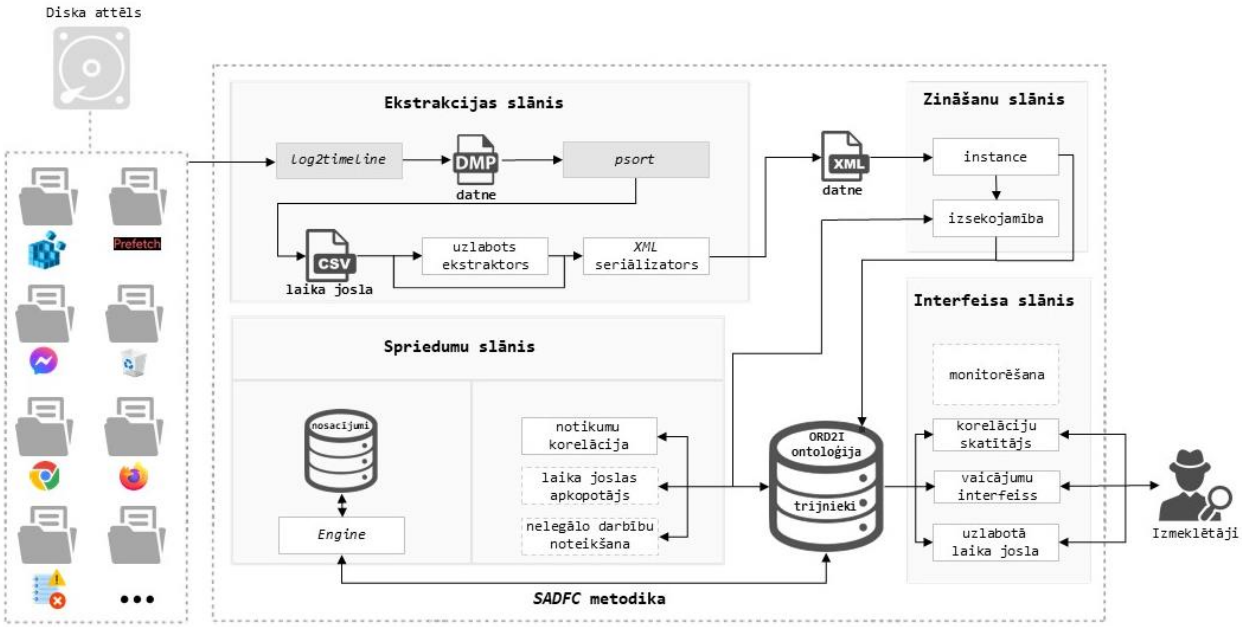
4.3.2. ORD2I un SPARQL

ORD2I ontoloģijā izmanto *SPARQL* vaicājumu valodu. *SPARQL* ir vaicājumu valoda, kas ir semantiska vaicājumu valoda datu bāzēm, tā paredzēta vaicājumu veikšanai un datu manipulēšanai no datubāzēm, kur tie glabājās *RDF* formātā, kas atšifrējās kā Resursu apraksta

sistēma (*Resource Description Framework*). [10] Pēc autoru J.Čabota, A. Beatriksas, K. Nikolas, un T. Kečadi domām [6], pateicoties *SPARQL*, ar ontoloģijām ir viegli manipulēt, jeb veikt vaicājumus par zināšanu diagrammām, un, attiecīgi, gala rezultātu vizuāli attēlot ontoloģijas kā grafus, kas ir vairāk intuitīvi, nekā datu attēlojums teksta veidā. Tā kā ontoloģijas pēc būtības ir izveidotas, lai palīdzētu uzbūvēt kopīgu skatu kādam domēnam, šī struktūra, it sevišķi, der izmeklēšanas rezultātu attēlošanas nolūkos.

4.3.3. Četru slāņu sistēma

Autori J.Čabots, A. Beatriksa, K. Nikola, un T. Kečadi savā darbā piedāvā datu virzības sistēmu, kura sastāv no četriem slāņiem. [6] Pirmais ir ekstrakcijas slānis, kas sastāv no rīkiem, kas izgūst zināšanas no datora diska attēla, kurā, lai izgūtu lielo datu apjomu autori piedāvā lietot jau iepriekš minēto, esošo rīku *Plaso*, taču der jebkāds līdzīgs datu izgūšanas rīks, kas ļautu izgūt automātiski informāciju no ļoti daudziem avotiem. Savukārt, zināšanas, kas tiek izgūtas ar šo rīku ir izmantotas, lai aizpildītu ontoloģiju. Otrais ir zināšanu līmenis, tas tiek izmantos, lai filtrētu iepriekš izgūtās zināšanas un, lai aktualizētu ontoloģijas tajās. Šis slānis arī nodrošina mehānismus, kas pārvalda informācijas izsekojamību. Pēc ontoloģiju aizpildīšanas, jeb procesa, kad izveidoti objekti, kas atbilst aizpildītā tipa noteiktajai definīcijai, atribūtiem, saistības lomām un ierobežojumiem, dati nonāk trešā slānī, jeb spriedumu slānī, kas tiek izmantots, lai analizētu un interpretētu zināšanas par incidentu, vienlaikus deducējot jaunus faktus par incidenta zināšanām, jeb notiek sava veida zināšanu uzlabošana, kā arī pēc tam notiek analīze un secinājumi. Visbeidzot, interfeisa slānī atrodas attēlošanas rīki un vaicājuma rīks. Pārskats par ontoloģijā balstīto pieeju, kas sastāv no četriem slāņiem ir redzams attēlā Nr. 2.



Attēls Nr. 2. **Pārskats par ontoloģijā balstīto pieeju, kas sastāv no četriem slāņiem.** [6]

Attēlā Nr. 2 [6] ir attēloti četri slāņi, kas iezīmēti kvadrātos gaiši pelēkā krāsā, četrpadsmit moduļi, kur izdalīti tie, kas ir operatīvi un iezīmēti līnijās. Savukārt, tie, kas iezīmēti ar pārtrauktu līniju autoru J.Čabota, A. Beatriksa, K. Nikola, un T. Kečadi darba rakstīšanas brīdi vēl nebija implementēti. Kvadrāti tumši pelēkā krāsā apzīmē datus, kas ir izgūti no *Plaso toolbox* rīka. Datu plūsma ir attēlota ar melnu līniju. Savukārt, spriedumu slānī nodefinētie nosacījumi tiek sasaistīti ar *ORD2I* ar *SADFC* metodiku jeb Digitālo kriminālistikas lietu semantiskās analīzes metodiku (*Semantic Analysis of Digital Forensic Cases*) [6].

4.3.4. ORD2I slāņi

ORD2I ir sadalīts trijos slāņos, kur katrā tiek modelētas dažāda veida zināšanas jeb informācija. Vispārējo zināšanu slānī, ontoloģijā tiek integrētas vispārīgas vienības jeb entītijas, specializēto zināšanu slāņa ontoloģijā tiek integrētas specializētas zināšanas un izsekojamības slānī - izsekojamība.

Autori J.Čabots, A. Beatriksa, K. Nikola, un T. Kečadi skaidroja [6], ka vispārēju zināšanu slānis sevī ietver informāciju par notikumiem, un to resursiem un informāciju par subjektiem, kas iesaistīti notikumos. Šis slānis nodrošina kanonisku notikumu attēlojumu, izmantojot polimorfismu. Tā kā, lielākoties, avoti ir neviendabīgi, tiem tiek piešķirts notikumiem specifisks atribūtu skaits un kopīgas vispārējās īpašības. Savukārt, polimorfisms izveido konsekventu visu notikumu attēlojumu nenoņemot specifiskus datus, kas raksturīga katram notikumiem, bet saglabā

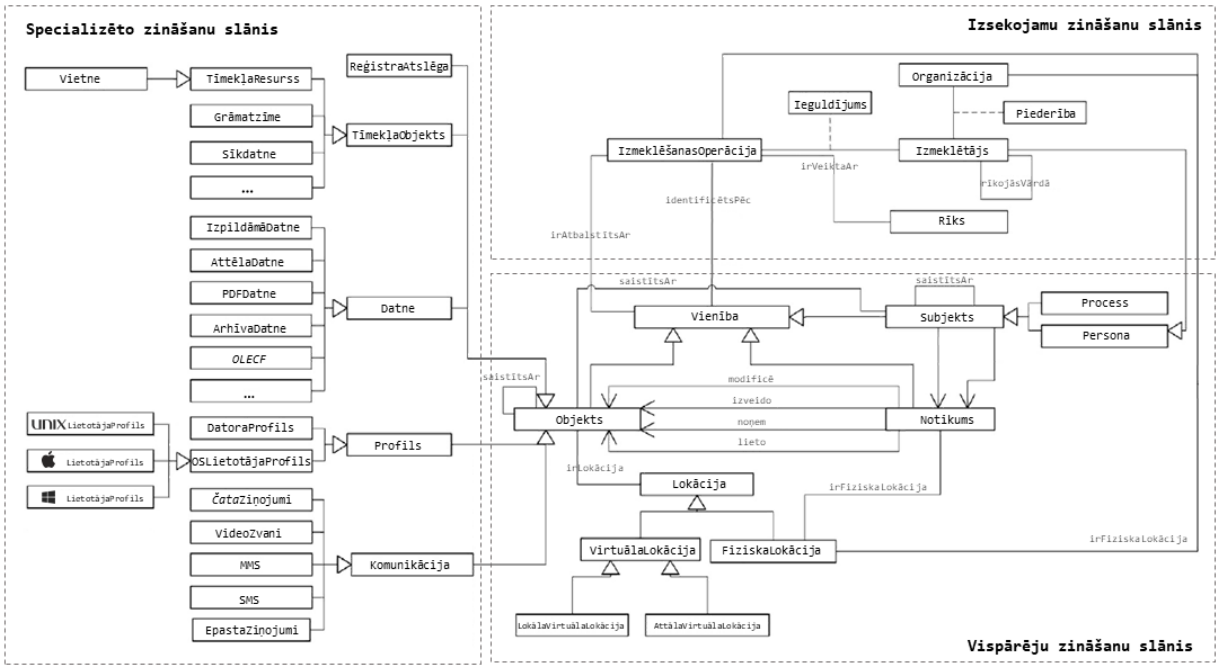
šos datus Speciālo zināšanu līmenī. Autori norāda, [6] ka šis vienotais attēlojums ļauj vienādi analizēt notikumus.

Vispārējo zināšanu līmenī entītija ir pakļauta notikuma klasei, subjekta klasei un objekta klasei, tādejādi notikuma klase ļauj modelēt jebkuru darbību, kas notiek uz konkrētas ierīces un to definē pēc darbības tipa, kas tiek veikts, piemēram, pēc datuma un laika, kad notikusi darbība.

Specializēto zināšanu slānī uzglabā specializētu objektu informāciju, kurā tiek modelētas tehniskās zināšanas par jebkura veida objektiem vai notikumiem, kā arī te, tās tiek strukturizētas. Autori savā darbā uzsvēra, ka šī slāņa mērķis ir nodrošināt struktūru tehniskām zināšanām, un slānis nav paredzēts būt pilnīgs, respektīvi, to ir konstanti jāuzlabo. Ontoloģijas lietošana šajā slānī ļauj integrēt jaunas klases un modificēt esošās. [6]

Izsekojamu zināšanu slānī glabājas informācija par izmeklēšanas procesā veiktajām darbībām, lietoto informāciju un iesaistītiem izmeklētājiem. Šis slānis tiek izmantots, lai iegaumētu kā katrs izmeklēšanas rezultāts tika producēts.

Katrs no šiem slāņiem ir attēlots Attēlā Nr. 3 [6], kurā redzamas klases, kas tos veido un īpašības, kuras šīs klases sasaista. Objektu īpašības ir attēlotas izmantojot bultas un mantojuma saites ir attēlotas izmantojot *UML* vispārināšanas attiecības. Attēlā nav uzrādīti klašu atribūti.

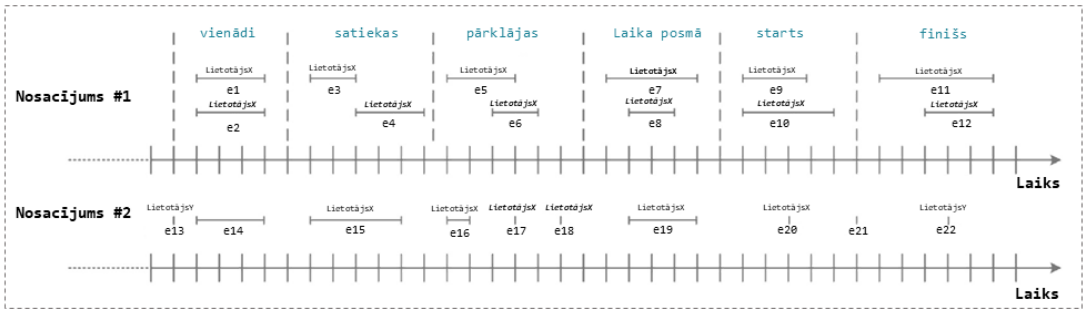


Attēls Nr. 3 **pārskats par *ORD2I*.** [6]

4.3.5. Sakarību izveidošana

Pēc tam, kad dati tiek izlaisti caur visiem *ORD2I* slāņiem. Autori J.Čabots, A. Beatriksa, K. Nikola, un T. Kečadi skaidro, [6] ka tālāk esošā informācija tiek papildināta un uzlabota, ar jauniem secinājumiem izmantojot rīku, kas identificē vēl papildus asociācijas starp lietotāju un notikumiem. Tajā tiek identificēti notikumi, kuri varētu sniegt informāciju par lietotāju, kas tajā iesaistīts. Piemēram, gadījumos kad zināšanu ierakstus ģenerē pārlūks, kuru vienā no apmeklēto vietņu vēstures *logiem* ir uzrādīts lietotāja vārds, lai korekti identificētu šī lietotāja iesaisti tiek izmantoti divi nosacījumi, kas pamatā lieto ontoloģijas laika un temporālos datus. Šie nosacījumi ir: 1. Ņemot vērā notikumu A, kura lietotājs nav zināms, un notikumu B, kas saistīts ar lietotāju P, ja A un B ir īslaicīgi saistīti ar īpašību p ∈ atribūtiem vienāds, pārklājas, laikā, sākas, tiekas, beidzas, tad lietotājs, kas saistīts ar A ir P. 2. Ņemot vērā divus notikumus A un B, kas saistīti ar lietotāju P un atrodas notikumu ķēdes abos galos, kas sastāv tikai no notikumiem, kuriem lietotājs nav zināms, katrs notikums, kas veido šo ķēdi, ir saistīts ar lietotāju P.

Šie divi nosacījumu rezultāti ir redzami attēlā Nr. 4. Lietotāji, kas attēloti treknrakstā un kursīvā ir uzlabojuma rezultāti. Attēla augšdaļa ilustrē pirmo nosacījumu, kurš tiek izmantots, lai secinātu par lietotāju, kurš ir saistīts ar notikumiem e2, e4, e6, e8, e10, e12. Attēla apakšdaļa ilustrē otro nosacījumu un parāda, kā lietotājs X ir iesaistīts notikumos e17 un e18.



Attēls Nr. 4 **Modeļi, kas izmantoti, lai secinātu jaunas attiecības starp lietotājiem un notikumiem** [6]

4.3.6. Papildus sakarību aprēķins

Par pēdējo posmu autori J.Čabots, A. Beatriksa, K. Nikola, un T. Kečadi skaidro , ka pēc augstākminēto procesu izpildes, tiek iniciēts laika analīzes rīks un pieejā piedāvātais laika skalas analīzes rīks ir domāts process, kas nosaka korelāciju starp notikumu pāri. Pamatā šo saistīto notikumu pāru identifikācija tiek veikta, izmantojot četrus kritērijus: abu notikumu mijiedarbību ar 1) kopīgiem objektiem vai 2) subjektiem, 3) laika tuvums un 4) noteikumu kopums, kas tiek vai netiek apmierināti. Katram notikumu pārim (e1, e2) korelācijas rezultātu (*Corr*) aprēķina, izmantojot attēlā Nr. 5. redzamo formulu.

$$Corr(e_1, e_2) = Corr_T(e_1, e_2) + Corr_S(e_1, e_2) + Corr_O(e_1, e_2)$$

Attēls Nr. 5 **Sakarību aprēķins** [6]

J.Čabots, A. Beatriksa, K. Nikola, un T. Kečadi skaidro, ka ja kopējais korelācijas rādītājs ir no 0,0 (tas nozīmē, ka abi notikumi nav savstarpēji saistīti) un 1,0 (kas nozīmē, ka abi notikumi ir savstarpēji cieši saistīti), kas ir vidējais rādītājs no trīs. Pirmie trīs kritēriji nav atkarīgi no notikumu veida, jo to pamatā ir vispārīgas pazīmes (laiks, objekts un subjekti).

Minētie nosacījumi, ļauj atklāt saistības, kurās iesaistīti notikumi no nezināmiem informācijas avotiem, un tāpēc izmeklētāji nav paredzējuši tos. Lai ņemtu vērā specifiskas zināšanas par katru objekta modeļa veidu *ORD2I* ir ieviests ceturtais mainīgais: *CorrEK (e1, e2).* Šis rādītājs tiek aprēķināts, izmantojot ekspertu definētu noteikumu kopumu jeb autori to definē par uz ekspertu zināšanām balstītiem noteikumiem. [6] Ja divi notikumi atbilst noteiktam noteikumam, kopējais korelācijas rādītājs starp tiem ir maksimāls.

4.3.7. Datu attēlošanas posms

Attēlošanas posmā galvenais mērķis ir ļaut izmeklētājiem vizualizēt datus, izmantojot intuitīvu un skaidru vizualizācijas rīku. Tas nodrošina uzraudzības saskarni, kas ļauj pārvaldīt sistēmas iestatījumus, tostarp korelācijas procesā. Tas nodrošina uzlabotu vaicājumu rīku lietotājiem, kurus esošie vizualizācijas rīki nespēj izpildīt. Ir ieviests rīks, lai vizualizētu korelācijas starp notikumiem. Šī vizualizācija tiek pārveidota par grafiku, kurā ir visi notikumi, kas veido incidentu, un korelācijas starp tiem. Saiknes starp notikumiem parāda saistīto notikumu ķēdes. Katras saites vizuālais stils (krāsa, līnijas tips, biezums) ir atkarīgs no korelācijas stipruma. Savukārt, lai redzētu informāciju, kas netiktu attēlota attēlošanā, autori piedāvā veikt *SPARQL* vaicājumus, lai vaicātu tieši datu bāzē esošās zināšanas. [6]

4.3.8. Ontoloģiju pielietojuma analīzes rezultāti

Autori J.Čabots, A. Beatriksa, K. Nikola, un T. Kečadi savā darbā [6] izdalīja dažādus datu viendabīguma klasificēšanas veidus - sintaktiskā, semantiskā un laika. Savā darbā viņi pētīja paņēmienus, kas ir saistīti ar sintaktisko un semantisko, jo viņuprāt šo abu dimensiju izmantošana ir saistītas ar vienu mērķi - rekonstruēt un parādīt skaidrāku un pilnīgāku datu attēlojumu. Klases izmanto vairākus dažādus atribūtus, kas tiek mantoti no notikumu klasēm, lai sniegtu informāciju par attiecīgā procesa lietotāju vai par to, kas šo procesu ir izveidojis. Šajā gadījumā ontoloģiju lietošana, lai attēlotu notikumus, ir efektīvs veids kā risināt datu neviendabīguma problēmu, vienlaikus, ontoloģijas ļauj atspēkot semantiski bagātus modeļus, kas spēj aptvert semantiku visām entītijām, kā arī to attiecībām. Šajā gadījumā precīzs un formāls modeļa komponentu apraksts var palielināt datora analītiskās iespējas datu nozīmīguma noteikšanā, tāpēc, lai gan šī pieeja ir automatizēta, tai ir nepieciešams diezgan liels izmeklētāju un analītiķu ieguldījums, lai gūtu rezultātus.

Savā darbā [6] autori izvēlējās *Plaso toolbox* kā rīku, kas spēj apstrādāt lielu daudzumu informācijas avotu. Tie atzīmē, ka *Plaso* ir spējīgs ģenerēt, tā saukto, laika super-laika joslu (*super-timeline*), kas ir notikumu laika josla, kura sevī var integrēt daudz notikumu avotus. Viens no Plaso rīkiem *log2timeline*, ļauj iegūt visus notikumus no spoguļkopijām. Otrs *Plaso* rīks ir *psort,* kas tiek izmantots lai pārveidotu iepriekš minēto izvadīto rezultātu kā *.csv* vai *.db* tipa datni. *Psort* izvadītais datņu formāts sastāv no ierobežota lauku skaita, kas sevī glabā datus par notikuma laiku un datumu. Tas tiek izmantots, lai iegūtu notikumu un ziņojumu, kas to raksturo avotus. Autori uzsvēra, ka šie izvades formāti neļauj precīzi un skaidri attēlot attiecības starp notikumiem un entītijām līdz ar to

nepieciešama to apstrāde, tam ir nepieciešams ieguldījums no izmeklētājiem, lai kategorizētu šos izgūtos datus.

4.4. Uz laika joslas balstīta pieeja kopā ar *DESO*

Autori K.Džeims un H.Dž. Patersons savā darbā [5] piedāvā automatizētu risinājumu izmeklētājiem un analītiķiem digitālo notikumu analīzē, kas automātiski rekonstruē tieši augsta līmeņa notikumus, piemēram, *USB* zibatmiņas pievienošana datoram. Viņu piedāvātais modelis ir uz *Python* bāzēts prototips, kas izgūst datumus un laikus no daudzām datnēm, kas atrodas spoguļkopijā un pēc tam producē atskaiti par augsta līmeņa notikumiem balstoties uz vienu vai vairāku zema līmeņa notikumu klātbūtni.

Šī kombinētā pieeja ļauj saglabāt detalizētu izcelsmi visiem automātiski iegūtiem secinājumiem, sākot no augsta līmeņa notikumiem līdz zema līmeņa notikumiem, kuri gan pastāvēja vai arī nepastāvēja, kā arī ļauj iegūt informāciju par pašu datni, kas šo notikumu izraisa. Lai gan šī pieeja izdala zema un augsta līmeņa notikumus, kas teorētiski atsijā nevajadzīgos datus, tā var nekorekti pieņemt, ka zema līmeņa notikumi tiek uzskatīti par nevajadzīgiem, kas dažreiz izrādās nepatiesi. Taču autori O. Bradlijs (*Owen Brady*), R. Overils (*Richard Overill*) un Dž. Keppens (*Jeroen Keppens*) savā darbā [11] secināja, ka, tā kā iepriekš minētā pieeja lielā mērā izdala procesu datus, tā grafiskais un vizuālais attēlojums ir trūcīgs, jo kopā ar korelācijas līmeņiem ir samazināts šīs pieejas darbība bez jebkādiem papildinājumiem liela apjoma digitālo notikumu un datu plūsmu gadījumos. Piemērs laika joslas attēlojam ar trīs notikumiem ir redzams zemāk Attēlā Nr. 6.

Attēls Nr. 6. **Laika joslas attēlojums ar 3 notikumiem.** [11]

4.4.1. DESO

O. Bradlijs, R. Overils un Dž. Keppens savā darbā [11] piedāvā *DESO* metodiku, kas ir atvasināta no autoru K.Džeims un H.Dž. Patersona piedāvātā risinājuma. *DESO* jeb Digitālo

pierādījumu semantiskā ontoloģija (*Digital Evidence Semantic Ontology*). Autori, norāda, ka tās galvenais mērķis ir darboties kā digitālo pierādījumu artefaktu krātuvei un klasifikatoram. Veids, kādā *DESO* savieno šos artefaktus, ļauj selektīvi skatīt digitālajās ierīcēs ietvertos datus un nodrošina to salīdzināšanu kopējā formātā. Pēc autoru domām, veicot artefaktu atlasi pirms mēģinājuma tos iegūt, tiek ietaupīts apstrādes laiks un uzglabāšanas vieta.

Saskaņā ar autoru domām, *DESO* nodrošina sistēmu digitālo pierādījumu artefaktu organizēšanai un kategorizēšanai, un, to darot, izmeklētājam tiek piedāvātas ievērojamas priekšrocības. Līdzīgi, kā iepriekš minētās apakšnodaļas risinājumā, šai ontoloģijai tika izmantots *RDF*. Autori šo izvēli pamato, ka tas tika izvelēts tādēļ, ka tas ir vispārēji pieņemts standarts un labi integrējas ar citām ontoloģijām, kuras autori sasaista ar *DESO*. [11]

4.4.2. DESO klases

Autori O. Bradlijs, R. Overils un Dž. Keppens, izdala trīs *DESO* galvenās klases: artefakts Atrašanās vieta (*Location*) – kur var atrast artefaktus, artefakts Tipa identifikators (*Type Identifier*) - kā tie tiek atlasīti un salīdzināti; un artefakts Atsauce (*Reference*) — artefakta izcelsme.

4.4.3. DESO artefakti

Atsauces klases artefakti dokumentē informācijas avotus, ko izmanto, lai atbalstītu konkrētu digitālo datu klasificēšanu kā artefaktu. Piemēram, tiek dokumentēts artefakts, kas aprakstīts tīmekļa vietnē ar kādu konkrētu nosaukumu. Pirmais uzdevums ir izveidot *indivīdu* šai atsaucei attiecīgajā artefakta *atsauce* (*Artefact Reference)* apakšklasē. Šīs apakšklases, piemēram, var būt: grāmatas, izstrādātāju lapas, žurnāli un tīmekļa ieraksti.

Atrašanas vietas klases artefakti ļauj izmeklētājiem vai analītiķiem saprast vietas, kur var atrast konkrēto artefaktu. Augstākajā līmenī tas ir iedalīts apakšgrupās “Ierīces”, “Failu sistēmas” un “Operētājsistēmas” ar tālākām robežlīnijām, lai ņemtu vērā dažādās kategorijas, ar kurām pārbaudītājs var saskarties.

Papildus autori skaidroja [11], ka atsauces un atrašanās vietas klases ir jāsaista, lai izmeklētāji varētu saskatīt, kādi pierādījumi ir pieejami no konkrētā avota, un iemeslu, kāpēc tas tiek apstiprināts. Šī informācija ir nepieciešama divu iemeslu dēļ: pirmkārt, lai pārbaudītājs var novērtēt, vai ir nepieciešama turpmāka pārbaude pirms paļaušanās uz šiem datiem; un, otrkārt, lai ikviens no aizstāvības puses jeb pārbaudītājs, kuram šie dati tiek sniegti, arī varētu ātri tos novērtēt.

Savukārt, tipa identifikators klases artefakti tiek veidoti kā atsauces uz artefaktu, kurus var pārveidot par indivīdu klases struktūrā, kas attēlo, kur šo artefaktu var atrast. Autori norādīja, ka tam ir nepieciešama artefaktu salīdzināšanas metode, lai neatkarīgi no to atrašanās vietas varētu identificēt tos, kas pārstāv vienu un to pašu jēdzienu, piemēram, *USB* sērijas numurs. Programmā *DESO* tā ir klase tipa identifikators (*type Identifier*). Šīs klases indivīdi ir konceptuālie artefaktu veidi, kurus var atrast un salīdzināt. Tie var būt tādi identifikatori kā "*IMEI*" un "*USB* ierīces sērijas numurs".

4.4.4. DESO klašu atribūtu īpašības

Autori skaidro, ka katram atribūtam ir arī šo datu īpašības (*DataProperties)*, kas nodrošina formātu, kuram ir jāattēlo šāda veida dati. Tas nodrošina, ka visi viena veida artefakti ir viegli salīdzināmi. Tipa identifikators tiek piešķirts atribūtam, iepriekš aprakstītajā atrašanās vietas klasē. Kad artefaktu dati tiek iegūti no dažādām atrašanās vietām var salīdzināt tās, kurām ir vienāds tipa identifikators.

4.4.5. Uz laika joslas balstīta pieeja kopā ar DESO pielietojuma analīzes rezultāti

Autori O. Bradlijs, R. Overils un Dž. Keppens skaidro, ka *DESO* koncepcijas pamatā ir tas, ka artefakti tiek pievienoti tikai tad, ja tam ir noteikta *atrašanās vieta, tipa Identifikators* un *atsauce* klase. Tādejādi, kopā ar autoru K.Džeims un H.Dž. Patersons ideju, par temporālo datu kā bāzi, kā nošķirt datorā procesus ir izveidota, ontoloģija, kas nav paredzēta kā visu datorsistēmu tehnisko detaļu modelis, bet gan tā attiecas tikai uz datiem, kas var palīdzēt izmeklēšanā. Šī iemesla dēļ, katram artefaktam ir jābūt skaidri identificējamiem un datiem izmantojamiem, pretējā gadījumā tas paildzina laiku, kurus izmeklētāji velta datu atlasei. [11]

4.5. Datu rekonstrukciju pieeju kopsavilkums

Analizējot dažādās pieejas, tika atzīmēti to risinājumi, ja tādi ir, attiecībā uz digitālās izmeklēšanas datu apstrādes problēmjautājumiem, kas tika apskatīti 3. nodaļā. Kopsavilkums ir attēlots tabulā Nr. 1. *digitālās izmeklēšanas datu apstrādes problēmjautājumi*. Ir secināms, ka risinājums datu problēmai katrai pieejai ir salīdzinoši vienāds t.i. lielais datu apjoms tiek izgūts ar jau esošiem rīkiem, kuros ir automatizēta datu izgūšana un, atkarībā no rīka, tiek iegūta informācija par konkrēto datni. Savukārt, attiecībā uz risinājumu datu neviendabīguma problēmai, ir secināms, ka divas no pētījumiem paļaujas uz pieskaņotām ontoloģijām, taču visos gadījumos datiem piešķir papildus identifikatorus, kam piedod lielāko būtiskumu – temporālo, semantisko un konkrētās

datnes būtiskumu. Saistībā ar risinājumu uz problēmjautājumu, kas attiecinās uz datu likumiskām prasībām, tad tās ir salīdzinoši dažādas.

	Digtālās izmeklēšanas datu apstrādes problēmjautājumi		
Pieejas veids	*Risinājums datu apjoma problēmai*	*Risinājums datu neviendabīguma problēmai*	*Risinājums datu likumiskām prasībām*
Uz Beiesa tīkliem balstīta pieeja	Automātiska datu izgūšana izmantojot jau esošus rīkus, piemēram, *Plaso toolbox, FACE, CyberForensic, TimeLab* u.c.	Tīklā veidojās cēloņsakarības starp datnēm, kas veido pierādījumu kopu, izdalot konkrētās atrastās datnes. Kvalitāte atkarīga no lietotāja nodefinētiem ievaddatiem jeb pierādījumiem piešķirtiem stipruma līmeņiem.	nav iekļauts
Uz ontoloģijām balstīta pieeja	Automātiska datu izgūšana izmantojot jau esošus rīkus, piemēram, *Plaso toolbox, FACE, CyberForensic, TimeLab* u.c.	Zināšanu modelis. *ORD2I* ontoloģija, kas formāli definē semantiskos konceptus un to starpatiecības. Savukārt, SPARQL tiek izmantota datu atlase. Semantiskā saikne tiek izmantota kā bāze.	*ORD2I* Izsekojamu zināšanu slānis, kurā glabājas informācija par izmeklēšanas procesā veiktajām darbībām, lietoto informāciju un iesaistītiem izmeklētājiem.
Uz laika joslas balstīta pieeja kopā ar DESO	Automātiska datu izgūšana izmantojot jau esošus rīkus *Plaso toolbox, FACE, CyberForensic, TimeLab* u.c.	*DESO* ontoloģija, kas darbojas kā digitālo pierādījumu artefaktu krātuve ar klasifikatoriem. Savukārt, SPARQL tiek izmantota datu atlase. Temporālā saikne tiek izmantota kā bāze.	Pie artefaktiem tiek piesaistītas datu īpašības, kuras ietver datņu kontrolsummas.

Tabula Nr. 1. **Digitālās izmeklēšanas datu apstrādes problēmjautājumi**

5. DATU ATTĒLOŠANA

Saistītie galvenie problēmjautājumi digitālā izmeklēšanā ir tādi, ka no vienas puses, kā dati tiek iegūti un apstrādāti, un, no otras puses, kā dati tiek attēloti. Abi šie jautājumi idejiski ir atšķirīgi, taču tie ir savstarpēji atkarīgi. Risinājumi šiem procesiem ir atšķirīgi, it sevišķi saistībā ar datu attēlošanu, jo ir pieejamas daudzas metodes kā nodibināt mijiedarbību starp lietotāju, telpu, tehnoloģiju, informāciju un attēlu. Piemēram, autori A. Bajarī (*Ahmed Bayyari*) un E. Turdanu (*Eduard Tudoreanu*) savā pētījumā [12] uzsvēra datu attēlošanas nozīmi, norādot, ka tā ir būtiskākā pēdējā sastāvdaļa IT projektos, kuri ir saistīti ar datiem un to attēlojumu. Tādejādi, datu vizualizācija ir būtiska datu komunikācijas procesa iezīme, kas tiks apskatīta šajā nodaļā. Saskaņā ar autoru Dž. Sjindžou (*Zhang Xinzhou*), efektīva datu attēlošana ir viena no būtiskām daļām izmeklēšanas procesā, jo tas ir, savā ziņā, tilts no kvantitatīva datu satura un cilvēka intuīciju jeb māku no redzētā izveidot likumsakarības, līdz ar to, tā ir būtiska sastāvdaļa ceļā no datiem uz informāciju - tās zināšanām un izpratni [13].

Ņemot vērā, ka attēlojumam ir būtiska nozīme datu analīzes procesā, jo tā palīdz identificēt nevajadzīgos datus no analīzes, tādejādi samazinot analīzei patērēto laiku un lietotāja noslogojumu, šajā nodaļā tiks aprakstīti vairāki datu attēlošanas veidi un to iepriekš veiktie pētnieku salīdzinājumi, lai atrastu efektīvāko veidu, kā panākt iepriekš nodefinēto vienu no šī darba mērķiem.

5.1. Pētījumu par datu attēlošanu atlase

Pētījumi tika apvienoti pēc to risinājuma metodoloģijas atsevišķās šīs nodaļas apakšnodaļās. Minētie pētījumi tika atlasīti posmā no 2021. gada oktobra līdz 2021. gada decembrim. Pētījumu atlasei salikumos tika izmantoti atslēgvārdi: lielo datu attēlošana (*big data visualization*), virtuālā realitāte (*virtual reality*), divu dimensiju datu attēlošana (*2D data visualization*), trīs dimensiju datu attēlošana (*3D data visualization*). Platformas kurās tika veikta atlase ir *Google Scholar*, *ResearchGate* un *Nature*.

5.2. Datu attēlošanas nozīmīgums

Savā darbā [14] autori B. Laha (*B. Laha*) un D.A. Boumans (*D. A. Bowman*) norādīja, ka datu vizuāla analīze un izpēte lielos apjomos ir ikdienas dažādu nozaru pētnieku uzdevums. Lielu datu analīze regulāri tiek lietota medicīnā, piemēram, smadzeņu, sirds vai plaušu magnētiskās rezonanses skenējuma dati, molekulārā bioloģijā, piemēram, dati, kas iegūti no konfokālās

mikroskopijas, ģeoloģijā, piemēram, iežu slāņu dati, digitālā noziegumu izmeklēšanā un citās jomās. Tradicionāli, dažādo nozaru zinātnieki un pētnieki izmanto attiecīgās un specifiskas programmatūras uz galddatoriem, lai vizualizētu un analizētu datus. Taču šīm sistēmām salīdzinājumā ar virtuālo realitāti ir monoskopiska renderēšana, mazs redzes lauks, mazs ekrāna izmērs un tām trūkst tālvadības atveidošanas. [14] Ņemot vērā iepriekš minēto, daudzi pētnieki ir rosinājuši virtuālo realitāšu izmantošanu, kā alternatīvu, kurai ne tikai piemīt augstāks iesaistes līmenis, bet arī tādēļ, ka pēc būtības virtuālā realitāte ir izveidota ar ieceri telpiskai attēlošanai, kur sarežģītas struktūras ir atveidotas saprotamā un izpētāma veidā. [14]

Ņemot vērā iepriekš minēto, daudzi pētnieki ir rosinājuši virtuālo realitāšu izmantošanu, kā alternatīvu, kurai ne tikai piemīt augstāks iegremdēšanas līmenis, bet arī dēļ tā, ka pēc būtības virtuālā realitāte ir izveidota ar ieceri telpiskai attēlošanai, kur sarežģītas struktūras ir atveidotas saprotamā un izpētāma veidā. [14]

5.3. Datu attēlošanas izaicinājumi

Autori Z.M. Kalīds (*Z. M. Khalid*) un S. R. M. Zēbarē (*S. R. M. Zeebaree*) savā darbā [15] norādīja, ka lielo datu vizualizācija ir sarežģīta, pamatojoties uz datu skaitu, dažādību un ātrumu. Lielākā problēma, strādājot ar lielajiem datiem, ir, kā pārvaldīt milzīgus datu apjomus un efektīvi parādīt datu attēlošana un analīzes praktiskos un izmantojamos rezultātus. Ir jāizveido jauns mehānisms, lai datus aplūkotu tā, lai palīdzētu politikas veidotājiem pamanāmi un ātri gūt ieskatu tajos, izmantojot grafikus un kartes. Savukārt, autori savā darbā E. Olešikova (*E. Oleshikova*), A. Ometovs (*A. Ometov*) un J. Kučeravijs (*Y .Koucheryavy*) [16] tradicionālie vizualizācijas rīki nespēj apstrādāt plašas datu kopas. Prezentācijas rīks nodrošinās zemāko iespējamo displeja latentumu. Paralēlizācija bieži ir nepieciešama, lai apstrādātu tik lielu datu apjomu, kas ir vizualizācijas uzdevums. Interesantas tendences var raksturot kā lielu datu vizualizācijas centrālo aspektu. Datu mērījumi ir jāizvēlas rūpīgi paraugu ieguvei. Ja tiek izvēlētas tikai dažas dimensijas, vizualizācijas kvalitāte var samazināties, un var tikt zaudēti vairāki aizraujoši modeļi; tāpat, ja tiek atlasīti visi mērījumi, tas var veicināt sarežģītu skatu, kas lietotājiem nav izmantojams.

5.4. Datu attēlošanas metodes

Autori Z.M. Kalīds un S. R. M. Zēbarē savā darbā [15] norādīja ka datu attēlošanas veidam, ir būtiska nozīme, jo ar tiem ātri iegūts veids, kā cilvēkam veidotu izpratni par liela daudzuma informāciju. Savukārt, cilvēki var atklāt lietas, ko viņi nezina (novirzes, slēptus modeļus vai

grupas), izmantojot ideālu datu vizualizācijas rīku. Šie instrumenti arī ļauj iedziļināties strauji mainīgās datu kopās.

Savā darbā [15] viņi analizēja un izdalīja metodes, kuras var izmantot lielo datu attēlošanai. Šīs metodes viņi vērtēja un pēc tam kategorizēja pēc datu lieluma, datu dažādības un datu dinamikas. Viņu izdalītās datu vizualizācijas metodes ir:

1) Koka karte. Šī metode aplūko veidu, kā skatīt hierarhiskos datus kā apvienotu taisnstūru kolekciju. Sākotnējais taisnstūris ir sadalīts apakš-taisnstūros, izmantojot sadalīšanas algoritmu. Parasti tiek izmantota trenēta metode. Taisnstūra apgabals nosaka kategorijai piešķirto numuru. Tāpēc nulles un negatīvo vērtību ierobežojums attiecas tikai uz koku kartēm. Tāpat, hierarhija ir sagrozīta ar vairāk pikseļiem.

2) Piepildītā apļa diagramma. Tā ir alternatīva koka kartes pieejai, kas izmanto apļus, lai attēlotu dažādus hierarhiskus slāņus. Apļa apgabals nosaka veida numuru. Tas izmanto arī vairākas krāsas dažādās grupās, tostarp koka kartē. Šī pieeja nav efektīva telpas ziņā, atšķirībā no koka kartes.

3) Paralēlas koordinātes. Šī metode ir lielu datu parādīšanas līdzeklis. Datu komponentes var tikt izvietotas atsevišķi, izmantojot dažādus izmērus; gan mežu, gan koku var redzēt paralēlās koordinātēs. Līniju tendences tiek uzzīmētas, lai apkopotu konsekventus rezultātus. Personu līnijas var būt ieskicētas, lai redzētu precīzu atsevišķu datu vienumu izvadi. Tomēr daudzi datu objekti veicina pārplānošanu. Šo metodi neizmanto kategoriskiem datiem.

4) Plūduma/Straumes grafiks (*stream graph*). Šo metodi izmanto, lai parādītu vērtību nobīdi vienlaikus ar citu centrālo laika skalu. Tas norāda uz vairāku kategoriju datu uzlabojumiem laika gaitā. Katras straumes formas lielums ir vienāds ar katras kategorijas vērtībām straumes diagrammā. Autori Z.M. Kalīds un S. R. M. Zēbarē uzskatīja, ka šis veids ir piemērots lielas datu kopas prezentēšanai.

5) Grafs. Autori norādīja, ka grafs ir labi lietojams nelineāru datu struktūras attēlošanas veids. Attēlojamos datus var šķelt vairākos grafos, vai apvienot vienā lielā grafā.

Balstoties iepriekš minēto datu attēlošanas veidu raksturojumiem, risinājumā ir izvēlēts veikt datu attēlojumu ar pēdējo metodi, kas ir grafs.

5.5. Salīdzinājums starp datu attēlošanu 2D vidē un virtuālā vidē

Autori P. Milais (*P. Millais*), S. L. Džouns (*S. L. Jones*) un R. Kelijs (R. Kelly), savā darbā [17] norādīja, ka virtuālā realitāte bieži tiek apspriesta kā daudzsološs līdzeklis visaptverošai datu attēlošanai un izpētei. Tomēr daži pētījumi ir izvērtējuši lietotājiem atvērto daudzdimensiju datu kopu izpēti, izmantojot virtuālā vidē, un salīdzinājuši rezultātus ar tradicionālo (*2D*) vizualizāciju rezultātiem. Izmantojot uz darba slodzi un uz ieskatu balstītu novērtēšanas metodoloģiju, tika veikta lietotāju izpēte, lai veiktu šādu salīdzinājumu. Tika atklāts, ka starp tradicionālajiem attēlojumiem un attēlojumiem virtuālā vidē nav vispārēju uzdevumu un darba slodzes atšķirību, taču ir atšķirības lietotāju gūto ieskatu precizitātē un dziļumā. Rezultāti arī liecināja, ka lietotāji jūtas apmierinātāki un veiksmīgāki, izmantojot virtuālo vides datu izpētes rīkus, tādējādi demonstrējot tā potenciālu kā saistošu līdzekli vizuālo datu analīzei. [17]

Autori K. Vārs (*C. Ware*) un G. Franks (*G. Franck*), jau 1994. gadā, savā pētījumā [18] ziņoja par eksperimenta rezultātiem, kurā viņi pārbaudīja vai datu tīkla sistēmas attēlošanas informācija tiek efektīvāk parādīta trīs dimensiju telpā vai divu dimensiju telpā. Eksperimenta dalībniekiem bija uzdevums izsekot ceļam tīklā. Pats eksperiments tika veikts 2D, 3D stereo skatījumā, 2D skatā ar perspektīvu, kas savienota ar galvu, un 3D stereo skatā ar perspektīvu, savienotu ar galvu. Autori atzīmēja, ka šis pēdējais nosacījums radīja lokalizētu virtuālās realitātes displeju. [18] Viņu pētījuma rezultāti liecināja [18], ka kustības paralakse (*motion parallax*) jeb perspektīvas novirzīšanās no novērojamā objekta [19], kas iegūta no galvas savienojuma ar perspektīvu, ir svarīgāka par stereopsi (*stereopsis*) jeb dziļuma un trīs dimensiju struktūras uztvere caur binokulāro redzi, [20] strukturālās informācijas atklāšanā. Secinājumos, autori norādīja, ka kopumā rezultāti liecināja, ka ar galvu savienotajā stereoskatā var uztvert trīs reizes vairāk informācijas nekā 2D skatā. [21]

Līdzīgu pētījumu bija veicis autors A. Sankārs (A. Sankar). Viņa pētījumā [21] tika pētīts kā izveidot iesaistošu (*immersive*) un interaktīvu, balstoties uz cilvēka galvas kustībām, trīs dimensiju datu attēlojumu kā analogu esošai 2D datu attēlojumam. Pētījumā tika izvirzīta hipotēze, kura secinājumos tika apstiprināta, balstoties uz lietotāju atsaucēm, ka iesaistošai datu attēlošanai ir priekšrocības, kuras nevar sasniegt ar tradicionālu ekrānu, un ka lietotājs spēj labāk izprast datus, gan no uztveres, gan no emocionālā viedokļa.

5.6. Iesaistes priekšrocības

Autori B. Laha un D.A. Boumans savā darbā [14], kā būtisku raksturīpašību virtuālai realitātei iezīmēja iesaisti. Viņi iesaisti interpretēja kā daudzdimensionālu nepārtrauktu kustību, kurā ir atsevišķas komponentes, kuras var novērtēt neatkarīgi. Savukārt, lai apstiprinātu iesaistes noderīgumu un tās priekšrocības, viņi veica empīriskus pētījumus, salīdzinot dažādu iesaistes līmeņu efektivitāti pie procesiem, kad tiek analizēti liela apjoma datu attēlojumi.

5.6.1. Pētījumu ierobežojumi saistībā ar iesaistes priekšrocībām

Autori B. Laha un D.A. Boumans savā darbā [14] norādīja, ka, pēc viņu novērojumiem, pētnieki, kas pētīja iesaistes priekšrocības, tradicionāli ir salīdzinājuši noteiktas sistēmas vispārīgā veidā, piemēram, galddatoru attēlojumu pret attēlojumiem virtuālajā realitātē. Viņi, uzsvēra, ka šie pētījumi ir ļoti vērtīgi virtuālās realitātes pētnieku kopienai, jo tie demonstrē iesaistes priekšrocības, kas pārsniedz virtuālās realitātes iespaidīgo vizuālo pievilcību, taču autori B. Laha un D.A. Boumans atzīmēja, ka šo eksperimentu rezultāti ir ierobežoti divos svarīgos veidos. [14]

Pirmais ierobežojums ir rezultātu vispārinātības trūkums attiecībā pret citām virtuālām realitātes sistēmām. Viņi norādīja, ka tad, kad visas virtuālās realitātes sistēmas tiek salīdzinātas savā starpā, vairākas iesaistes komponentes (piemēram, stereoskopija un galvas izsekojamības atveidošana) vienlaicīgi atšķiras starp nosacījumiem. Viņuprāt, ja šāds pētījums identificē iesaistes priekšrocības, tad nav zināms, kuri komponenti vai iesaistes komponentu kombinācija radīja šīs priekšrocības. Tādejādi, B. Laha un D.A. Boumans secināja [14], ka šīs neskaidrības rezultātā nav iespējams vispārināt rezultātus virtuālās realitātes sistēmām, izņemot konkrēti pētītās sistēmas.

Otrais autoru B. Laha un D.A. Boumana norādītais ierobežojums [14], ir tas, ka nav zināms, vai virtuālās realitātes sistēmas ar vidēju iesaistes līmeni varētu sniegt tādas pašas priekšrocības kā virtuālā realitātes sistēma ar augstu iesaistoša sistēma. Viņuprāt, šo smalkāko atšķirību nozīme ir saistīta ar virtuālās aparatūras (piemēram, *CAVE* vai *headmounted* displeju (*HMD*)) dārdzību. Viņi norādīja, ka ņemot vērā lētākas virtuālās realitātes aparatūras pieejamību, kas piedāvā vidēju dažu komponentu iesaistes līmeni, ir nepieciešami precīzāki empīriski rezultāti, lai noteiktu, vai šādām sistēmām varētu būt labvēlīgāka izmaksu un ieguvumu attiecība.

5.6.2. Pētījumu par iesaisti rezultāti

Autori B. Laha un D.A. Boumans savā darbā [14] veica kontrolētu pētījumu ar mērķi mainīt konkrētas iesaistes komponentes un novērtēt visas tās līmeņu kombinācijas, vienlaikus saglabājot

visas pārējās sastāvdaļas nemainīgas. Lai īstenotu visus nosacījumus, autori izmantoja vienu virtuālās realitātes sistēmu.

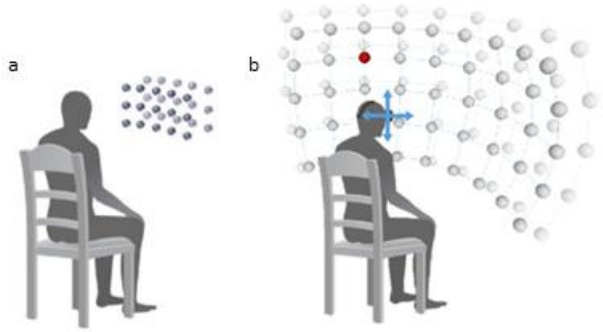
Autori pētījumā laikā mainīja vairākas komponentes, piemēram, ieslēdzot un izslēdzot stereoskopiju un ieslēdzot un izslēdzot galvas izsekojamības atveidi, kamēr lietotāji veica uzdevumus ar divām liela apjoma datu kopu attēlojumiem.

Rezultātā, autori secināja, ka šādu kontrolētu eksperimentu rezultāti palīdz ne tikai identificēt atsevišķu komponenšu ietekmi, bet arī mijiedarbības efektu, kas ietver divas vai vairākas komponentes. Pētījuma rezultātā, tika atklāti pierādījumi par labāku uzdevumu izpildi un augstāku uztveramo lietojamību apstākļos ar galvas izsekošanu un ieslēgtu stereoskopiju, salīdzinot ar pārējiem diviem šo komponentu kombinācijām. [14]

5.7. Priekšrocības datu attēlošanai virtuālā realitātē

Saskaņā ar K. Donaleka (*Christopher Donalek*) u.c. autoru novērojumu, funkcionāli, datu vizualizācija virtuālā realitātē, līdzīgi kā jebkura cita datu attēlošana trīs dimensiju vidē, piedāvā iespēju attēlot datus, un kalpot kā instrumentam, kas ļauj lietotājam mijiedarboties ar objektiem. [22] Atšķirībā no divu dimensiju projekcijas, šīs tehnoloģijas ļauj atspējot vizuālo reprezentāciju, tā teikt, tālāk par divdimensionālām projekcijām uz plaknes ar fiksētu redzes lauku, taču atšķirībā no attēlošanas divu dimensiju vai trīs dimensiju vides uz plakanā ekrāna, kurām, lai mainītu skatu ir nepieciešami papildus kontroliera (piemēram, tastatūras) iespējas, virtuālā realitātē ir iespēja darboties tikai ar dabiskām cilvēka ķermeņa kustībām. [22]

Piemēram, pētījumā [12], par datu attēlojuma izpratni, kurā tika salīdzināti datu projekcijas veidi parastā trīs dimensiju vidē uz plakanā ekrāna ar attēlojumu virtuālā vidē, tika secināts, ka, apstākļos, kad lietotājs atrodas telpiski dabiskā stāvoklī, ko sekmē virtuālo realitāšu atspējošanas ierīces, tas palīdz lietotāja attēlojuma izpratnei par datu projekciju un vizuālā informācija tiek labāk uztverta. Piemērs cilvēka redzes lauka uztveramai zonai izmantojot plakanu ekrānu un virtuālās realitātes brilles ir apskatāms attēlā Nr. 7. Ņemot vērā iepriekš minēto, autori A. Bajarī un E. Tuderanu secināja [12], ka virtuālā realitāte piedāvā dabisku saskarni starp cilvēku un datoru, kas vienkāršos sarežģītas manipulācijas ar datiem. Tas arī sniedz iespēju paļauties uz kombinēto sajūtu mijiedarbību, nevis uz vienu vai pat dominējošu sajūtu.



Attēls Nr.7. **Piemērs cilvēka redzes lauka uztveramai zonai.**

Savukārt, B. Laha un D. Boumans savā darbā [14] secina, ka visefektīvākā un intuitīvāka datu vizualizācija ir jāveic virtuālā vidē. Šī dedukcija ir balstīta uz izpratnes par cilvēka uztveres ierobežojumiem kopsakarā ar projicējamo lielo informācijas daudzumu. Balstoties uz iepriekšminēto pētījumu rezultātiem, tika konstatēts, ka virtuālās realitātes iespējas var pielietot lielo datu attēlošanā, jo būtībā tajā būtiskākos datus var daudz intuitīvāk novietot cilvēka redzes lauka centrālajās zonā, kas ļauj uztvert uzrādīto informāciju īsā laika periodā bez būtiskiem informācijas zudumiem cilvēka uztveres problēmu dēļ.

Tā kā galvenais maģistra darba mērķis ir, izmantojot datu attēlošanu, padarīt nesakārtotus, neskaidrus un abstraktus datus par strukturētu informāciju, kas būtu intuitīvi un viegli uztverama, tad izvēlēts veikt iepriekšējo nodaļu minētos datus attēlot virtuālajā realitātē.

6. UZ PĒC LAIKA GRUPĒTO DATU KOPU BALSTĪTA PIEEJA

Lai pārvarētu iepriekš konstatētās problēmas, kā viens no risinājumiem problēmjautājumiem, kas nodefinēti 3. nodaļā, tiek piedāvāts izveidot analīzes rīku , kas uz digitālo izmeklēšanas datiem skatās kā uz virkni datu kopām, kas, galvenokārt, saistītas uz datņu piekļuves laiku. Autors piedāvā risinājumu, kas ietver strukturētu, standartizētu un formālu datu attēlošanu, kas ietvertu sakarību, balstoties uz laika vienībām, kā prezentējamo informāciju gala lietotājam. Ja iepriekš minētie pētījumi par pieejām un metodikām, kas saistītas ar digitālo izmeklēšanu, sniedza risinājumus datu bāzes izveidei, neatkarīgi no tā, vai tā ir uz ontoloģijām balstīta vai uz *SQL* datu bāzes tabulām, attiecīgais risinājums ir balstīts uz datu klasteru ķēdi, kas savstarpēji tiek pakārtota neatkarīgi no avota, no kā tā tiek iegūta t.i. vai no datu bāzes tipa datnes, vai no noteiktās ierīces datnēm.

Galvenā atšķirība starp šī rīka pieeju un iepriekš minētiem rīkiem ir tā, ka iepriekšējās pieejas darbojas kā neatkarīgs galējs produkts, un paredz, ka lietotāji, šajā gadījumā – izmeklētāji un analītiķi - darbojās primāri ar tiem, lai veiktu digitālo izmeklēšanu. It sevišķi pieejās, kas iesaista ontoloģijas. To gadījumā, tie ir ļoti lietderīgi un spējīgi rīki, kas paredz, ka galējie lietotāji izmantos *OWL2, SPARQL* vai citus vaicājuma valodas, kas balstās uz *RDF*. Taču šī risinājuma, galvenais mērķis ir izveidot rīkus, kas būtu kā sākotnējie analīzes rīki turpmākam darbam ar citiem. Šo rīku galvenā funkcija ir rast izmeklētājiem un analītiķiem zināšanas un priekšstatu par datiem, kurus tie turpmāk analizēs, tādā veidā samazinot nepieciešamo izmeklēšanas laiku.

Ņemot vērā iepriekš minēto, šajā rīkā ir izlaista daļa, kurā izmeklētājiem ir jāveic pierādījumu meklēšana neatkarīgi no tā, vai tie ir vaicājumi pēc *OWL2* [6] ontoloģijas valodas vai *SQL* [5] vaicājumi, jo risinājums veic meklēšanu automātiski, vienīgie dati, kas jāievada lietotājam ir periods, un, ja ir zināms, tam interesējošie atslēgvārdi. Tādejādi, otra šī risinājuma galvenā būtība ir tāda, ka tas rada svarīgāko notikumu attēlojumu ar, pēc iespējas mazāku, ievadi no lietotāja puses, bet tajā pašā laikā nodrošina precīzu informācijas attēlojumu par notikumiem un datu plūsmu, kas izgūta no izmeklēšanas laikā iegūtiem datiem.

6.1. Risinājums datu apjoma problēmai

Līdzīgi kā iepriekš analizētām pieejām, datu izgūšana notiek, izmantojot jau esošus rīkus. Datu atdalīšana un nošķiršana notiek, balstoties uz konkrēto izgūto datņu tipiem, kurā tiek ielādēti diska attēla dati. Autora piedāvātā risinājuma netiek veidota datu bāze. Balstoties uz datņu tipu,

metadati tiek ielasīti no konkrētās datnes nolasot tās no spoguļkopijas izvilktiem datiem, un, atsevišķiem datu tipiem, izmantojot pieejamos analīzes rīkus tiek sagatavoti *.csv* datnes, kas satur izgūšanai nepieciešamo informāciju. Savukārt, izmeklēšanai nevajadzīgie dati, netiek izmantoti.

6.2. Datu analīze un sasaiste

Atlasītas datnes, pēc kritērijiem, kas tiek definēti nākamā šī darba nodaļā, neatkarīgi no avota, tiek ierakstīti *.json* datnē, kas satur informāciju, par pašām datnēm, lai veidotu datu kopas. Ieraksts par vienu datni satur vispārīgu informāciju par izmantotajiem datu avotiem, savukārt, savienojumi starp datnēm satur informāciju par notikumiem un datu plūsmām, kas nodrošina, ka pašas datnes ir sagrupētas *klasteros* pēc to sasaistes. Detalizēts apraksts, kā tiek veidotas sasaistes ir aprakstīts nākamā nodaļā.

Pašas datņu sasaistes pēc analīzes līmeņa sevī ietver atribūtu, kas pamatā ir definētu nosacījumu kopa. Tie ir nosacījumi, saskaņā ar kuriem tiek izveidota un grupēta datne ar pārējām datnēm. Datņu kopa pēc vērtībām tiek noteikta, pamatojoties uz svarīgākajiem notikumiem datu plūsmā, piemēram, e-pasta ziņojumu, kurā ir atslēgvārds. Nosacījumu ķēde, pie piepildīšanās sevī ietver sasaistes līmeni, ar kuru tiek savienoti dati, kas ir iepriekš definēts, pamatojoties uz četru līmeņu taksonomiju, kurus savā darbā [5] definēja autori K.Džeims un H.Dž. Patersons. Viņi izdalīja četru līmeņu taksonomiju, kurā datu plūsmā tiek šķirta nebūtiska informācija, ko analītiķim var nodot paši izsekošanas pierādījumi – centrālais avota fails (1. līmenis), atsauces paraugi – faili, kas atvasināti no avota faila vai paralēli izveidoti (2. līmenis), lietas informācija - atslēgvārdi (3. līmenis) un ar organizāciju un sistēmu saistīti procesi (4. līmenis).

6.3. Risinājums datu neviendabīguma problēmai

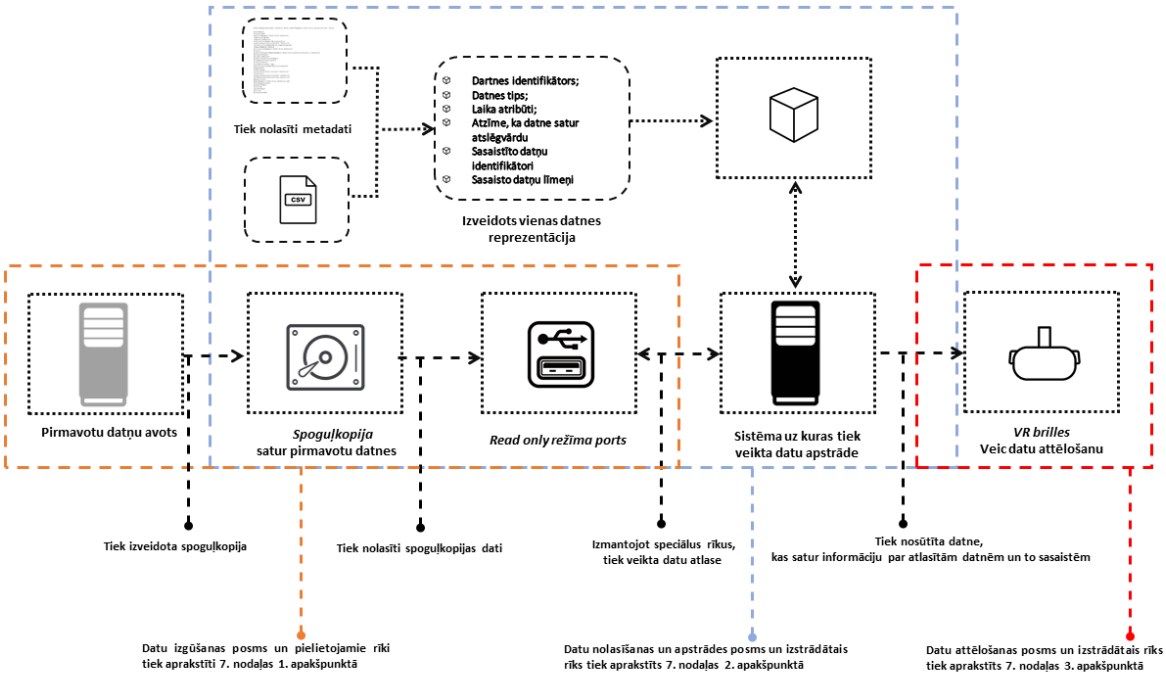
Augstākminētais datu strukturēšanas veids nodrošina datu kanonisku attēlojumu, vienlaikus veidojot polimorfisma ķēdi, kuru savā darbā [6] kā nozīmīgu būtību atzīmēja autori J.Čabots, A. Beatriksa, K. Nikola, T. Kečadi. Un tā kā avoti ir neviendabīgi, šāda veida datņu sasaistes struktūras izveide no datu bāzes viedokļa rada līdzīgu datu struktūras kopu, jo to aizpilda tikai vienāda atribūta ieraksti. Tā kā avoti ir neviendabīgi, datnēm tiek piešķirti laika notikuma intervāla specifiski atribūti, kā arī datu kopīgās vispārīgās īpašības. Piemēram, kā autori K.Džeims un H.Dž. Patersons savā darbā [5] atzīmēja, ja diskā ir *Apache* serveris un pārlūkprogramma, ir grūti analizēt šādus notikumus un manuāli meklēt savienojumu starp tiem, jo katrai datu plūsmai ir specifiskas īpašības. *Apache* serveris savienojuma laikā satur informāciju par IP adresēm, savukārt pārlūkprogramma veic pārlūkprogrammai specifiskas darbības, t.i., *Apache* servera

darbības nav saistītas ar pārlūkprogrammu. Taču šiem notikumiem ir ar informāciju saistītas īpašības, piemēram, notikuma laiks. Līdz ar to autori norāda, ka teorētiski visi notikumi pieder vienam un tam pašam jēdzienam, kas ir vispārīgs notikuma jēdziens. Polimorfisms rada konsekventu visu notikumu attēlojumu, nenoņemot konkrētos datus, kas ir raksturīgi katram notikumam, bet saglabā sasaistēs šos datus laika līmenī.

Šāda veida pieeja rada datu analīzes funkciju, kas tiek veikta soli pa solim un, vienlaikus, nodrošina minēto datu attēlošanu tādā veidā, ka izmeklētājiem un analītiķiem faktiski nav nepieciešamība ieskatīties konkrētās datnes saturā, bet tie var gūt vispārīgu izpratni par datu plūsmu, tikai aplūkojot klasteru, ko izveido tajā iesaistītās datnes. Turklāt, šāda veida datu segregācija automātiski atmet liekos datus (t.i. sistēmas procesus vai programmas, kas tika palaistas, bet ne vienmēr ir svarīgas, piemēram, videospēles (izņemot, ja pastāv .log dati spēles iekšējai tērzēšanai). Kā jau minēts iepriekš, šī pieeja tiek veidota tā, lai, lietotājam aplūkojot datus virtuālās realitātes vidē, viņš/viņa redzētu skaidru priekšstatu par datnēm, kuriem tika piekļūts un kā tie tika saistīti bez ielūkošanās pašās avotu datnēs.

6.4. Risinājums likumiskām prasībām

Lai izpildītu stingrās juridiskās prasības, datnes, kuras nepieciešamas analīzei tiek nolasītas izmantojot speciālu, digitālai izmeklēšanai, paredzētu iekārtu izmantojot tikai lasāmo (*read only)* režīmu, tādejādi nodrošinot datu nemainīgumu un vienlaicīgi saglabājot autoru J.Čabots, A. Beatriksa, K. Nikola un T. Kečadi minēto uzraudzības ķēdi [3], t.i., ka datnes saturs netiek mainīts. Šāda metodika nodrošina ātru veidu, kā tiesā pierādīt, ka diska spoguļkopijas pirmavota datne nav bijusi labota, jo šajā gadījumā konkrētās datnes ticamību var viegli pamatot, salīdzinot pirmavota datnes kontrolsummu no spoguļkopijas ar oriģinālu.



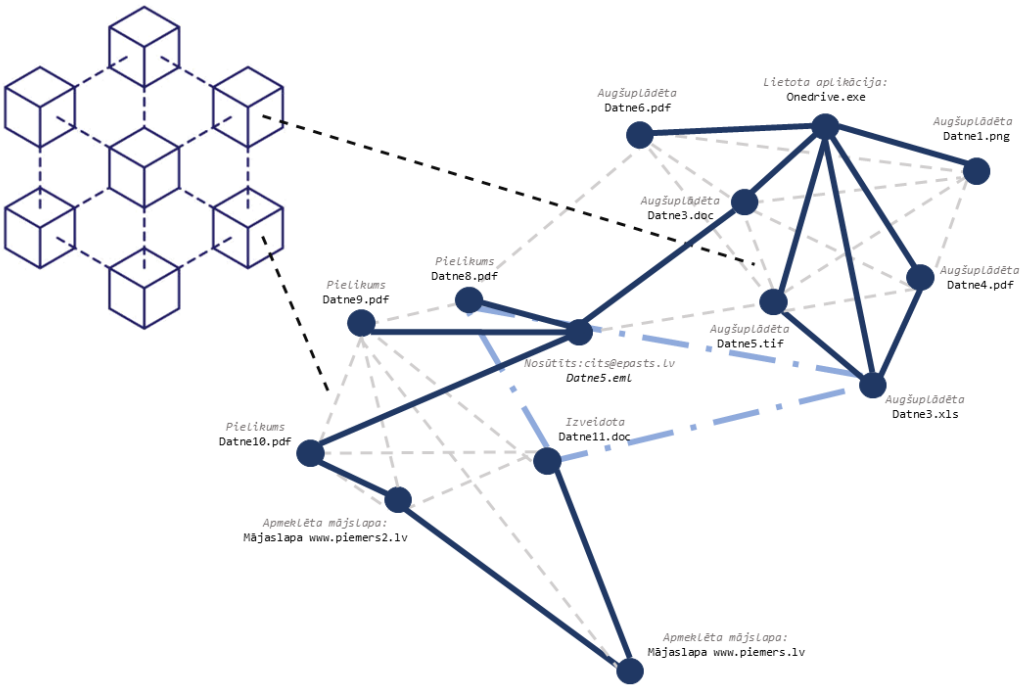
Attēls Nr.8. **Datu virzības diagramma starp ierīcēm**

6.5. Datnes ieraksta izveidošana

Diagrammu par to, kuri dati izveido datnes atribūtus attēlošanai kā arī ierakstu izveidošanas procesu diagrammu, var apskatīt augstāk attēlā Nr. 8. Lai attēlotu vienu datni kā virsotni grafā, tai ir jāsatur sekojošus ievades datus. Tajā tiek ierakstītas konkrētās datnes metadati. Pēc tam tiek ierakstīti saistīto datņu identifikatori, kas ir atribūti, kuri tiek izveidoti balstoties uz nosacījumiem, kas tiks aprakstīti nākamajā nodaļā un kas ir vērtība, pēc kuras tiek izveidotas sasaistes ar citām datnēm. Tiek ierakstīta vērtība, kas norāda vai datnē figurē atslēgvārds. Papildus vienā blokā tiek ierakstīts grupēšanas nosacījuma jeb sasaistes līmeņa vērtība, kas tika definēta autoru I. E. Drors (*Itiel E. Dror*), V.K. Tompsons (*William C. Thompson*), K. A. Meisners (*Christian A. Meissner*) u.c. darbā. [23]. Šie dati kopā ar sasaistes datiem kalpo kā ievades vērtības notikumu laika joslas un datu plūsmas attēlošanai.

6.6. Datu attēlošana

Balstoties uz iepriekšējā nodaļā secināto un tā kā datu plūsma un notikumi veido tīklu, kas pats par sevi sastāv no datnēm, kuras ir grupētas klasteros, pamatojoties uz grupēšanas nosacījumu, kas ir definēts starp bloku sakarībām (detalizēts nosacījumu kopums tiek aprakstīts nākamajā nodaļā), kas veidotas balstoties uz metadatu atribūtiem, datu vizualizācija tiek veikta, izmantojot grafa tipa attēlojumu. Saistīto divu datu grupu diagrammas attēlojuma paraugs ir skatāms attēlā Nr. 9.



Attēls Nr.9. **Saistītu datu attēlojums divu datņu grupu kontekstā.**

Augstāk redzamā piemēra attēlā ir redzama datu kopu attēlošana divām datu grupām. Pelēkā krāsā ir norādīta informācija, kuru izmeklētājs var potenciāli secināt, jo tā ir netieša. Piemēram, grupā, kas ietver bloku par datni *Onedrive.exe* var secināt, ka tika izmantots *Microsoft* serviss, kas veica rezerves kopijas mākonī. Savukārt otrā blokā ir redzams, ka figurē *Datne5.eml*, pēc kuras var secināt, ka uz ierīces pastāv e-pasts ar saistītām datnēm, kas visdrīzāk liecina par to, ka tās ir šī e-pasta pielikumi. Papildus, blokā ir redzams mājaslapu apmeklējums, ar kurām ir saistīta vēl viena pielikuma datne, līdz ar to var secināt, ka lietotājs ir vai nu skatījies mājaslapā informāciju ar kuru piepildīt dokumentu, vai arī to ir lejuplādējis no vienas no uzrādāmām mājaslapām.

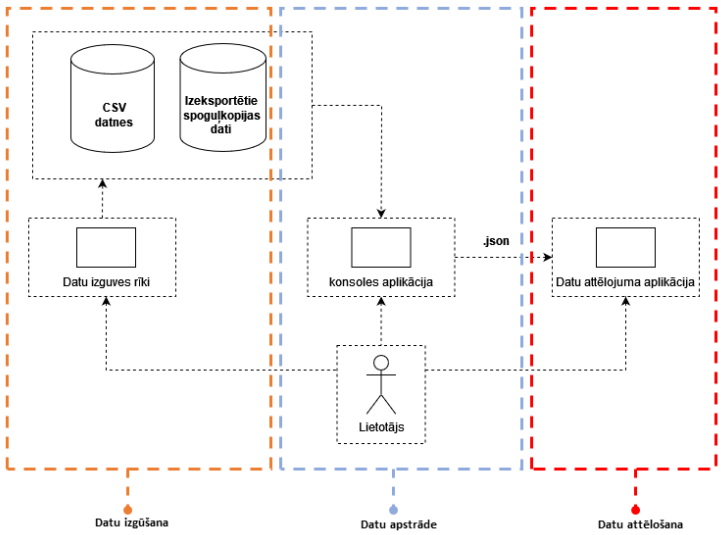
Attēlā redzamo divu datu kopu izkārtojums darbojas balstoties uz spēka virzīta grafu (*force-directed graph*) algoritmu, kas detalizētāk tiek aprakstīts nākamā nodaļā. Balstoties uz iepriekš minēto algoritmu, saistītās datņu kopas tiek novirzītas viena otrai tuvāk, savukārt nesaistītās datņu kopas tiek novirzītas tālāk, balstoties uz sasaistē esošo sasaistes līmeņa atribūtu.

Attēlā redzamās, stiprās sasaistes (attēlā apzīmētas ar tumši zilām līnijām) ir tās, balstoties uz kurām, darbojas pievilkšanās spēks, savukārt vidēja tipa sasaistes pievilkšanās spēku neietekmē (attēlā apzīmētas ar punktētām, pelēkām līnijām), taču starp tām tiek novilktas līnijas. Gaiši zilā

krāsā tiek attēlotas sasaistes, kur datnes sevī satur atslēgvārdus. Šis vizualizēto saistību līmeņu projicējums ir definēts autoru M. Kvāna, R. Overila, K.P. Čova, H. Tse, F. Lau un P. Lai darbā, kurā tiek pielietota pielāgota *FORE* metodoloģija līmeņu noteikšanā. [24]

7. IZSTRADĀTO RĪKU APRAKSTS

Šajā nodaļā tiek aprakstīts izstrādātais rīks, balstoties uz pētījumu analīzi, kas tika veikta iepriekšējās nodaļās. Rīks sastāv no divām izstrādātām programmām un jau esošo programmu izmantošanu datu ieguvei. Izstrādātās programmas ir izstrādātas uz *Microsoft Visual Studio* integrētās izstrādes vides *C#* valodā. Pirmā izstrādātā programma ir atbildīga par datu apstrādi. Programmā tiek izmantotas iebūvētās *Microsoft* bibliotēkas. Šajā programmā tiek nolasīti dati, noteiktas sasaistes starp tiem un izveidota *.json* datne, kuru ielādē otrā programmā. Otrā izstrādātā programma ir izstrādāta uz *Unity* izstrādes platformas (2020.3.23f1 versijas) priekš *Oculus Quest 2* virtuālās realitātes brillēm, izmantojot *Oculus* programmatūras izstrādes komplektu. Datu ieguve tiek veikta izmantojot jau esošus rīkus. Izmantoto un izstrādāto rīku procesu diagramma ir attēlota zemāk attēlā Nr. 10.



Attēls Nr. 10. **Izmantoto un izstrādāto rīku procesu diagramma**

Šīs nodaļas struktūra ir sekojoša. 7.1. apakšnodaļā ir aprakstīti esošie datu izgūšanas rīki. 7.2. apakšnodaļā tiek aprakstīta izstrādātā datu analīzes programma. 7.3. apakšnodaļā tiek aprakstīta izstrādātā datu attēlošanas programma.

7.1. Datu izgūšana

Dati tika izgūti izmantojot jau esošos rīkus. Programma *FTK Imager* tiek izmantota cieto disku spoguļkopiju izveidošanai un datu eksportam. *FTK Imager* veic spoguļkopijas izveidi *.E01* formātā, pēc tam no esošās spoguļkopijas tiek eksportēti dati, tādejādi garantējot oriģinālo datu integritāti pret visām iespējamām manipulācijām. *FTK Imager* izeksportētie dati glabājās mapē pēc

sistēmas kādas tās tika glabātas pirmavotu datņu datorā. Eksportētās datnēs tiek apstrādāti tikai konkrētie datņu tipi. Datņu tipi, kuri tiek apstrādāti ir aprakstīti šīs nodaļas 2.2. apakšpunktā.

Programma *Vound Intella* tiek izmantota atslēgvārdu meklēšanai un pārlūkprogrammas datu, kā arī e-pastu datņu sarakstu izgūšanai. Programmā tiek ielādētas ar *FTK Imager* izveidotās spoguļkopijas vai spoguļkopija. Papildus, ar programmu *Vound Intella* spoguļkopijās tiek meklētas datnes pēc atslēgvārdiem. Spoguļkopijām, kurās tika atrastas datnēs saturošos atslēgvārdus tiek izveidots saraksts ar to metadatiem un izeksportēts *.csv* datnes formātā, kas satur informāciju par šīm datnēm. Informācija, kas tiek izgūta katrā atsevišķā *.csv* datnē ir attēlota attēlā Nr.11.

email.csv un ***emailKeywords.csv***					
Item ID	Subject	File Name	From	To	Primary Date

browser.csv un ***browserKeywords.csv***			
Item ID	Subject	File Name	Primary Date

Keywords.csv
Location

Attēls Nr. 11. **Datu atribūti ar kuriem tiek izveidotas *.csv* datnes *Vound Intella* programmā**

Email.csv un *emailKeywords.csv* datnes satur sekojošus atribūtus:

- *Item ID* ir unikāls identifikators konkrētam ierakstam *Vound Intella* programmā.
- *Subject* ir e-pasta tēma.
- *File name* ir datnes, kurā dabūts ieraksta nosaukums. Tas tiek izmantots gadījumā, ja neeksistē atribūts *Subject*;
- *From* ir e-pasta nosūtītāja e-pasta adrese;
- *To* ir e-pasta saņēmēja e-pasta adrese;
- *Primary Date* ir datums un laiks, kad e-pasts tika atvērts, gadījumā, ja analizētās ierīces lietotājs ir saņēmējs vai *Primary Date* ir datums un laiks, kad e-pasts tika nosūtīts, gadījumā, ja analizētās ierīces lietotājs ir nosūtītājs.

browser.csv un *browserKeywords.csv* datnes satur sekojošus atribūtus:

- *Item ID* ir unikāls identifikators konkrētam ierakstam *Vound Intella* programmā;
- *Subject* ir mājaslapas nosaukums;
- *File name* ir datnes, kurā dabūts ieraksts nosaukums. Tas tiek izmantots gadījumā, ja neeksistē atribūts *Subject*;

- *Primary Date* ir datums un laiks, kad konkrētā mājaslapa tika apmeklēta.

Keywords.csv datne satur sekojošus atribūtus:

- *Location* ir datnes atrašanās vieta pirmavota datorā.

Tālāk ar *FTK Imager* izveidoto datņu eksportu, kā arī *Vound Intella* izeksportēto *.csv* datņu apstrādi veic izstrādātais datu analīzes un saistību izveidošanas programma.

7.2. Datu analīzes un saistību izveidošanas programma

Šī izstrādātā programma ir atbildīga par izgūto datņu analīzi un apstrādi. Programma ir konsoles lietotne. Programma ir paredzēta, lai nolasītu no datnēm metadatus, izveidotu starp datnēm sasaistes, piešķirtu tiem sasaistes stipruma līmeņus un sagatavotu visu šo informāciju, lai to varētu nolasīt Otrā programma. Lai darbinātu programmu, lietotājam ir nepieciešams to novietot blakus interesējošai failu sistēmas mapei un norādīt periodu, un ja eksistē atslēgvārdi norādīt to atrašanās vietu.

7.2.1. Izgūto datu nolasīšana un atslēgvārdu salīdzināšana

Konsoles lietotne atlasa un nolasa tikai noteiktos datņu tipus *FTK Imager* spoguļkopijas izeksportētos datos. Konkrētie atlasāmie datņu tipi ir uzskaitīti nākamajā apakšpunktā. Papildus, ja eksistē atslēgvārds *Vound Intella* izeksportētā *Keywords.csv* datnē, tad tiek salīdzināta atlasīto datņu atrašanās vieta ar datnē *Keywords.csv* kolonnas *Location* vērtībām.

Savukārt, pārlūkprogrammu un e-pastu atribūti tiek nolasīti no *Vound Intella* programmā izveidotām datnēm – *browser.csv* un *email.csv*. Līdzīgi, kā ar iepriekš minētām noteiktām datu kategorijām, ja eksistē datnes *emailKeywords.csv* un *browserKeywords.csv,* tad tiek salīdzinātas kolonnu *Item Id* vērtības starp *.csv* datnēm – e-pastiem starp *emailKeywords.csv* un *email.csv* kā arī starp *browser.csv* un *browserKeywords.csv*. Šīs salīdzināšanas darbības nozīme, kā arī no tās iegūto vērtības ir aprakstītas 7.2.0. apakšnodaļā.

7.2.2. Datu sadalījums pēc tipiem

Pirms risinājuma izstrādāšanas tika veikta aptauja, kurā piedalījās 27 izmeklētāji. Aptaujas izmeklētājiem tika lūgts identificēt visvairāk tiem interesējošās datnes. Balstoties uz izmeklētāju atbilžu apkopojuma, tika izdalīti konkrētie datņu tipi, kurus nolasīs šī programma, tādejādi nodrošinot, ka tiktu atlasītas tikai izmeklētājiem interesējošie datņu veidi, vienlaikus, izslēdzot

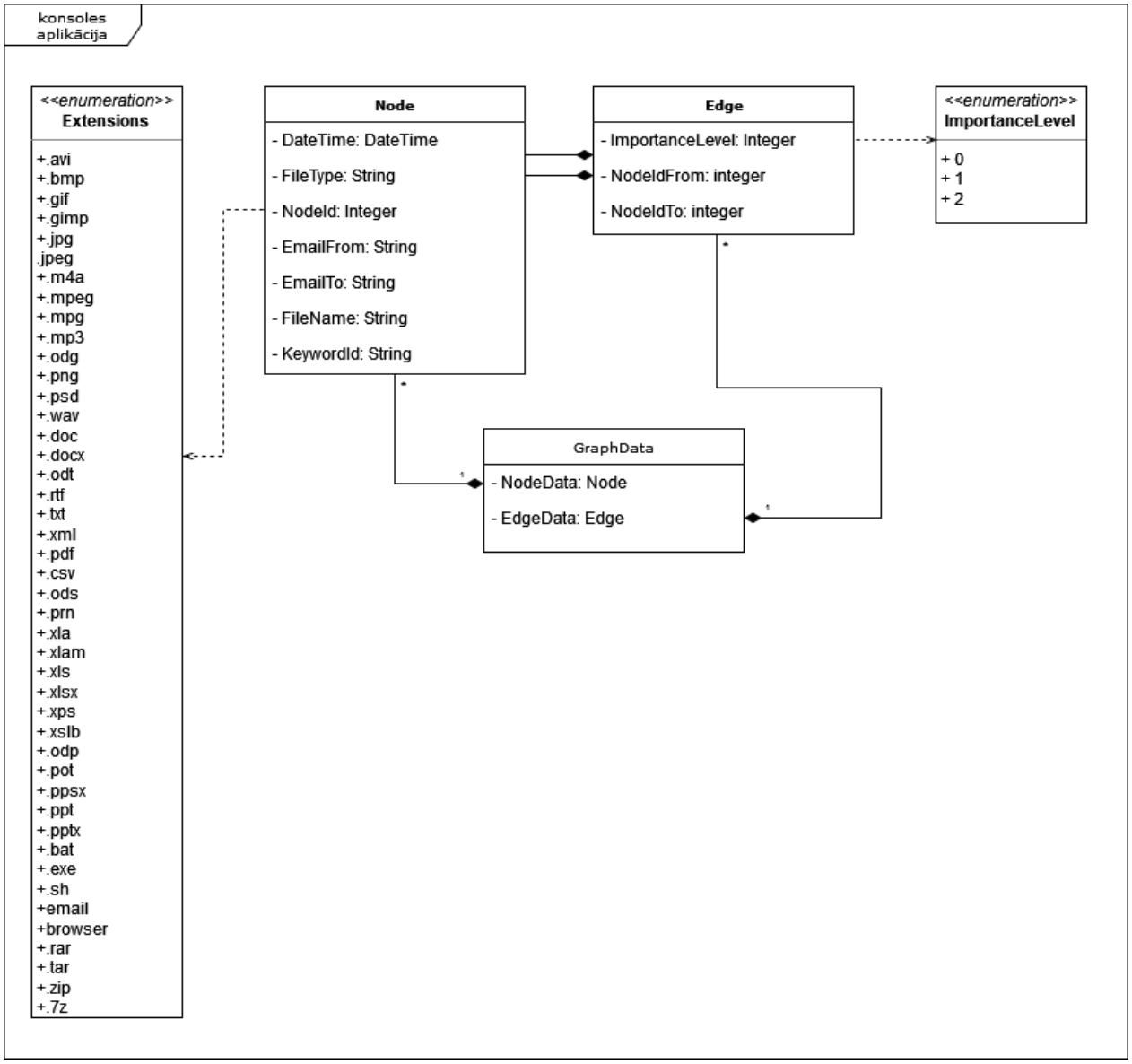
izmeklētājus neinteresējošo informāciju. Izgūtās datnes, tika sadalītas 9 grupās balstoties uz to pagarinājumiem:

1) Datnes ar pagarinājumiem *.avi, .bmp, .gif, .gimp, .jpg, .jpeg, .m4a, .mpeg, .mpg, .mp3, .odg, .png, .psd, .wav,* tiek grupētas zem mēdiju tipa datņu (*mediaNodeTypes*) kategorijas;
2) Datnes ar pagarinājumiem *.doc, .docx, .odt, .rtf, .txt, .xml,* tiek grupētas zem teksta tipa datņu (*textNodeTypes*) kategorijas;
3) Datnes ar pagarinājumiem *.pdf,* tiek grupētas zem pdf tipa datņu (*pdfNodeTypes*) kategorijas;
4) Datnes ar pagarinājumiem *.csv, .ods, .prn, .xla, .xlam, .xls, .xlsx, .xps, .xslb*, tiek grupētas zem tabulu tipa datņu (*excelNodeTypes*) kategorijas;
5) Datnes ar pagarinājumiem *.odp, .pot, .ppsx, .ppt, .pptx* tiek grupētas zem prezentāciju tipa datņu (*presentationNodeTypes*) kategorijas;
6) Datnes ar pagarinājumiem *.bat, .exe, .sh* tiek grupētas zem izpildprogrammu tipa datņu (*executableNodeTypes*) kategorijas;
7) E-pastu datnes tiek nolasītas no *email.csv*, kur tām piešķirta e-pastu (*emailNodeTypes*) kategorijas;
8) Pārlūkprogrammu datnes tiek nolasītas no *browser.csv*, kur tām piešķirta pārlūkprogrammue-pastu (*browserNodeTypes*) kategorijas;
9) Datnes ar pagarinājumiem *.rar, .tar, .zip, .7z* tiek grupētas zem arhīva tipa datņu (*archiveNodeTypes*) kategorijas;

Visas datņu kategorijas izņemot e-pastu datnes (*emailNodeTypes*) un pārlūkprogrammu datnes (*browserNodeTypes*) tiek nolasītas no *FTK Imager* izeksportēto spoguļkopiju datiem, savukārt, augstāk minētās atsevišķās kategorijas tiek nolasītas no attiecīgās *.csv* datnes.

7.2.3. Programmas datu struktūru apraksts

No augstāk aprakstīto datņu metadatiem, programmā tiek izveidota grafa datu struktūra, kura ir aprakstīta zemāk attēlotā klašu struktūras diagrammā (*Attēls* Nr. 12).



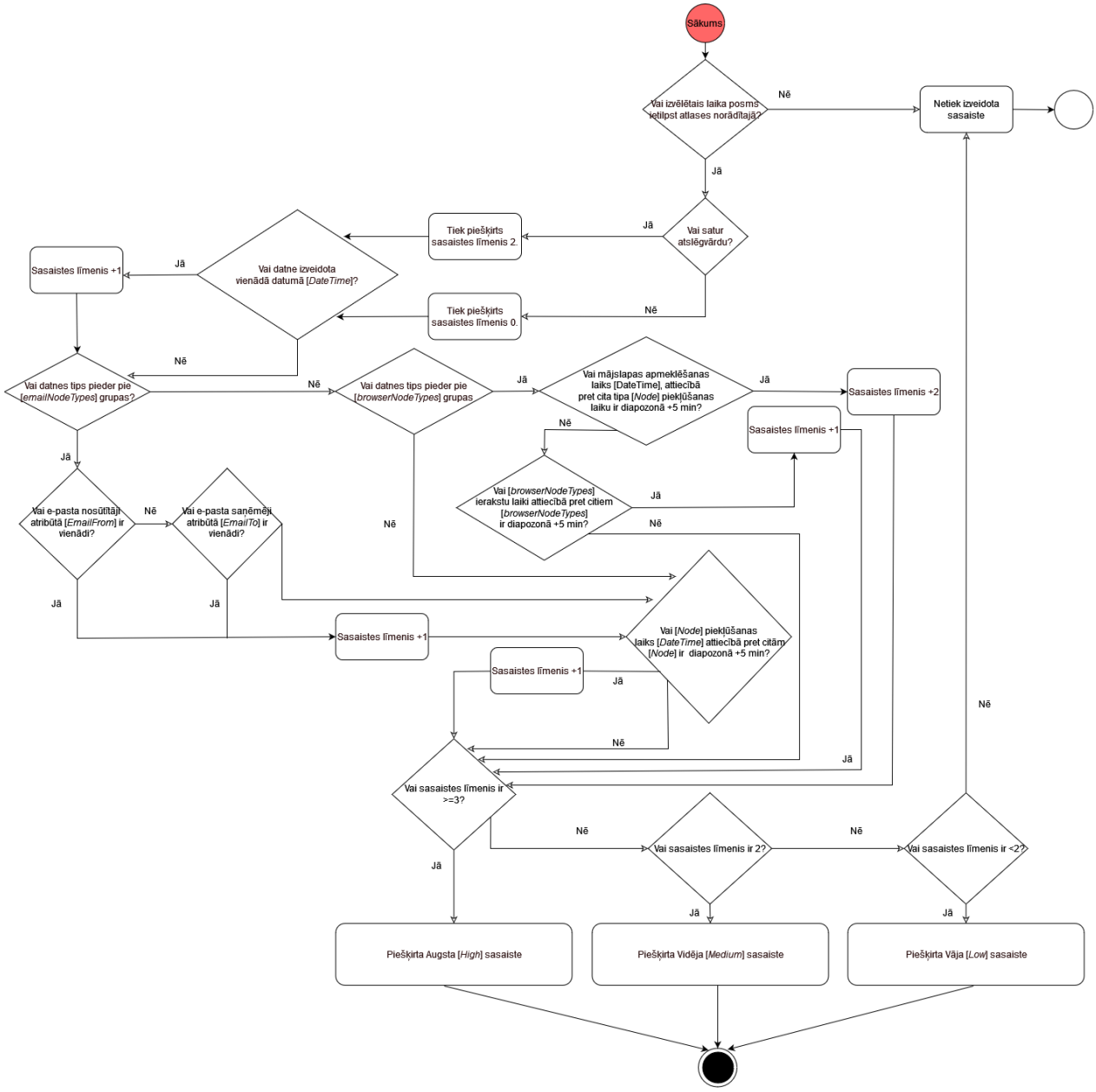
Attēls Nr. 12. **Attēlojamo datņu un šķautņu klašu diagramma**

7.2.4. Saistības līmeņu piešķiršana

Lai izveidotu sasaistes starp datnēm tika nodefinēti vairāki kritēriji, balstoties uz kuriem tiek veikta iepriekš minēto nolasīto datņu salīdzināšana starp visām datnēm, to ievietošanai grafā, kurā tās tiek attēlotas kā virsotnes. Papildus, tiem tiek piešķirts sasaistes līmenis balstoties uz nosacījumu atbilstību, kas grafā tiek attēlotas kā šķautnes. Atkarībā no šī līmeņa, tiek piešķirts attiecīgais sasaistes līmenis. Nosacījumi sasaistes izveidei ir attēloti datņu salīdzināšanas diagrammā – attēlā Nr. 13.

Nosacījumu definēšana tika veikta balstoties uz šīs nodaļas 2.2. apakšpunktā minēto izmeklētāju aptauju, kuras rezultātā, balstoties uz atbilžu variantiem, tika secināts, ka viens no

primāriem izmeklētāju atlasāmiem kritērijiem, ir iespēja redzēt iesaistītos datus noteiktā, viņiem interesējošā posmā. Balstoties uz iepriekš minēto, kā pirmais kritērijs datņu atlasei ir posma izvēle. Savukārt, turpmāko sasaistes līmeņu piešķiršana notiek tikai tām datnēm, kuras atbilst izvēlētajam laika perioda intervālam.



Attēls Nr. 13. **Datņu salīdzināšanas diagramma**

Balstoties uz augstāk norādītajiem kritēriju atbilstības rezultātiem tiek izveidots datņu ierakstu saraksts un to saistību saraksts, atbilstoši klases diagrammai, kas minēta 7.2.3.

apakšnodaļā. Grafā datnes tiek attēlotas kā virsotnes un sasaistes tiek attēlotas kā šķautnes. No tām tiek izveidota objektu *.json* datne, kura aprakstīta šīs nodaļas 2.6. apakšpunktā.

7.2.5. Pārlūkprogrammu datņu apstrāde

Ar mērķi, lai samazinātu lielo daudzumu pārlūkprogrammas datu attēlojumu, kā atsevišķo attēlojamo datņu Otrā programmā, tika iestrādāts nosacījums, ka, ja pārlūkprogrammas apmeklētās lapaspuses pirmo 25 simbolu virkne sakrīt ar nākamās apmeklētās lapaspuses pirmo 25 simbolu virkni un tās secīgi atrodas laika diapazonā līdz 5 minūtēm, tad tās tiek grupētas zem vienas datnes attēlojumā. Papildus, lai atstātu saprotamu datnes nosaukuma attēlojumu un samazinātu attēlojamo simbolu virkni otrajā programmā, pārlūkprogrammas datnēm tika samazināts nosaukums uz līdz 25 simboliem. Piemērs pārlūkprogrammas tipa datņu apstrādei ir apskatāms zemāk attēlā Nr. 14.

Attēls Nr. 14. **pārlūkprogrammas datņu tipa apstrāde**

7.2.6. Datnes .json apraksts

Datnes *.json* shēma ir attēlota attēlā Nr. 15. Datnes *.json* shēma ir veidota pēc "*json schema*" vadlīnijām. [25] Savukārt divu datņu un vienas sasaistes ierakstu piemērs *.json* datnē ir attēlots attēlā Nr. 16, kurā ir redzami divi datņu ieraksti, kas izveido divas virsotnes (*NodeData* grupa) grafa attēlojumā un viens sasaistes ieraksts, kas izveido starp augstāk minētām datnēm šķautni grafa attēlojumā (*EdgeData*).

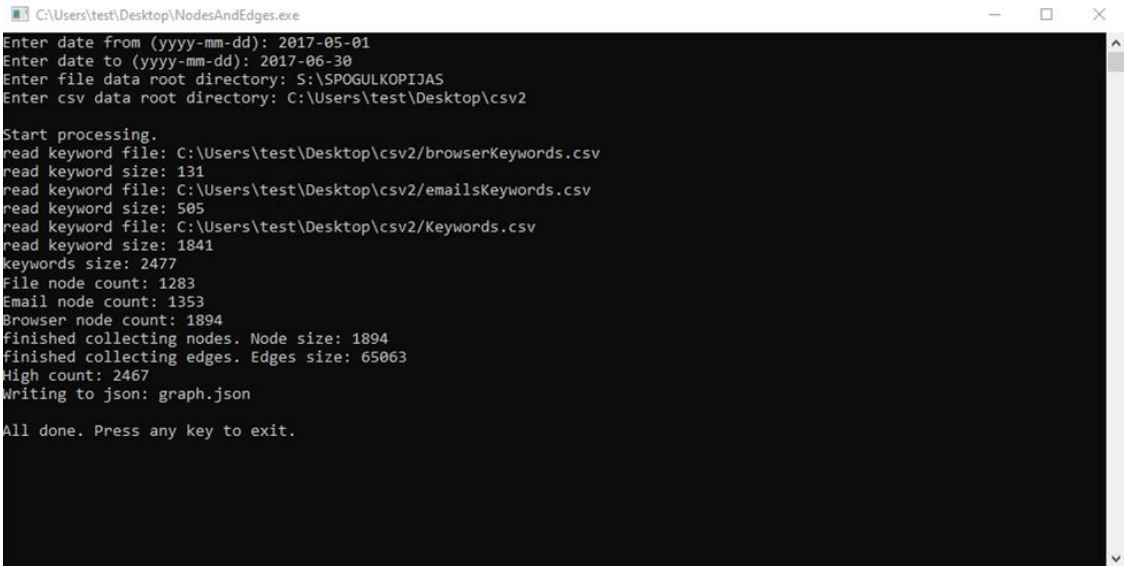
Attēls Nr. 15. **Datnes *.json* shēma**

Attēls Nr. 16. **Datnes .json ierakstu piemērs**

Konsoles lietotnes izveidoto *.json* datni var uztvert kā integrāciju starp konsoles lietotnes un datu attēlošanas lietotni. Pēc tam, kad tā tiek uztaisīta, to ir nepieciešams ievietot nākamās programmas *Unity* projektā, kurā tā tiek nolasīta un tiek veikta datu attēlošana.

7.2.7. Konsoles lietotnes darbināšana

Konsoles lietotne darbināšanas laikā tiek lūgts ievadīt meklējamo periodu norādot meklējamā perioda sākuma un beigu datumus, kā arī tiek lūgts norādīt vietu, kur datnes tiks meklētas un norādīt mapi, kurā atrodas *.csv* datnes priekš e-pastu un pārlūkprogrammu metadatu nolasīšanas, kā arī atslēgvārdu *.csv*, ja tādi eksistē, datņu nolasīšanas. Veiksmīgas programmas izpildes gadījuma piemērs ir attēlots attēlā Nr. 17.



Attēls Nr. 17. **Konsoles lietotnes darbināšanas piemērs**

7.3. Datu attēlošanas programma

Šī izstrādātā programma ir atbildīga par iepriekš analizēto datņu rezultātu attēlošanu. Programma ir izstrādāta priekš lietošanas ar *Oculus Quest 2* virtuālās realitātes brillēm. Tā ir paredzēta, lai nolasītu otrās programmas izveidoto *.json* datnes virsotņu un šķautņu īpašības un balstoties uz tām izveidotu to attēlojumu.

7.3.1. Vidē atrodamie objekti

Lietotāja reprezentācijai virtuālā telpā tiek izmantoti *Oculus* programmatūras izstrādes komplektā [26] atrodamie iepriekš konfigurētie un atkārtoti lietojamie *Unity* objekti (*prefab*) *OVRCameraRig* [26] – kamerai un *PlayerController* [26] – lietotāja modelim. [26] Projektā izmantotie materiāli (*assets*) galds un grīda, kā arī to tekstūra tika izveidoti *Blender* grafikas programmatūras rīku komplektā un ieimportēti *Unity* projektā (datne *interjers.blend*). Virsotnes, līnijas un to materiāli (krāsas un gaisma) tika izveidoti *Unity*. Trešo pušu izmantotie materiāli projektā ir debesu (*Skybox*) materiāls *Customizable skybox* [27].

7.3.2. Vidē attēlojamo datu nošķiršana

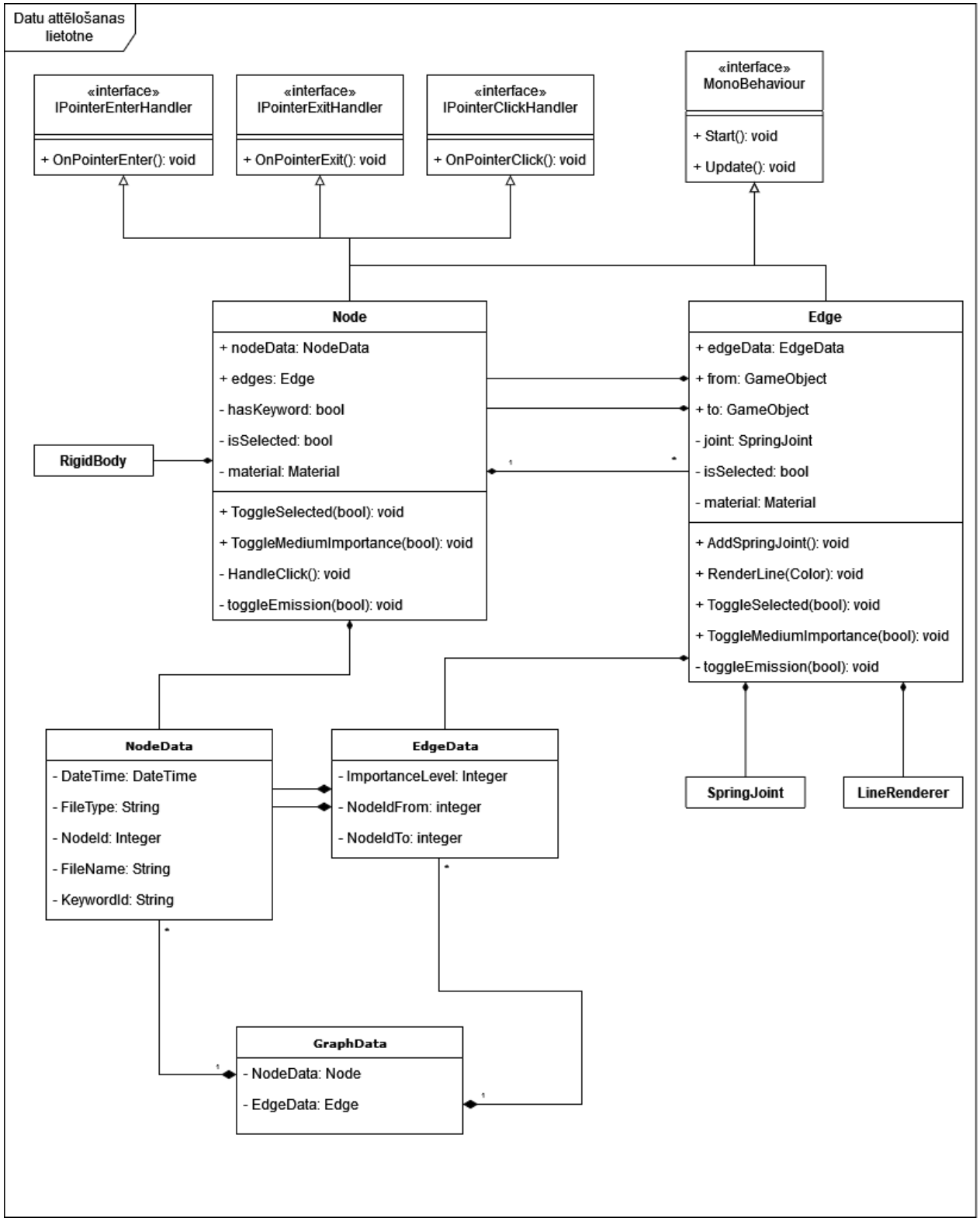
Vidē virsotnes tiek attēlotas dažādās krāsās. Virsotņu krāsas, tiek attēlotas balstoties uz konkrēto datņu tipu kategoriju, kas tika definētas 7.2.2. nodaļā:

1) Mēdiju tipa datņu kategorijas datnes tiek attēlotas ar virsotnēm dzeltenā krāsā;
2) teksta tipa datņu kategorijas datnes tiek attēlotas ar virsotnēm tumši zilā krāsā;
3) *Pdf* tipa datņu kategorijas datnes tiek attēlotas ar virsotnēm sarkanā krāsā;
4) Tabulu tipa datņu kategorijas datnes tiek attēlotas ar virsotnēm tumši zaļā krāsā;
5) Prezentācijas tipa datņu kategorijas datnes tiek attēlotas ar virsotnēm gaiši zaļā krāsā;
6) Izpildprogrammu tipa datņu kategorijas datnes tiek attēlotas ar virsotnēm violetā krāsā;
7) E-pastu kategorijas tipa datnes tiek attēlotas ar virsotnēm gaiši zilā krāsā;
8) Pārlūkprogrammu tipa datnes tiek attēlotas ar virsotnēm oranžā krāsā;
9) Arhīva tipa datņu kategorijas datnes tiek attēlotas ar virsotnēm brūnā krāsā;

Vidē šķautnes tiek attēlotas dažādās krāsās. Šķautņu krāsas, tiek attēlotas balstoties uz konkrēto sasaistes līmeni:

1) Stiprs sasaistes līmenis tiek attēlots gaiši dzeltenā krāsā;
2) Vidējais sasaistes līmenis tiek attēlots gaiši zilā krāsā;

7.3.3. Programmas datu struktūru apraksts



Attēls Nr. 18. **Attēlojamo virsotņu un šķautņu klašu diagramma**

Grafa zīmēšana tiek uzsākta nolasot *.json* objektu. Virsotnes tiek zīmētas pēc nejaušības principa fiksēta izmēra trīs dimensiju telpas daudzskaldnī. Starp virsotnēm tiek zīmētas šķautnes. Ja starp šķautnēm ir izveidojusies stipra līmeņa sasaiste, tad starp tām tiek izveidots *SpringJoint Unity* objekts un uzzīmēta *LineRender* līnija. Par *SpringJoint Unity* objekta darbību tiek detalizēti aprakstīts nākamā apakšnodaļā. Savukārt, ja ir izveidojusies vidēja līmeņa sasaiste, starp virsotnēm tiek uzzīmēta *LineRender* līnija, taču netiek izveidots *SpringJoint Unity* objekts. Šī līnija tiek izveidota, bet ir neaktīvā formā un tā parādās tikai tad, kad lietotājs nospiež uz konkrēto virsotni. Stringrā saisaišu gadījumā, tiek uzzīmētas visas saistītās šķautnes un virsotnes, savukārt, vidēja līmeņa sasaistēm šķautnes tiek uzzīmētas tikai no konkrētās virsotnes. Vāja sasaistes līmeņa šķautnes netiek zīmētas.

7.3.4. Spēka virzīts grafs

Unity iespējas spēka virzītam grafa attēlojumam ir apzinātas M. Dakvīna (*M. d'Aquin*) rakstā. [28] Izmantojot rakstā minētās *Unity* izstrādes platformas iespējas, tika implementēta datu attēlošana spēku virzītā grafā. M. Dakvīns savā rakstā [28] aprakstīja, ka grafa izkārtojums tiek panākts balstoties uz to, ka katrai virsotnei, kas tiek attēlota vidē, tiek piešķirtas divas fiziska objekta raksturīpašības:

1) Virsotnes, kas savā starpā ir savienotas ar šķautni pievelk viena otru;

2) Virsotnes, kas ir tuvu vienai otrai, atgrūž viena otru.

Unity izstrādes platforma piedāvā iespēju piešķirt objektiem noteiktas fiziskas īpašības. Balstoties uz iepriekš minēto, katrai šķautnei, tiek piešķirts *Unity SpringJoint* objekts, kas savieno divas virsotnes kopā, bet ļauj attālumam starp tām mainīties atdarinot mijiedarbību it kā tās būtu savienotas ar atsperi. Virsotnēm, tiek piešķirts *Unity Rigidbody*. *Rigidbody* ir *Unity* piedāvātā iespēja, kas ļauj virsotnēm uzvesties atbilstoši vidē esošai fizikai. Tādā veidā viena virsotne, kas saistās ar otru pievelkas viena otrai tuvāk caur škautnē esošo *SpringJoint* objektu. Šī veida fizika tiek piešķirta tikai tām šķautnēm, kurām ir augsts sasaistes līmenis, saskaņā ar 7.2.4. apakšpunktā attēloto datu salīdzināšanas diagrammu(Attēls Nr. 13).

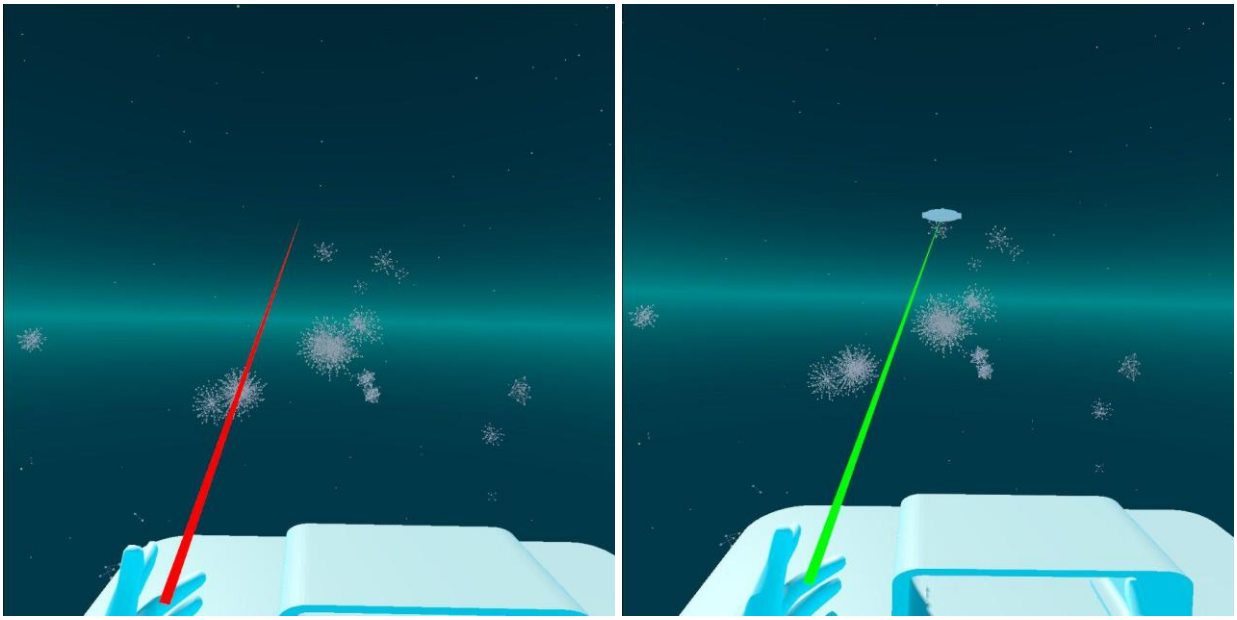
Savukārt, lai virsotnes nepārklājās, tām tiek piešķirti *Unity* piedāvātie kolīzijas ietvari (c*ollider*), kuru pamatfunkcija ir noteikt, vai divi un vairāki objekti sadurās savā starpā. Kolīzijas ietvari tiek padarīti lielāki par pašu virsotnes objektu, kas nodrošina to, ka virsotnes ir atdalītas un rada ilūzija, ka tās atrodas atstatus. Izmantojot abus *Unity* objektus *RigidBody* un *SpringJoint*, pēc

īsa laika intervāla grafs sasniedz līdzsvara stāvokli. Grafam atrodoties šajā stāvoklī, lietotājs var sākt ar to darboties.

7.3.5. Lietotāja iesaiste un vadīklas

Lietotājam tiek piedāvātas vairākas iespējas kā mijiedarboties ar vidē attēlojamiem datiem. Pārvietošanās iespējas ir implementētas izmantojot *Oculus* programmatūras izstrādes komplektā [26]atrodamos iepriekš konfigurētos un atkārtoti lietojamos *Unity* objektus (*prefab*).

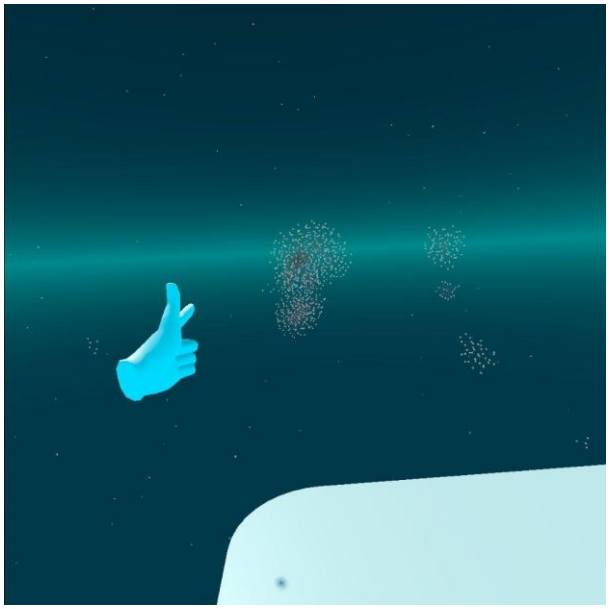
Lai pārvietoties vidē, lietotājs var izvēlēties starp divām pārvietošanas iespējas – teleports un lidošana. Teleports tiek implementēts izmantojot *LocomotionController* [26] iepriekš konfigurēto un atkārtoti lietojamo *Unity* objektu (*prefab*), kurš tiek vadīts ar kreisās pults analogo pogu. Teleportācija darbojās, uz datu kopu virsmām. Ar sarkano staru lietotājs meklē teleportējamo virsmu, kad stars paliek zaļš un tā galā parādās indikācija uz kuras virsmas lietotājs teleportēsies, atlaižot analogo pogu tiek veikta teleportācija. Attēlots attēlā Nr. 19. Savukārt, lidošanai tiek izmantota kreisās pults rādītājpirksta poga. Pogu pieturot, un ar roku norādot pārvietošanas virzienu, lietotājs pārvietojās. Attēlots attēlā Nr. 20.



(a) (b)

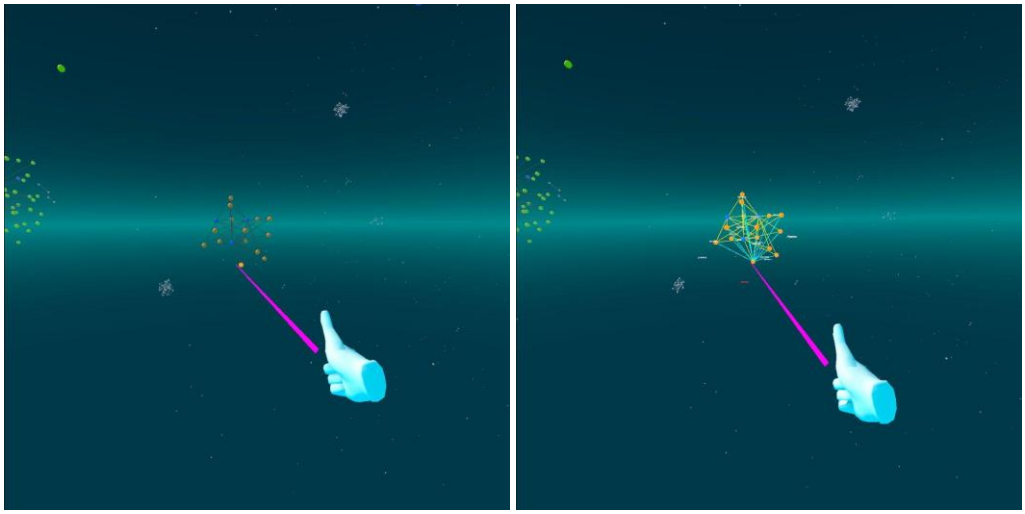
Attēls Nr. 19. **Pārvietošanās iespējas: teleportācija**

(a) Teleportācijas vietas meklēšana (b) Piemērots teleportācijas mērķis ir atrasts



Attēls Nr. 20. **Pārvietošanās iespējas: lidošana**

Papildus lietotājs var apskatīt datu kopas saturošos datņu nosaukumus izvēloties vienu no datu kopas reprezentējošām virsotnēm un uz to nospiežot ar labās pults *B* pogu. Pieturot labās pults rādītājpirksta pogu, tiek attēlots violets stars ar kuru var izvēlēties interesējošo virsotni. Pēc pogas nospiešanas darbības baltā krāsā tiek attēloti visi nosaukumi datu kopu reprezentējošām virsotnēm, savukārt izvēlētās virsotnes datnes nosaukums tiek parādīts sarkanā krāsā. Vienlaicīgi, ar dzeltenu krāsu tiek iezīmētas visas stipra līmeņa sasaistes izvēlētai datu kopai, savukārt, ar zilu krāsu tiek iezīmētas vidēja stipruma līmeņa sasaiste konkrētai izvēlētai virsotnei. Attēlots attēlā Nr. 20.



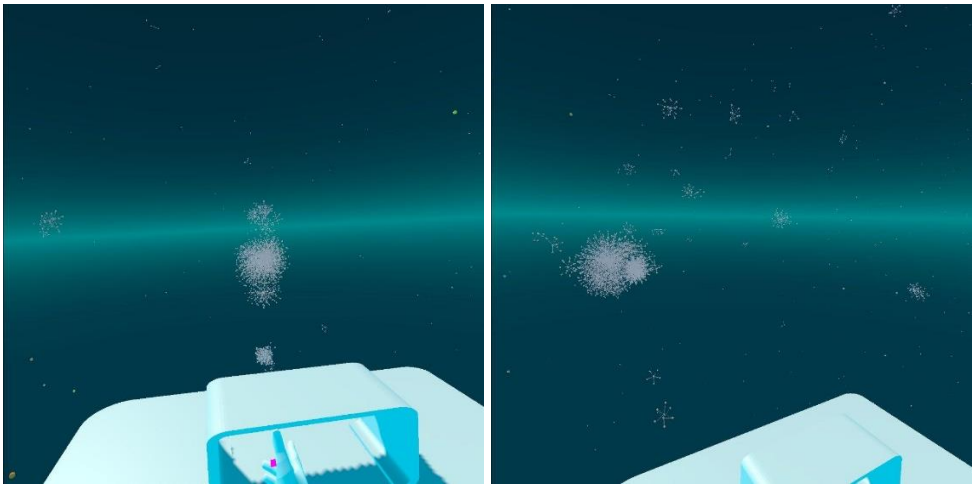
(a) (b)

Attēls Nr. 21. **Apskatāmās datnes izvēlēšanās**

(a) Apskatāmā datne izvēlēta (b) Nospiesta poga uz izvēlētās virsotnes

7.3.6. Saskarnē attēlojamo datu piemēri

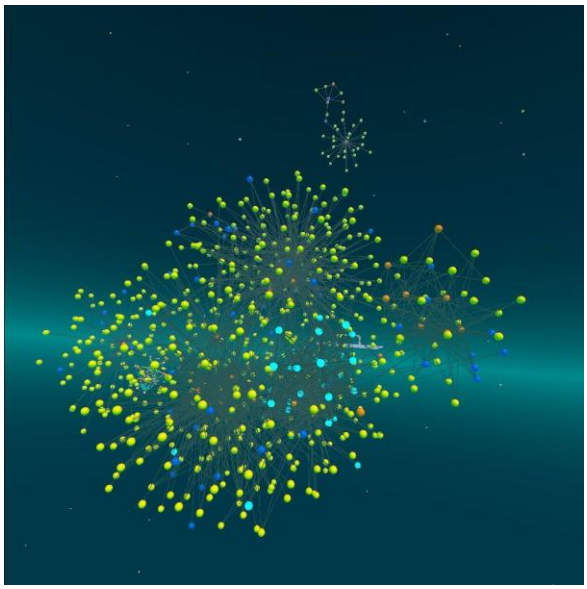
Šajā nodaļas sadaļa tiek demonstrētas dažu spoguļkopiju datu attēlošanas piemēri. Apgalvojumi, kas tiek balstīti uz datu kopu saišu attēlojumu, kuri ir aprakstīti zem piemēriem, tika pārbaudīti un salīdzināti ar pirmavota spoguļkopiju datņu saturu. Virsotnēm datu nosaukumi tika anonimizēti.



(a) (b)

Attēls Nr. 22. **Datu grafu kopskats**

Attēlā Nr. 22. ir redzami divi datu grafu kopskati. (a) ir redzamas 2574 virsotnes, un (b) ir redzamas 6762 virsotnes.



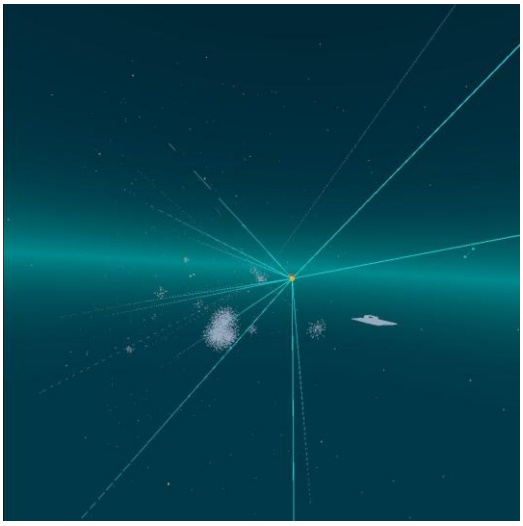
Attēls Nr. 23. **Trīs cieši saistītu datu grupu attēlojums, kuros figurē atslēgvārdi**

Attēlā Nr. 23 ir redzami vairāki grafi, kuros ir sasaistīti pēc tā, ka tika aktivizēts Microsoft *One drive* serviss, kas veica attēloto mēdiju un dokumentu rezerves kopiju izveidi. Papildus, daži

no do dokumentiem un mēdiju datnēm tika nosūtīti caur e-pastiem, kuri saturēja atslēgvārdus, grafā tās ir attēlotas kā izgaismotās gaiši zilas virsotnes.

Attēls Nr. 24. **Divu dažādu nosūtīta un saņemtā e-pasta attēlojumi, kuri satur atslēgvārdus**

Attēlā Nr. 24 ir redzamas divas savā starpā nesaistītu e-pastu apmaiņas attēlojums. Apskatot e-pastu saņēmējus un nosūtītājus, izmeklētāji var ātri secināt, vai šie e-pasti ir nepieciešami kriminālprocesam, šādā veidā, ātri atsijājot informāciju, kas tos neinteresē.

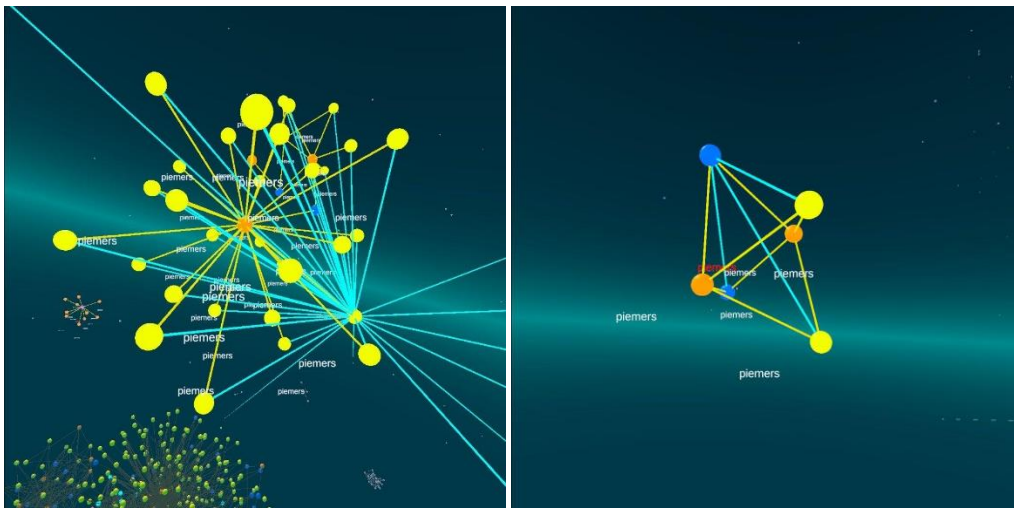


Attēls Nr. 25. **Vienas datnes attēlojums ar vidēja līmeņa sasaistēm**

Attēlā Nr. 25 ir redzama pārlūkprogrammas datnes attēlojums, kurai nav stipru saišu ar citām virsotnēm, taču tai ir vidēja līmeņa sasaiste ar citām virsotnēm, kurām ir stipra līmeņa sasaistes ar citām virsotnēm. Šajā gadījumā, konkrētā datne ir izmeklētāju interesējošā un ar vidēja līmeņa saistēm tā noved izmeklētājus uz citām datnēm, kas tos interesē.

Attēls Nr. 26. **Lietotnes, kas lieto interneta servisus, attēlojums**

Attēlā Nr.26 ir redzams lietotnes attēlojums, kas nav pārlūkprogramma, bet veic darbības internetā. Šajā gadījumā ir attēlots kriptovalūtas maks.



(a) (b)

Attēls Nr. 27. **Attēlojum mediju datņu un dokumentu lejuplādes no pārlūkprogrammas**

Attēlā Nr.27 ir redzamas no mājaslapām lejuplādētas datnes, kas (a) gadījumā, ir tieši saistītas ar vienu pārlūkprogrammas datnes reprezentāciju, un (b) gadījumā ir netieši saistītas ar vienu pārlūkprogrammas datnes reprezentāciju.

8. REZULTĀTI

Tika izpētīti vairāki pētījumi par dažādām datu apstrādes metodikām digitālā izmeklēšanā, balstoties uz kuriem tika secināts:

- Uz digitālo izmeklēšanas datiem ir jāskatās kā uz virkni ar datu kopām;
- Lai izveidotu saistības starp datiem ir nepieciešami skaidri un pamatoti kritēriji, kas datus vieno;
- Sakarības starp datnēm ir jāveido balstoties uz laika vienībām
- Lai izveidotu stingras sakarības ir nepieciešams ieviest papildus nosacījumus datu sasaistei;

Pēc secinājumu veikšanas, tika izstrādāts rīks, kas veic datu apstrādi un sasaisti, balstoties uz noteiktu nosacījumu kopu, kas tika definēta, balstoties uz atgriezenisko saiti no izmeklētājiem un informāciju, kas iegūta no analizētiem pētījumiem.

Tika izpētīti vairāki pētījumi par dažādiem datu attēlošanas paņēmieniem ar mērķi attēlot iegūto informāciju saprotamā un viegli uztveramā veidā, balstoties uz kuriem tika secināts:

- Attēlojumam ir būtiska nozīme datu analīzes procesā;
- Informācija virtuālā realitātē tiek uztverta salīdzinoši īsākā laika periodā nekā citos attēlojuma veidos;
- Cilvēks daudz intuitīvāk uztver liela apjoma informāciju atrodoties virtuālā realitātē;
- Viens no intuitīviem veidiem kā attēlot datu sasaistes ir grafs;
- *Unity* izstrādes platformas iespējas piedāvā efektīvu veidu kā veikt datu sasaistu attēlojumu grafā;

Pēc secinājumu veikšanas, tika izstrādāts rīks, kas veic apstrādāto datu attēlojumu virtuālā realitātē.

Lai novērtētu rīka lietderību, tika veikts pētījums, kurā piedalījās 23 izmeklētāji. Pētījuma laikā izmeklētāji salīdzināja datu attēlojumu virtuālā realitātē ar datu attēlojumu tiem pieejamos rīkos.

Rezultātā tika secināts, ka 17 no 23 izmeklētājiem šāda veida datu attēlojums samazinātu kriminālprocesā ietvaros veltāmo laiku datu apstrādei, jo datu attēlojums ir viegli uztverams, līdz ar to, šis risinājums tika raksturots kā lietderīgs.

12 no 23 izmeklētājiem trūka papildus iesaistes ar datiem, piemēram, iespēja pārkārtot izveidotos grafus it sevišķi lielo datu kopu attēlojamā grafā, iztrūkst atzīme ka noteiktās datu kopas ir apskatītas, interesējošo datņu satura izskatīšana virtuālās realitātes vidē.

5 no 23 izmeklētājiem risinājums šķita grūti saprotams, līdz ar to, tiem būtu nepieciešams veltīt vairāk laika, lai ar to darboties un veikt secinājumus par tā efektivitāti.

Izstrādāto rīku pirmkods ir pieejams mākonī [29]

Balstoties uz atgriezenisko saiti par izstrādātiem rīkiem tiek secināts, ka iegūtie rezultāti atbilst darbā izvirzītajam mērķim.

SECINĀJUMI

Digitālo datu izmeklēšanā ir iepriekš ir piedāvāti vairāki risinājumi, kuru mērķis ir apstrādāt liela apjoma datus, pamatojoties uz notikumu korelāciju vai laika līnijām un artefaktiem. Maģistra darbā piedāvātā datu apstrāde kā pierādījumu sadale, ir šo pētījumos piedāvāto risinājumu apkopojums, kas varētu atrisināt problēmjautājumus saistībā ar datu daudzveidību un to lielo apjomu, kurā papildus tiek pievienots drošības slānis, kas netraucē izmeklētāju un analītiķu darbam, bet darbojās kā stingrs drošības līdzeklis, lai saglabātu pierādījumu derīgumu.

Balstoties uz pētījumu, rezultātiem par cilvēka uztveres ierobežojumiem attiecībā uz liela informācijas daudzuma projicējumu, ir secināms, ka virtuālās realitātes iespējas var pielietot lielo datu attēlošanai, jo, būtiskākos datus var daudz intuitīvāk novietot cilvēka redzes lauka centrālajā zonā, kas secīgi ļauj tam izprast uzrādīto informāciju īsā laika periodā bez būtiskiem datu zudumiem cilvēka uztveres problēmu dēļ. Šāda veida datu attēlošana potenciāli var ievērojami samazināt patērēto laiku, kas tiek izmantots digitālo pierādījumu meklēšanā, vienlaikus ļaujot lietotājiem iespēju intuitīvi aplūkot datu kopas kā informāciju. Šis risinājums, iespējams, var ievērojami samazināt izmeklētāju un analītiķu darba slodzi, kā arī nodrošināt precīzākus, drošākus, kā arī vieglāk un ātrāk pārbaudāmus rezultātus.

Lai sekmētu izstrādātā risinājuma uzlabojumu un pārvērstu to par neatkarīgu izmeklēšanas rīku, turpmākā izstrādē ir nepieciešams ieviest semantiskās vērtības datu sasaistēm un secīgi veikt datu reprezentēšanu kā ontoloģiju. Savukārt, datu attēlošanas risinājumā ir jāievieš papildus mijiedarbošanās iespējas lietotājam.

PATEICĪBAS

Paldies prof. Vladislavam Fominam un Dr. jur. Karenai Ričmondai par padomiem, atbalstu un ieinteresētību, kā arī kolēģiem par ieteikumiem rīku uzlabošanā un testēšanā.

IZMANTOTĀ LITERATŪRA UN AVOTI

[1] S. L. Garfinkel, «Digital forensics research: The next 10 years,» Digital Investigation, Volume 7, Supplement, 64 - 73.lpp, (2010). [Tiešsaiste]. – [atsauce – 15.1.2022] Pieejams: https://www.sciencedirect.com/science/article/pii/S1742287610000368

[2] M. Politt, K. P. Chow un S. Shenoi, «A History of Digital Forensics,» Advances in Digital Forensics VI, 3-15.lpp., (2010). [Tiešsaiste]. – [atsauce – 25.1.2021] Pieejams: https://link.springer.com/chapter/10.1007/978-3-642-15506-2_1

[3] Y. Chabot, a. Bertaux, C. Nicolle un T. Kechadi, «A Complete Formalized Knowledge Representation Model for Advanced Digital Forensics Timeline Analysis,» Conference: DFRWS 2012, (2012). [Tiešsaiste]. – [atsauce – 17.1.2022] Pieejams: https://www.researchgate.net/publication/261507681_A_Complete_Formalized_Knowledge_Representation_Model_for_Advanced_Digital_Forensics_Timeline_Analysis

[4] K. Chen, A. Clark, D. Vel un G. O. Mohay, «ECF - Event Correlation for Forensics,» First Australian Computer, Network and Information Forensics Conference, 1-10.lpp, (2003). [Tiešsaiste]. – [atsauce – 22.03.2022] Pieejams: https://www.researchgate.net/publication/27478337_ECF_-_Event_correlation_for_forensics

[5] C. James un H. J. Patterson, «An automated timeline reconstruction approach for digital forensic investigations,» Digital Investigation 9:S69–S79, (2012). [Tiešsaiste]. – [atsauce – 22.4.2022] Pieejams: https://www.researchgate.net/publication/257687900_An_automated_timeline_reconstruction_approach_for_digital_forensic_investigations

[6] Y. Chabot, A. Bertaux, C. Nicolle un T. Kechadi, «An ontology-based approach for the reconstruction and analysis of digital incidents timelines,» Digital Investigation Vol. 15, 83-100. lpp., (2015). [Tiešsaiste]. – [atsauce – 15.5.2022] Pieejams: https://www.researchgate.net/publication/280648329_An_Ontology-Based_Approach_for_the_Reconstruction_and_Analysis_of_Digital_Incidents_Timelines

[7] M. Kwan, K. P. Chow, F. Law un P. Lai, «Reasoning About Evidence Using Bayesian Networks,» Advances in Digital Forensics IV. DigitalForensics (2008). IFIP — The International Federation for Information Processing, vol 285, 2008. [Tiešsaiste]. – [atsauce – 25.10.2022] Pieejams: https://doi.org/10.1007/978-0-387-84927-0_22

[8] M. Kwan, R. Overill, K. P. Chow, H. Tse, F. Law un P. Lai, «Sensitivity Analysis of Bayesian Networks Used in Forensic Investigations,» Advances in Digital Forensics VII. IFIP Advances in Information and Communication Technology, vol 361., (2011). [Tiešsaiste]. – [atsauce – 30.10.2021] Pieejams: https://www.researchgate.net/publication/221352826_Sensitivity_Analysis_of_Bayesian_Networks_Used_in_Forensic_Investigations

[9] M. Smith, I. Horrocks un M. Krotzsch, «OWL 2 Web Ontology Language: Conformance (Second Edition),» Birte Glimm, eds. W3C Recommendation, (2012). [Tiešsaiste]. – [atsauce – 18.5.2022] Pieejams: http://www.w3.org/TR/owl2-conformance/

[10] R. Angles un C. Gutierrez, «The Expressive Power of SPARQL,» The Semantic Web - ISWC, Lecture Notes in Computer Science, vol 53.18., (2008). [Tiešsaiste]. – [atsauce – 18.5.2022] Pieejams: https://doi.org/10.1007/978-3-540-88564-1_8

[11] O. Brady, R. Overill un J. Keppens, «DESO: Addressing volume and variety in large-scale criminal cases,» Digital Investigation (15.2015). [Tiešsaiste]. – [atsauce – 15.4.2022] Pieejams: https://www.researchgate.net/publication/284084838_DESO_Addressing_volume_and_variety_in_large-scale_criminal_cases

[12] A. Bayyari un E. Tudoreanu, «The impact of immersive virtual reality displays on the understanding of data visualization,» VRST '06: Proceedings of the ACM symposium on Virtual reality software and technology, (2006). [Tiešsaiste]. – [atsauce – Pieejams: https://dl.acm.org/doi/abs/10.1145/1180495.1180570

[13] Z. Xinzhou, «Virtual Reality Application in Data Visualization and Analysis,» Journal of Computer Science, (2017). [Tiešsaiste]. – [atsauce – 21.4.2022] Pieejams: https://www.semanticscholar.org/paper/Virtual-Reality-Application-in-Data-Visualization-Zhang/1fbee5534f3bda51387f709cf83b9c6139a93fa8

[14] B. Laha un A. Bowman, «Identifying the Benefits of Immersion in Virtual Reality for Volume Data Visualization,» (2012). [Tiešsaiste]. – [atsauce – 29.04.2022] Pieejams: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.452.3078&rep=rep1&type=pdf

[15] Z. M. Khalid un S. R. Zeebaree, «Big Data Analysis for Data Visualization: A Review,» International Journal of Science and Business Volume: 5, Issue: 2, 64-75.lpp, (2021). [Tiešsaiste]. – [atsauce – 10.5.2022] Pieejams: https://ijsab.com/wp-content/uploads/671.pdf

[16] E. Olshannikova, A. Ometov, Y. Koucheryavy un et al., «Visualizing Big Data with augmented and virtual reality: challenges and research agenda,» Journal of Big Data (2.22.2015). [Tiešsaiste]. – [atsauce – 15.03.2022]. Pieejams: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-015-0031-2#citeas

[17] P. Millais, S. L. Jones un R. Kelly, «Exploring data in virtual reality: Comparisons with 2D Data Visualizations,» CHI EA '18: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, April Paper No.: LBW007, 1 – 6.lpp., (2018). [Tiešsaiste]. – [atsauce – 17.3.2022]. Pieejams: https://dl.acm.org/doi/abs/10.1145/3170427.3188537

[18] C. Ware un G. Franck, «Viewing a graph in a virtual reality display is three times as good as a 2D diagram,» Proceedings of 1994 IEEE Symposium on Visual Languages, 182-183. lpp., (1994). [Tiešsaiste]. – [atsauce – 17.3.2022]. Pieejams: https://ieeexplore.ieee.org/abstract/document/363621/authors#authors

[19] Wikipedia, «Parallax,» Wikimedia Foundation, (24.4.2022). [Tiešsaiste]. – [atsauce – 18.5.2022] Pieejams: https://en.wikipedia.org/wiki/Parallax

[20] Wikipedia, «Stereopsis,» Wikimedia Foundation, (10.4.2022). [Tiešsaiste]. – [atsauce – 18.5.2022] Pieejams: https://en.wikipedia.org/wiki/Stereopsis

[21] A. Sankar, «Immersive Data Visualization with Virtual Reality,» 2015. [Tiešsaiste]. – [atsauce – 15.3.2022]. Pieejams:https://www.semanticscholar.org/paper/Immersive-Data-Visualization-with-Virtual-Reality-Sankar/eef92baf04498eb4040c44f5b43bb44b8acc6e01#citing-papers

[22] D. C. un et al., «Immersive and collaborative data visualization using virtual reality platforms,» 2014. [Tiešsaiste]. – [atsauce – 22.04.2022]. Pieejams: https://ieeexplore.ieee.org/abstract/document/7004282.

[23] I. Dror, W. Thompson, C. A. Meissner un et al., «Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making",» Journal of Forensic Sciences 60(4), 1111-1112.lpp, (2015). [Tiešsaiste]. – [atsauce – 15.4.2022]. Pieejams: https://www.researchgate.net/publication/277556511_Context_Management_Toolbox_A_Linear_Sequential_Unmasking_LSU_Approach_for_Minimizing_Cognitive_Bias_in_Forensic_Decision_Making

[24] T. L. Kim, K. Kim, C. Choi un et al., «FOPR test: a virtual reality-based technique to assess field of perception and field of regard in hemispatial neglect,» J NeuroEngineering Rehabil 18, 39, 2021. [Tiešsaiste]. – [atsauce – 19.04.2022]. Pieejams https://www.researchgate.net/publication/27478337_ECF_-_Event_correlation_for_forensics

[25] M. Droettboom, «Understanding JSON Schema,» Space Telescope Science Institute, (02.2022). [Tiešsaiste]. – [atsauce – 17.5.2022]. Pieejams: https://json-schema.org/understanding-json-schema/

[26] Oculus, «Oculus Integration,» asset Version 39.0, (27.4.2022). [Tiešsaiste]. – [atsauce – 27.4.2022]. Pieejams: https://assetstore.unity.com/packages/tools/integration/oculus-integration-82022

[27] Key Mouse, «Customizable skybox,» asset version 1.0, (13.7.2020). [Tiešsaiste]. – [atsauce – 15.5.2022]. Pieejams: https://assetstore.unity.com/packages/2d/textures-materials/sky/customizable-skybox-174576

[28] M. d'Aquin, «3D Force-Directed Graphs with Unity,» (27.10.2020). [Tiešsaiste]. – [atsauce – 20.5.2022]. Pieejams: https://towardsdatascience.com/3d-force-directed-graphs-with-unity-587ad8f7dff

[29] R. Rodiks, «Izstrādāto rīku pirmkods un izpildkods,» (2022). [Tiešsaiste]. – [atsauce – 23.5.2022]. Pieejams: https://drive.google.com/file/d/1fJP_xr15Q-VakINlghZu2e-uLQkeBzM7/view?usp=sharing

Maģistra darbs “Strukturētu datu grafu vizualizācija virtuālā realitātē kā rīks lai asistētu noziegumu digitālā izmeklēšanā” izstrādāts LU Datorikas fakultātē.

Darba teksta galīgā versija izgatavota 23.05.2022**.**

Ar savu parakstu apliecinu, ka pētījums veikts patstāvīgi, izmantoti tikai tajā norādītie informācijas avoti un iesniegtā darba elektroniskā kopija atbilst izdrukai.

Autors: Ronalds Rodiks

Ar savu parakstu apliecinu, ka esmu lasījis augstāk minēto maģistra darbu un atzīstu to par **piemērotu** aizstāvēšanai Latvijas Universitātes datorzinātņu maģistrantūrā.

Darba vadītājs: Dr.dat. Leo Seļāvo

Darbs iesniegts **maģistratūras sekretariātā** 23.05.2022.

Ar šo es apliecinu, ka darba elektroniskā versija ir augšupielādēta LU informatīvajā sistēmā.

Studiju metodiķe: Ella Arsa

Recenzents: Dr. dat. Kārlis Čerāns

Darbs aizstāvēts maģistra gala pārbaudījuma komisijas sēdē

_____ prot. Nr. _____

(Darba aizstāvēšanas datums)

Komisijas sekretārs: _____

(Sekretāra paraksts)