

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond
Tarkvarateaduse instituut

Mattias Jõesaar 154887IAPB

**RUUMIANDMETE VÕRDLEVA ANALÜÜSI
TULEMUSTE AUTOMAATNE
VISUALISEERIMINE**

bakalaureusetöö

Juhendaja: Riina Maigre
PhD

Tallinn 2020

Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Mattias Jõesaar

24.05.2020

Annotatsioon

Selle bakalaureusetöö eesmärk on automatiseerida ruumiandmete võrdleva analüüsi tulemuste esitamise protsess, mis vähendaks võrdlemise teostamise töömahtu.

Töös on kirjeldatud ruumiandmeid võrdlevat meetodit ning analüüsitud nõudeid loodavale tarkvarale. Töö tulemusena valmis veebirakenduse prototüüp ruumiobjektide võrdlemiseks ning tulemuste esitamiseks. Võrrelda saab mistahes kahe andmekogu samas piirkonnas asuvaid sama tüüpi ruumiobjekte. Rakendus esitab võrdluse tulemuste statistika tabelite ja diagrammidena.

Valminud prototüüp on kasutuses sihtasutus Archimedese nutika spetsialiseerumise rakendusuringus, mille ühe osana võrreldakse ruumiandmeid.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 19 leheküljel, 6 peatükki ja 8 joonist.

Abstract

Automatic visualization of the results of comparative spatial analysis

The aim of this thesis is to automatize the process of visualizing the results of comparative spatial analysis. The intent is to reduce the manual work of comparing spatial data.

This thesis first describes the method of comparative spatial analysis and reviews the requirements for the proposed application. Other applications with similar purposes were reviewed before the development. The thesis provides a description of the application's architecture, the programming languages and frameworks used for the development.

A working web application prototype was developed as a result. This application can compare same types of spatial features in the same area from two different datasets. The comparison is applicable for point objects and line objects such as poles and curbstones. The application's interface contains a form for comparing spatial data with specified parameters. Results of comparative spatial analysis are displayed in tables and diagrams. Visualized results give domain experts insights into the accuracy of compared spatial data.

The developed prototype is being used in a Archimedes Foundation smart specialization study where spatial data comparison is being carried out.

The thesis is in Estonian and contains 19 pages of text, 6 chapters, 8 figures.

Lühendite ja mõistete sõnastik

<i>3D</i>	<i>Three-dimensional</i> , kolmemõõtmeline
<i>Front-end</i>	andmete teisendamine graafiliseks liideseks, et kasutajad saaksid neid andmeid vaadata ja nendega suhelda
GIS	<i>Geograafiline infosüsteem</i> , omavahel seotud kogum tarkvarast ja andmetest, mida kasutatakse geograafilise info vaatamiseks ja haldamiseks, ruumiliste seoste analüüsimiseks ning ruumiliste protsesside modelleerimiseks [1]
Klaster	kogum ühesuguseid või omavahel seotud objekte
LiDAR	<i>Light Detection And Ranging</i> , laserskaneerimisseade
<i>Shapefile</i>	georuumiliste vektorite andmeformaad, mida kasutatakse geograafilistes infosüsteemides
SQL	<i>Structured Query Language</i> , andmebaasi päringukeel
XML	<i>Extensible Markup Language</i> , standardne üldotstarbeline märgistuskeel, mille eesmärk on struktureeritud info jagamine infosüsteemide vahel

Sisukord

1 Sissejuhatus	8
1.1 Taust	8
1.2 Probleem	9
1.3 Eesmärk	9
2 Nõuded	10
2.1 Funktsionaalsed nõuded	10
2.2 Mittefunktsionaalsed nõuded	11
3 Metoodika	12
3.1 Ruumiandmete võrdlemise meetod	12
3.2 Varasemad lahendused	13
3.3 Rakenduse arhitektuur	13
3.4 Kasutatavad keeled ja raamistikud	14
4 Rakenduse kirjeldus	15
4.1 Serveri päringud	16
4.1.1 Ruumiandmetega tabelid	16
4.1.2 Ruumiobjektide võrdlus	17
5 Kasutajaliides	18
5.1 Vorm	19
5.2 Tulemuste esitamine	20
5.2.1 Klasterdumise statistika	21
5.2.2 Parima kaugusläve statistika	22
5.2.3 Graafikud joonobjektide nihke hindamiseks	23
5.3 Testimine	24
6 Võimalikud edasiarendused	25
6.1 Ruumiandmete sisestamine läbi kasutajaliidese	25
7 Kokkuvõte	26
Kasutatud kirjandus	27
Lisa 1 – Ruumiandmete võrdluse päringu vastus	28
Lisa 2 – Excel'i dokument ruumiandmete võrdluse tulemustega	29

Jooniste loetelu

Joonis 1. Rakenduse järgnevusskeem	15
Joonis 2. Kasutajaliidese eskiis	18
Joonis 3. Vorm ruumiandmete võrdluse käivitamiseks	19
Joonis 4. Rakenduse võimalikud kuvatavad graafilised elemendid	20
Joonis 5. Graphic komponendi sisendi näidis sektordiagrammi kuvamiseks	20
Joonis 6. Ruumiandmete võrdluse klasterdumise baasstatistika.	21
Joonis 7. Parima kaugusläve detailne statistika.	22
Joonis 8. Joonobjektide võrdluse tulemusi väljendavad diagrammid.....	23

1 Sissejuhatus

3D linnamudelite ning teiste ruumiandmeid kasutatavate rakenduste loomine nõuab täpseid andmeid. Andmete täpsustamiseks kasutatakse üldjuhul andmete joondamist maapealsete kontrollpunktide (GCP) järgi. Maapealsete kontrollpunktide kogumine on ajakulukas ja kallis, seetõttu on oluline uurida alternatiivseid võimalusi.

1.1 Taust

Käimas on Archimedese nutika spetsialiseerumise rakendusuring: Kuluefektiivse ühildatava geodeetilise täpsusega 3D ruumiandmete taristu loomise rakendusuring [2]. Projekti ühe osana tegeletakse ruumiandmete võrdleva analüüsiga, mis võimaldab hinnata ühe ruumiandmete andmebaasi täpsust võrreldes teise täpsema andmebaasiga. Välja töötatud protsessi abil saab võrrelda mistahes kahest andmekogust ruumiandmeid samas piirkonnas asuvate sama tüüpi ruumiobjektide kohta. Selles töös toodud näidetes kasutatakse LiDARi andmeid, mis on kogutud Eesti ettevõtte Reach-U AS poolt ning GIS andmebaasi kihte Eesti topograafia andmekogust (ETAK) ja Tallinna ruumiandmete registrist (TAR). Võrdlev analüüs võimaldab statistiliselt hinnata võimalusi ruumiandmete joondamiseks olemasolevate andmebaaside järgi. [3]

Võrdluseks valitud ala kohta käivad mõlema andmekogu kihtide andmed laetakse PostgreSQL/PostGIS [4], [5] andmebaasi. PostgreSQL'i abil liidetakse omavahel mõlemad kihid. Peamine eesmärk on leida võrreldavate andmehulkade kattuvate objektide hulk. Selleks, et leida kattuvad objektid, rakendatakse ruumilist objektide klastriteks rühmitamise ehk klasterdamise algoritmi DBSCAN [6].

Lisaks kattuvatele objektidele leitakse ka muud tüüpi vastavused objektide vahel ning kummastki andmehulgast puuduvad objektid uuritava alal. Klastrid ning nende sisu analüüsitakse, et määrata, milliseid objekte saab paaritada, milliseid ei saa ning mis põhjustel objektid ei kattu. [3]

1.2 Probleem

Selleks, et esitada valdkonna eksperdile võrdluse statistika, peab loendama erinevatesse kategooriatesse langevad klastrid ning arvutama suhtelise jaotuse andmebaasi päringute ja Exceli abil. Võrdluse statistika visualiseerimiseks kantakse tulemused Exceli dokumenti manuaalselt. Kopeeritud andmetega tehakse Excelis lisaarvutusi ning luuakse diagrammid. Sellist toimingut tuleb korrata iga võrdluse korral. Käsitsi andmete tabelarvutussüsteemi dokumenti kandmine on töömahukas ja toob endaga kaasa vigu.

1.3 Eesmärk

Töö eesmärk on luua veebirakendus, mis automatiseerib võrdlusandmete visualiseerimise protsessi ning likvideerib ajakulu ja vead, mis kaasnevad praegusel kujul ruumiandmete võrdleva analüüsi teostamisega. Lõputöö oodatav tulemus on rakendus, mille kasutajaliides võimaldab kasutajal valitud parameetritega pärida serverist võrdlusandmed ning nende andmete põhjal kuvatakse vastavalt tulemustele dünaamilised tabelid ja diagrammid. Andmebaasi päringud ruumiandmete võrdlemiseks on töö juhendaja poolt kirjutatud ning ei ole osa töö mahust.

2 Nõuded

Järgnevalt on kirjeldatud nõuded, mida realiseeritav rakendus peab täitma. Nõuded on jaotatud funktsionaalseteks ja mittefunktsionaalseteks. Funktsionaalsed nõuded määratlevad, kuidas tarkvara käitub konkreetsetes olukordades. Mittefunktsionaalsed nõuded hõlmavad tarkvara iseloomustavaid omadusi, näiteks jõudlus, turvalisus ja kasutatavus. [7]

2.1 Funktsionaalsed nõuded

Rakendus peab:

1. Kuvama andmebaasis olevate tabelite nimed, mis sisaldavad andmeid ruumiobjektide kohta
2. Võimaldama kasutajaliideses võrdlusandmete pärimiseks ette anda järgnevad parameetrid:
 - a. Kaks andmebaasi tabeli nime, milles sisalduvaid ruumiobjektide andmeid soovitakse võrrelda.
 - b. Nimetused valitud tabelitele, mida kasutatakse statistika kuvamisel tabelitele viitamiseks.
 - c. Minimaalne kauguslävi (EPS)
 - d. Maksimaalne kauguslävi (EPS)
 - e. Aste, mida kasutatakse järjestikuste kauguslävede arvutamisel
 - f. Parim kauguslävi, mille kohta kuvatakse täiendav statistika
3. Kuvama tabelite ja diagrammide võrdluse statistika kohta olemasoleva Exceli dokumendi näitel [Lisa 2]

2.2 Mittefunktsionaalsed nõuded

Rakenduse mittefunktsionaalsed nõuded on järgmised:

1. Rakenduse kasutajaliides on ingliskeelne.
2. Kasutajaliides on lihtne ja kergesti kasutatav.
3. Võimaldab ligipääsu erinevatelt sisevõrku ühendatud arvutitelt.
4. Rakendust saab kasutada veebilehitseja abil.

3 Metoodika

Selles osas on kirjeldatud, kuidas töös kasutatud ruumiandmete võrdlemise meetod töötab. Lisaks antakse ülevaade varasematest sarnaste eesmärkidega lahendustest ning hinnatakse võimalusi nende kohandamiseks ja kasutamiseks käesolevas töös. Viimasena kirjeldatakse loodava rakenduse arhitektuuri ning esitatakse arendamiseks kasutatavad programmeerimiskeeled ja raamistikud.

3.1 Ruumiandmete võrdlemise meetod

Võrreldavaid ruumiandmeid hoiustatakse PostGIS laiendusega PostgreSQL relatsioonilises andmebaasis. Ruumiobjektid peavad omama andmebaasis unikaalset identifikaatorit (gid) ning geomeetria välja (geom).

PostGIS 2.5.3 implementeerib ruumiandmete võrdluseks vajalikku DBSCAN algoritmi `ST_ClusterDBSCAN` funktsioonis [8]. Lisaks ruumiandmetele nõuab funktsioon sisendiks kaugusläve meetrites (EPS), mille vahetus kauguses peavad objektid klasterdumiseks asuma, ning tiheduse (MinPts), mis piiritleb minimaalse objektide arvu klastris.

Funktsiooni rakendatakse tihedusega `MinPts=2` eesmärgiga leida 1:1 klastrid, mis sisaldavad mõlemast andmehulgast ühte ruumiobjekti. Sellistesse klastritesse sattunud objektid loetakse samaks. Klasterdamise tulemusena leiame ka teist tüüpi klastreid, kuna need võivad sisaldada rohkem kui kahte objekti ning objekte ainult ühest andmehulgast. Objekte, mis ei kuulu ühtegi klastrisse, käsitletakse mürana (*noise*). [3]

Sobiva kaugusläve väärtuse valimine on keeruline, sest see sõltub sellest, kui hästi võrreldavates andmebaasides olevad andmed omavahel kokku lähevad. Selles meetodis rakendatakse `ST_ClusterDBSCAN` funktsiooni ruumiandmetele erinevate kaugusläve väärtustega, et anda ülevaadet kogutud andmetele erinevate kauguslävede puhul. Tulemuste kohta esitatakse baasstatistika, mille põhjal saab valdkonna ekspert otsustada parima kaugusläve üle, millega edasist statistikat arvutada. [3]

Praeguseks väljatöötatud võrdlusmeetodiga on võimalik võrrelda punktobjekte ja joonobjekte. Võrreldavateks punktobjektideks võivad olla näiteks tänavapostid, joonobjektideks võivad olla näiteks äärekivid. Punktobjektide puhul leitakse eukleidiline kaugus punktide vahel. Joonte puhul selline lähenemine ei sobi, sest nende puhul võib minimaalne kaugus olla null ehk jooned ristuvad, kuid see ei ütle kasutajale midagi. Joonobjektide nihke hindamiseks kasutatakse erinevaid statistilisi meetodeid ning nende põhjal loodud graafikuid.

3.2 Varasemad lahendused

Enne rakenduse arendamist uuriti, kas leidub lahendus, mis juba ülaltoodud nõudeid rahuldab. 2014. aastal Rootsi geoinformaatika bakalaureusetöö käigus hinnati Rootsi kaldajoone ruumiantmete täpsust võrreldes omavahel andmehulki [9]. Selline meetod on sarnane käesoleva töö võrdlusele. Rootsi geoinformaatika bakalaureusetöö käigus ei arendatud tarkvara automaatselt võrdluse tulemuste esitamiseks, mis vastaks käesoleva töö nõuetele.

2015. aastal avaldatud töös loodi veebiteenus, mis võimaldab võrrelda kahte ruumiantmete andmehulka [10]. Sellel on sarnasusi ja erinevusi käesoleva töö meetodile. Veebiteenus võtab sisendiks kaks SHP faili (Shapefile) ning tagastab XML ja PDF kujul raporti tulemuste hinnanguga. Detailseid tulemusi tabelite ja graafikute kujul kasutajale ei esitata ning võrreldavates andmehulkades peavad vastavate objektide identifikaatorid olema samad. Antud töös sellist kitsendust ei ole, klasterdamise eesmärk on leida 1:1 klastrid ning nendes sattuvad objektid loetakse samaks.

Ühtegi olemasolevat tarkvara, mida oleks võimalik kohandada ja kasutada Archimedese projekti ruumiantmete võrdluse tulemuste esitamiseks, ei leitud.

3.3 Rakenduse arhitektuur

Loodav veebirakendus põhineb klient/teenindaja (*Client/Server*) mudelil. Selline lahendus võimaldab mitmel ruumiantmete võrdlemisega tegeleval isikul ligipääsu ühisele andmebaasile võrdlusstatistika pärimiseks. Realiseeritav rakendus arendatakse ühelehelisena (*SPA – Single-Page Application*), mis võimaldab dünaamiliselt muuta kuvatavat veebilehte vastavalt kasutaja päringutele.

3.4 Kasutatavad keeled ja raamistikud

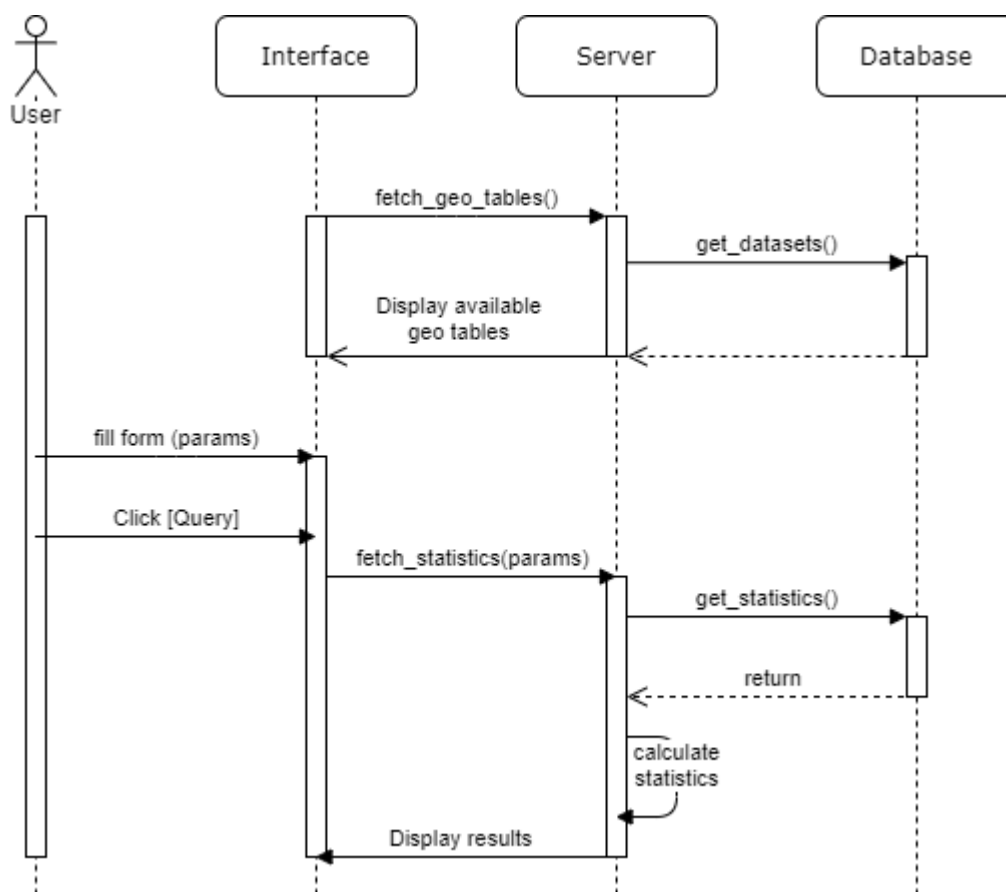
Ruumiandmeid võrdlev skript on kirjutatud Python programmeerimiskeeles. Seetõttu on programmikoodi ümberkirjutamise vältimiseks valitud serverirakenduse realiseerimiseks Python'il baseeruv Django veebiraamistik [11]. Django võimaldab lühikese ajaga luua tervikliku veebirakenduse.

Rakenduse veebiliidese loomisel kasutatakse programmeerimiskeelt JavaScript ning React raamistikku [12], mida haldab Facebook ja üksikarendajate kogukond. React on kõige populaarsem *front-end* JavaScripti raamistik [13]. React sobib üheleheliste veebirakenduste arendamiseks ning on komponendipõhine, mis võimaldab antud projektis lihtsalt ehitada taaskasutatavad komponendid vajalike tabelite ja diagrammide kuvamiseks.

Diagrammide koostamiseks kasutatakse Reacti komponentidele ehitatud Recharts raamistikku [14]. Recharts'i abil saab kiirelt ehitada erinevat tüüpi diagramme ning kohandada neid vastavalt projekti nõuetele.

4 Rakenduse kirjeldus

Joonisel 1 on esitatud järgnevusskeem, mis joonestab lahti stsenaariumi, kuidas kasutaja pärib rakenduses ruumiandmete võrdluse tulemusi. Rakenduse avamisel veebilehitsejas päritakse serverist ruumiobjekte sisaldavad andmebaasi tabelid. Nende tabelite nimed on valitavad kasutajaliidese vormis. Kui kasutaja on vormi täitnud ning seejärel esitanud, käivitatakse serveris ruumiandmete võrdlemise protsess. Protsessi lõppedes kuvatakse tulemused kasutajale.



Joonis 1. Rakenduse järgnevusskeem

4.1 Serveri päringud

Siin on kirjeldatud päringuid, mida liides teeb serverisse. Päringute juures on välja toodud päringu parameetrid, vastus, meetod ja aadress.

4.1.1 Ruumiandmetega tabelid

See päring võimaldab küsida kõik andmebaasis olevad tabelid, mis sisaldavad ruumiobjekte. PostGIS tekitab tabeli *geometry_columns*, kus *f_table_name* väli sisaldab ruumiobjektidega tabelite nimesid. Sellest tabelist saab pärida kõik vajalikud tabelite nimed.

- **URL**

/gis/datasets/

- **Meetod**

GET

- **Parameetrid**

Puuduvad

- **Õnnestunud vastus:**

Kood: 200 OK

Sisu: Loend tabelite nimedega

4.1.2 Ruumiobjektide võrdlus

Selle päringu abil võrreldakse ruumiobjekte kahest erinevast andmebaasi tabelist ning tagastatakse vastavad tulemused.

- **URL**

`/gis/statistics/`

- **Meetod**

POST

- **Parameetrid**

```
eps_min=[integer]
eps_max=[integer]
eps_step=[integer]
first_dataset=[string]
second_dataset=[string]
first_dataset_name=[string]
second_dataset_name=[string]
best_eps=[string]
```

- **Õnnestunud vastus:**

Kood: 200 OK

Sisu: võrdluse tulemused [Lisa 1]

5 Kasutajaliides

Lõputöö käigus arendatud kasutajaliideses on üks vaade, mis on võimalikult lihtne. Joonisel 2 on näha kasutajaliidese esialgne eskiis. Vaade koosneb vormist ruumiandmete võrdluse käivitamiseks ning visualiseeritud kujul võrdluse tulemustest. Võrdluse tulemuste nägemiseks tuleb täita vorm ning see esitada.



Joonis 2. Kasutajaliidese eskiis

5.1 Vorm

Võrdluse käivitamiseks on rakenduse päises vorm (vt. Joonis 3). Vorm võimaldab kasutajal sisestada nõutud parameetrid. Kasutaja saab valida kaks andmebaasi tabelit, milles sisalduvaid ruumiobjekte soovitakse võrrelda. Kuna tabelite nimed andmebaasis võivad olla pikad ja keerulised, siis tulemustes nendele tabelitele viitamiseks on võimalik kasutajal ette anda kohandatud nimed. Kui kohandatud nimede väljad jäetakse täitmata, kasutatakse tulemustes tabelite algseid nimesid.

Lisaks tabelite valikule tuleb ette anda minimaalne kauguslävi, maksimaalne kauguslävi ning kaugusläve aste. Vaikimisi on parima kaugusläve valik automaatne, kuid vorm võimaldab soovi korral valida minimaalse ja maksimaalse kaugusläve vahemikus sobiva väärtuse.

Vormi esitamiseks peavad olema täidetud järgmised tingimused:

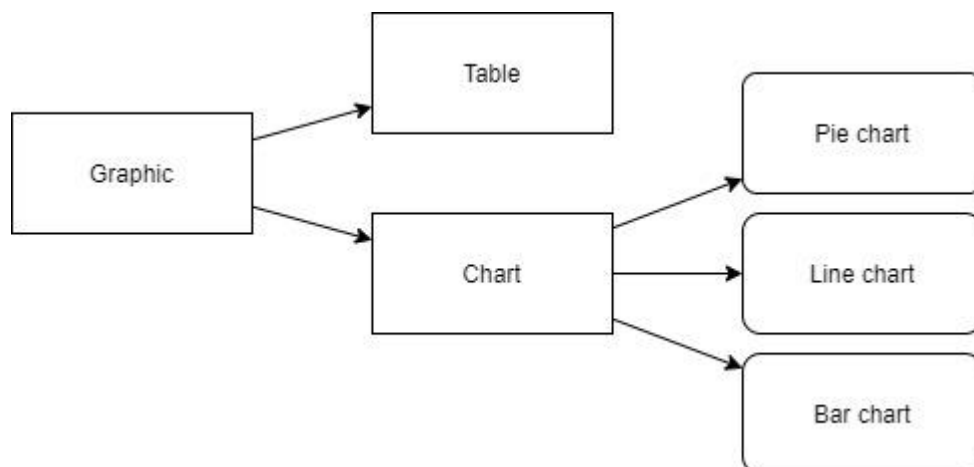
- Valitud on kaks erinevat andmebaasi tabelit
- Minimaalse kaugusläve väärtus on suurem kui 1
- Maksimaalse kaugusläve väärtus on väiksem kui 1000
- Kaugusläve aste on võrdne või suurem kui 5
- Ei ole pooleli eelnev võrdlusandmete päring

1st dataset	2nd dataset	Minimum EPS	Maximum EPS	EPS step	Best EPS	
Choose option ▾	Choose option ▾	10	240	10	Automatic ▾	Query
name	name					

Joonis 3. Vorm ruumiandmete võrdluse käivitamiseks

5.2 Tulemuste esitamine

Võrdluse tulemuste statistika esitamiseks kasutatakse tabeleid ja diagramme. Töö käigus loodi taaskasutatav komponent *Graphic*, mis tegi lihtsaks kõikide nõutud tabelite ning diagrammide esitamise. *Graphic* komponent võimaldab kuvada tabelit, tulp-, joon- või sektordiagrammi (vt. Joonis 4).



Joonis 4. Rakenduse võimalikud kuvatavad graafilised elemendid

Ühe tabeli või diagrammi välja joonistamiseks nõuab *Graphic* komponent sisendina objekti, mis sisaldab andmeid, mida soovitakse kuvada, ning täpsustusi kuvatava graafika kohta (vt. Joonis 5). Määrata saab graafilise elemendi tüüpi ning pealkirja.

```
{
  "component": "chart",
  "data": {
    "type": "pie-chart",
    "title": "Ratio (%) of features of different clusters in
TAR dataset",
    "dataKey": "percentage",
    "chartData": [
      {"name": "TAR matching", "percentage": 72},
      {"name": "Clusters type 2", "percentage": 3},
      {"name": "Clusters type 3", "percentage": 3},
      {"name": "Clusters type 4", "percentage": 3},
      {"name": "TAR noise", "percentage": 17}
    ]
  }
}
```

Joonis 5. *Graphic* komponendi sisendi näidis sektordiagrammi kuvamiseks

5.2.1 Klasterdumise statistika

Esimesena esitatakse kasutajale baasstatistika kasterdumise kohta valitud kauguslävete puhul (vt. Joonis 6). Tabelitena kuvatakse mõlema andmehulga objektide üldarv vaadeldaval alal, erinevat tüüpi klastrite kogus ning nende suhteline jaotus. Joondiagrammina esitatakse esimest tüüpi klastrite suhteline jaotus. Joonisel 6 näidatud joondiagramm võimaldab kasutajal hinnata, milline on parim kaugusläve väärtus. Antud näites ei suurene alates 80 cm kokkulangevate objektide ehk esimest tüüpi klastrite arv oluliselt.

Clustering statistics

Statistics of clustering

EPS (given distance for clustering) in cm	200	180	160	140	120	100	80	60	40	20
Number of clusters containing one feature from TAR dataset and one feature from LIDAR dataset (type 1)	132	132	134	133	133	132	130	120	101	60
Number of clusters containing one feature from TAR dataset and more than one feature from LIDAR dataset (type 2)	0	0	0	0	0	0	0	0	0	0
Number of TAR features in these clusters	0	0	0	0	0	0	0	0	0	0
Number of LIDAR features in these clusters	0	0	0	0	0	0	0	0	0	0
Number of clusters containing one feature from LIDAR dataset and more than one features from TAR dataset (type 3)	0	0	0	0	0	0	0	0	0	0
Number of LIDAR features in these clusters	0	0	0	0	0	0	0	0	0	0
Number of TAR features in these clusters	0	0	0	0	0	0	0	0	0	0
Number of clusters containing more than one feature from both TAR and LIDAR dataset (type 4)	1	1	0	0	0	0	0	0	0	0
Number of TAR features in these clusters	2	2	0	0	0	0	0	0	0	0
Number of LIDAR features in these clusters	2	2	0	0	0	0	0	0	0	0
Number of clusters with only TAR features (type 5)	3	3	3	0	0	0	0	0	0	0
Number of TAR features in such clusters	9	9	7	0	0	0	0	0	0	0
Number of clusters with only LIDAR features (type 6)	0	0	0	0	0	0	0	0	0	0
Number of LIDAR features in such clusters	0	0	0	0	0	0	0	0	0	0
Number of features that are in TAR dataset but are out of EPS distance of any LIDAR feature	29	29	31	39	39	40	42	52	71	112
Number of features that are recognized in LIDAR dataset but are out of EPS distance of any TAR feature	4	4	4	5	5	6	8	18	37	78

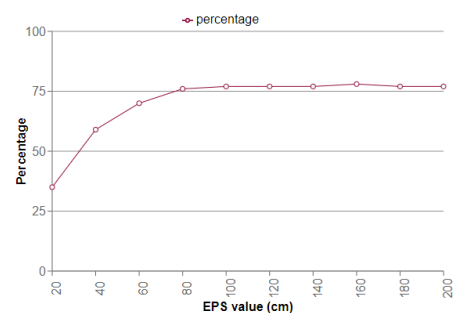
Number of features

Number of TAR features in the area	172
Number of LIDAR features in the area	138

Statistics of clustering (%)

EPS (given distance for clustering) in cm	200	180	160	140	120	100	80	60	40	20
Ratio (%) of matching features in TAR dataset (features in clusters of type 1 in TAR dataset)	77	77	78	77	77	77	76	70	59	35
Ratio (%) of TAR features of clusters of type 2 in all TAR features	0	0	0	0	0	0	0	0	0	0
Ratio (%) of TAR features of clusters of type 3 in all TAR features	0	0	0	0	0	0	0	0	0	0
Ratio (%) of TAR features of clusters of type 4 in all TAR features	1	1	0	0	0	0	0	0	0	0
Ratio (%) of LIDAR features of clusters of type 2 in all digitized LIDAR features	0	0	0	0	0	0	0	0	0	0
Ratio (%) of LIDAR features of clusters of type 3 in all digitized LIDAR features	0	0	0	0	0	0	0	0	0	0
Ratio (%) of LIDAR features of clusters of type 4 in all digitized LIDAR features	1	1	0	0	0	0	0	0	0	0
Ratio (%) of features in TAR dataset that are out of EPS Distance of any LIDAR feature	17	17	18	23	23	23	24	30	41	65
Ratio (%) of features in LIDAR dataset that are not in 3rd party database (out of EPS distance of any TAR feature)	3	3	3	4	4	4	6	13	27	57
Ratio (%) of features in LIDAR dataset that are not in 3rd party database plus those of clusters type 6 (ratio of LIDAR noise)	3	3	3	4	4	4	6	13	27	57
Ratio (%) of features in database that are not in TAR dataset plus those of clusters type 5 (ratio of TAR noise)	22	22	22	23	23	23	24	30	41	65
Ratio (%) of matching features in LIDAR dataset (features in clusters of type 1 in LIDAR dataset)	96	96	97	96	96	96	94	87	73	43

Percentage of correspondences (%)



Joonis 6. Ruumiandmete võrdluse klasterdumise baasstatistika.

5.2.2 Parima kaugusläve statistika

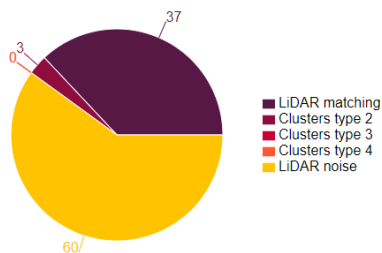
Selles osas esitatakse detailsem ülevaade ühe kaugusläve võrdlustulemuste kohta (vt. Joonis 7). Esitatav kauguslävi on määratud kasutaja poolt võrdluse käivitamisel või määramata jätmisel arvutatud meetodi poolt tulemuste põhjal. Uuesti on välja toodud tabelina andmed klastrite suhtelise jaotuse kohta. Lisaks on samad andmed esitatud sektordiagrammina mõlema andmehulga suhtes.

Meetod arvutab kaugused esimest tüüpi klastrites olevate objektide vahel. Kasutajale esitatakse maksimaalne ja minimaalne kaugus. Kõik kaugused jaotatakse kaugusklassidesse, et leida absoluutne ja suhteline kaugusjaotus. Selline statistika esitatakse tulp- ning sektordiagrammina.

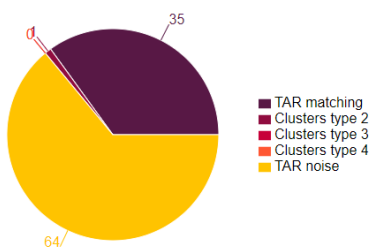
Statistics for EPS 180 cm

Clustering statistics for EPS value Ratio (%) of features of different clusters in LiDAR dataset

LiDAR matching	37	TAR matching	35
Clusters type 2	3	Clusters type 2	1
Clusters type 3	0	Clusters type 3	0
Clusters type 4	0	Clusters type 4	0
LiDAR noise	60	TAR noise	64



Ratio (%) of features of different clusters in TAR dataset



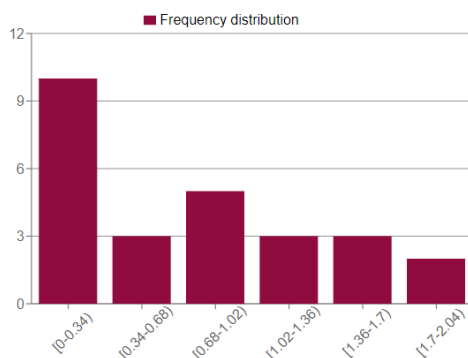
Distance frequency distribution

Count 180:	26
MAX 180:	1.7303
MIN 180:	0

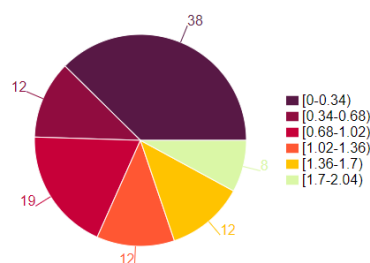
Relative frequency distribution (%)

Distance classes	Frequency distribution	Relative frequency distribution (%)
[0-0.34]	10	38
[0.34-0.68]	3	12
[0.68-1.02]	5	19
[1.02-1.36]	3	12
[1.36-1.7]	3	12
[1.7-2.04]	2	8

Distance frequency distribution



Relative frequency distribution (%)



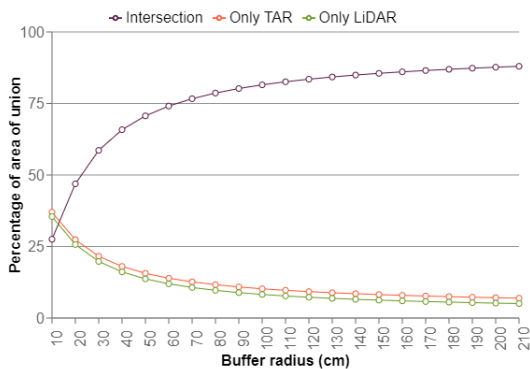
Joonis 7. Parima kaugusläve detailne statistika.

5.2.3 Graafikud joonobjektide nihke hindamiseks

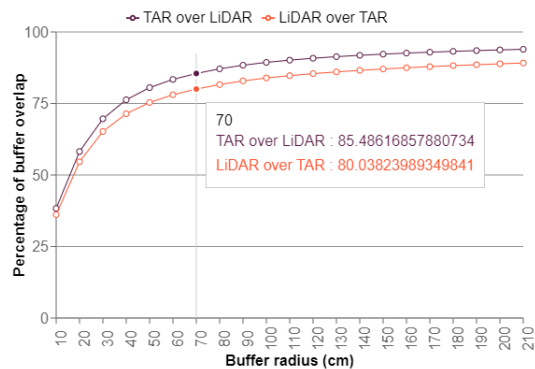
Järgnevad diagrammid esitatakse ainult joonobjektide võrdlemisel (vt Joonis 8). Joonte võrdlemisel esitatakse kasutajale kolm erinevat meetodil arvatud graafikut. Esimese kahe graafiku statistika aluseks on mõlema andmehulga kõigi paaritunud joonte ümber vastavalt kauguslääve väärtusele puhvri loomine. Esimene diagramm – puhvrite ülekatte statistika (*buffer overlay statistics (BOS)*) – väljendab mõlema puhvri ühisosa ning omavahelise vahe suhtelist jaotust nende ühendi suhtes [15]. Teine diagramm (BOM) väljendab, kui suur osa ühe andmehulga joonte puhvril kattub teise andmehulga puhvriga. Kolmanda graafiku (*single buffer method*) saamiseks luuakse puhver ümber esimese andmehulga joonte ning arvutatakse mitu protsenti teise andmehulga pikkusest meetrites langeb sellesse puhvrisse.

Line statistics

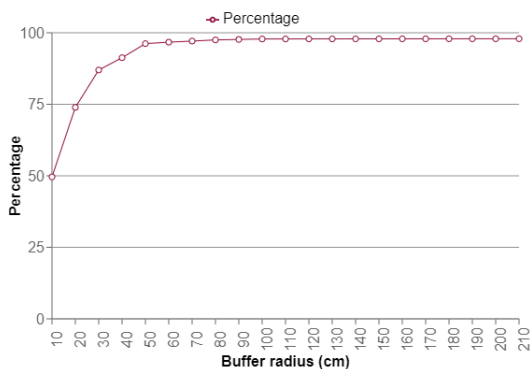
Buffer overlay statistics (BOS)



Overlap percentage of multilines (BOM)



Single buffer method



Joonis 8. Joonobjektide võrdluse tulemusi väljendavad diagrammid.

5.3 Testimine

Valminud tarkvara kasutajaliidesele kirjutati automaatseid teste. Testide kirjutamiseks kasutati JavaScript'i testimise raamistikku Jest [16], mida haldab Facebook. Jest'i fookus on testide kirjutamise lihtsusel ning sobib projektidele, mis kasutavad React raamistikku.

Testitud on tabelite ja diagrammide korrektne kuvamine ning vormi kasutamine. Testide eesmärk on veenduda kasutajaliidese korrektse töötamises ning tuvastada vead. Lisaks automaatsetele viidi läbi manuaalne kasutajaliidese testimine.

6 Võimalikud edasiarendused

Lõputöö käigus loodud prototüüp kiirendab ruumiandmete võrdluse tulemuste visualiseerimist, kuid sellele lisaks tuleks uurida, kuidas teha mugavamaks ja kiiremaks andmete ettevalmistamine võrdluseks.

6.1 Ruumiandmete sisestamine läbi kasutajaliidese

Valminud rakendus eeldab, et võrreldavad andmehulgad ruumiobjektidega on PostgreSQL/PostGIS andmebaasi tabelites. Uute andmete lisamiseks peab olema otse ligipääs andmebaasi. Võrreldavate ruumiobjektidega SHP failid (Shapefile) tuleb PostGIS'i *shp2pgsql* käsklusega teisendada SQL päringulauseteks ning seejärel laadida päringulausete abil ruumiandmed andmebaasi.

Ruumiandmete võrdlemise protsess oleks lihtsam, kui andmed, mida soovitakse võrrelda, oleks võimalik sisestada andmebaasi läbi antud töös arendatud rakenduse. Selline funktsionaalsus võiks käia läbi uue vaate, kus kasutaja saab lisada vormi kaudu SHP faile ning need esitada. Failid laetakse serverisse, teisendatakse SQL päringulauseteks ning laetakse andmebaasi. Uue funktsionaalsuse arendamise teeb raskeks sisendfailide valideerimine ja kontrollimine, kas sisestatud failid on SHP formaadis.

7 Kokkuvõte

Käesolevas töös loodi veebirakendus, mis kuvaks kasutajale ruumiandmete analüüsi statistilisi tulemusi. Rakenduse eesmärk oli asendada tulemuste manuaalne töötlemine ja diagrammide loomine Excel tabelarvutussüsteemis lahendusega, mis teeb seda automaatselt. Kuna võrdluse tulemuste tabelite suurused varieeruvad vastavalt võrdluse parameetritele, siis eelnev meetod tulemuste visualiseerimiseks oli töömahukas ning tõi endaga kaasa vigu.

Töö esimeses osas tehti kindlaks veebirakenduse nõuded ning uuriti, kas leidub olemasolevat tarkvara, mida saaks kohandada antud töö vajadustele. Vastavalt nõuetele määrati kindlaks rakenduse tehniline lahendus.

Tehnilise lahenduse põhjal arendati tarkvara, mis aitab ruumiandmete valdkonna eksperdil kiiresti võrrelda kahte andmehulka soovitud parameetritega. Võrreldavad objektid võivad olla punktid või jooned. Veebirakendus kuvab kasutajale võrdluse kohta statistikat tabelite ja diagrammide näol, mille põhjal on võimalik hinnata andmehulkade kokkulangevust. Rakendus aitab Archimedese rakendusuringus hoida kokku aega ning vältida vigu võrdlustulemuste esitamisel.

Valminud veebirakenduse prototüüp on lihtne ja kasutatav. Planeeritud eesmärk sai täidetud, kuid loodud tarkvara on võimalik täiendada, lisades ruumiandmete andmebaasi lisamise funktsionaalsuse, mis teeks võrdlusprotsessi terviklikumaks.

Kasutatud kirjandus

- [1] “Mis on GIS?,” [Online]. Available: <http://www.gispaev.ee/avaleht/mis-on-gis/>. [Accessed 25 05 2020].
- [2] “Kuluefektiivse ühildatava geodeetilise täpsusega 3D ruumiandmete taristu loomise rakendusuuring,” [Online]. Available: <https://www.etis.ee/Portal/Projects/Display/e344ff8c-ce7d-4b03-8db0-ed85cb4a8170>. [Accessed 19 05 2020].
- [3] R. Maigre, H.-M. Haav, R. Lillements, K. Julge and G. Anton, “A Method of Comparative Spatial Analysis of a Digitized (LiDAR) Point Cloud and the Corresponding GIS Database,” in *14th International Baltic Conference on Databases and Information Systems*, 2020, [].
- [4] “PostgreSQL,” [Online]. Available: <https://www.postgresql.org/>. [Accessed 19 05 2020].
- [5] “PostGIS,” [Online]. Available: <https://postgis.net/>. [Accessed 19 05 2020].
- [6] M. Ester, H.-P. Kriegel, J. Sander and X. Xu, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp. 226-231, 1996.
- [7] K. Wiegers and J. Beatty, *Software Requirements*, Microsoft Press, 2013, pp. 8-9.
- [8] “ST_ClusterDBSCAN,” [Online]. Available: https://postgis.net/docs/manual-dev/ST_ClusterDBSCAN.html. [Accessed 20 05 2020].
- [9] I. Hast, “Quality Assessment of Spatial Data: Positional Uncertainties of the National Shoreline Data of Sweden,” 2014.
- [10] E. Xavier, F. Ariza-Lopez and M. Ureña-Camara, “WPS for positional quality control applying the method proposed in UNE 148002,” 2015.
- [11] “Django,” [Online]. Available: <https://www.djangoproject.com/>. [Accessed 19 05 2020].
- [12] “React,” [Online]. Available: <https://reactjs.org/>. [Accessed 19 05 2020].
- [13] “Front-end frameworks popularity (React, Vue and Angular),” [Online]. Available: <https://gist.github.com/tkrotoff/b1caa4c3a185629299ec234d2314e190>. [Accessed 19 05 2020].
- [14] “Recharts,” [Online]. Available: <https://recharts.org/en-US/>. [Accessed 19 05 2020].
- [15] H. Tveite, “An accuracy assessment method for geographical line data sets based on buffering,” *International Journal of Geographical Information Science*, vol. 13, no. 1, pp. 27-47, 1999.
- [16] “Jest,” [Online]. Available: <https://jestjs.io/en/>. [Accessed 24 05 2020].

Lisa 1 – Ruumiandmete võrdluse päringu vastus

```
▼ {eps_ranges_table: [,...],...}
  best_eps_value: "50"
  ▼ distance_statistics: {stats_table: [{"Count 50:"}, 42], [{"MAX 50:"}, 0.44], [{"MIN 50:"}, 0]},...}
    ▼ distribution_classes: [{"[0-0.07)"}}, 37, 88], [{"[0.07-0.14)"}}, 0, 0], [{"[0.14-0.21)"}}, 2, 5], [{"[0.21-0.28)"}}, 1, 2],...]
      ▶ 0: [{"[0-0.07)"}}, 37, 88]
      ▶ 1: [{"[0.07-0.14)"}}, 0, 0]
      ▶ 2: [{"[0.14-0.21)"}}, 2, 5]
      ▶ 3: [{"[0.21-0.28)"}}, 1, 2]
      ▶ 4: [{"[0.28-0.35)"}}, 1, 2]
      ▶ 5: [{"[0.35-0.42)"}}, 0, 0]
      ▶ 6: [{"[0.42-0.49)"}}, 1, 2]
    ▼ stats_table: [{"Count 50:"}, 42], [{"MAX 50:"}, 0.44], [{"MIN 50:"}, 0]}
      ▶ 0: [{"Count 50:"}, 42]
      ▶ 1: [{"MAX 50:"}, 0.44]
      ▶ 2: [{"MIN 50:"}, 0]
  ▼ eps_ranges_table: [,...],...
    ▶ 0: [{"EPS (given distance for clustering) in cm"}, "240", "230", "220", "210", "200", "190", "180", "170",...]
    ▶ 1: [,...],...
    ▶ 2: [,...],...
    ▶ 3: [{"Number of taltech_curbstones_tar features in these clusters"}, "2", "2", "2", "3", "3", "3", "3", "3",...]
    ▶ 4: [{"Number of curbstones_taltech_ids features in these clusters"}, "4", "4", "4", "6", "6", "6", "6", "6",...]
    ▶ 5: [,...],...
    ▶ 6: [{"Number of curbstones_taltech_ids features in these clusters"}, "2", "2", "2", "1", "1", "1", "1", "1",...]
    ▶ 7: [{"Number of taltech_curbstones_tar features in these clusters"}, "4", "4", "4", "2", "2", "2", "2", "2",...]
    ▶ 8: [,...],...
    ▶ 9: [{"Number of taltech_curbstones_tar features in these clusters"}, "9", "9", "7", "3", "3", "3", "3", "2",...]
    ▶ 10: [{"Number of curbstones_taltech_ids features in these clusters"}, "8", "8", "6", "2", "2", "2", "2", "2",...]
    ▶ 11: [{"Number of clusters with only taltech_curbstones_tar features (type 5)"}}, "0", "0", "0", "0", "0", "0",...]
    ▶ 12: [{"Number of taltech_curbstones_tar features in such clusters"}, "0", "0", "0", "0", "0", "0", "0",...]
    ▶ 13: [{"Number of clusters with only curbstones_taltech_ids features (type 6)"}}, "1", "0", "0", "0", "0", "0",...]
    ▶ 14: [{"Number of curbstones_taltech_ids features in such clusters"}, "2", "0", "0", "0", "0", "0", "0",...]
    ▶ 15: [,...],...
    ▶ 16: [,...],...
  first_dataset_name: "curbstones_taltech_ids"
  ▼ line_stats: {bos: {10: [27.550859309268212, 36.96703846351488, 35.48210222721652]},...},...}
    ▶ bos: {10: [27.550859309268212, 36.96703846351488, 35.48210222721652]},...}
    ▶ overlap: {10: [38.33319548936484, 36.1460989856645], 20: [58.19863868848325, 54.61511042257564]},...}
    ▶ singlebuffer: {10: [49.618188481591346], 20: [73.90966759769377], 30: [86.97857893815342], 40: [91.28362675062135]},...}
  ▼ number_of_features: [{"Number of taltech_curbstones_tar features in the area"}, 58],...}
    ▶ 0: [{"Number of taltech_curbstones_tar features in the area"}, 58]
    ▶ 1: [{"Number of curbstones_taltech_ids features in the area"}, 56]
  second_dataset_name: "taltech_curbstones_tar"
```

