TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Maile Mäesalu 176250IDDR

Andmekvaliteedi hindamiseks näidisrakenduse loomine R keele näitel

Diplomitöö

Juhendaja: Meelis Antoi
Magistrikraad

Tallinn 2021

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: Maile Mäesalu

16.05.2021

Annotatsioon

Käesoleva diplomitöö eesmärgiks on luua rakendus andmekvaliteedi erinevate dimensioonide hindamiseks kasutades selleks peamiselt R keele vahendeid.

Rakenduse loomise käik vastab tavapärasele arendusmustrile. Rakenduse disainimise etapile järgneb prototüübi loomise etapp, mille tulemusena valmimad eraldiseisvana kasutajaliidese ja serveriloogika. Järgneva etapina ehitatakse rakendus ühtseks tervikuks ning sellele järgnevalt testitakse ja tugevdatakse valminud rakendust, et see vastaks paremini kasutaja ootustele ja pakuks paremat kasutajakogemust. Lõpptulemusena valmib töövahend andmekvaliteedi hindamiseks.

Rakenduse loomisel on eesmärgiks see, et selle kasutajaskond saaks olla võimalikult lai. Seega keskendub rakendus põhiandmetele, mis on omased valdavale osale asutustest ning pakub võimaluse analüüsimaks Eesti registris olevate juriidiliste isikute andmete andmekvaliteeti. Valminud rakenduses saab kasutaja üles laadida oma andmestiku ning selles sisalduvatele isikute andmetele teostatakse analüüs erinevate andmekvaliteedi dimensioonide (täielikkus, ühekordsus ja ajakohasus/õigsus) osas. Kasutajale kuvatakse nii analüüsi koondtulemusi kui ka mittevastavaid kirjeid. Samuti on kasutajal võimalik soovi korral mittevastavad kirjed CSV andmestikuna alla laadida.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 26 leheküljel, 6 peatükki, 14 joonist, 2 tabelit.

Abstract

Creating a Sample Application for Evaluating Data Quality Using the R Language

The aim of this thesis is to create a sample application using tools of the R programming language. This application uses several packages of the R language to show how a data driven application can be built using mainly tools in R.

The thesis presents different phases of creating the application and follows the suggested workflow. The first step, designing the application, is followed by prototyping. Prototyping includes building the front-end and the back-end, but doing it separately. Next the application is built by combining the business logic with the front-end. Finally, the working application is tested and strengthened to meet expected behaviour and increase conformity with user expectations. As the end result, the application provides a tool for evaluation data quality.

The aim of the application is to be of universal use, thus the application is focusing on main data which is common across almost all legal entities (i.e. the data of legal entities registered in Estonian). The application lets user upload a user specific dataset and performs analyses of different data quality dimensions – completeness, uniqueness and timeliness/validity. Both the summary results of the analysis and the problematic data rows are displayed to the user. The user also has the possibility to download problematic data rows as a CSV dataset.

The thesis is in Estonian and contains 26 pages of text, 6 chapters, 14 figures, 2 tables.

Lühendite ja mõistete sõnastik

Andmekvaliteedi dimensioonid	Aspektid, läbi mille on võimalik andmete kvaliteeti kirjeldada ja mõõta
Andmekvaliteet	Andmete vastavus andmete kasutajate ootustele
CRAN	*The Comprehensive R Archive Network* – R keele paketihaldur
CSS	*Cascading Style Sheets* – märgistuskeel, mida kasutatakse veebilehtede kujundamisel
CSV	*Comma-separated values* - failivorming, kus andmed on üksteisest eraldatud semikoolonitega, komadega või tabeldustega
DOM	*Document Object Model* – dokumentide objektimudel
ggplot2	R keele lisapakett graafika loomiseks
golem	Raamistik R Shiny rakenduste loomiseks
HTML	*Hyper Text Markup Language* – keel, milles märgendatakse veebilehti
IDE	*Integrated development environment* – integreeritud arenduskeskkond
plotly	R keele lisapakett kaasaegse graafika loomiseks
PyPI	*The Python Package Index* – python keele paketihaldur
Python	Programmeerimiskeel
R	Kõrgtaseme objektorienteeritud interpreteeritav programmeerimiskeel
renv	R keele lisapakett sõltuvuste haldamiseks projektis
RStudio	R keele IDE
Shiny	Pakett (veebi)rakenduste loomiseks R keeles

Sisukord

Jooniste loetelu

Tabelite loetelu

1 Sissejuhatus

Andmehalduse ja andmekvaliteedi teema avaliku sektori andmekogude kontekstis on hetkel väga aktuaalne ning käesoleval hetkel on riik ellu viimas riiklikele andmekogudele suunatud andmehalduse tegevuskava [1]. Andmekogudesse on aastate jooksul talletatud teenuse osutamiseks vajalikku infot, kuid andmete kvaliteedireegleid ei ole kõikide andmekogude osas formaliseeritud ning samal ajal on ka kasutajarakenduste funktsionaalsus selles osas pideva täiendamise protsessis. Samas tunnetatakse, et puudused andmete kvaliteedis raskendavad teenuse osutamist. Andmete kvaliteet on oluline ka erasektori ettevõtete kontekstis, kuna on leitud, et andmete kvaliteeditaseme probleemidega kaasnevad ettevõtetele märkimisväärsed kulud.

Käesolevas diplomitöö eesmärk on luua rakendus, mida on võimalik kasutada konkreetse andmekogu või andmestiku osa, Eesti registris olevate juriidiliste isikute, andmekvaliteedi hindamiseks. Võrdlusandmestikuna on seejuures võimalik kasutada Registrite ja Infosüsteemide Keskuse poolt avaandmetena pakutavat ja iganädalaselt uuenevat äriregistri andmeid sisaldavat andmestikku. Loodavat rakendust saab kasutada ka andmekvaliteedi parendamise protsessis andmete seisundi kordushindamisteks, et jälgida muudatuste ellu viimise tulemuslikkust. Töö teostamiseks kasutan viimastel aastatel TIOBE indeksi [2] kohaselt populaarsust kogunud vabavaralist R keelt [3], mis võimaldab efektiivselt nii vaatlusalust andmestikku kui ka võrdlusandmestikku töödelda. Kasutades lisapaketti Shiny on võimalik R keelt kasutades luua ka kasutajaliidest sisaldav ning analüüsi tulemusi visualiseeriv rakendus.

Töö on jagatud viieks peatükiks. Esimeses peatükis analüüsitakse andmekvaliteedi valdkonda ning andmete kvaliteeditasemega seonduvaid probleeme ja nendest lähtuvaid kulusid. Teine peatükk toob välja antud töö raames loodava rakenduse funktsionaalsed ja mittefunktsionaalsed nõuded. Kolmandas peatükis analüüsitakse rakenduse loomiseks kasutatavate tehnoloogiate valikut ning rakenduse loomiseks valitud R keele raames kasutamiseks sobivaid vahendeid. Töö neljas peatükk annab ülevaate rakenduse loomisese käigust hõlmates rakenduse disaini, prototüübi loomise etappi ja üheks

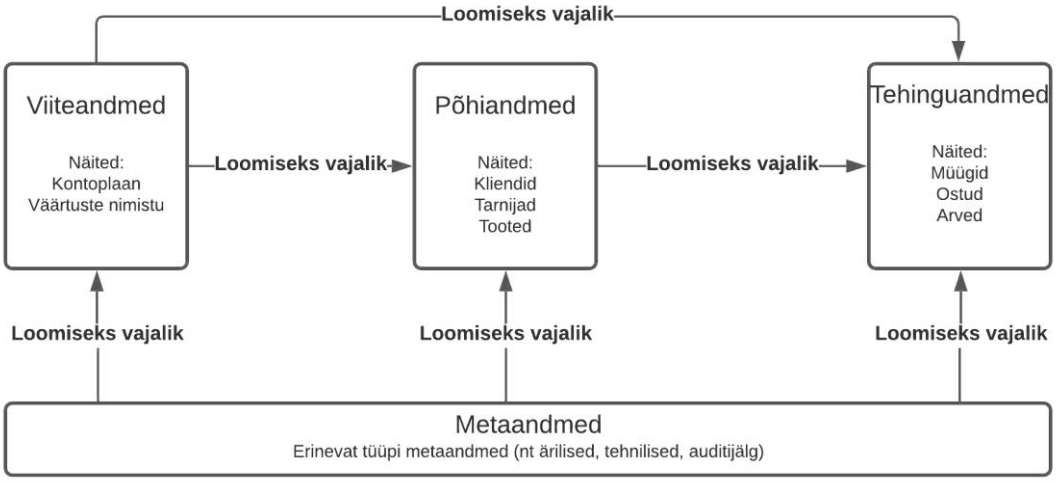
tervikuks ehitamist. Viiendas peatükis leiab kajastamist rakenduse testimine ja tugevdamine.

2 Andmekvaliteedi valdkonna ja probleemi kirjeldus

Käesolev peatükk annab ülevaate andmete ja andmekvaliteedi olemusest, probleemidest, mida tingib madal andmekvaliteet ning andmekvaliteedi hindamise põhimõtetest.

2.1 Andmete mõiste, kategooriad ja andmekvaliteet

Sebastian-Coleman on andmed defineerinud järgnevalt – andmed on abstraktne päriselus olemas olevate objektide, sündmuste ja kontseptsioonide valitud tunnuste esitamine. Tunnuseid esitatakse ja mõistetakse läbi selgelt määratletud kokkulepete, mis on seotud andmete tähenduse, kogumise ja salvestamisega [4].



Joonis 1. Andmekategooriate vahelised seosed

MacGilvray on andmed jaganud kategooriatesse järgnevalt [5]:

- Põhiandmed (*master data*) – andmed isikute (sh kliendid, töötajad, tarnijad), kohtade ja asjade kohta organisatsioon tegevusvaldkonnas
- Tehinguandmed (*transactional data*) – andmed sündmuste ja tehingute kohta
- Viiteandmed (*reference data*) – andmegrupid või klassifikatsioonid, millele rakendustes, protsessides, aruannetes jne põhi- või tehinguandmete raames viidatakse.

- Metaandmed (*metadata*) – andmed andemete kohta

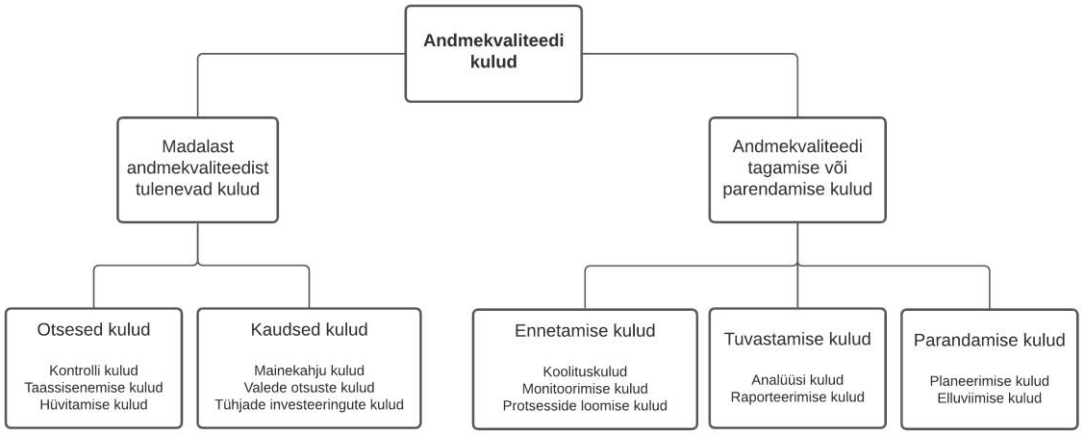
Nimetatud andmekategooriatest on põhiandmed need, mida kasutavad olulised äriprotsessid, mistõttu just põhiandmete formaat ja ajakohasus on kriitilised. Andmekategooriate vahelisi seoseid kajastab Joonis 1 [5].

Andmekvaliteeti defineerib Sebastian-Coleman sarnaselt paljude teiste autoritega. Andmete kvaliteeditase näitab seda, mil määral andmed vastavad andmete kasutajate ootustele [4]. Seega sõltub hinnang kvaliteeditasemele konkreetsetest andmetest ja ka sellest, millisel eesmärgil neid andmeid kasutatakse.

2.2 Andmete kvaliteeditasemega seotud kulud

Kirjanduses on laialdaselt leitud, et madala andmekvaliteediga on seotud kõrged kulud ning täpsemalt on hinnatud, et ettevõtted võivad seetõttu kaotada üle 10% oma võimalikest tuludest [7]. Andmete kvaliteeditasemega seoses tekib organisatsioonides kulusid kahelt erinevalt poolelt: ühelt poolt andmete kvaliteedi soovitud taseme tagamise kulud ja teisalt andmete kvaliteeditaseme puudustest tulenevad kulud [6] [7].

Andmete kvaliteeditaseme puudused viivad kliendirahulolu vähenemiseni, jooksvate kulude suurenemiseni, ebaefektiivsete juhtimisotsusteni ja ka töötajate rahulolu vähenemiseni [6]. Eppler ja Helfert poolt on koostatud andmekvaliteedi probleemidega seotud kulude jaotus, mida kajastab Joonis 2 [8].



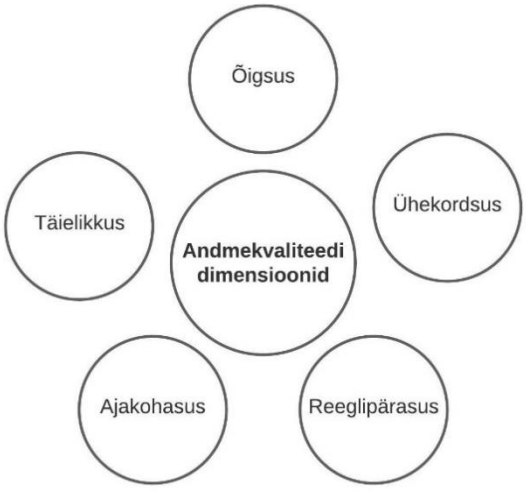
Joonis 2. Andmekvaliteedi kulude liigitus

Kuna käesoleva töö eesmärgiks on universaalse kasutusega töövahendi (rakenduse) loomine, ei lähtuta siinjuures ühegi konkreetse organisatsiooni andmetest ja ootustest andmetele. Sellest tulenevalt ei ole käesoleva töö raames ka võimalik välja arvutada loodava töövahendi rakendamisest saadavat rahalist kasu. Küll aga võib käesolevas peatükid käsitletud allikatele tuginedes eeldada, et põhiandmete kvaliteeditaseme parendamisel olukorras, kus andmekvaliteedi lähtetase on allpool oodatavat taset, oleks organisatsiooni kulusid vähendav mõju ning seeläbi positiivne mõju edukusele.

2.3 Andmekvaliteedi dimensioonid ja juhtimise protsess

Andmekvaliteedi valdkonnas kasutatakse hinnangu andmiseks andmekvaliteedi dimensioone. Mõiste andmekvaliteedi dimensioonid kirjeldab aspekte, läbi mille on võimalik andmete kvaliteeti kirjeldada ja mõõta [4].

Erinevatel autoritel on palju erinevaid andmekvaliteedi dimensioonide käsitlusi. Näiteks Sebastian-Coleman[4] toob välja viis dimensiooni, Eesti andmehalduse metoodikaprojekti raames koostatav andmekvaliteedi juhis samuti viis [7], kuid ka need kaks näidet on oma sisult veidi erinevad. Siinkohal toon välja nimetatud andmekvaliteedi juhises kasutatud viis dimensiooni, milleks on õigsus, täielikkus, ajakohasus, reeglipärasus ja ühekordsus (Joonis 3). Antud käsitlusest lähtun käesolevas töös ka edaspidi.



Joonis 3. Andmekvaliteedi dimensioonid

MacGilvray [5] toob välja 10etapilise protsessi, mille abil hinnata, parendada ja luua info ja andmete kvaliteeti. Esimeseks etapiks on andmekvaliteedi mõõtmiseks vajalike äri- ja andmekvaliteedi reeglite tuvastamine. Sellele järgneb infokeskkonna analüüs ja seejärel andmekvaliteedi hindamine koos mõju analüüsiga organsatsiooni tegevuse kontekstis. Mõõtmistulemuste baasilt on järgnevalt võimalik välja tuua andmekvaliteeti mõjutavad peamised põhjused ning seejärel luua tegevuskava andmekvaliteedi parendamiseks.

O'Brien *et al* [6] toovad välja, et andmekvaliteedi parendamise protsess ei seisne ainuüksi andmestiku andmete parandamises, vaid terviklik lähenemine peaks lisaks andmetele hõlmama ka inimesi ja protsesse organisatsioonis. Seejuures ei saa olla ühest lahendust kõigile olukordadele, vaid sobivad meetmed sõltuvad olukorrast ja asjaoludest. Nii akadeemiline kui ka ärimaailm on leidnud tõendeid, et lahendused andmekvaliteedi parendamiseks on olemas [6].

Järgnevas alapeatükis analüüsitakse, millist rolli võiks täita ja milliseid andmekvaliteedi dimensioone ning andmekvaliteedi hindamise ja parendamise protsessi etappe haarata käesoleva töö raames loodava rakenduse funktsionaalsus.

2.4 Loodava rakenduse skoop

Käesoleva töö eesmärgiks on luua rakendus sellele osale organisatsioonide andmestikes, mida saaks lugeda organisatsioonide üleseks. Fookus on seega sellel, et loodav lahendus ei oleks ühe organisatsiooni keskne, vaid selle kasutajaskond võiks ja saaks olla laiem. Organisatsioonide ülese olemusega on ennekõike põhiandmete hulka kuuluvate isikute (klientide, tarnijate) andmed. Isikute andmeid leidub nii avaliku sektori andmekogudes kui ka erasektori ettevõtete andmestikes ning nende õigsust võib lugeda paljude organisatsioonide seisukohalt oluliseks. Selleks, et oleks võimalik hinnata andmete kvaliteeti, peab leiduma korrektsete andmetega võrdlusandmestik. Eestis on siinjuures sobivaks andmekoguks äriregister.

Vastavalt äriseadustiku §22 lõikele 1 on äriregister „riigi infosüsteemi kuuluv andmekogu, mille pidamise eesmärk on koguda, säilitada ja avalikustada teavet Eestis asuvate füüsilisest isikust ettevõtjate ettevõtete, äriühingute ja välismaa äriühingu filiaalide kohta. Äriregistrit peab Tartu Maakohtu registriosakond." Registrite ja Infosüsteemide Keskus annab avaandmetena välja regulaarselt uuenevaid riigi

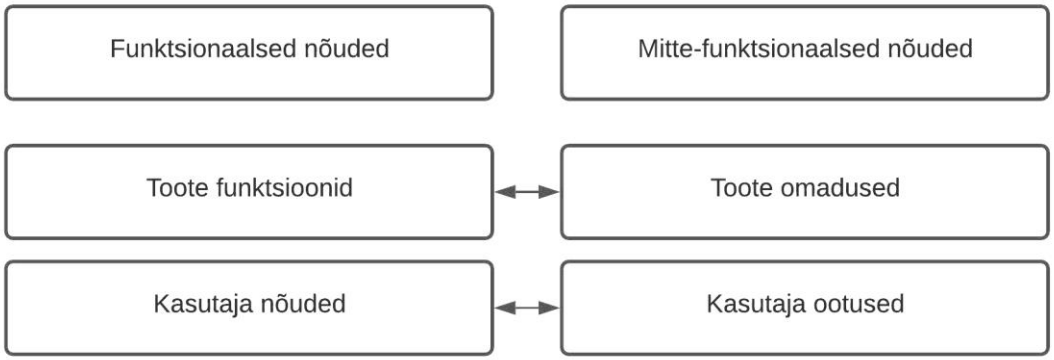
andmekogude andmeid, sh ka äriregistri andmeid, mida uuendatakse üks kord nädalas [10]. Seega on tasuta kättesaadavad äriregistrisse kantud ajakohased andmed kõikide Eesti registrisse kantud juriidiliste isikute kohta ning käesoleva töö raames loodava rakenduse seisukohalt võib seda lugeda sobivaks andmestikuks, millega kasutaja andmeid võrreldes on võimalik hinnata kasutaja andmete õigsust.

Ka eelnevalt selgitatud kontrolli ulatuse puhul on võimalik vaadelda andmekvaliteeti erinevate dimensioonide raames. Eelnevas punktis nimetatud andmekvaliteedi dimensioonidest on sellise piiritletusse korral võimalik hinnata kasutaja andmestikus olevate andmete osas täielikkust (kuivõrd on kõik kohustuslikud väljad sisustatud) ja ühekordsust (kas iga äriregistri kood ja isiku nimi esineb andmestikus vaid ühel korral). Võrdlusandmestiku ja kasutaja andmestiku andmete baasilt saab anda hinnangu kasutaja andmete õigsusele (kasutaja andmestikus toodud äriregistri kood ja isiku nimi vastavast võrdlusandmestikus toodule), ajakohasusele (see on antud juhul täidetud juba läbi õigsuse dimensiooni) ja ka reeglipärasusele (kas äriregistri kood on numbriline).

Loodav rakendus keskendub küll kitsalt isikute andmekvaliteedi tuvastamisele, kuid peab võimaldama hilisemat andmekogu või andmestiku eripärast tulenevat kontrolli skoobi laiendamist. Sellest tulenevalt peab loodava rakenduse ülesehitus olema selline, et seda oleks võimalik äri- ja andmekvaliteedi reeglite osas suuremale kontrolli ulatusele laiendada.

3 Loodava rakenduse nõuete analüüs

Käesolev peatükk käsitleb loodava rakenduse nõudeid. Olgu siinkohal ülevaatlikult selgituseks välja toodud, et nõuded jagunevad funktsionaalseteks ja mitte-funktsionaalseteks nõueteks.



Joonis 4. Funktsionaalsed ja mittefunktsionaalsed nõuded tarkvaraarenduses

Funktsionaalsed nõuded hõlmavad toote funktsionaalsust ja kasutaja poolseid nõudeid. Mitte-funktsionaalsed nõuded seevastu toote omadusi ja kasutaja ootusi (Joonis 4 [11]).

3.1 Funktsionaalsed nõuded

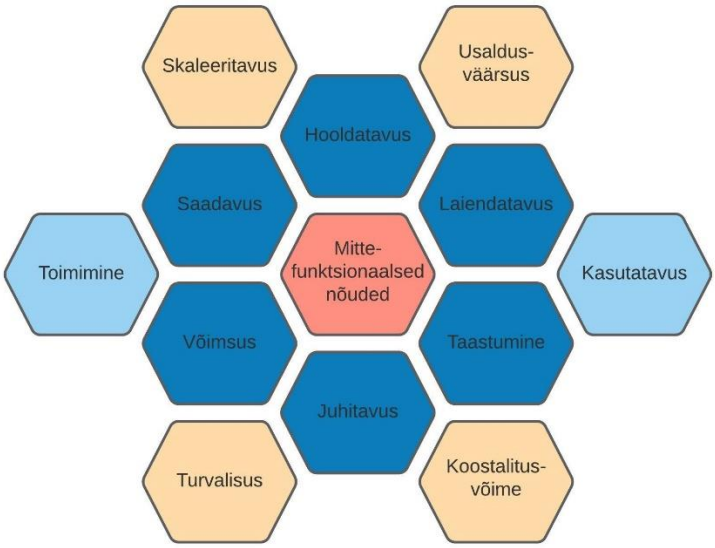
Funktsionaalsed nõuded seega kirjeldavad, mida loodav rakendus peab tegema. Kasutaja poolsed funktsionaalsed nõuded loodavale rakendusele kirjeldatuna läbi kasutajalugude oleksid:

- Kasutajana soovin võrguühenduse puudumisel analüüsi teostamiseks kasutada eellaaditud andmestikku
- Kasutajana soovin analüüsi teostamiseks alla laadida kõige värskema võrdlusandmestiku

- Kasutajana soovin võrdlusandmestiku andmete vaatamise võimalust
- Kasutajana soovin üles laadida oma andmestiku
- Kasutajana soovin oma andmestiku vaatamise võimalust peale selle üles laadimist
- Kasutajana soovin määrata, millised andmeväljad minu andmestikus vastavad kontrollitavatele väärtustele võrdlusandmestikus
- Kasutajana soovin andmeid analüüsida registrikoodi ja isiku nime osas
- Kasutajana soovin analüüsi tulemuste vaatamist võimalust rakenduses
- Kasutajana soovin analüüsi tulemused CSV failina alla laadida

3.2 Mitte-funktsionaalsed nõuded

Mitte-funktsionaalsed nõuded hõlmavad nõudeid sellele, kuidas funktsionaalseid nõudeid täitma peaks. Paradkar [12] poolt välja toodud mitte-funktsionaalseid nõudeid kajastab Joonis 5.



Joonis 5. Tähtsamad mitte-funktsionaalsed nõuded

Antud rakenduse osas järgin ka e-teenuse disaini põhimõtet, mille kohaselt peaks teenus olema lihtne kasutada. See tähendab, et see peaks olema lihtne nii funktsionaalsuselt kui ka keeleliselt ja kujunduslikult. [13]

Loodava rakenduse mitte-funktsionaalseteks nõueteks on:

- Rakendust peab olema võimalik kasutada ka ilma võrguühenduseta
- Rakendus peab kasutaja andmestikku lugeda suutma
- Rakendus peab kuvama kasutajale infot andmestiku lugemise õnnestumise kohta
- Rakendus peab olema korrektses eesti keeles
- Rakenduse kasutajaliides peab olema etteaimatav
- Rakenduse kasutajaliides peab olema arusaadav ilma andmekvaliteedi valdkonnas põhjalikke teadmisi omamata
- Rakendus peab olema lihtsa kujundusega
- Rakendus peab vaadete vahetamisega kasutajat läbi vajalike etappide suunama
- Rakendus peab olema skaleeritav

4 Vahendite ja tehnoloogiate valik

Antud peatüki raames analüüsitakse rakenduse loomiseks arenduskeele, selle lisapakettide ja raamistiku valikut.

4.1 Arenduskeele valik

Antud rakenduse näol on tegemist andmeanalüüsi töövahendiga. Üks rakenduse sisendandmestik, Registrite ja Infosüsteemide Keskuse poolt avaandmetena pakutav ja iganädalaselt uuenev andmestik, sisaldab üle 300 tuhande andmekirje [10]. Seega on otstarbekas antud rakenduse teostamiseks valida arenduskeel, mis võimaldab efektiivset ja mugavat andmete töötlemist. Samuti on antud töös teostatava rakenduse seisukohalt oluline kasutajaliidese loomiseks vajalike vahendite olemasolu.

Nimetatud eeldusi võiksid arenduskeeltest täita Python, R ja Matlab. Järgnevalt analüüsin Tabel 1 nende andmeanalüüsiks sobivate arenduskeelte plusse ja miinuseid käesoleva töö raames loodava rakenduse seisukohalt. Matlab ja R on loodud suunitlusega teadusele (esimene matemaatiliste probleemide lahendamiseks ja graafikute visualiseerimiseks, teine pigem kitsama fookusega statistiliseks andmeanalüüsiks). Python seevastu on üldkasutatav programmeerimiskeel, mis on ulatuslikuma kasutuse saanud läbi täiendavate pakettide lisandumise.

R ja Python võimaldavad mõlemad luua nii töölaua- kui ka veebirakendusi. R keelel on selle jaoks lisapakett Shiny ning Python puhul on võimalik kasutada Django raamistikku. Matlabi puhul on samuti võimalik veebirakendusi luua, kuid need on turvalisuskaalutlustest tulenevalt mõeldud kasutamiseks ainult asutuse sisevõrgus [14]. Funktsionaalsuselt sobiksid antud töö raames loodava rakenduse puhul kasutamiseks kõik kolm käsitletavat arenduskeelt. Olulise erinevusena on Matlabi näol tegemist litsentseeritud tarkvaraga, samal ajal kui nii Python kui ka R on vabavara. Nimetatud põhjustel jääb ka Matlab antud töös kõrvale. Litsentseeritud tarkvarast tulenevalt on ka

Matlabi kasutajaskonna näol tegemist suletud ringkonnaga, samal ajal kui Pythoni ja R keeli arendab laia osalejaskonnaga avatud kasutajatering.

Tabel 1. Arenduskeelte võrdlus

	Python	**R**	**Matlab**
Loomise aasta	1989 [15]	1993 [16]	1984 [17]
Keele tüüp	Interpreteeritav programmeerimiskeel	Interpreteeritav programmeerimiskeel	Kõrgkeel
Peamine kasutusvaldkond	Üldkasutus	Andmeanalüüs, statistiline analüüs	Matemaatika, statistika, masinõpe
Arendus-keskkond	IDLE, PyCharm jne	RStudio, IntelliJ jne	Integreeritud arenduskeskkond
Veebirakenduse loomise võimalus	Lisapakett Django	Lisapakett Shiny	Web App Compiler (ainult intraneti rakendused)
Sobivus andme-analüüsiks	Väga hea; Lisapaketid Pandas, Plotly	Väga hea; Lisapaketid dplyr, ggplot jne	Väga hea
Paketihaldus	Python Package Index (PyPI) [18]	The Comprehensive R Archive Network (CRAN) [19]	Matlab Package Manager [20]
Kasutajaskond	Lai avatud kasutajaskond	Lai avatud kasutajaskond (ajalooliselt pigem akadeemiline)	Lai suletud kasutajaskond, suures osas akadeemiline
Tarkvara tüüp	Vabavara	Vabavara	Litsentseeritud tarkvara
Isiklik kasutamis-vajadus tulevikus	Hetkel vajadust ei prognoosi	Vajalik tööalaselt	Hetkel vajadust ei prognoosi

Käesoleva diplomitöö autori isikliku huvi seisukohalt on R eelistatud Pythonile, kuna seda keelt kasutab autor tööalaselt tõenäoliselt ka edaspidi. Samas on R keel autorile ka väljakutseks, sest sellega põhjalikum kokkupuude eelnevalt puudub.

4.2 Arenduskeskkond ja R keele pakettide valik

R keele näol on tegemist statistiliste arvutuste ja graafika teostamiseks mõeldud keele ja keskkonnaga. See on sarnane S keelele, üks selle implementatsioone. R ei ole ainuüksi statistikaks mõeldud süsteem, vaid keskkond, mille raames saab rakendada statistika tehnikaid [21]. Lisaks põhipakettidele on R keele paketihalduri CRAN abil kättesaadavad üle 17 tuhande lisapaketi [19].

Arenduskeskkonnana on antud töö raames kajastatud rakenduse kirjutamiseks otstarbekas kasutada tarkvara RStudio. RStudio on R keele jaoks loodud integreeritud arenduskeskkond (ingl k *integrated development environment (IDE)*), mis hõlmab konsooli, redaktorit ning tööriistu koodi silumiseks ja töölaua haldamiseks. RStudio on olemas nii vabavaralise kui ka kommertsverioonina [22].

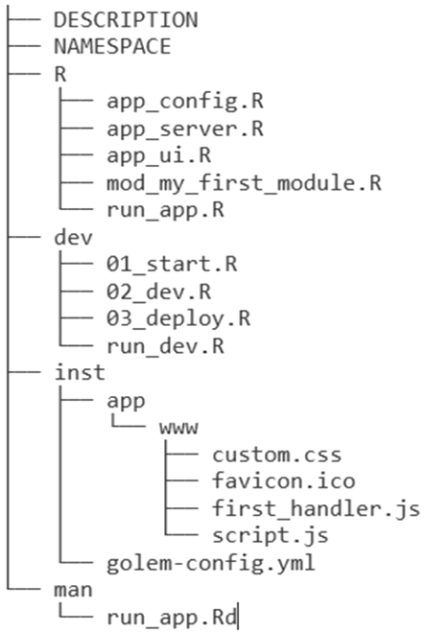
```
library(shiny)
ui <- fluidPage(
  "Hello, Shiny!"
)
server <- function(input, output, session) {
}
shinyApp(ui, server)
```

Joonis 6. R Shiny baasrakenduse koodinäide

Antud töö raames loodava rakenduse seisukohalt on olulisel kohal R Shiny lisapakett, mis võimaldab R keele abil luua tänapäevaseid interaktiivseid (veebi)rakendusi [23]. Shiny rakenduse moodustab R keele skript nimega app.R, mis koosneb kasutajaliidese objektist ja serverifunktsioonist. Kasutajaliidese objekt kirjeldab kasutajaliidese ja sisendite ning väljundite elemente. Serverifunktsiooni põhiülesandeks on sisendite ja väljundite vaheliste seoste defineerimine. Serverifunktsiooni raames loetakse sisendväärtused ning peale arvutuste teostamist omistatakse reaktiivsed väärtused kasutajaliideses loodud väljunditele. Lihtne töötav Shiny baasrakendus on näidatud Joonis 6.

4.3 {golem} arendusraamistik

Selleks, et loodav rakendus vastaks hilisema hallatavuse ja skaleeritavuse nõudmisele, tuleb selle loomise juures pöörata tähelepanu rakenduse arhitektuurile. Stabiilsete ja hallatavate R Shiny rakenduste ehitamiseks on loodud {golem} raamistik.



Joonis 7. {golem} raamistikuga loodud R Shiny rakenduse baasarhitektuur

{golem} raamistiku esimene väljalase valmis aastal 2019 ja selle autoriks on R keele eksperte koondav prantsuse ettevõte ThinkR [24]. {golem} raamistiku abil loodud rakenduse baasarhitektuuri kajastab Joonis 7. Võrreldes lihtsa ühefailise Shiny rakendusega on see mugav töövahend mahukama rakenduse loomise protsessi ajaks. Samuti soosib see moodulitel põhineva rakenduse ülesehituse loomist, mis tähendab, et {golem} raamistiku kasutamine lihtsustab skaleeritava rakenduse loomist.

4.4 Rakenduse loomise käik

{golem}raamistiku abil R Shiny rakenduse loomisel on soovitav lähtuda Tabel 2 kajastatud etappidest [25].

Tabel 2. Rakenduse loomise etapid

Etapp	**Täpsem sisu**
Disainimine	Kasutaja ootuste selgitamine ja nõuete tõlkimine tehniliseks spetsifikatsiooniks
Prototüübi loomine	Kasutajaliidese ja serveripoolse lahenduse eraldiseisev ehitamine
Ehitamine	Äriloogika (serveripoolse lahenduse) ühendamine kasutajaliidesega
Tugevdamine	Testimine, koodi refaktoreerimine

Disainimise etapis tuleb paika panna, milline peaks olema rakenduse funktsionaalsus. Funktsionaalsus omakorda peaks lähtuma kasutaja ootustest. Prototüübi loomise etapp hõlmab rakenduse kasutajaliidese ja serveriloogika ehitamist, kuid seda eraldiseisvatena. Kasutajaliidese lahendus ja serveriloogika liidetakse järgmises, rakenduse ehitamise etapis. Tugevdamise etapp peaks tagama rakenduse töökindluse läbi mitmekülgse testimise ja testimise käigus tuvastatud probleemkohtade lahendamise.

Järgnev peatükk annab ülevaate kolmandas peatükis välja toodud nõuetele vastava rakenduse loomisest kasutades antud peatükis välja toodud vahendeid.

5 Rakenduse loomine

Antud peatükis tuuakse välja rakenduse loomise käik alates disainimisest ja prototüübi loomisest kuni rakenduse tervikuks ehitamise ja selle tugevdamiseni. Rakenduse lähtekood on kättesaadav aadressil https://github.com/mamaesalu/theShinyApp.

5.1 Disainimine

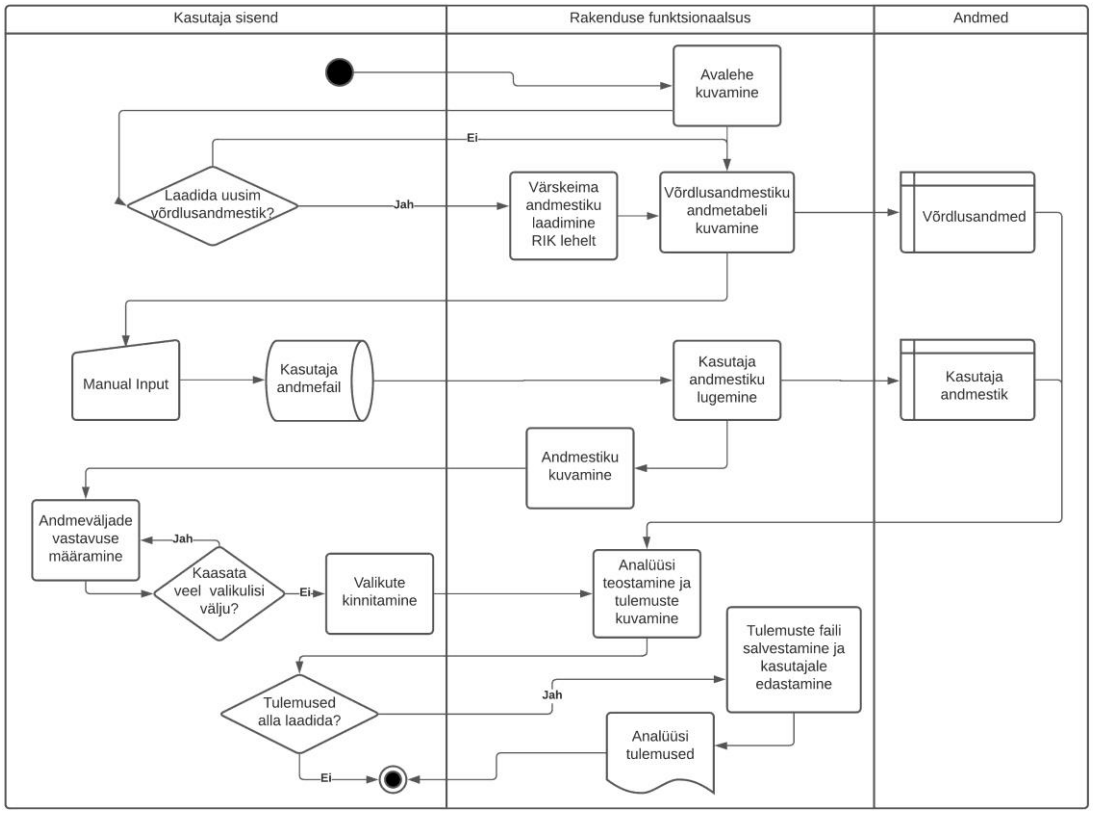
Rakenduse loomist on otstarbekas alustada kasutaja ootuste analüüsist ning tuvastatud nõuete baasilt panna kokku tehniline spetsifikatsioon. Kasutaja ootusi olen analüüsinud käesoleva töö punktis 2.4 ning nõudeid loodavale rakendusele töö 3. peatükis. Teostatud analüüsi põhjal saan kirjeldada rakenduse funktsionaalsuse (Joonis 8).

Loodava rakenduse kasutajaks on tõenäoliselt kogenud arvutikasutaja, kuna rakenduse kasutamise vajaduse juurde jõudmine eeldab mõningast teadlikkust andmekvaliteedi valdkonna ja olulisuse kohta. Samuti on eelduseks see, et kasutajal on olemas CSV fail andmetega, mille osas ta hinnangut soovib. Seega ei ole tõenäoline, et rakenduse kasutajaks on algaja arvutihuviline. Vaatamata eeltoodule järgin disainipõhimõtetest rakenduse loomisel seda, et rakendus oleks võimalikult lihtne ja intuitiivse kasutajateega.

Kujunduse osas otsustasin mitte kasutada lisapakette (nt Shinydashboards vms), vaid kogemuse saamiseks luua rakenduse kujundus iseseisvalt. Kuna eesmärgiks on luua universaalse kasutusulatusega informatiivne rakendus ning võrdlusandmestik on kättesaadav avalikul veebilehel, ei ole antud rakenduse puhul vajalik luua andmebaasiühendusi. Samuti ei pea ma vajalikuks kasutajakontode loomise funktsionaalsust – see muudaks rakenduse veebiversiooni kasutamise keerukaks ja ajamahukaks.

Kasutaja poolt rakenduses võrdluseks üles laetud andmestiku sisu piiritletuse eest vastutab kasutaja ning võrdluseks üles laetavad kasutaja andmed ei ole oma olemuselt tundlikud andmed (tegemist on avalikult kättesaadavate äriregistri andmete kliendipoolse versiooniga). Samas võivad need siiski muutuda tundlikeks teades täiendavalt andmete

päritolu (millise ettevõtte või asutuse andmetega on tegemist) ja konteksti (kelle andmetega on tegemist, nt klientide või tarnijate). Nimetatud riskide maandamiseks on alternatiivina loodavat rakendust võimalik versioonihaldusest alla laadida ja kasutada seda näiteks sisevõrgus.



Joonis 8. Loodava rakenduse funktsionaalsus

5.2 Prototüübi loomine

Shiny rakenduste loomisel on prototüübi loomiseks võimalik kasutada {shinipsum} lisapaketti, mis võimaldab R Shiny rakenduse kirjutamise käigus kasutajaliidese ja serverifunktsiooni eraldiseisvat arendamist. Koodinäidet{shinipsum} paketi kasutamisest serverifunktsioonis, mis võimaldab kasutajaliidese eraldiseisvat disainimist, kajastab Joonis 9.

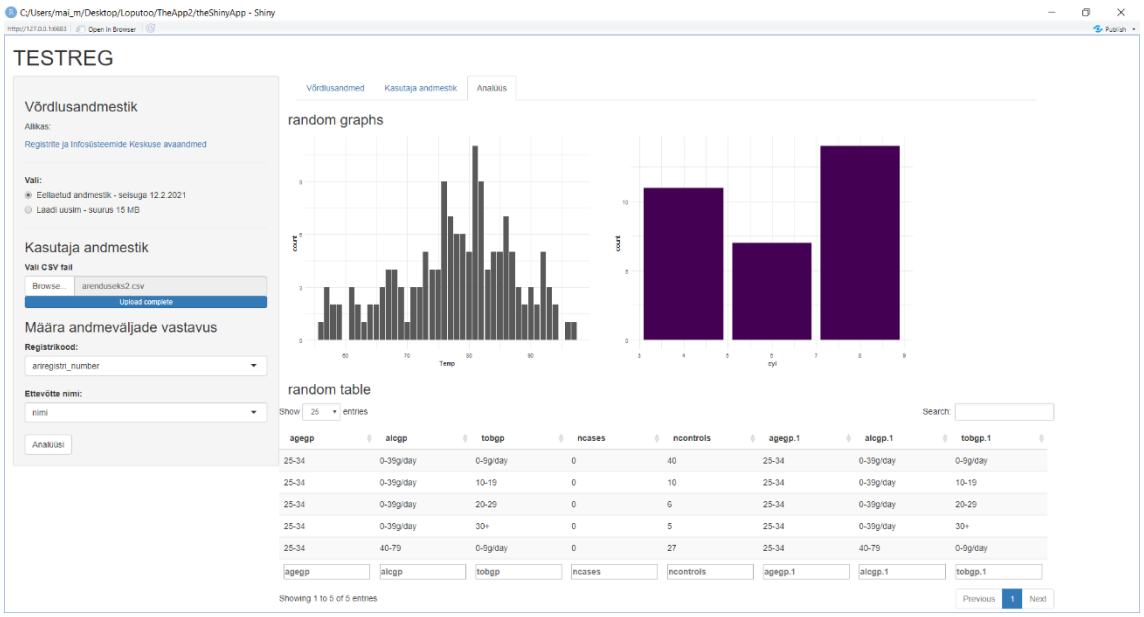
```
output$numberPie <- renderPlot({
  shinipsum::random_ggplot("bar")
})

output$namePie <- renderPlot({
  shinipsum::random_ggplot("bar")
})

output$compResult <- renderDataTable({
  shinipsum::random_table(5, 8)
})
```

Joonis 9. Rakenduse prototüübi loomise kasutades {shinipsum}paketti

Loodava rakenduse kasutajaliidese prototüüpi kajastab Joonis 10. Rakenduse vasakus ääres on ala kasutaja valikutega ja võimalus andmestiku üleslaadimiseks. Rakenduse paremal poolel on erinevatel vahelehtedel testandmestik ja kasutaja andmestik. Eraldi vahelehel kuvatakse analüüsi tulemusi.



Joonis 10. Rakenduse prototüüp

5.3 Rakenduse terviklikuks ehitamine

Loodud rakenduse ülesehitus koosneb moodulitest, st et peamine serverifunktsioon on oma olemuselt lühike, sisaldades pöördumisi alammoodulite poole (Joonis 11). See sisaldab erinevaid alammooduleid võrdlusandmestiku valimiseks ja kasutaja andmestiku üleslaadimiseks. Samuti on rakenduses eraldiseisvad moodulid kasutaja andmestiku analüüsimiseks ning võrdluseks võrdlusandmestikuga.

Moodulite vahelise andmete edastamise jaoks on kolm võimalust. Esimene neist on reaktiivse funktsiooni tagastamine ühest moodulist ja parameetrina teise moodulisse edastamine. Selle strateegia miinuseks on rohkemate reaktiivsete sisendite ja väljundite korral nende üle arvestuse pidamise keerukus. Samuti võib rohkete reaktiivsete funktsioonide kontrollimisel tekkida probleeme. Teise võimalusena saab kasutada „väikse r-i strateegiat“, mille puhul luuakse globaalne reactiveValues tüüpi list r. Sisuliselt luuakse selle strateegia korral rakenduse sisene andmebaas ning ühes moodulis sinna lisatud muutujad on automaatselt kättesaadavas kõigis moodulites, mille sisendiks on r.

```
#' The application server-side
#'
#' @param input,output,session Internal parameters for {shiny}.
#'     DO NOT REMOVE.
#' @import shiny
#' @noRd

app_server <- function( input, output, session) {
  # List the first level callModules here

  #used for passing values to and from modules
  r <- reactiveValues()

  callModule(mod_choose_referencedata_server, "choose_referencedata_ui_1", r=r)

  output$table <- renderDataTable({
    if(!is.null(r$new_data)){
      r$reference_data <- r$new_data
    }
    else {
      r$reference_data <- my_dataset
    }
    if(!is.null(r$reference_data)){
      r$reference_data[,1:5]
    }
  })

  callModule(mod_choose_userdata_server, "choose_userdata_ui_1", r=r, parent_session = session)

  output$table2 <- renderDataTable({
    if (!is.null(r$userdata)){
      r$userdata[,1:5]
    }
  })

  observeEvent(input$analyzeButton, {
    updateTabsetPanel(session, "theTabs",
                      selected = "analysis")
    callModule(mod_analysis_server, "analysis_ui_1", r=r)
    callModule(mod_analysis2_server, "analysis2_ui_1", r=r)
  })
}
```

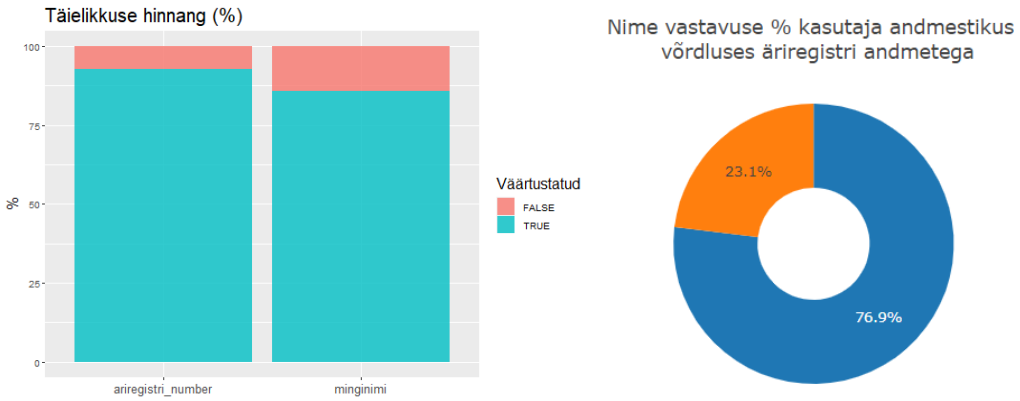
Joonis 11. Rakenduse serverifunktsioon

„Väikese r-i strateegia“ puhul on oluline väärtuse r sisu dokumenteerimine koodis. Kolmandaks võimaluseks moodulite vahel andmete edastamisel on „suure R6 strateegia“, mille puhul luuakse R6 tüüpi (mitte reaktiivne) objekt, mis võimaldab kontrollida liigset reaktiivsust. Antud rakenduse jaoks kasutasin neist kolmest teist valikut ehk „väikese r-i strateegiat“, kuid sellest oluliselt mahukamate rakenduste jaoks oleks otstarbekas kasutada R6 strateegiat [25].

5.4 Rakenduse lõplik ülesehitus

Võrdlusandmestiku puhul on rakenduses kaks võimalust. Esimene neist on kasutada rakenduses andmete võrdlemiseks eellaaditud andmestikku. Selle eelis on see, et kasutaja ei pea andmestikku eraldi alla laadima. Samas ei ole tegemist värskeima andmete seisuga. Andmaks kasutajale võimalus kasutada uusimat andmestikku, on rakenduses olemas ka viimase uuenduse alla laadimise võimalus. Selle valiku tegemisel laadib rakendus värskeima andmestiku lehelt https://avaandmed.rik.ee/andmed/ARIREGISTER/ ning edasise analüüsi käigus toimub võrdlus nende andmete alusel.

Kui veel prototüübi loomise faasis oli plaan kuvada kasutaja andmestiku analüüsi ja võrdlustulemusi samal vahelehel, siis rakenduse loomise käigus selgus, et ülevaatlikum on kuvada neid eraldi vahelehtedel. Kasutaja andmestiku analüüsi puhul on diagrammide loomisel kasutatud lisapaketi ggplot2 võimalusi. Võrdlustulemuste visualiseerimiseks kasutati lisapaketi plotly vahendeid. Plotly puhul on tulemus silmale kaasaegsem ja haaravam. Visualiseerimaks erinevate lisapakettide võimalusi, on mõlemad variandid jäetud ka näidisrakenduse lõplikku verisooni (Joonis 12).



Joonis 12. Ggplot2 ja plotly pakettide abil kuvatud analüüsi tulemused

Rakenduse analüüsi tulemuste vahelehed on mõlemad üles ehitatud sarnaselt: kasutajale kuvatakse vastavalt kas kasutaja andmestiku analüüsi või andmestike võrdleva analüüsi tulemuste kokkuvõtet graafiliselt. Samuti on võimalik mittevastavaid kirjeid kuvada nende tüüpide lõikes, et oleks võimalik neid täpsemalt analüüsida. See on vajalik selleks, et teha kindlaks mittevastavuste põhjused ja alati ei pruugi tegemist olla andmekvaliteedi probleemiga.

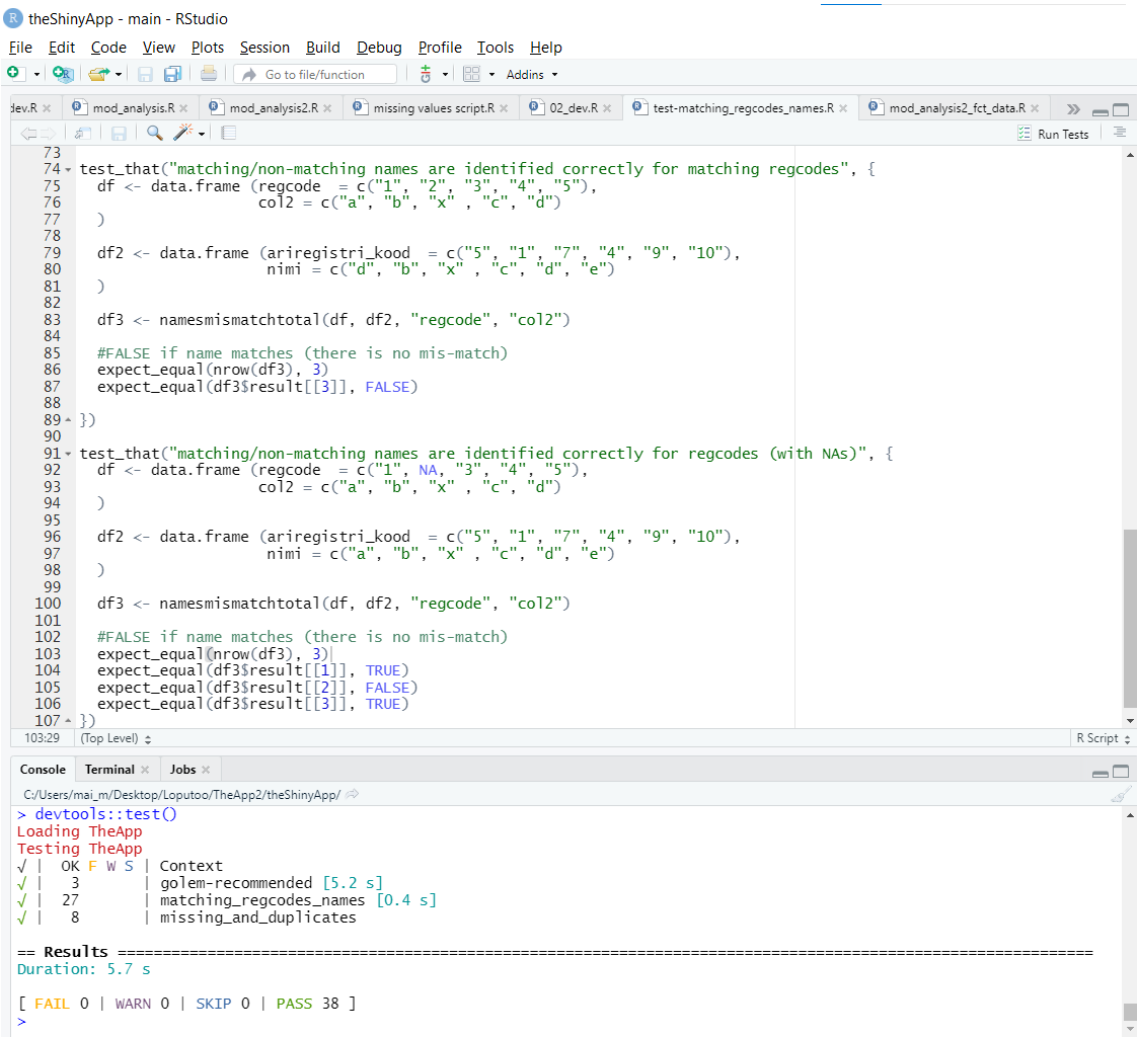
Analüüsides näitena võrdlusandmestikku ehk äriregistri andmeid, on seal üle 100 kirje, mille näol on tegemist kirjetega, kus sama nimega juriidiline isik on andmestikus vähemalt kahel korral. Põhjuseks näiteks see, et tegemist on sama nimega taluga, mida omavahel eristab aadress erinevas maakonnas. Samas on selliste kirjete osakaal vaid 0,04% kõigist kirjetest. Soovi korral on kasutajal võimalik ka mittevastavad andmekirjed tüüpide lõikes CSV formaadis alla laadida, et selle baasilt asuda näiteks andmeid parandama.

6 Rakenduse testimine ja tugevdamine

Rakendust testiti nii komponenttestidega (nn unit testid) kui ka manuaalselt. Manuaalse testimise käigus tuvastatud puudused lahendati rakenduse tugevdamise käigus.

6.1 Rakenduse testimine

Komponentide testimist sai teostada nende funktsioonide osas, mis ei sõltu reaktiivsetest sisenditest. Sellest kasutati test_that paketti (Joonis 13) ning {golem} raamistiku abil oli võimalik kõik testid käivitada mugavalt ühe käsuga devtools::test().



Joonis 13. Komponenttestide kasutamine

6.2 Rakenduse tugevdamine

Rakenduse kasutamise seisukoht on kriitiline kasutaja andmestiku lugemise õnnestumine. Võrdlusandmestiku uuendamine ei ole niivõrd kriitiline, kuna on võimalik kasutada ka eellaaditud andmeid. Selleks, et rakendus oleks võimalikult töökindel, kasutati fread() funktsiooni, mis on kiirem kui read.csv() [26]. Siiski võivad kasutaja andmed olla kujul, kus nende lugemine ei õnnestu. Selleks lisati andmestike lugemisele veakäsitlus koos kasutaja juhendamisega veateates, juhuks kui andmestiku lugemine ei õnnestu. Rakendust täiendati ka muus osas nii R keelel põhinevate teavituste kui ka JavaScripti põhiste teavitustega. Mõlemaid on võimalik Shiny rakenduses edukalt kasutada.

Rakenduse kasutajaliidese manuaalse testimise käigus ilmnes, et mahuka andmestiku puhul võtab analüüsi teostamine ja graafikute kuvamine teatud aja ning kasutajakogemus oleks parem, kui sel ajal kuvataks protsessi edenemise kohta kasutajale teavet. Selle lahendamiseks lisati graafikutele laadimisele viitav graafika. Samuti selgus testimise käigus, et kasutaja andmestiku analüüsi vahelehel rippmenüü valikutest mittevastavuse tüübi valimise järgselt uuendas rakendus kogu lehe, mis ei vastanud oodavale kasutajakogemusele. Soovitud viis oleks olnud selline, et valiku tegemisel uueneb vastavalt kasutaja poolt tehtud valikule ainult tabeli sisu.

Antud puudus lahendati R keele lisapaketi DT [27] kasutuselevõtuga. Selle abil on võimalik R keele andmeobjekte kuvada HTML tabelitena kasutades JavaScripti teeki „DataTables“. Funktsiooniga dataTableProxy() luuakse proxy objekt, mille abil on võimalik muuta olemasolevat DataTables objekti. Antud juhul oli täidetud ka eeltingimus, et andmeid on antud lahendusega võimalik muuta ainult sel juhul, kui uued ja asendatavad andmed on sama veergude arvuga. Antud muudatuse täiendav kasutegur ilmneb rakenduse mälukasutuse poolelt – varasema mitme tabeli objekti loomise asemel luuakse antud lähenemise korral tabeli objekt vaid ühel korral.

Selleks, et pakettide uuendamisel tekkivate konfliktide korral oleks võimalik taastada rakenduse varasem seis, on otstarbekas fikseerida projekti hetkeseisu renv paketi abil (Joonis 14).

Joonis 14. Osa renv abil loodud sõltuvuste hetktõmmisest

7 Kokkuvõte

Käesoleva diplomitöö raames analüüsisin andmehalduse ja andmekvaliteedi valdkonna olulisust ning lõin R keelt ja selle lisapakette kasutades rakenduse, mida saab kasutada andmekvaliteedi hindamiseks ning andmekvaliteedi parendamise protsessis Eesti registris olevate juriidiliste isikute andmete seisundi kordushindamiseks. Rakenduse funktsionaalsusele seab piirid võrdlusandmestiku kättesaadavus.

Töö raames analüüsisin andmekvaliteedi valdkonda, sellega seotud kulusid ning andmekvaliteedi juhtimise protsessi. Samuti käsitlesin loodava rakenduse funktsionaalseid ja mittefunktsionaalseid nõudeid ning vahendite ja tehnoloogiate valikut, mille abil tuvastatud nõudeid täitev rakendus teostada. Rakenduse loomist puudutav peatükk andis ülevaate rakenduse disainimisest, prototüübi loomisest ning loodud kasutajaliidese ja serveripoole loogika terviklikuks rakenduseks ehitamisest. Samuti on rakenduse loomise peatükis välja toodud rakenduse testimise ja tugevdamise aspektid.

Loodud rakenduses on võimalik üles laadida kasutaja andmestik ning rakendus teostab selles sisalduvate kirjete täielikkuse ja ühekordsus analüüsi. Samuti analüüsib rakendus kasutaja andmete ajakohasust ning teeb seda võrreldes kasutaja andmeid Registrite ja Infosüsteemide Registri poolt avaandmetena pakutava äriregistri andmestikuga. Rakendus teostab antud töö raames tuvastatud andmekvaliteedi dimensioonidele (täielikkus, ühekordsus, ajakohasus) vastavuse kontrolli ning kuvab kokkuvõtliku tulemuse. Täiendavalt on rakendusest võimalik vaadata nõuetele mittevastavate kirjete nimekirja ning alla laadida CSV formaadis andmefail hindamise tulemusena leitud mittevastavate kirjetega.

Töö teostamise järeldusena võib öelda, et R keele ja selle lisapakettide abil on võimalik ehitada terviklik kasutajarakendus, mille abil saab automatiseerida määratletud ulatuses andmekvaliteedi analüüsi teostamist. Põhiandmete, nagu seda on isikute (klientide, tarnijate) andmetest nimi ja registrikood, hindamise funktsionaalsusega on rakendus

universaalse kasutusega. Laiema funktsionaalsuse lisamisel muutuks see konkreetse andmestiku ja asutuse spetsiifiliseks.

Kasutatud kirjandus

[1] „Andmehalduse tegevuskava,“ [Võrgumaterjal]. Loetud aadressil: https://digiriik.ee/index.php/andmehalduse-tegevuskava/. [Kasutatud 08.02.2021].

[2] „TIOBE Programming Community Index Definition,“ [Võrgumaterjal]. Loetud aadressil: https://www.tiobe.com/tiobe-index/programming-languages-definition/. [Kasutatud 08.02.2021].

[3] „The R Programming Language,“ [Võrgumaterjal]. Loetud aadressil: https://www.tiobe.com/tiobe-index/r/. [Kasutatud 08.02.2021].

[4] L. Sebastian-Coleman, Measuring Data Quality for Ongoing Improvement, Morgan Kaufmann, 2013.

[5] D. McGilvray, Executing Data Quality Projects, Morgan Kaufmann, 2008.

[6] T. O'Brien, M. Helfert ja A. Sukumar, „Classifying costs and effects of poor Data Quality – examples and discussion,“ %1 *Annual Conference of Irish Academy of Management*, Maynooth, 2012.

[7] A. Haug, F. Zachariassen ja D. van Liempd, „The costs of poor data quality,“ *Journal of Industrial Engineering and Management,* kd. 4, nr 2, pp. 168-193, 2011.

[8] M. Eppler ja M. Helfert, „A classification and analysis of data quality costs,“ %1 *9th MIT International Conference on Information Quality*, Boston, 2004.

[9] Statistikaamet, „Eesti andmehalduse metoodikaprojekt. Andmekvaliteedi juhis (kavand),“ August 2020.

[10] „Avaandmed,“ Registrite ja Infosüsteemide Keskus, [Võrgumaterjal]. Loetud aadressil: https://www.rik.ee/et/avaandmed. [Kasutatud 7.03.2020].

[11] U. Eriksson, „Why is the difference between functional and Non-functional requirements important?,“ ReQtest AB, [Võrgumaterjal]. Loetud aadressil: https://reqtest.com/requirements-blog/functional-vs-non-functional-requirements/. [Kasutatud 28.03.2021].

[12] S. Paradkar, Mastering Non-Functional Requirements, Packt Publishing, 2017.

[13] „Kuidas disainida head e-teenust?,“ *Äripäev,* 5.05.2014.

[14] „Web Apps,“ The MathWorks, Inc., [Võrgumaterjal]. Loetud aadressil: https://www.mathworks.com/help/compiler/web-apps.html. [Kasutatud 28.03.2021].

[15] G. v. Rossum, „Foreword for "Programming Python" (1st ed.),“ [Võrgumaterjal]. Loetud aadressil: https://www.python.org/doc/essays/foreword/. [Kasutatud 28.03.2021].

[16] R. D. Peng, „R Programming for Data Science,“ 03 09 2020. [Võrgumaterjal]. Loetud aadressil: https://bookdown.org/rdpeng/rprogdatascience/. [Kasutatud 28.03.2021].

[17] „A Brief History of MATLAB,“ The MathWorks, Inc., [Võrgumaterjal]. Loetud aadressil: https://www.mathworks.com/company/newsletters/articles/a-brief-history-of-matlab.html. [Kasutatud 28.03.2021].

[18] „Python Package Index,“ Python Software Foundation, [Võrgumaterjal]. Loetud aadressil: https://pypi.org/. [Kasutatud 28.03.2021].

[19] „The Comprehensive R Archive Network,“ [Võrgumaterjal]. Loetud aadressil: https://cran.r-project.org/. [Kasutatud 28.03.2021].

[20] „File Exchange,“ The MathWorks, Inc., [Võrgumaterjal]. Loetud aadressil: https://www.mathworks.com/matlabcentral/fileexchange/54548-mpm. [Kasutatud 28.03.2021].

[21] „What is R?,“ The R Foundation, [Võrgumaterjal]. Loetud aadressil: https://www.r-project.org/about.html. [Kasutatud 03.04.2021].

[22] „RStudio,“ RStudio, 2021. [Võrgumaterjal]. Loetud aadressil: https://www.rstudio.com/products/rstudio/. [Kasutatud 03.04.2021].

[23] „Shiny from RStudio,“ RStudio, 2020. [Võrgumaterjal]. Loetud aadressil: https://shiny.rstudio.com/. [Kasutatud 03.04.2021].

[24] „Documentation. Blog posts, podcasts, videos, other documentation of the {golemverse},“ ThinkR, 2020. [Võrgumaterjal]. Loetud aadressil: https://golemverse.org/documentation/. [Kasutatud 03.04.2021].

[25] C. Fay, S. Rochette, V. Guyader ja C. Girard, „Engineering Production-Grade Shiny Apps,“ 02 04 2021. [Võrgumaterjal]. Loetud aadressil: https://engineering-shiny.org. [Kasutatud 03.04.2021].

[26] D. Cook, „Speeding up Reading and Writing in R,“ 20 10 2019. [Võrgumaterjal]. Loetud aadressil: https://www.danielecook.com/speeding-up-reading-and-writing-in-r/. [Kasutatud 26.04.2021].

[27] Y. Xie, „Package ‘DT’,“ 14 04 2021. [Võrgumaterjal]. Loetud aadressil: https://cran.r-project.org/web/packages/DT/DT.pdf. [Kasutatud 17.04.2021].

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks[1]

Mina, Maile Mäesalu

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Andmekvaliteedi hindamiseks näidisrakenduse loomine R keele näitel“, mille juhendaja on Meelis Antoi
   1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
   1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

16.05.2021

1 Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtestamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud kaks või enam isikut oma ühise loomingulise tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtajaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.

Lisa 2 – Rakenduse vaated