

TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Theodor Taimla

**Nutikas otsing parimate pakkumiste
leidmiseks muutmälu ja SSD-ketta näitel**

Bakalaureusetöö (9 EAP)

Juhendaja(d): Anne Villems
Aleksi Kelli

Tartu 2020

Nutikas otsing parimate pakkumiste leidmiseks muutmälu ja SSD-ketta näitel

Lühikokkuvõte:

Lõputöö raames luuakse tarkvaralahendus, mille eesmärk on võimaldada mugaval ja kiirel viisil leida SSD ja muutmälu parimaid pakkumisi ehk teostada nutikat otsingut. Analüüsitakse Eesti hinnavaatlusportaalide otsimisvõimekust ja autoriõigusega seonduvaid küsimusi. Seejärel põhjendatakse informatsiooni otsimis- ja kogumisviiside valikut, kirjutatakse tarkvaralahenduses kasutatud tehnoloogiatest, teکیدest, raamistikest, arhitektuurist ja saavutatud jõudlusest.

Võtmesõnad:

Andmekoorimine, veebikoorimine, veebiproovimine, PostgreSQL, Python, Micronaut, Angular, autoriõigus

CERCS: P175 Informaatika, süsteemiteooria

Clever Search for Best Offers Using RAM and SSD as Examples

Abstract: The result of the thesis will be the creation of a software solution, which purpose is to enable a fast and comfortable way to find the best possible offers for SSD and RAM products. Estonian price comparison websites and possible copyright issues are analyzed. Then the reasoning for the choice of information retrieval methods is explained, selection of the chosen technologies, libraries and frameworks are introduced, the software architecture is elaborated upon and performance metrics are determined.

Keywords:

Data extraction, web scraping, web crawling, PostgreSQL, Python, Micronaut, Angular, copyright

CERCS: P175 Informatics, systems theory

Sisukord

1.	Sissejuhatus	5
2.	Mõisted ja terminid	6
3.	Olemasolevad sarnased lahendused	7
3.1	Hinnavaatlus.ee	7
3.2	Hind.ee	11
4.	Ülevaade tarkvaralahenduse nõuetest	13
5.	Tarkvaralahenduse seaduslikkus Eesti Autoriõiguse kontekstis	15
5.1	Tarkvaralahenduse kasutamine hariduslikel eesmärkidel	15
5.2	Tarkvaralahenduse kasutamine ärielistel eesmärkidel	15
6.	Informatsiooni otsimis -ja kogumisviisid	17
6.1	Veebiroomamise eeskirjad	18
6.1.1	Valimiseeskiri	18
6.1.2	Sünkronisatsioonieeskiri	19
6.1.3	Viisakuseeskiri	19
6.1.4	Paralleelsuseeskiri	20
7.	Tarkvaralahenduse arhitektuur	21
7.1	Üldine struktuur	21
7.1.1	Andmemudel	21
7.1.2	Veebiliides	22
7.1.3	Veebiserver	23
7.1.4	Andmekoguja	24
7.2	Protsessist lähemalt	24
7.2.1	Informatsiooni kogumise erisused ja probleemilahendus	26
8.	Kasutatud raamistikud, keeled ja teegid	28
8.1	Andmete haldus	28
8.2	Andmekoguja tehnoloogilised valikud	28
8.2.1	Veebiroomamise teegid	28
8.2.2	Andmekoorimise teegid	29
8.2.3	PDF-failist andmete ekstraheerimise teegid	29
8.2.4	Andmete pärimise ja talletamise teek	30
8.3	Veebiserver	30
8.4	Veebiliides	31
9.	Saavutatud jõudlus ja nõuetele vastavus	33
9.1	Veebiserveri jõudlus	33

9.2	Andmekoguja andmete ekstraheerimise ajakulu	33
9.3	Veebilehe jõudlus	34
9.4	Nõuetele vastavus	35
10.	Kokkuvõte	36
11.	Viidatud kirjandus	37
	Lisad	38
I.	Tarkvaralahenduse paigaldamine, seadistamine ja käivitamine operatsioonisüsteemis Windows 10	38
a.	Andmebaasi haldamise süsteem	38
b.	Andmekoguja	39
i.	Python'i ja selle teekide paigaldamine	39
ii.	Andmebaasi ühenduse seadistamine	40
iii.	Selenium-nimelise teegi sõltuvuste paigaldamine	41
iv.	Optilise tekstituvastuse lubamine	41
c.	Veebiserver	45
i.	Java JDK paigaldamine ja seadistamine	45
ii.	Andmebaasi ühenduse seadistamine	46
iii.	Käivitamine	46
d.	Veebiliides	46
i.	Sõltuvuste paigaldamine ja seadistamine	46
ii.	Käivitamine	46
II.	Google Drive API ja Tesseract tekstituvastuse võrdlus	47
III.	Valminud veebiliides	51
IV.	Litsents	54
	Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks	54

1. Sissejuhatus

Järjest rohkem ostetakse kaupu internetist ning seetõttu on ka suurenenud vajadus keskonna järele, mis võimaldab seda efektiivselt teha. Statistikaameti 2019. aasta artiklis¹ kirjutati, et 75% inimestest kasutab e-kaubanduse võimalusi ning 2017. aasta artiklis² selgub, et populaarseimate internetitegevuste hulka kuulub toote või teenuse kohta info otsimine – sõltuvalt vanuserühmast tegeles sellega 85% kuni 90% inimestest.

Lõputöö eesmärk on luua veebipõhine lahendus, mis võimaldab kiirelt ja mugavalt leida parimaid muutmälu (ingl *Random Access Memory*) ja SSD (ingl *Solid State Drive*) pakumisi. Lõputöö autori isiklik kogemus näitas, et parima hinna-jõudluse suhtega toote leidmine on raskendatud, kui tootel on palju tehnilisi parameetreid. Olemasolevad lahendused seda raskust oluliselt ei leevenda. Muutmälu ja SSD parameetrite rohkus võimaldab demonstreerida tarkvaralahenduse võimekust lisada tugi sarnastele toodetele nagu näiteks kõvakettad ehk HDD-d (ingl *hard disk drive*) ja emaplaadid (ingl *motherboard*). Kui jõuda sellisesse seisu, et tootevalikus on enamus arvutis sisalduvaid komponente, siis avaneb uniikaalne võimalus hinnata arvuti jõudlust tervikuna.

Lõputöö raames valmib tarkvaralahendus, mis loob aluse järgmiste probleemide lahendamiseks:

- Eesti e-arvutipoodide rohkuse tõttu on iga poe käsitsi läbi käimine ajamahukas.
- Raske on järke pidada, mis toode on parima hinna ja kvaliteedi suhtega, sest lisaks poodide ja toodete rohkusele on palju tootemadusi ja nende vahelisi seoseid.
- Üldjuhul e-poed ei võimalda täppisotsinguid parameetrite järgi. Tavaliselt saab sorteerida hinna, tarneaja ja tootja järgi. Kolmest populaarsest e-poest, millelt info kogumist valminud tarkvaralahendus toetab (Arvutitark³, Galador⁴, 1a.ee⁵), on üks hiljuti lisanud toe sorteerimiseks taktisageduse, latentsi, mahu jt. andmete järgi, kuid endiselt ei ole võimaldatud otsimist näiteks hinna-mahu, tegeliku võimsuse (latent-suse ja taktisageduse suhe) ning tegeliku võimsuse ja hinna suhte järgi.

Lõputöö teoreetilises osas kirjutatakse olemasolevate lahenduste otsimisvõimekuse analüüsi peatükis 3, tarkvara nõuetest peatükis 4, tarkvaralahenduse seaduslikkusest peatükis 5, informatsiooni otsimis- ja kogumisviisidest peatükis 6 (sh. veebiprogrammide eeskirjadest), tarkvaralahenduse arhitektuurist ja sellega seonduvalt informatsiooni kogumisel tekkinud erisustest ning probleemilahendusest peatükis 7, kasutatud raamistikest, keeltest ja tekidest peatükis 8, nõuetele vastavusest ja jõudlusest peatükis 9 ning kõigest kokkuvõtvalt peatükis 10. Esimeses lisas on tarkvaralahenduse paigaldamise, seadistamise ja jooksutamise juhend, teises lisas tuuakse välja Google Drive API⁶ ja Tesseract'i⁷ abil tehtud andmete ekstraheerimise tulemusi, kolmandas on ekraanitõmmised valminud veebiliidesest.

¹ <https://www.stat.ee/internetti-kasutatakse-uha-enam-ostlemiseks>

² <https://blog.stat.ee/2017/10/26/noored-it-seadmete-ja-interneti-maailmas/>

³ <https://arvutitark.ee/>

⁴ <https://www.galador.ee/>

⁵ <https://www.1a.ee/>

⁶ <https://developers.google.com/drive/api/v2/about-sdk>

⁷ <https://github.com/tesseract-ocr/tesseract>

2. Mõisted ja terminid

Veebiroomaja (ingl *web crawler*) on programm, mis käib süstemaatiliselt läbi erinevaid veebiressursse informatsiooni leidmise eesmärgil [1].

Andmekoorimine (ingl *data scraping*) on andmete ekstraheerimine teise programmi inimloetavast väljundist, eeskätt veebilehelt⁸.

Veebikoorija (ingl *web scraper*)⁹ on programm, mis ekstraheerib veebilehekülgedelt andmeid.

Esitusoleku siire (ingl *Representational State Transfer* ehk REST) on veebiteenuste programmeerimise paradigma [2].

Javascript on objektorienteeritud programmikeel, mida peamiselt kasutatakse dünaamilise veebisisu loomiseks¹⁰.

HTML (ingl *HyperText Markup Language*) on hüpertexti märgistuskeel ehk keel, milles märgendatakse veebilehti¹¹.

API (ingl *Application Program Interface*) ehk rakendusprogrammiliides on tarkvara -või süsteemide vaheline suhtlust võimaldav reeglistik¹².

SQL (ingl *Structured Query Language*) ehk struktureeritud päringute keel on deklaratiivne interaktsiooni- ja programmikeel¹³.

⁸ <https://akit.cyber.ee/term/6201-data-scraping>

⁹ <https://akit.cyber.ee/term/6203-web-scraping>

¹⁰ <https://akit.cyber.ee/term/3662-javascript>

¹¹ <http://www.keelevaab.ee/dict/speciality/aks/dict.cgi?word=HTML&lang=en>

¹² <https://akit.cyber.ee/term/3088-rakendusliides-api>

¹³ <https://akit.cyber.ee/term/8501-sql>

3. Olemasolevad sarnased lahendused

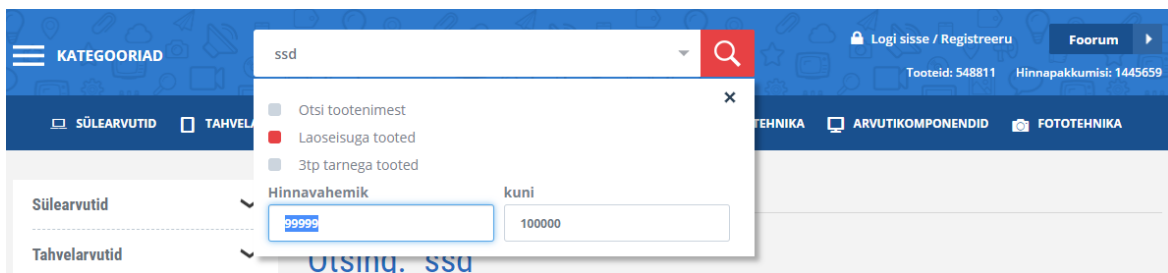
Selles peatükis analüüsitakse otsimisvõimekust hinnavaatlusportaalidel, mis pakuvad võrdlemiseks Eesti e-poodides müügil olevaid tehnikatooteid. Teadaolevalt on nendeks hinnavaatlus¹⁴ ja hind.ee¹⁵.

3.1 Hinnavaatlus.ee

Populaarses Hinnavaatluse otsingumootoris on kümme erinevat tehnoloogia- ja tootekategooriat¹⁶ ning keskmiselt umbes 250 000 unikaalset külastajat kuu kohta. Parimaid pakumisi võimaldatakse otsida märksõna ja toote kategooriasse kuuluvuse alusel. Tehnoloogia- ja tootekategooriatest on võimalik võrrelda süle- ja tahvelarvuteid, mobiiltelefone, kodu- ja suurelektroonikat, fototehnikat ning arvutikomponente. Analüüsis keskendutakse viimasele, kuna sellesse kategooriasse kuuluvad ka lõputöös käsitletavad SSD ja muutmälu.

Esimene võimalus pakkumiste otsimiseks on kirjutada märksõnu Hinnavaatluse üldotsingusse, mis asub lehekülje ülemises osas. Suurendamisklaasi ikooniga punase otsingunupu kõrval on noolekese kujuga nupp, millele vajutades ilmub filtrikomponent (vt Joonis 1), kus saab märkeruutudega (ingl *checkbox*) valida, kas filtreerida tootenime järgi või piirata otsingut nii, et oleksid vaid laos olevad või ainult kuni kolme päevase tarnega tooteid. Samuti kaks tekstikasti, mis on mõeldud hinnavaahemiku määramiseks, kuid sisendina saab anda ka mittenumbriilist teksti. Filtrit ei rakendata kui tegemist on vigase sisendiga või ei leitud ühtegi toodet. Märkeruutudega filtrite puhul ei ole nende kõrval oleva teksti järgi aru saada, mida need täpselt teevad. Katsetamise tulemusena tundub nii:

- tootenimest otsimine tähendab seda, et otsitakse ainult tootenimest (Sellisel juhul võib arvata, et vaikimisi otsitakse ainult või lisaks tootekirjeldusest)
- Laoseisuga toodete otsimise puhul leitakse vaid laos olevaid tooteid
- 3tp tarnega tooted filtri puhul kuvatakse vähemalt 3 päeva jooksul tarnitavaid kaupu



Joonis 1. Üldotsing ja otsingualune filtrikomponent hinnavaatluses.

Teine võimalus on otsida konkreetsest kategooriast samuti märksõnade abil, kuid siis juba kitsendatult valitud kategooriast ning täpsustavate filtritega. Näiteks, kui valida arvutikomponentide alt *kõvakettad* – SSD, siis ilmub otsingu komponent vasakule ülemisse serva ning selle alla järgmised filtrid:

¹⁴ <https://www.hinnavaatlus.ee/>

¹⁵ <https://www.hind.ee/>

¹⁶ <https://www.hinnavaatlus.ee/firms/all/>

- Tootjad – Märkeruutes saab märkida linnukene soovitud tootja nime juurde. Antud näites on neid 64. Õige tootja leidmist lihtsustab see, et nad on tähestikulises järjekorras, kuid nii paljude võimalike variantidega on sobivam kasutada filtreeritavat automaatjätkamisega (ingl *autocompletion*) mitmevalikulist (ingl *multiselect*) ripploendit (ingl *dropdown list*).
- Hinnavahemik – Analoogselt üldotsingu alla tekkiva filtriga, saab hinnavahemikku ilma igasuguse tõrketa kirjutada ka mittenumbrilisi märke. Filtri nimest paremal asub kahte suunda näitav nooleotsa ikooniga nupp, mida vajutades ilmub kahe tekstkasti asemel uus komponent, millega saab kahte palli lohistada, et hinnavahe- mik paika sättida.
- Maht – Kõikvõimalike (olemasolevate) SSD mahtude märkeruute filter. Selle kohta käivad samad probleemid, mis tootjate filtriga.
- Lisavalikud – Siin valikutes leidub kujutegureid (ingl *form-factor*), liideseid kui ka NAND-mälu tüüpe. Tegemist on samuti märkeruutudega.
- Firmad – Loetelu e-poodide firmadest, mida saab filtreerida vastavasse märkeruutu klõpsutades.

Muutmälu ehk *Mälud* kategoorias on sarnaselt kõik filtrid välja arvatud hinnavahemiku määraja märkeruudud. Ühised on firma, tootja ja mahu filtrid. Filtreeritavatest tootepara- meetritest on muutmälutüüp (näiteks DDR4), taktisagedus ja latentsus (ingl *CAS latency*).

Filtrite valik sõltub sellest, millised tooted on olemas. Samas, kui filtreerimise tulemusena toodete valik väheneb, siis filtrites seda ei kajastata. Võimalik, et lehekülje arendajad arvasid, et oleks segadusse ajav, kui võimalikud filtreerimisvalikud peaksid jäljetult kaduma, aga ei arvestatud, et saaks ka näiteks sellised valikud visuaalselt eristada ning kohtspikriga (ingl *tooltip*) seda seisu selgitada.

Kategooriasisesel otsingul on samuti otsingualune filter. Sellel on lisaks ühisele laoseisu filtrile ka kaks märkeruutu nimega *Otsi ka tooteinfost* ning *Otsi ka pakkumise infost*. Katsetamise käigus ei suutnud lõputöö autor selgusele jõuda, mis tulemust need annavad ning mis vahe on tooteinfo ja pakkumise info ja kus neid kuvatakse. Mõne toote puhul toodete vaates olev tarneaeg ehk saadavus ei klapi sellega, mis pakkujate vaates on (vt Joonis 2 ja Joonis 3).

Kui võrrelda ühe toote kirjeldust sellega, mis vastava toote pakkuja kodulehel on, siis selgub, et andmed on sisuliselt samal kujul (vt Joonis 2 ja Joonis 4) ning kui panna Google otsingumootori märksõnaks hinnavaatlusest saadud toote kirjeldust, siis selgub, et see samuti pärineb mõnest e-poe veebileheküljest (vt Joonis 5 ja Joonis 6).

Asjaolu, et muutmälusid ja SSD-kettaid on võimalik filtreerida kas piiramata kujul märksõnadega või mõne üksiku toote nimes või -kirjelduses oleva tehnilise parameetri järgi viitab sellele, et Hinnavaatlus ei oma informatsiooni toote kõigi parameetrite kohta ning seega ei saa ka võimaldada nii-öelda nutikat otsingut.



Crucial SATA 2.5" 120GB BX500

Tootekood: CT120BX500SSD1 Kategooria: KÕVAKETTAD - SSD 1-3TP

SSD | CRUCIAL | BX500 | 120GB | SATA 3.0 | Write speed 500 MBytes/sec |
Read speed 540 MBytes/sec | 2,5" | MTBF 1500000 hours | CT120BX500SSD1

27,43 €
33 pakkumist



Joonis 2. Näide tootest toodete vaates.



PCSTO

Tasuta tarne üle Eesti alates 300€
ostusummast... DPD pakipoe, ... Näita rohkem



89 kommentaari

27,43 €

↓ -0,07 €

Saadavus: 1-5tp

POODI



Joonis 3. Näide tootest pakujate vaates.

Pealeht > ARVUTI KOMPONENDID JA LISAD > ARVUTIKOMPONENDID > KÕVAKETTAD /HDD/ > SSD > SSD|CRUCIAL|BX500|120GB|SATA 3.0|Write speed 500 MBytes/sec|Read speed 540 MBytes/sec|2,5"|MTBF 1500000 hours|CT120BX500SSD1

SSD|CRUCIAL|BX500|120GB|SATA 3.0|WRITE SPEED 500 MBYTES/SEC|READ SPEED 540 MBYTES/SEC|2,5"|MTBF 1500000 HOURS|CT120BX500SSD1



27,50 EUR

Kogus


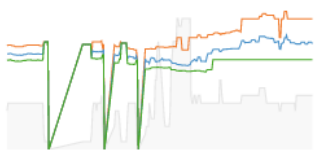
Garantii (kuud): 36
10 kogus laos
Part number: CT120BX500SSD1
Tarneaeg: 1-3 tp

Lisada ostukorvi

Printi
View full size

Joonis 4. BX500 SSD-ketta pakkumine pcsto.ee veebileheküljel.

SSD Gigabyte 1TB 2,5 SATA3 550/500MB/s 7mm

Toode lisatud: 2019-06-16	
Toote kommentaarid 0tk	
Tootekood GP-GSTFS31100TNTD	
GTIN 4719331804565	
	Model: GIGABYTE SSD HDD capacity: 1 TBW: 600 Reading speed: 550 Writing speed: 500 Random read IOPS (max): 75 Random write IOPS (max): 85 Height: 7 Width: 69.85 Depth: 100
LISA VÖRDLUSESE	ARVUSTUSED: https://www.google.com/search?q=Gigabyte+1TB+2%2...

Joonis 5. Tooteinfo hinnavaatluses.

Tootekood	GP-GSTFS31100TNTD
Model	GIGABYTE SSD
Hdd capacity	1
Tbw	600
Reading speed	550
Writing speed	500
Random read iops (max)	75
Random write iops (max)	85
Height	7
Width	69.85
Depth	100
Tootjagarantii	3 aastat

arvutitark.ee > est > tootekataloog > Arvutikomponendi... ▼

Gigabyte SSD 1TB 2,5 SATA3 550/500MB/s 7mm - Arvutitark

Tootekood, GP-GSTFS31100TNTD. Model, GIGABYTE SSD. Hdd capacity, 1. Tbw, 600. Reading speed, 550. Writing speed, 500. Random read iops (max), 75. Random write iops (max), 85. Height, 7. Width, 69.85. Depth, 100. Tootjagarantii, 3 ...

Joonis 6. Google otsingumootori täpne vaste paksus kirjas (all) ja sellele vastav informatsioon e-poe koduleheküljel (üleval).

Järgmises alampeatükis analüüsitakse hind.ee hinnavõrdlusportaali otsimisvõimekust.

3.2 Hind.ee

Veebipõhises hinnavõrdlusportaalis hind.ee on võimalik otsida rohkem kui 100 000 toote seast, mida pakuvad müüjiks üle 50 e-poe. Selle domeen registreeriti aastal 2015¹⁷.

Otsida saab kas märksõnadega veebilehe ülemises osas (vt Joonis 7) filtreerimisvõimaluseta või valides täpsema tootekategooria nagu näiteks *sülearvutid*.



Joonis 7. Märksõnaotsing hinnavõrdlusportaalis hind.ee.

Esimene võimaldab sorteerida tähestiku ja hinna järgi, viimase puhul enam märksõnadega otsingut teostada ei saa, aga võimaldatakse filtreerida kaubamärgi, hinna ja müüja järgi (kõik on märkeruudu tüüpi filtrid) ning sorteerida järgmiselt: populaarsed esimesena, uusimad esimesena, odavamad esimesena, kallimad enne, tähestikulises järjekorras, müüja number ehk müüjate kogus toote kohta (vt Joonis 8).

Kaubamärk	Hind	Müüjad
ASUS 298	Kuni 400.00 €	Arvutitark.ee 1004
LENOVO 212	400.00 - 600.00 €	Rde.ee 265
DELL 200	600.00 - 800.00 €	Kaup24.ee 238
HP 191	800.00 - 1000.00 €	01.ee 234
ACER 116	1000.00 - 1200.00 €	
MSI 88	1200.00 - 1400.00 €	
HIRO 77	Alates 1400.00 €	
FUJITSU 36		
TOSHIBA 22		
APPLE 15		
MICROSOFT 9		
HUAWEI 5		
PRESTIGIO 5		
SAMSUNG 3		
HEWLETT PACKARD 3		

TOP	Product	Price	Seller
4	Lenovo IdeaPad L340	alates 658.85 €	Eelvaade
	Dell Latitude 3301		
	Lenovo ThinkPad E495	alates 658.85 €	Eelvaade
1	Lenovo IdeaPad S145	alates 239.00 €	Eelvaade
50	Lenovo ThinkPad L590	alates 863.00 €	Eelvaade
15	Lenovo ThinkPad X390	alates 1263.00 €	Eelvaade

Joonis 8. Otsinguvõimalused spetsiifilise kategooria vaates hinnavõrdlusportaalis hind.ee.

Pealehel, ega mujal ei paista olevat arvutikomponentide kategooriat. Kõige sarnasemas ehk arvutitarvikute liigitusrühmas ei leidu SSD-kettaid ja muutmälusid. Samas, kui kirjutada märksõnaotsingusse SSD või RAM (ingl *Random Access Memory* ehk muutmälu), siis

¹⁷ <https://www.internet.ee/domeenid/kuidas-registreerida-ee-domeeni?domain=hind.ee>

paistab, et sellised kategooriad eksisteerivad, sest lehekülj oskab liigitada vastavaid tooteid kõvaketasteks või operatiivmäludeks ehk muutmäludeks (vt joonis 9).

Bränd	Model	Maht	Liik	Hind	Kauplus
Kingston	SSD 2.5" to 3.5" Bracket with screws		Kõvakettad (SSD)	1.99 €	RDE.ee
Intenso	SSD 2.5" 128GB	128GB	Kõvakettad (SSD)	26.99 €	RDE.ee
Integral	SSD V Series 3D 120GB SATA 3 2.5	120GB	Kõvakettad (SSD)	27.49 €	RDE.ee
Silicon Power	SSD S55 120GB 2.5 SATAIII 6Gb/s SP120GBSS55S525	120GB	Kõvakettad (SSD)	27.99 €	RDE.ee
Plextor	Plextor MV8 Series SSD 128GB M.2 6.0 GB/s (PX-128M8VG)	128GB	Sisemised kõvakettad (HDD, SSD, Hybrid)	30.00 €	kaup24.ee
Intenso	Sisemine kõvaketas Intenso High Performance internal SSD, 120GB M.2 SATA III	120GB	Sisemised kõvakettad (HDD, SSD, Hybrid)	34.99 €	kaup24.ee
SanDisk	SSD Plus 120GB SATA3 (SDSSDA-120G-G27)	120GB	Sisemised kõvakettad (HDD, SSD, Hybrid)	35.99 €	kaup24.ee
Transcend	SSD 220S 120GB SATAIII TS120GSSD220S	120GB	Sisemised kõvakettad (HDD, SSD, Hybrid)	36.49 €	kaup24.ee

Joonis 9. Märksõnaotsingu tulemused hinnavõrdlusportaalis hind.ee, kus märksõna vastab peidetud liigitusrühmale.

Võrreldes hinnavaatlusega on hind.ee keskkonnas tootekirjelduse ja tarne võrra vähem andmeid ning filtreerida ei saa ühegi tootele spetsiifilise tehnilise parameetri järgi. Hinnavaatluses on puudu mõned olulised filtreerimisparameetrid nagu muutmälu komplekti suurus ja SSD vastupidavus (palju mahtu lubatakse kokku kirjutada enne SSD olulise kulumiseni). Mitte kummaski hinnavõrdlusportaalis ei ole võimalik otsida, sorteerida või filtreerida toote hinnangulise jõudluse järgi. Kokkuvõttes tundub, et mõlema lehekülje otsimisvõimekus piirdub sellega, mis toote puhul on parasjagu nimetuses või kirjelduses – see seab nende potentsiaalile olulise piirangu, kuid võimaldab neil väiksema vaevaga indekseerida paljusid tooteid. Järgmises peatükis kirjutatakse tarkvaralahenduse nõuetest.

4. Ülevaade tarkvaralahenduse nõuetest

Selles peatükis antakse ülevaadet tarkvaralahenduse eesmärkidest ehk sellest, millised on nõuded. Eelkõige käsitletakse veebiliidesega seonduvat, sest see mõjutab kasutajakogemust ning läbi selle väljendub ka töö tulemus kõige selgemini.

Tehnilist parameetrit loetakse oluliseks, kui ta antud tootetüübi puhul parandab märgatavalt kasutaja võimekust leida parimat toodet. Muutmälu jõudlust defineeritakse kui taktisageduse ja latentsuse suhet. SSD jõudlus on seotud tema lugemise -ja kirjutamise kiirusega, kuid kasutajale jäetakse valida, mis tüüpi kiirused (järjestikune või suvaline) on jõudluse arvutamisel määravad. Tarkvaralahenduse otsingu nutikus esiteks seisneb selles, et kombineeritakse erinevaid tootespetsiifilisi andmeid, et arvutada välja jõudlust, hinda mahu kohta, hinna-mahu ja jõudluse vahelist suhet ning võimaldatakse kasutajal nende järgi sorteerida. Teiseks, toodete tehniliste parameetrite teadmine võimaldab identsed välja filtreerida ning jätta vaid parimad pakkumised (identsete omaduste korral parima hinna-mahu suhtega).

Järgnevalt on välja toodud funktsionaalsed nõuded:

1. Mõlema tootekategooria puhul on kasutajal võimalus filtreerida iga olulise tehnilise parameetri järgi.
2. Kasutaja saab sorteerida hinna, jõudluse ja hinna-jõudluse suhte järgi.
 - 2.1. Hind kasvavas järjestuses.
 - 2.2. Jõudluse ja hinna-jõudluse suhe paremuse mõttes kahanevas järjestuses.
3. Kasutaja saab kõiki filtreid korraga tühistada.
4. Kasutaja saab filtreid peita ja näidata.
 - 4.1. Tervet filtrite nimekirja peab saama peita.
 - 4.2. Iga tehnilise parameetri tüüp (näiteks eluiga, kujutegur, liides) peab olema organiseeritud ja peidetav.
5. Kõik identsete tehniliste parameetritega tooted, mis on halvema hinna-mahu suhtega, peavad olema välja filtreeritavad
6. Informatsiooni kuvamiseks peab olema kaks vaadet.
 - 6.1. Kompaktsel kujul nimekirja vaade.
 - 6.2. Tabel.
 - 6.2.1. Olulised parameetrid peavad olema individuaalselt sorteeritavad.
7. Iga toote kohta peavad olema nähtavad selle olulised parameetrid.
8. Pakkumisele vajutades avaneb uus vaheleht (ingl *tab*), mis viib vastava e-poe tootespetsiifilisele lehele.

Mittefunktsionaalsed nõuded:

1. Veebiliides peab korrektselt toimima vähemalt brauserites Firefox¹⁸ ja Chrome¹⁹.
2. Akna suuruse muutmisel peavad komponendid olema reageerivad ehk dünaamiliselt kohalduvad (ingl *responsive*).
3. Veebiliides peab olema kasutajasõbralik.

¹⁸ <https://www.mozilla.org/et/firefox/new/>

¹⁹ <https://www.google.com/intl/et/chrome/>

4. Kasutajale kuvatakse ainult selliseid tooteid, kus on kõik olulised tehnilised parameetrid leitud.
5. Lisaks hinnale kuvatakse ka hinna-mahu suhet (hind/GB kohta).
6. Pääringuvastused saavad normaaltingimustes mitte kauem kui pool sekundit.
7. Rohkete andmetega vaate vahetamine toimub mitte kauem kui pool sekundit.

Kokkuvõttes eelmainitud nõuded võimaldavad kasutajal mugavalt ja väikese vaevaga leida endale sobiv toode. Järgmises peatükis käsitletakse võimalikke autoriõigusega seonduvaid küsimusi.

5. Tarkvaralahenduse seaduslikkus Eesti Autoriõiguse kontekstis

Käesoleva bakalaureuse töö autor pöördus Tartu Ülikooli Tech Startup Legal Clinic' u poole, et saada abi seadust puudutavates küsimustes ning jõuda selgusele, kas tarkvaralahenduse jooksutamine ehk veebiroomajaga informatsiooni automaatne kogumine ja kuvamine on lubatud²⁰. Järgnevalt käsitletakse kahte juhtu: hariduslikel eesmärkidel tarkvara kasutamist ning kommertseesmärgil tarkvaralahenduse kasutamist. Seejuures on fookus intellektuaalsel omandil, kuid tuleb meeles pidada, et ka hariduslikel eesmärkidel loodud tarkvaraga kahju tekitamisel tuleb see hüvitada.

5.1 Tarkvaralahenduse kasutamine hariduslikel eesmärkidel

Tartu Ülikooli Tech Startup õigusabi kliinikumi üliõpilaste arvamuse kohaselt hariduslikel eesmärkidel autoriõigusega kaitstud teoste kasutamine on lubatud juhul, kui on näidatud autorit (olemasolul) ja andmete päritolu (AutÕS²¹ § 19 punkt 3 (1) ja § 75⁶ punkt 2). Tarkvaralahendus ei kogu pilte, videosid ja muid sarnaseid autoriõigusega kaitstavaid teoseid. Kogutavale tehnilisele informatsioonile Eesti Autoriõiguse seadust ei kohaldata (AutÕS § 5 punkt 7). AutÕS kohaldatakse andmebaasidele (AutÕS § 4 lõikes 1 punkt 22), kuid hariduslikel eesmärkidel lubatakse siiski kasutada ka autori nõusolekuta ning autoritasu maksimiseta ning seega seaduslikkuse küsimust hariduslikel eesmärkidel kasutamisel ei teki.

5.2 Tarkvaralahenduse kasutamine ärilistel eesmärkidel

Eelnevas alampeatükis sai mainitud, et Eesti Autoriõiguse seadust kohaldatakse andmebaasidele, kuid eelmainitud õigusabi kliinikumi üliõpilaste arvamusel tuleb andmebaasi sisu vaadelda eraldi. Nende sõnul arvestades, et andmebaasi sisu mõttes kogub tarkvaralahendus vaid tehnilisi parameetreid ja hinda (mis ei ole autoriõigusega kaitstud), siis oluline on jõuda selgusele, kas andmebaas on kaitstud *sui generis* õigusega. Kirjavahetusest selgus, et kui andmebaasi pole märkimisväärselt panustatud või andmebaas on organisatsiooni peamise tegevuse kõrvalsaadus, siis see ei ole *sui generis* õigusega kaitstud. Kui selline õigus kohaldub, siis keelatud on tegevused, mis on vastuolus andmebaasi tavapärase kasutamisega või kahjustavad põhjendamatult andmebaasi tegija seaduslikke huve (§ 75⁵ lõiked 1-3). Mõlemal juhul tuleb veenduda, et lehekülgedelt, kust informatsiooni kogutakse, pole kasutustingimustes seda keelatud²². Õiguskliinikumi soovitusel tuletatakse meelde, et alati on võimalik poodidelt otse küsida ja pakkuda mingisugust teenust.

Viimaks tuleb kontrollida, kas tarkvara lahenduse kirjutamiseks kasutatud tehnoloogiate litsentsid lubavad neid ärilistel eesmärkidel kasutada. Samuti tarkvaralahenduses kuvatakse kaubamärke (Corsair, Adata, Goodram jne) – seni, kuni neid kuvatakse seoses tootega, on see seadusega kooskõlas.

Kokkuvõttes saab lõputöös valminud tarkvaralahendust kasutada hariduslikel eesmärkidel luba küsimata ja tasu maksmata, kui on näidatud andmete päritolu ning olemasolul autorit.

²⁰ Tartu Ülikooli Tech Startup õigusabi kliinikumi (<https://oigus.ut.ee/en/admissions/tech-startup-legal-clinic>) üliõpilaste arvamus väljendati lõputöö autoriga kirjavahetuse teel.

²¹ <https://www.riigiteataja.ee/akt/128122011005?leiaKehtiv>

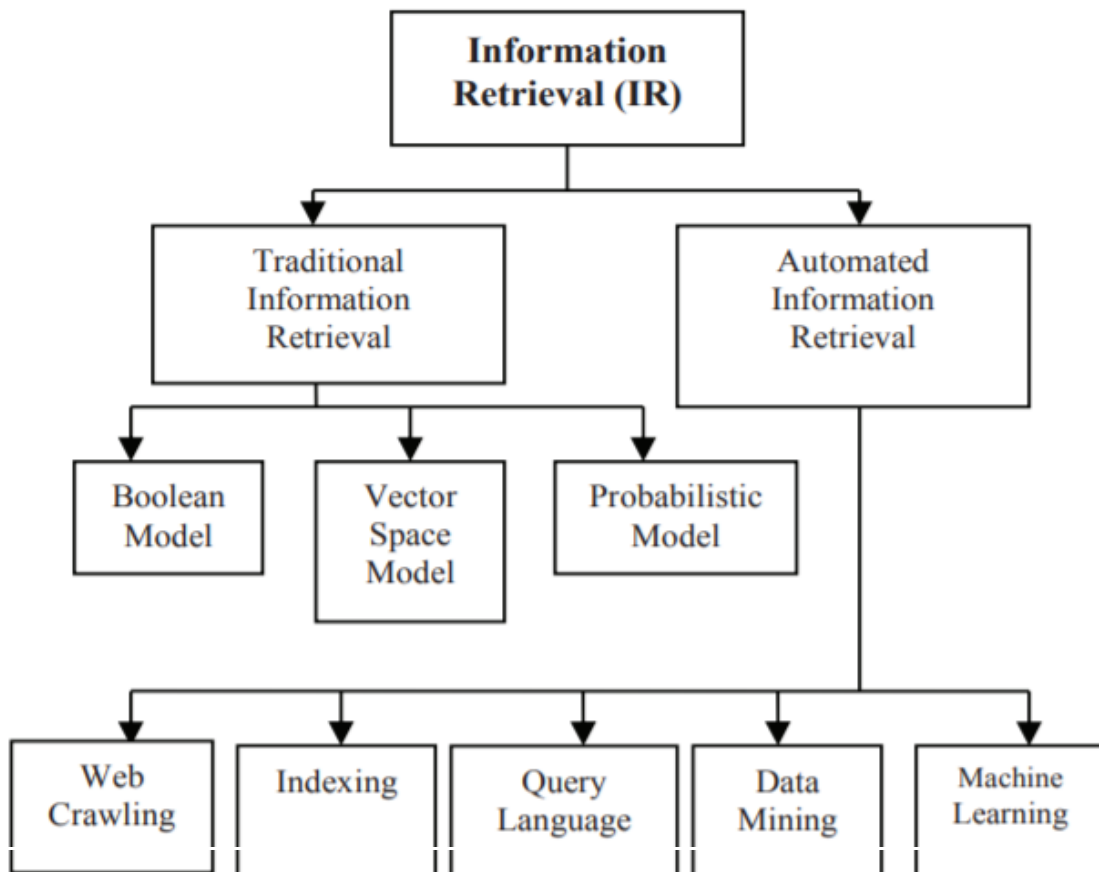
²² <https://parissmith.co.uk/blog/web-crawling-screen-scraping-legal-position/>

Ärulistel eesmärkidel tuleb veenduda esiteks, kas andmebaas on kaitstud *sui generis* õigusega. Teiseks, kas veebilehekülje kasutustingimused lubavad andmeid ekstraheerida ja kolmandaks, kas tarkvara loomisel kasutatud teekide litsents lubab ärulistel eesmärkidel kasutamist. Järgmises peatükis kirjutatakse informatsiooni otsimis- ja kogumisviisidest.

6. Informatsiooni otsimis- ja kogumisviisid

Tarkvara üheks osaks on pakkumiste kohta informatsiooni kogumine Eesti veebipoodide kodulehekülgedelt ning konkreetse toote omaduste ekstraheerimine vastavalt tootja koduleheküljelt.

Saini ja Arora [3] on osutanud enda artiklis „Information retrieval in web crawling: A survey” sissejuhatuses automatiseeritud tööriistade tähtsusele. Nad toovad välja, et otsingumootorid, mida kasutatakse väärtusliku informatsiooni leidmiseks, koosnevad kahest osast: veebroomamine ja andmete ekstraheerimine ehk andmekoorimine. Joonises 10 tuuakse välja erinevad andmete otsimisviisid.



Joonis 10. Erinevate informatsiooni otsimisviiside liigitamine [3].

Traditsiooniline informatsiooniotsingu (ingl *traditional information retrieval*) eesmärk on vastavalt otsija päringule leida kõige asjakohasemat informatsiooni. Automatiseeritud informatsiooniotsingut (ingl *automated information retrieval*) kasutatakse, et vähendada informatsiooni üleküllust (ingl *information overload*). [3]

Lõputöö raames valminud tarkvaralahenduses kasutatakse järgmiseid automatiseeritud otsinguviise: veebroomamine kitsendatud kujul ehk veebikoorimine, indekseerimine (teisisõnu leitud andmete korrastamine²³) ja päringukeel SQL. Neid käsitletakse täpsemalt peatükis 7. Andmekaevandamist ehk suurest andmebaasist uute seoste tuvastamist läbi automaatse analüüsi [3] ei tehta, sest esiteks puudub eelnimetatud suur andmebaas ja teiseks,

²³ <https://www.techopedia.com/definition/7705/indexing>

valminud tarkvara veebiroomamise käigus ekstraheeritud andmete peidetud seoste otsimine ei ole käesoleva lõputöö eesmärk. Automaatseks informatsiooniotsinguks masinõpet ei rakendata, sest veebilehed, millest andmeid ekstraheerida on ette teada ja sellises piiratud kogustes on lihtsam ja kiirem läheneda igale veebileheküljele individuaalselt.

Järgmises alampeatükis kirjutatakse, milleks saab veebiroomajaid kasutada, seletatakse, mis funktsioon on veebiroomamise eeskirjadel ja seejärel tuuakse eeskirja kaupa välja, mis üks või teine eeskiri tähendab valminud tarkvaralahenduse kontekstis (täpsemalt implementatsioonist peatükis 7).

6.1 Veebiroomamise eeskirjad

Gold ja Latonero [4] kirjutasi enda artiklis, et veebiroomajaid üldjuhul kasutatakse otsingumootorite võimaldamiseks, kuid tänu andmekandjate mahu suurenemisele ja arvutite võimsuse kasvule, saab neid rakendada informatsiooni kogumiseks ja selle hilisemaks analüüsiks.

Castillo [5] tõi välja erinevaid veebiroomamise eeskirju, mida valmivas tarkavara arvestatakse. Need eeskirjad on järgmised:

- Valimiseeskiri (ingl *A selection policy*), mis määrab, milliseid lehekülgi alla tõmata.
- Taaskülastamiseeskiri (ingl *A re-visit policy*), mis määrab, millal uuendada andmeid.
- Viisakuseeskiri (ingl *A politeness policy*), mis määrab, kuidas vältida lehekülgede ülekoormamist
- Paralleelsuseeskiri (ingl *A parallelization policy*), mis määrab, kuidas koordineerida paljude veebiroomajate tööd.

Järgmisena kirjutatakse eeskirja kaupa sellest, millised on veebiroomamise eeskirjad tarkvaralahenduse kontekstis. Tarkvaralahenduses üldjuhul on andmekoorimiseks vajalikud veebiaadressid ette teada, seetõttu on veebiroomamine teostatud piiritletud kujul ehk korrektsem on nimetada seda veebikoorimiseks. Sellegi poolest on järgnevad eeskirjad asjakohased, sest külastatakse paljusid lehekülgi andmete ekstraheerimise eesmärgil.

6.1.1 Valimiseeskiri

Vaadeldakse kolme Eestis üldtuntud e-poodi (Galador, Arvutitark, 1a), mis müüb arvutikomponente, sealhulgas muutmälu ja SSD-kettaid. Mitmel põhjusel automaatset otsingut usaldusväärsete arvutikomponentide müüjate leidmiseks ei teostata. Esiteks on Eesti usaldusväärsete arvutikomponente müüvate e-poodide arv piiratud ja lühema aja lõikes drastiliselt ei muutu. Teiseks, lehekülgede ette valimine võimaldab neid personaalselt käsitleda ehk andmekoorimise etapis saab suurema tõenäosusega asjakohased andmed eraldada ning teha seda efektiivseimal moel. Kolmandaks, äsja loodud ning vähepopulaarse e-poe andmete kogumine pole niivõrd ajakriitiline tegevus, et valminud tarkvaralahenduse veebiliidese kasutaja sellest märgatavalt võidak.

6.1.2 Sünkronisatsioonieskiri

Cho ja Garcia-Molina [6] uurisid, kuidas hoida andmeid värskena. Nad töid välja, et lokaalse informatsiooni uuendamise protsess koosneb neljast osast: sünkroniseerimissagedus, ressursijaotus, uuendamisjärjekord ning sünkronisatsiooni alustamisajad.

Valminud tarkvara lahenduse kontekstis defineeritakse uuendamise protsessi jaotust nii:

- Sünkroniseerimissagedus ehk kui tihti e-poode külastatakse pakkumiste uuendamiseks. Cho ja Garcia-Molina käsitlesid seda kui N andmebaasis oleva elemendi (pakkumise) I ajaühiku kohta uuendamist. Valminud lahenduse kontekstis tuleb kasuks eristada, millisest e-poest saadi pakkumine, sest see võimaldab saata e-poe kohta päringuid valikuliselt ning seeläbi tagada aja efektiivsema kasutuse viisakuseeskirja silmas pidades (vt peatükk 6.1.3).
- Ressursijaotus tähendab seda, et kui on otsustatud uuendada veebilehti näiteks kaks korda päevas ehk kolme lehekülje peale kokku kuus korda, siis kas uuendamine toimub ühtlaselt kaks korda e-poe kohta päevas või esimest e-poodi uuendatakse rohkem ja teist vähem, juhul kui esimene veebilehe haldaja uuendab andmeid tihedamini.
- Uuendamisjärjekord käsitleb e-poodide uuendamisjärjekorda. Kitsamalt pakkumiste uuendamisjärjekorda ei ole mõistlik täpselt paika sättida, sest puuduvad andmed selle kohta, millal üht või teist pakkumist tõenäoliselt uuendatakse.
- Sünkronisatsiooni alustamisajad on kellaajad, millal e-poe andmete sünkroniseerimisega alustatakse.

Mida tihedamini andmebaasi uuendada, seda värskemad on andmed [6], kuid tulenevalt ressurssi-, ajapiirangutest ning viisakuseeskirjast, tuleb leida optimaalne sagedus. Cho ja Garcia-Molina võrdlesid kahte lihtsamat ressursijaotuse viisi: ühtlast- (ingl *uniform*) ja proportsionaalset eeskirja. Nende sõnul ühtlane eeskiri tähendab andmete muutumissagedusest sõltumatut uuendamist ja proportsionaalne eeskiri sellest proportsionaalselt sõltuvat. Uurimuse tulemusena jõuti järeldusele, et ühtlane eeskiri toomis paremini, sest kui uuendati mõne väga palju muutuva lehe andmeid teise lehe arvelt rohkem, siis üldine andmete värskus oli oluliselt kehvem. Sellest tulenevalt valminud lahenduses uuendatakse pakkumisi muutmis-sagedusest sõltumatult.

6.1.3 Viisakuseeskiri

Võrguroomajad on võimelised läbima veebilehte palju kiiremini kui seda on võimelised tegema inimesed ning seega on nende potentsiaal teha pahandusi oluliselt suurem. M. Koster [7] nendib, et robotid ehk võrguroomajad nõuavad rohkelt võrguressursse ning nende halvad implementatsioonid süvendavad seda probleemi. Ta toob välja, et mõned sellised vead on mälu (ajaloo) puudumise tõttu lehe korduv taaskülastamine, suvaliste dokumenditüüpide alla laadimine ning ekvivalentsete aadresside mitte eristamine. Samuti toob ta välja, et on reaalne oht serverit üle koormata, kui teha sagedaid päringuid.

Veebilehe hooladajate ja külastajate kaebuste vältimiseks, on käesoleva töö autor otsustanud implementeerida veebiprogrammi järgmiselt:

1. Päeva esimene külastus pakkumiste üldinfo ja veebiaadressite saamiseks toimub hommikul.
2. Kui pakkumise URL on varem külastatud (tootekood veebiaadressi kohta on teada) ja üldinfos on saadaval teabe hinna ja saadavuse kohta, siis vajadusel need uuendatakse.
3. Vastasel juhul lisatakse pakkumise veebiaadress järjekorda, et hankida puuduv informatsioon.
4. Järjekorrast võetakse aadress ja sooritatakse päring mitte tihedamini kui 1 kord sekundis.
5. Sammud 2-4 korratakse mõni minut peale keskööd uuesti, et päeva jooksul muutunud andmed saaksid talletatud ning hommikul selle võrra vähem e-poode koormata.

Sellisel viisil saab vältida serverite üle koormamist ja pärida vaid seda, mis on vajalik.

6.1.4 Paralleelsuseeskiri

Päringuid saadetakse ühe e-poe kategooria kaupa ning enne käsil oleva e-poe järgmise kategooria ette võtmist, kogutakse teiste e-poodide sama kategooria andmeid. Näiteks päritakse enne ühe poe kaupa muutmälu andmeid, seejärel samas järjekorras SSD andmed. Selle eeliseks on esiteks see, et on väiksem oht, et muutmälu saab otsa, kuna peale ühe poe andmete andmebaasi salvestamist saab andmed mälust kustutada. Teiseks, viisakuseeskirja silmas pidades on üks e-pood korraga koormatud lühemat aega.

Kokkuvõtvalt informatsiooni otsimis- ja kogumisviiside peatükis sai ülevaadet otsimisviiside liigitusest ja sai põhjendatud nende valikulist kasutamist tarkvaralahenduses. Samuti on kirjutatud veebiroomamise eeskirjadest ja andmete kogumisest eeskirju silmas pidades. Järgmisena kirjutatakse tarkvaralahenduse arhitektuurist.

7. Tarkvaralahenduse arhitektuur

Selles peatükis kirjutatakse lahenduse üldisest struktuurist ning toimimise protsessist lähemalt. Tarkvaralahendus koosneb järgmistest osadest:

- Andmebaasi haldamise süsteem (sh. andmebaas) – võimaldab andmeid hoiustada, uuendada ja neid pärida.
- Veebiserver – võtab vastu veebiliidese HTTP päringud, saadab andmebaasi haldamise süsteemile vastava SQL päringu, teisendab andmebaasist saadud andmed sobivale kujule ning seejärel edastab veebiliidesele.
- Veebiliides – saadab vastavalt kasutaja soovile päringu veebiserverile, võtab saadud päringuvastuse vastu ja töötleb sobival kujul.
- Andmekoguja – kogub andmeid ettemääratud veebilehtedelt, viib nad sobivale kujule ning saadab salvestamise -või uuendamise päringu andmebaasi haldamise süsteemile.

Tarkvaralahenduse paigaldamise, seadistamise ja jooksumise õpetus on lisas 1. Järgmisena kirjutatakse nende osade üldisest struktuurist.

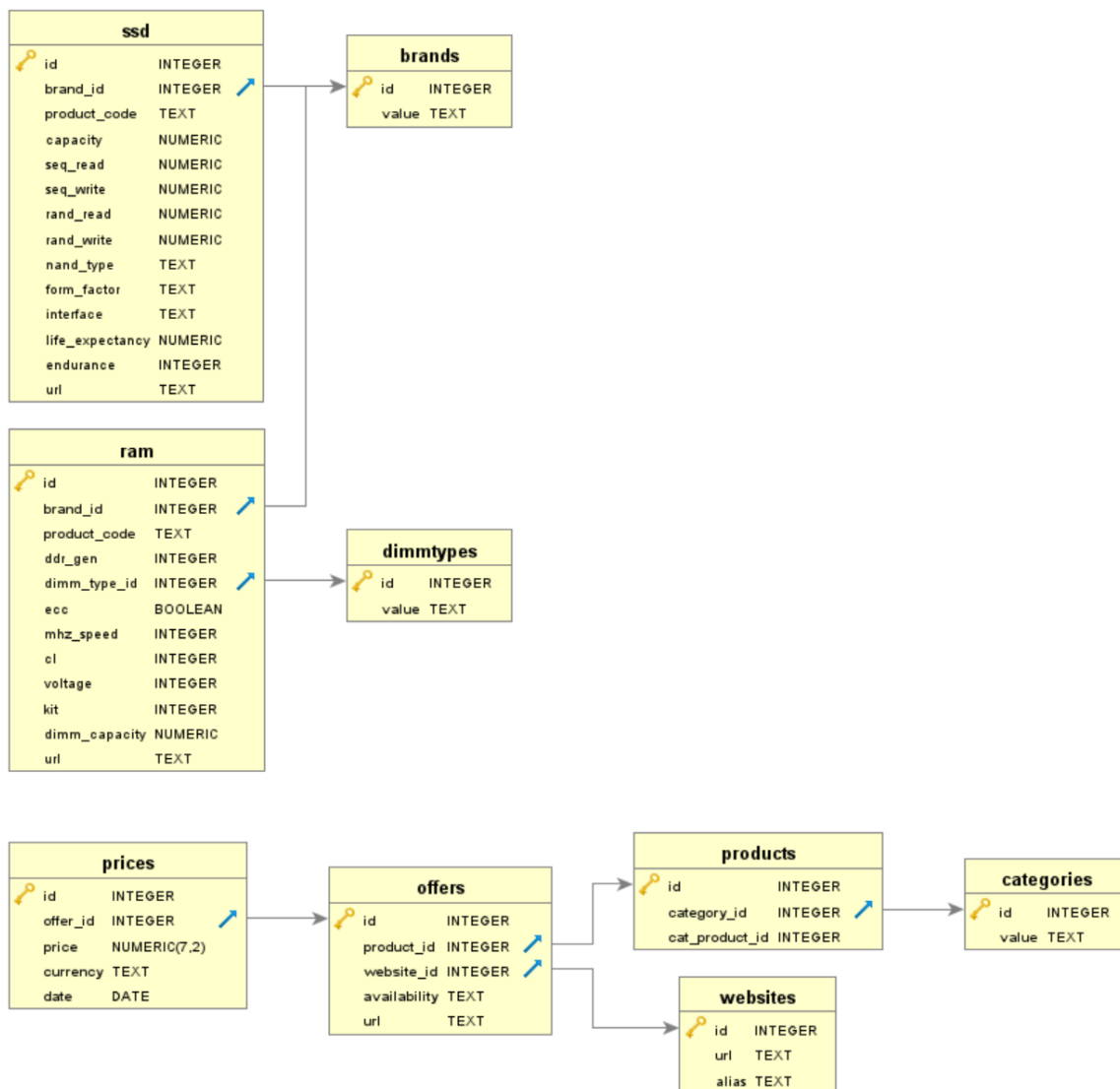
7.1 Üldine struktuur

Selles peatükis kirjeldatakse, milline on andmebaasis olevate andmete struktuur ehk andmemudel. Veebiliidese, andmekoguja ja veebiserveri puhul antakse ülevaadet sellest, mis funktsionaalsuse eest vastutab fail või kataloog ning milline on nende hierarhiline seos.

7.1.1 Andmemudel

Andmemudel koosneb üheksast tabelist (vt Joonis 11), mille koostamisel arvestati võimalusega lisada juurde uusi tootekategoriaid ja talletada hinda pakkumise kohta kuupäeva lõikes. Tabelid *brands*, *websites* ja *dimmtypes* on selleks, et dubleerivat informatsiooni oleks vähem. Selliseid andmeid, mida esineb piiratud ja kindlalt määratletud kujul on veel (näiteks SSD liidesed ja -kujutegurid), kuid lihtsuse mõttes pole eraldi tabelitesse paigutatud. Järgnevalt on välja toodud tabelite eesmärk:

- a) Ssd – tabel SSD-kettaste omaduste ning informatsiooni kättesaamise aadressi talletamiseks.
- b) Ram – muutmälu omaduste ja informatsiooni algallika aadressi hoiustamiseks.
- c) Products – üldine tabel kõikide tootekategoriate kokku koondamiseks.
- d) Brands – erinevate muutmälu ja SSD-ketaste firmanimed talletamiseks. Kasutatakse tabelites Ssd ja Ram, et dubleerivat informatsiooni oleks vähem.
- e) Dimmtypes – sellesse tabelisse salvestatakse erinevad muutmälu tüübid.
- f) Prices – kuupäeva lõikes pakkumiste hinna talletamiseks.
- g) Offers – sellesse tabelisse salvestatakse pakkumise päritolu, tarneaeg ning viide tootele tabelist Products, mis omakorda viitab muutmälule või SSD-ketale.
- h) Websites – e-poodide nimede ja leheküljeaadressite talletamiseks.
- i) Categories – kategooriate nimed (muutmälu ja SSD-kettas), millele viidatakse Products tabelist.



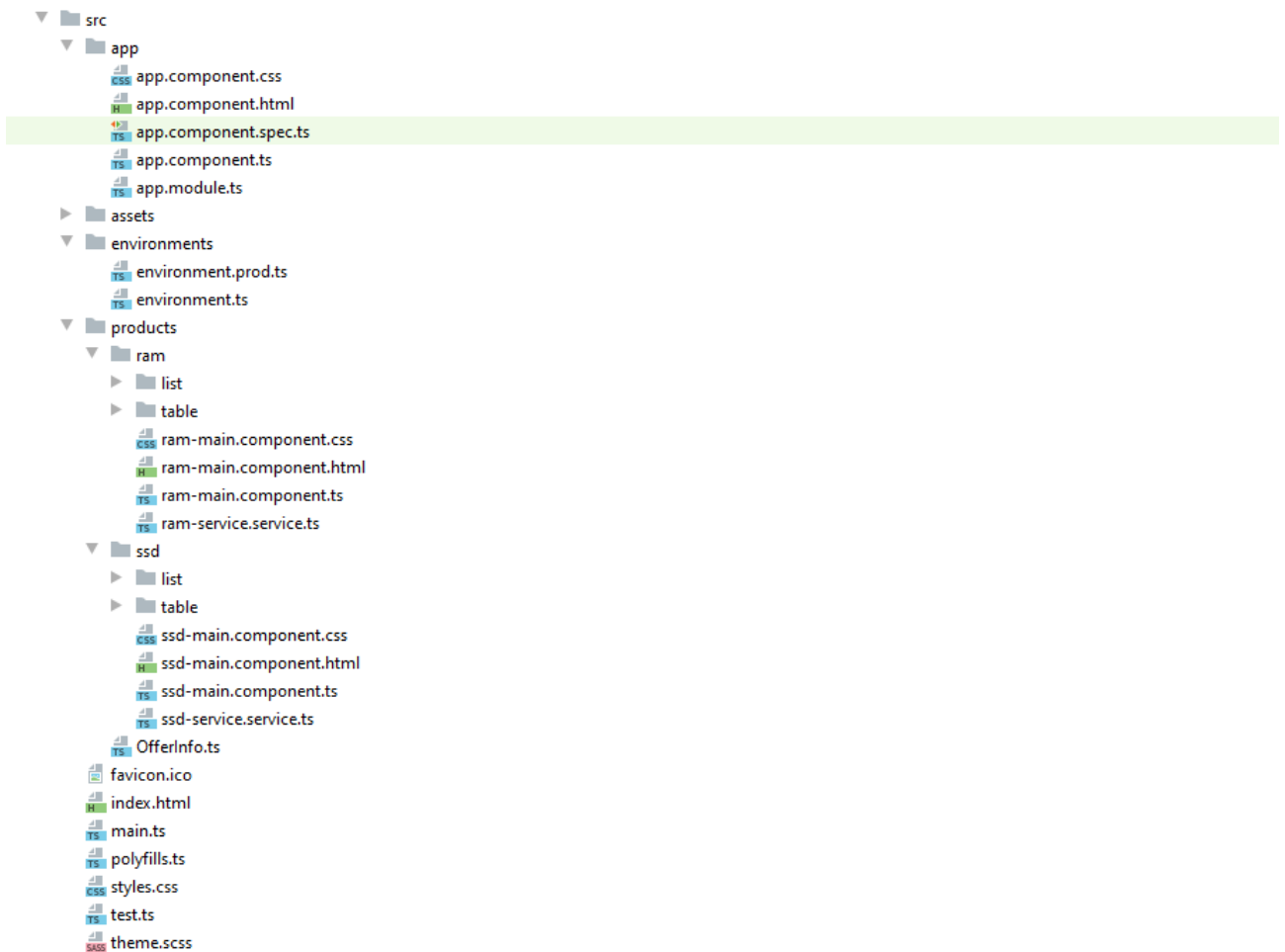
Joonis 11. Andmebaasi struktuur nagu kujutatud programmis DbVisualizer²⁴ 11.0.

Järgmisena kirjutatakse veebiliidese struktuurist.

7.1.2 Veebiliides

Src-nimelises kaustas asuvad koodifailid, mis vastutavad veebiliidese funktsionaalsuse ja väljanägemise eest. Üldist kujundust on paika sätitud ka juurkaustas olevas *styles.css* failis. Joonisel 12 on kujutatud *src* kausta hierarhilist ülesehitust. Kaustas *products* on eraldatud muutmälude ja SSD-ketaste peamised (ingl *main*) komponendid, mis omakorda kasutavad nimekirja -ja tabeli vaate kuvamiseks vastavalt *list* ja *table* kaustades olevaid komponente. Tootekategooriate peamisi komponente kasutatakse projekti *app*-nimelises komponendis. *Environment.ts* failis on defineeritud muutuja nimega *apiUrl*, mis määrab ära *service*-nimelistes failides andmete pärimiseks kasutatavat aadressi.

²⁴ <https://www.dbvis.com/>



Joonis 12. Ekraanitõmmis veebiliidese failistruktuurist programmis IntelliJ IDEA²⁵.

Järgmises alampeatükis kirjeldatakse veebiserveri failistruktuuri ja kirjutatakse failide eesmärkidest.

7.1.3 Veebiserver

Src/main/resources kaustas olevas *application.yml* failis on fikseeritud, mis andmebaasist päritakse andmeid. *Src/main/java/price.watcher.backend* kataloogis asuvad *Controller*, *Service* ja *Repository* lõpuga failid, mille eesmärgid on järgmised:

- *Controller* – Siin on defineeritud API otspunktid (ingl *API endpoint*), kuhu veebiliidese saab päringuid.
- *Service* – Selle kihi eesmärk on koondada kokku igasugune loogika, mis ei ole otsest seotud API otspunktide defineerimisega ja andmebaasist pärimisega. Antud tarkvaralahenduse raames teisendatakse päritud andmed JSON formaati, et neid veebiliidesele edastada.
- *Repository* – Nendes klassides (ingl *class*) päritakse andmebaasist vastava tootekategooria andmed.

²⁵ <https://www.jetbrains.com/idea/>

Src/main/java/price.watcher.backend kataloogis *model*-nimelises pakis (ingl *package*) olevad *Params* ja *Info* lõpuga klassid fikseerivad vastavalt, mis viisil on võimalik andmeid sorteerida, filtreerida ning milliseid andmeid talletatakse muutujatesse, mida hiljem teised datakse JSON formaati ning edastatakse veebiliidesele.

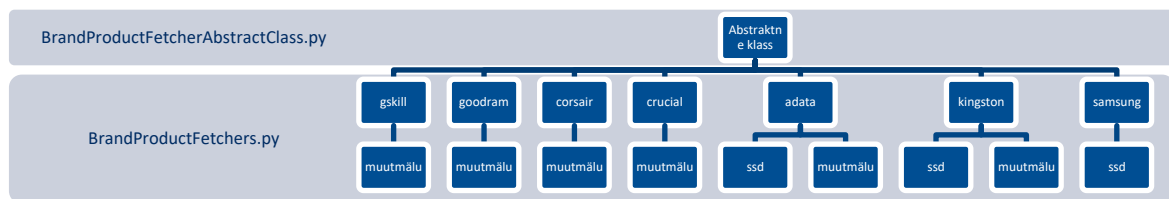
Järgnevalt käsitletakse andmekoguja failistruktuuri ning failide eesmäärke.

7.1.4 Andmekoguja

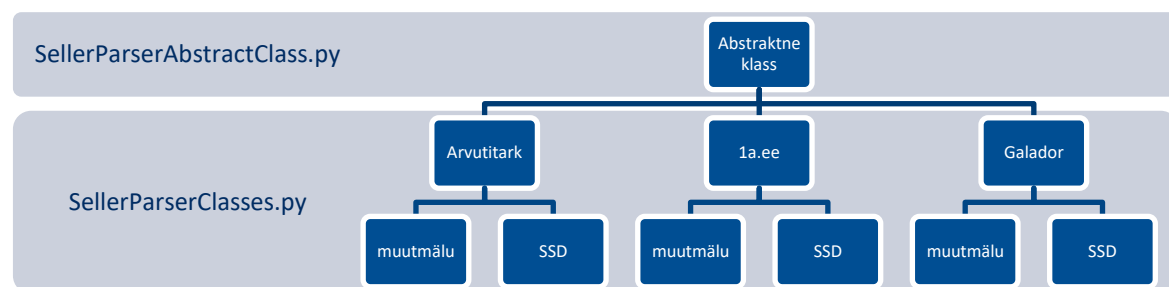
Andmekoguja saab ülesannete osas jaotada kolmeks osaks:

- Andmebaasiga suhtlemine – salvestamise ja pärimise kood asub *database*-nimelises pakis (ingl *package*).
- Tooteandmete kogumine tootja koduleheküljelt – vastav kood asub juurkaustas kahes failis (vt Joonis 13).
- E-poodidest pakkumiste hinna, tarneaja ja tootekoodide informatsiooni kogumine – vastav kood asub samuti juurkaustas kahes failis (vt Joonis 14).

Andmete teisendamise ja tootekoodi regulaaravaldisega kontrollimise abifunktsioonid asuvad failis *parserHelpers.py*.



Joonis 13. Muutmälu ja SSD-ketaste tootja poolsete andmete kogumise hierarhiline struktuur ja koodi asukoht.



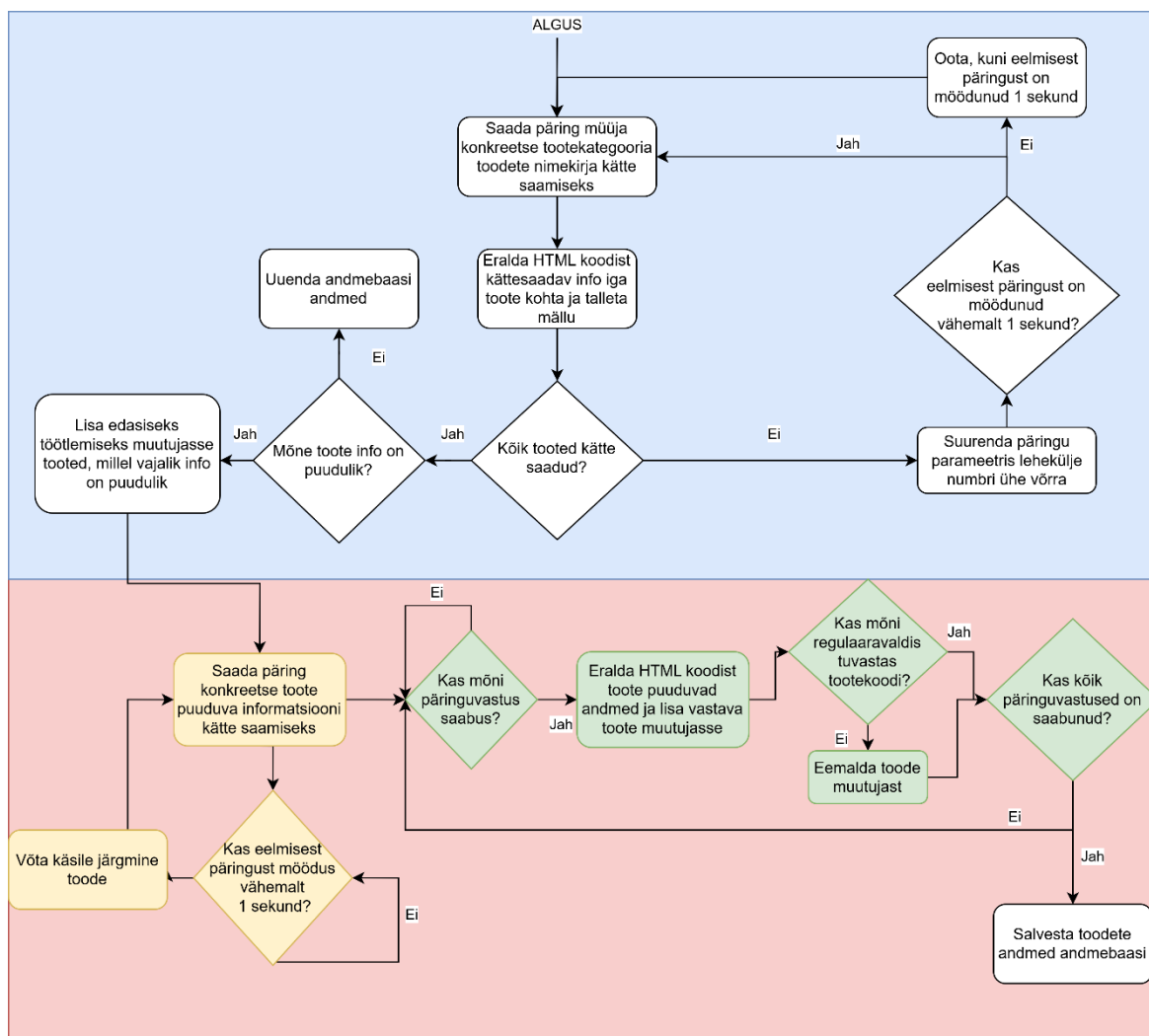
Joonis 14. E-poodide andmete kogumise koodi üldine struktuur ja asukoht.

Järgnevalt analüüsitakse põhjalikumalt, kuidas tarkvaralahenduse erinevad osad üksteisega suhtlevad ning milline näeb välja andmekoguja informatsiooni käsitlemise voog.

7.2 Protsessist lähemalt

Joonisel 15 on kujutatud, kuidas tootjatelt ja e-poodidelt informatsiooni kogutakse ning käsitletakse. Sinises kastis olevad tegevused toimuvad e-poe konkreetse tootekategooria toodete nimekirjas, kus on tavaliselt vaja mitu lehte läbi käia, et kõik andmed kätte saada. Punases kastis toimub tegevus andmetega, mille puhul e-poe nimekirja vaates ei õnnestunud kogu vajalik informatsioon kätte saada ning seetõttu tekib vajadus külastada suurema

andmesaldusega tootespetsiifilist lehte. Kollast värvi kastides toimuv on eraldi lõimes ning jookseb roheliste kastide tegevusega samaaegselt ehk rohelise värviga ülesanded täidetakse niipea kui saabub mõnele väljasaadetud päringule vastus. Iga uue toote puhul kontrollitakse regulaaravaldisega selle tootekoodi. Kui tootekood ei ole regulaaravaldisega tuvastatav, siis oletatakse, et ta on vigane või tegemist on tootega, mis sattus valesse kategooriasse. Sellisel juhul eemaldatakse see salvestamise nimekirjast.

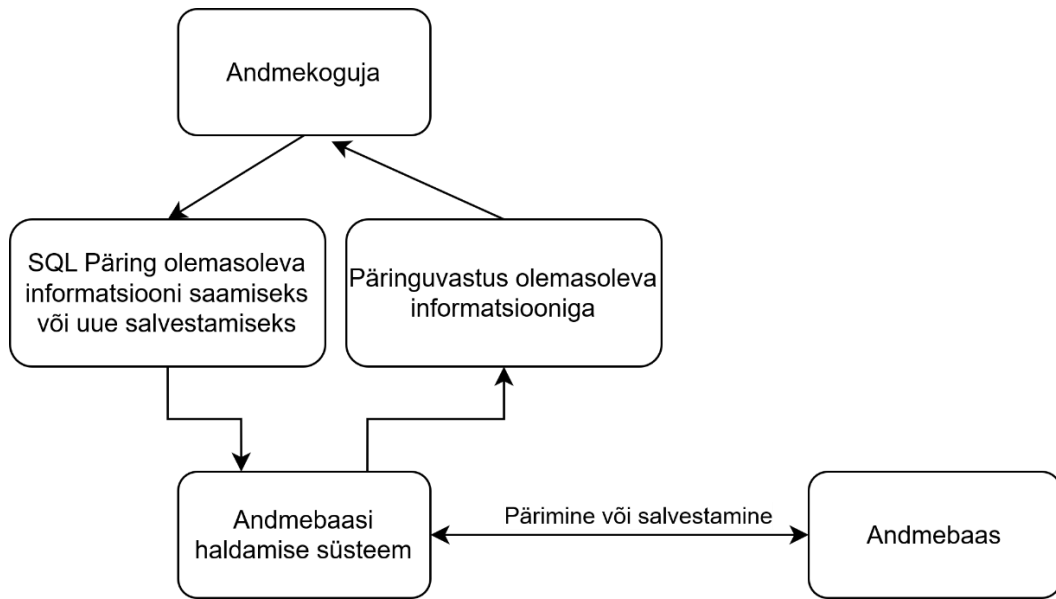


Joonis 15. Informatsioonikogumise voog tootja -ja e-poe lehekülgedelt.

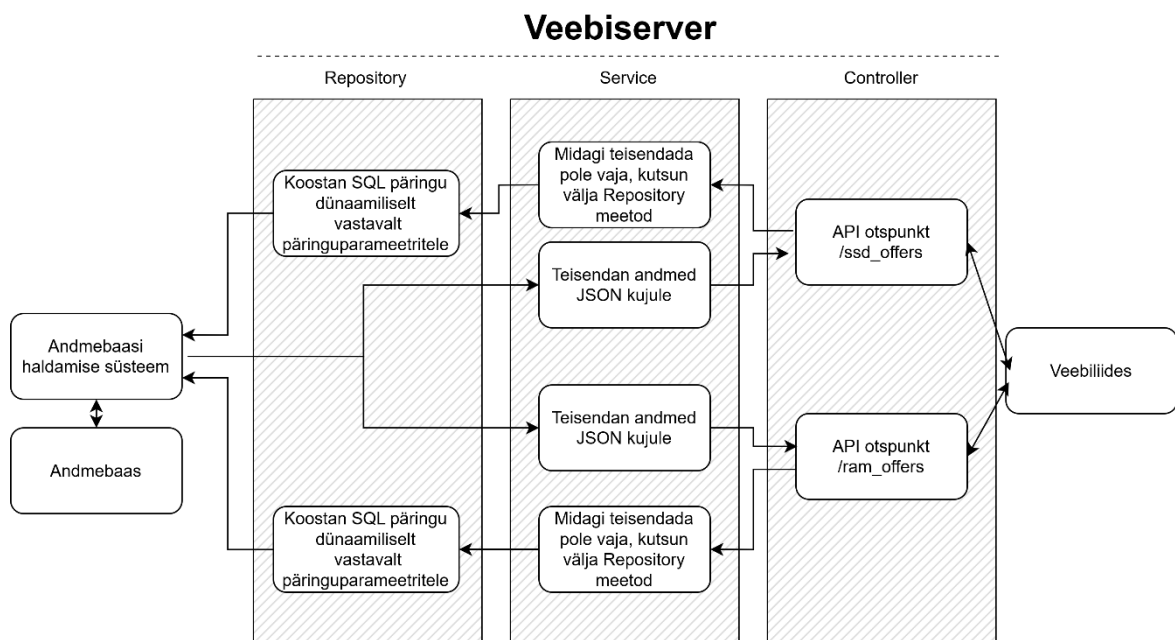
Andmekoguja suhtlust toodete andmete salvestamisel või olemasolevate andmete pärimisel on kujutatud joonisel 16. Andmebaasis olemasolevat informatsiooni päritakse näiteks siis, kui on vaja teada, kas mõne toote informatsioon on puudulik, sest kui puudu on tootekood, siis võib selguda, et tootespetsiifiline aadress on minevikus juba külastatud ning seega ka tootekood teada.

Kui andmebaasis on vajalikud andmed olemas, siis saab neid veebiliidese kaudu veebiserverilt pärida. Veebiliidese poolt päringu saatmisel võetakse päring vastu veebiserveri kontrolleris (ingl *controller*) asuvas API otspunktis, seejärel suunatakse päring edasi teenuskihti, mis kutsub repositooriumikihis (ingl *repository*) oleva funktsiooni välja. Seal koostatakse SQL päring vastavalt päringu parameetritele ning vastus päringule antakse tagasi

teenuskihti, kus andmed teisendatakse JSON kujule ja lõpuks saadetakse veebiliidese läbi API otspunkti tagasi (vt joonis 17).



Joonis 16. Andmekoguja suhtlus andmebaasiga andmete pärimisel või salvestamisel.



Joonis 17. Veebiserveri suhtlusvoog veebiliidese ja andmebaasi haldamise süsteemiga.

Seoses sellega, et iga leht on isemoodi, tekkis vajadus mõne lehe puhul teha lisatoiminguid, et andmed edukalt kätte saada. Sellest kirjutatakse järgmises alampeatükis.

7.2.1 Informatsiooni kogumise erisused ja probleemilahendus

E-poodidest rääkides, Arvutitarga koduleheküljelt saab toodete nimekirja vaates kätte toote hinna, tarneaja ja aadressi, mis viib tootespetsiifilisele lehele, kus asub täpsem informatsioon (seal hulgas tootekood). See tähendab seda, et kui järgneval korral koguda andmeid, siis saab teha oluliselt vähem päringuid, kuna eelneval korral saadud tooteaadressile vastav

tootekood on juba teada. Kahjuks e-poes 1a.ee ei ole toodete nimekirja vaates tarneaega, seega uuendamisel tuleb paratamatult iga toote puhul külastada spetsiifilise informatsiooni lehte. Arvutitarga spetsiifika seisneb selles, et aadress, mis viitab tootekategoriale (näiteks muutmälud) muutub. Selleks, et selliste muutustega hakkama saada, tehakse päring vastu Arvutitarga pealehekülge ja vastuse peal rakendatakse andmekoorimisvõtteid. Galadori puhul puudub mõne toote tootekood – sellisel juhul võib ta leiduda toote pealkirjas, mida eraldatakse muust infost kasutades regulaaravaldist. Igas e-poes võib juhtuda, et tootekoodile on kogemata midagi ette või taha lisatud, seega isegi tootekood pole täiesti usaldusväärne. Seetõttu on iga tootja tootekoodide tuvastamiseks kirjutatud regulaaravaldis. Selline lähenemine võimaldab valel kujul olevat koodi õigeks parandada ja tuvastada, mis tootjaga on tegu. Lisaks on tuvastamise eelis see, et andmebaasi ei talletata tooteid, mis on e-poe poolt valesse kategooriasse paigutatud või kategooria ise hõlmab endas muid tooteid, mille informatsiooni kogumist parajasti vaja pole (näiteks SSD-ketaste tarvikud).

Tootjate kodulehekülgedel on samuti olnud probleeme ja erisusi. Corsair-nimelise brändi lehekülje puhul keelustatakse enamuse päringuid, kui päringusaatjal pole teatud küpsiseid (ingl *cookie*). See sai lahendatud nii, et enne päringute saatmist külastatakse üks kord peaveebilehekülge kasutades Selenium teeki ning edastatakse küpsised Requests teegi objektile. Crucial brändi puhul laetakse andmed dünaamiliselt Javascriptiga, mis tähendab, et HTTP päringus need ei kajastu. Seetõttu päringute saatmise asemel käiakse veebileht läbi kasutades vaid Seleniumit ning iga lehe puhul oodatakse kuni tooteinformatsioon on HTML koodis kajastatud. Paljudel Adata SSD-ketastel ei ole veebilehel kajastatud suvalise lugemise -ja kirjutamise kiirused (ingl *random read/write*) ning on kättesaadavad PDF failis. Selleks, et neid seal viisakal kombel kätte saada, kasutatakse Google Drive API teenust nii:

1. Laetakse toote andmetega PDF-fail alla.
2. Laetakse seesama fail Google Drive'i ülesse nimega *ocr_datasheet.pdf*.
3. Teisendatakse üles laetud fail Google dokumendi formaati.
4. Laetakse teisendatud fail .txt formaadis alla.
5. Kustutatakse pilvest *ocr_datasheet.pdf*.

Seejärel saab tekstifailist andmed kätte kasutades regulaaravaldist. Selline Google teenust kasutav meetod võimaldab PDF-failist andmed kätte saada ka siis, kui nad on pildi kujul. Järgmises peatükis kirjutatakse tarkvaralahenduses kasutatud raamistikest, keeltest ja teeki-dest.

8. Kasutatud raamistikud, keeled ja teegid

Käesolevas peatükis kirjeldatakse erinevaid tehnoloogilisi vahendeid ja põhjendatakse nende valikut. Esimesena käsitletakse andmete haldust.

8.1 Andmete haldus

Andmete halduseks valiti relatsioonilise andmebaasi haldamise süsteemi nimega PostgreSQL. See on avatud lähtekoodiga, tasuta, hea dokumentatsiooniga ning üldtunnustatud²⁶. Samuti on tal hea tugi Python keelele Psycopg2 teegi abil. Teisi üldtuntuid relatsioonilisi andmebaasi mootoreid nagu SQLite²⁷ ja MySQL²⁸ ei valitud, sest autoril oli varasem kogemus PostgreSQL haldamise süsteemiga ning selle funktsionaalsus antud tarkvaralahenduse jaoks tundus enam kui piisav.

Järgmisena põhjendatakse andmekoguja kirjutamisel tehtud tehnoloogilisi valikuid.

8.2 Andmekoguja tehnoloogilised valikud

Andmete koguja arendamiseks sai valitud programmeerimiskeel Python järgmistel põhjustel:

- Keelel Python on palju teeke, mis oluliselt lihtsustavad veebiroomamist ja andmekoorimist.
- Võrreldes näiteks Java või C++ keeltega on Python keeles mugav manipuleerida tekstiga.
- See on esimene keel, mida Tartu Ülikoolis õpetatakse bakalaureuse õppeastmel õpetatakse ning seega lõputöö autorile tuttav.

Järgnevalt on välja toodud teekide valikud, nende võimalikud alternatiivid ja põhjendatus ülesande kaupa.

8.2.1 Veebiroomamise teegid

Veebi roomamisel on kasutatud järgmiseid teeke:

- Requests – Teek toetab HTTP päringute saatmist, parameetrite automaatset õigele kujule teisendamist (kodeerimist) ning sellel on ka paljusid muid veebiroomamist lihtsustavaid omadusi [8]. Alternatiiviks on urllib3²⁹, kui on soovi saavutada maksimaalset efektiivsust lihtsuse arvelt. Requests kasutab osa funktsionaalsuse tagamiseks urllib3 teeki [8]. Lõputöö autor otsustas Requests teegi kasuks selle lihtsuse ja populaarsuse tõttu.
- Requests-futures – See teek võimaldab lihtsal ja loetaval viisil saata HTTP päringuid asünkroonselt³⁰. Alternatiivina on võimalik teegi funktsionaalsus implementeerida käsitsi, kuid lõputöö autor ei tundnud selleks vajadust.

²⁶ <https://www.postgresql.org/about/>

²⁷ <https://www.sqlite.org/index.html>

²⁸ <https://www.mysql.com/>

²⁹ <https://urllib3.readthedocs.io/en/latest/>

³⁰ <https://github.com/ross/requests-futures>

- Selenium – Teatud veebilehekülgede puhul võib juhtuda, et andmeid serveritakse dünaamiliselt Javascripti abil. Sellises olukorras HTML'i pärimine (näiteks Requests teegiga) ei aita, sest Javascripti jooksub brauser. Selenium võimaldab kasutades Pythoni keelt kontrollida brauseri tööd ja tagastada HTML'i juba töödeldud kujul ehk teisisõnu andmetega, mida laeti dünaamiliselt [9]. Python-spidermonkey³¹ ja Windmill³² on võimalikud alternatiivid. Esimene ei ole enam aktiivselt arenduses ja viimase dokumentatsioon annab võrreldes väljavalitud teegiga soovida.
- Brotli – See teek võimaldab selle nimelise pakkimisalgoritmiga pakitud HTTP päringuvastused lahti pakkida³³. Brotlipy³⁴ on teine teek, mida saaks kasutada. Seda uuendati viimati aastal 2017. Brotlit haldab Google, mis kindlustab tihedamat uuendamist ja paremat tuge probleemide korral.

Järgmises alampeatükis käsitletakse teeke, mida kasutatakse andmekoorimisel.

8.2.2 Andmekoorimise teegid

Järgmisena on välja toodud andmekoorimise teegid:

- BeautifulSoup4 – See võimaldab lihtsal ja kiirel viisil ekstraheerida andmeid HTML koodist [10]. Üks alternatiiv on kasutada lihtsalt Lxml'i. BeautifulSoup4 võimaldab kasutada HTML'i töötamiseks erinevaid teeke, mille hulka kuulub ka eelmainitud Lxml, seetõttu on ta paindlikum [10].
- Lxml – teisendab HTML koodi puu kujule ja võimaldab puuelementidega manipuleerida³⁵.

Seoses sellega, et Adata SSD-ketaste puhul on mõned andmed saadaval vaid spetsifikatsioonifailis, on järgmises alampeatükis välja toodud PDF-failist andmete ekstraheerimise teegid.

8.2.3 PDF-failist andmete ekstraheerimise teegid

Google-api ja Google-auth teegid võimaldavad Google Drive API'ga suhelda³⁶. Optiliseks tekstivastuseks saab kasutada Tesseract'i³⁷, kuid seda kasutusele ei võetud, sest Tesseract'i versioon 4.1.1 (Leptonica-1.78.0) väljundis esineb olulisi kirjavigu ning Google teenust kasutades tuginetakse ka tekstil, mis ei ole pildi kujul (vt lisa 2). Google Vision API on teine alternatiiv, mis ei sobinud, sest teenus muutub tasuliseks peale tuhandet pilti teisendust kuu jooksul [11], samas kui seisuga 15.04.2020 lubab Google Drive API kuni 1 000 000 000 päringut päevas, 10 000 päringut 100 sekundi jooksul ning 1000 päringut 100 sekundi jooksul kasutaja kohta (vt Joonis 18).

³¹ <https://github.com/mozilla/python-spidermonkey>

³² <https://github.com/windmill/windmill/wiki>




³³ <https://github.com/google/brotli>

³⁴ <https://pypi.org/project/brotlipy/>

³⁵ <https://lxml.de/>

³⁶ <https://developers.google.com/drive/api/v3/about-sdk>

³⁷ <https://github.com/tesseract-ocr/tesseract>

Quota Name	Limit	
Queries per day	1,000,000,000	
Queries per 100 seconds per user	1,000	
Queries per 100 seconds	10,000	

Joonis 18. Google Drive API kasutuspiirangud³⁸.

Järgmises alampeatükis vaadeldakse teeki, mis võimaldab keeles Python suhelda andmebaasi mootoriga ja selle alternatiivi.

8.2.4 Andmete pärimise ja talletamise teek

Psycopg2 võimaldab suhelda PostgreSQL andmebaasi mootoriga, sealhulgas saata SQL päringuid informatsiooni pärimiseks ja talletamiseks. Asyncpg võimaldab seda kõike teha asünkroonselt³⁹. Autor otsustas Psycopg2 kasuks, sest sellel on mugav ja turvaline viis liisada päringuparameetrid⁴⁰ ning otsest vajadust andmebaasiga suhtluse optimeerimiseks töö käigus ei ilmnenu.

Kokkuvõtvalt sai selles peatükis teada, mis teeki kasutatakse veebiroomamise võimaldamiseks, HTML koodist ja PDF-failist andmete ekstraheerimiseks ning kirjutati ka teegist, mis võimaldab PostgreSQL andmebaasi mootoriga suhelda, et andmeid pärida ja salvestada. Järgmisena kirjutatakse veebiserveris kasutatavatest tehnoloogiatest, keeltest ja teekidest.

8.3 Veebiserver

Veebiserveri kirjutamisel kasutati Micronaut nimelist raamistikku, mis põhineb Java virtuaalmasinal ja mida saab kasutada Java, Apache Groovy ja Kotlin keeltes programmeerides⁴¹. Seda kasutatakse näiteks järgmise põlvkonna haigla infosüsteemi (HEDA) mikroteenuste arendamisel. Lõputöö autor otsustas selle raamistikku kasuks lihtsuse ja loetavuse tõttu (vt Joonis 19). Võrreldes näiteks Spring raamistikuga, Micronaut käivitub kiiremini ja võtab vähem muutmälu⁴². Python'i keeles kasutatavat Django raamistikku ei kaalutud, sest sellega sunnitakse peale objekt-relatsioonvastendust⁴³ (ingl *ORM*) (vt Joonis 20), käsitsi SQL päringuid kirjutada pigem ei soovitata⁴⁴.

³⁸ <https://console.developers.google.com/apis/>

³⁹ <https://magicstack.github.io/asyncpg/current/usage.html>

⁴⁰ <https://www.psycopg.org/docs/sql.html>

⁴¹ <https://github.com/micronaut-projects/micronaut-core>

⁴² <https://medium.com/@walkingtreetech/micronaut-the-potential-poster-boy-for-microservices-7c9a008b1e57>

⁴³ <https://akit.cyber.ee/term/8489-hibernate-orm>

⁴⁴ <https://docs.djangoproject.com/en/3.0/topics/db/sql/>

```

import io.micronaut.http.annotation.*;

@Controller("/hello")
public class HelloController {

    @Get("/")
    public String index() {
        return "Hello World";
    }
}

```

Joonis 19. Micronaut teegi API ülespaneku lihtsuse näide teegi koduleheküljelt⁴⁵.

```

>>> queryset = Event.objects.filter(years_ago__gt=5)
>>> str(queryset.query)
SELECT "events_event"."id", "events_event"."epic_id", "events_event"."details",
"events_event"."years_ago" FROM "events_event"
WHERE "events_event"."years_ago" > 5

```

Joonis 20. Django raamastiku objekt-relatsioonvastendus⁴⁶.

Kasutatud teekidest mainimisväärsed on Spring⁴⁷ raamistik JDBC (*Java Database Connectivity*), mis võimaldab andmebaasiga suhelda ja Jackson Databind⁴⁸, millega teisendatakse andmed veebiliidesele sobivale kujule. Järgmisena kirjutatakse veebiliideses kasutatavaid teekidest ja raamistikest.

8.4 Veebiliides

Veebiliideses kasutatavad tehnoloogiad, teegid ja raamistikud on järgmised:

- Node.js on avatud lähtekoodiga Javascript'i käitusüsteem, mis on loodud skaleeritavate veebirakenduste loomiseks. Tal on asünkroonne sündmustjuhitav (ingl *event-driven*) arhitektuur, mis võimaldab mitmeid ühendusi teenindada korraga. [12]
- Npm (Node Package Manager) on ühelt poolt repositoorium, kus hoiustatakse avatud lähtekoodiga Node.js projekte ning teiselt poolt ka pakihaldussüsteem, mis aitab

⁴⁵ <https://micronaut.io/>

⁴⁶ Ekraanitõmmis koodinäitest leheküljel <https://books.agiliq.com/projects/django-orm-cookbook/en/latest/query.html>

⁴⁷ <https://spring.io/projects/spring-framework>

⁴⁸ <https://github.com/FasterXML/jackson-databind>

paigaldada pakke, hallata sõltuvusi ja versioone [13]. Veebiliidese projekti juurkaustas on *package.json*, mis defineerib kasutatavaid sõltuvusi, mida saab Npm'i kasutades ühe käsuga paigaldada. Sellest räägitakse täpsemalt lisas 1.

- TypeScript on Javascript keeleks kompileeruv keel, mis võimaldab kasutada staatilist tüübikirjeldust rakenduste loomisel. Tüübikontroll võimaldab kirjutada veakindlamat koodi.
- Angular raamistik on vahend veebiliideste loomiseks⁴⁹. Sellega võimaldatakse jaotada koodi loogilisteks osadeks või teisisõnu komponentideks ning neid hierarhiliselt siduda [14]. Koostöös Node.js käitussüsteemi ja Npm pakihalduriga saab kirjutada komponentide dünaamilist poolt TypeScript keeles, mida kompileeritakse veebilehe serveerimiseks Javascript'i [15].

Angular'i alternatiivina veebiliidese arendamiseks nähtakse teeki nimega React⁵⁰. Nende kahe vahe on selles, et esimene pakub terviklikumat lahendust, samas kui teine julgustab iseseisvalt valida teeki soovitud lisafunktsionaalsuse tagamiseks. React-nimeline teek suunab kirjutama komponendi dünaamilist (Javascript või TypeScript) ja staatilist (HTML) koodi ühes failis, Angular'is hoitakse neid eraldi .ts ja .html failides. [16]

- Angular'is kasutatavatest teekidest on mainimisväärne Material⁵¹. See sisaldab endas paljusid veebiliidese arendamise jaoks kasulikke komponente⁵². PrimeNG⁵³ ja NgZorro⁵⁴ on samuti asjalikud, kuid autor otsustas just Material kasuks, sest võrreldes esimesega on komponente rohkem ning võrreldes teisega tundus komponentide kujundus sobilikum olevat.

Järgmises peatükis kirjutatakse tarkvaralahenduse jõudlusest ja nõuetele vastavusest.

⁴⁹ <https://angular.io/>

⁵⁰ <https://reactjs.org/>

⁵¹ <https://material.angular.io/>

⁵² <https://material.angular.io/components/categories>

⁵³ <https://www.primefaces.org/primeng/>

⁵⁴ <https://ng.ant.design/docs/introduce/en>

9. Saavutatud jõudlus ja nõuetele vastavus

Kõik testid on sooritatud autori arvutis, kus operatsiooni süsteemiks on Windows 10, protsessor Intel i9-9900k ja muutmälu on 64 GB.

9.1 Veebiserveri jõudlus

Lokaalset veebiserverit testiti tööriistaga nimega *Apache HTTP server benchmarking tool*⁵⁵. Kontrolliti, milline on veebiserveri võimekus juhul, kui saata 100 kaupa 10 000 päringut vastu muutmälu API otspunkti, mis tagastas testimise hetkel 2712 pakkumist (umbes 1.3 MB päringu kohta). Testimiseks valiti muutmälu pakkumisi, sest testimise hetkel oli neid oluliselt rohkem. Tulemus on järgmine:

Kulunud aeg kokku: 73.101 sekundit

Ebaõnnestunud päringuid 0

Kokku alla tõmmatud 12995710000 baiti (umbes 13 GB)

Päringuid sekundis: keskmiselt 136.80

Keskmine aeg päringu kohta: 731 ms

Sellest võib järeldada, et kui interneti kiirus ei ole piirav faktor, siis veebiserver on võimeline mitme tuhande pakkumisega suure koormuse all kõigile päringutele vastata ja teha seda seejuures ühegi ebaõnnestumiseta ning (arvestades koormust) ka mõistliku aja jooksul päringu kohta.

9.2 Andmekoguja andmete ekstraheerimise ajakulu

Viisakuseeskirja tõttu seatud ühe sekundiline piirang päringute vahel on andmekoguja suurim ajaline kadu. Selenium'iga lehtede lahti laadimine on samuti ajakulukas. Sellegipoolest on hea teada, kui kiiresti saab erinevatest veebilehekülgedest andmed päringutest ekstraheeritud. Võetud on keskmine päringu kohta, väikeste erinevuste puhul võetakse SSD ja muutmälu tulemustest keskmine ja eraldi välja ei tooda. Tulemused on järgmised:

- a) Arvutitark 0.041 s
- b) Galadoris 0.038 s
- c) 1a.ee 0.032 s
- d) Gskill 0.012 s
- e) Goodram 0.0156 s
- f) Adata 0.01 s
- g) Crucial 0.077 s
- h) Kingstoni muutmälu 0.009 s, SSD 0.028 s
- i) Samsung 0.026 s

Kingstoni puhul võtab SSD andmete ekstraheerimine võrreldes muutmäludega oluliselt kauem aega selle tõttu, et SSD-ketaste andmed on keeruliste seostega leheküljele

⁵⁵ <https://httpd.apache.org/docs/2.4/programs/ab.html>

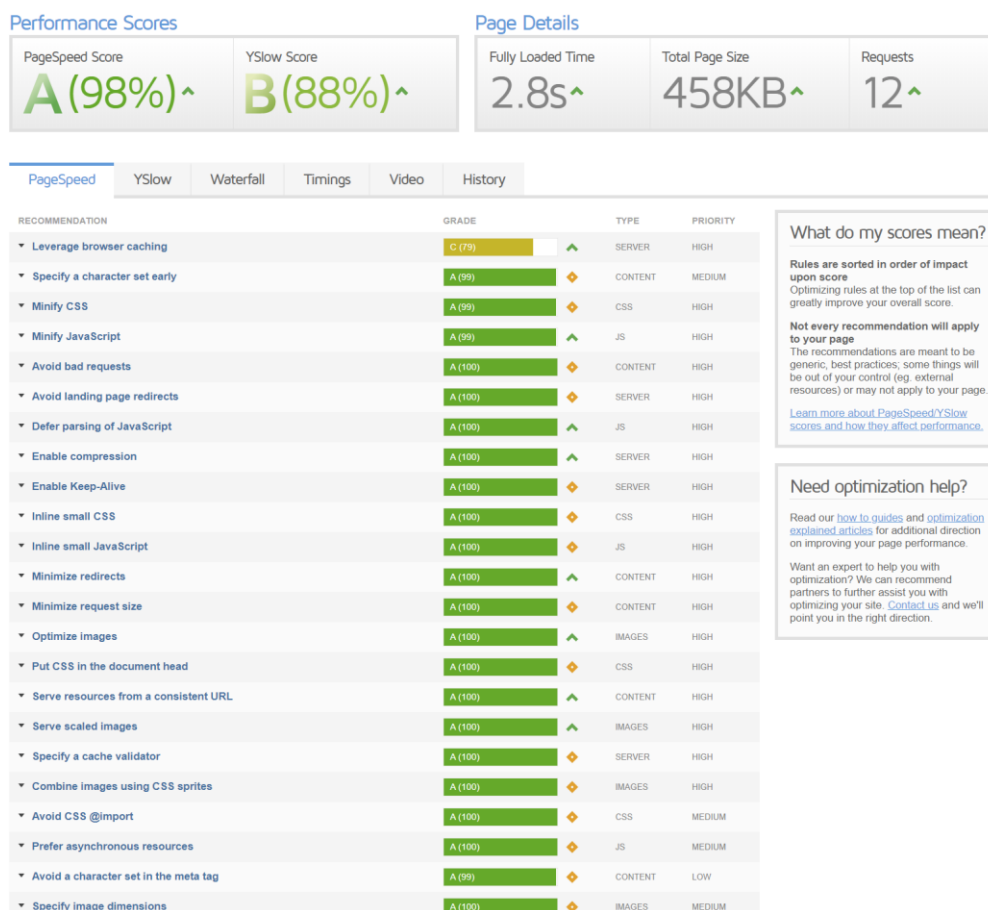
paigutatud⁵⁶. Ekstraheerimiskoodi saab optimeerida, kuid arvestades, et hetkel on pudelikaelaks (ingl *bottleneck*) päringute vaheline viisakuspaus, pole selleks vajadust veel tekkinud.

9.3 Veebilehe jõudlus

Mõõdeti veebilehe vaadete lahti laadimise kiirusi – võeti 5 mõõtmistulemust, nupuvajutusest vaate täieliku lahti laadimiseni. Järgmisena on välja toodud 5 korra keskmine, iga kord leht värskelt lahti laetud:

- SSD nimekiri umbes 5 ms
- SSD tabel, 24.6 ms
- Muutmälu tabel 27.2 ms
- Muutmälu nimekiri 7.6 ms

Leht pandi katsetamiseks avalikult ülesse, et selle jõudlust saaks GTmetrix automatiseeritud tööriistaga mõõta. Veebilehe serverimiseks kasutati Nginx⁵⁷ serveritarkvara. Selle tulemus on joonisel 21.



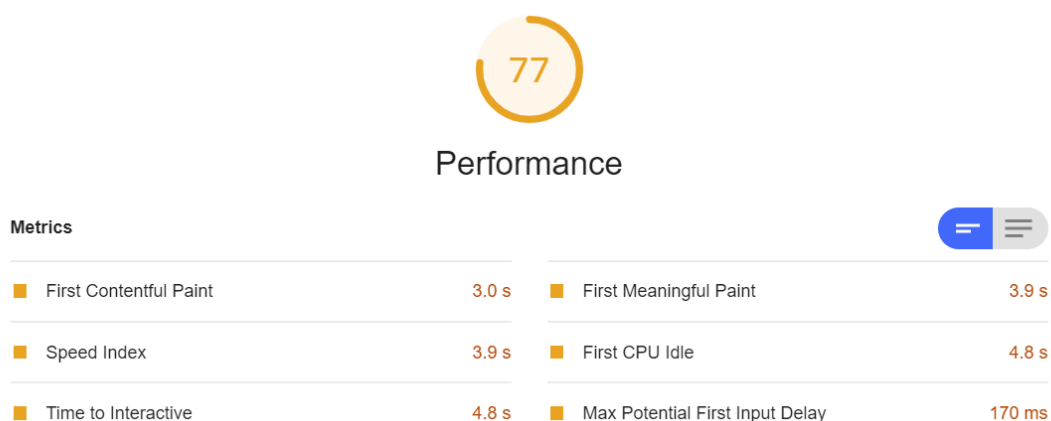
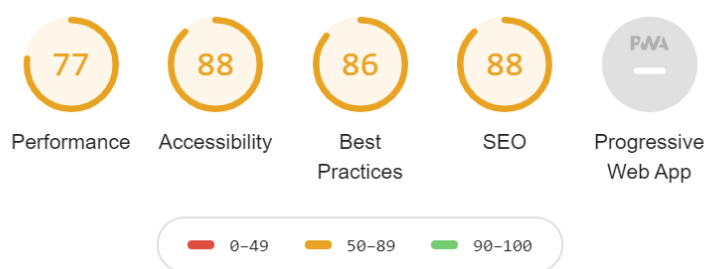
Joonis 21. GTmetrix automatiseeritud tööriista hinnang veebiliidese kohta⁵⁸.

⁵⁶ <https://www.kingston.com/en/ssd/dc1000b-data-center-boot-ssd#specifications>

⁵⁷ <https://www.nginx.com/>

⁵⁸ <https://gtmetrix.com/>

Joonisel 22 on välja toodud Google Chrome brauseris sisseehitatud Lighthouse tööriista tulemus.



Joonis 22. Google Chrome'i sisseehitatud Lighthouse tööriista tagasiside⁵⁹.

Seejuures tuleb meeles pidada, et esimese tööriista päringuid teostati Kanadast ja teine tööriist seadis maksimaalseks kiiruseks 1.64Mbps ja vähendas arvuti jõudlust neli korda. Kõige odavam kättesaadav pakett Telias lubab kiiruseks 10Mbps⁶⁰.

Kokkuvõttes võib järeldada, et optimeerimiseks on ruumi küll, aga keskmise kasutaja jaoks on veebileides kiiresti reageeriv ja andmed laetakse alla vähem kui sekundiga ka väga suure koormuse all. Järgmises alampeatükis kirjutatakse nõuetele vastavusest.

9.4 Nõuetele vastavus

Kõik funktsionaalsed nõuded said täidetud (vt lisa 3). Mittefunktsionaalsed nõuded on enamjaolt täidetud – komponendid on dünaamiliselt kohalduvad, kuid mobiilseadmetega ei ole eraldi arvestatud, seetõttu filtrid ei mahu ära ja kujundus on paigast ära. Kasutajasõbralikkust saab parandada sellega, et jätta kasutaja filtrite olek ja vaatevalik meelde, samuti võib võimaldada navigeerida otse soovitud arvutikomponendi leheküljele läbi veebilehekülje aadressi. Vaate vahetamine toimub üldjuhul viitega vähem kui 30 ms ja realistliku koormuse puhul keskmise internetiga saab mitu tuhat pakkumist alla laetud alla vähem kui poole sekundiga.

⁵⁹ <https://developers.google.com/web/tools/lighthouse>

⁶⁰ <https://www.telia.ee/era/koduinternet>

10. Kokkuvõte

Lõputöö eesmärk oli luua tarkvaralahendus, mis võimaldab kiirelt ja mugavalt leida parimaid muutmälu ja SSD pakkumisi. Analüüsi Eesti hinnavõrdlusportaalide võimekust mugavalt otsida parimat pakkumist toodete puhul, kus on palju tehnilisi parameetreid. Selle põhjal koostati tarkvaralahenduse nõuded ning selgitati, millised võivad olla autoriõigusega seonduvad küsimused nii hariduslikel kui ka kommertseesmärkide puhul. Vaadeldi erinevaid informatsiooni otsimis- ja kogumisviise, põhjendati nende valikut ning kirjutati veebiroomamise eeskirjadest tarkvaralahenduse kontekstis. Seejärel kirjeldati tarkvaralahenduse eraldiseisvate osade üldist struktuuri. Samuti kirjutati üksikasjalikult sellest, milline on andmekoguja informatsioonikogumise voog ja kuidas kõik tarkvaralahenduse komponendid omavahel suhtlevad. Põhjendati raamistike, keelte ja teekide valikut. Lõpuks jõuti järeldusele, et püstitatud eesmärgid said saavutatud ehk lõputöö raames sai loodud tarkvaralahendus, mis võimaldab muutmälu ja SSD-ketaste puhul mugavalt leida parimat toodet.

Tulevikus lisatakse võimekus otsida ülejäänud arvutikomponentide (kõvakettad, emaplaadid, protsessorid jne) parimaid pakkumisi. Siis saab välja arvutada terve arvuti hinnangulist jõudlust ja anda kasutajale võimaluse (sarnaselt SSD jõudlusega) valida, mis komponentide alusel seda hinnatakse. Samuti alustatakse pakkumiste kohta informatsiooni kogumist rohkematest e-poodidest.

11. Viidatud kirjandus

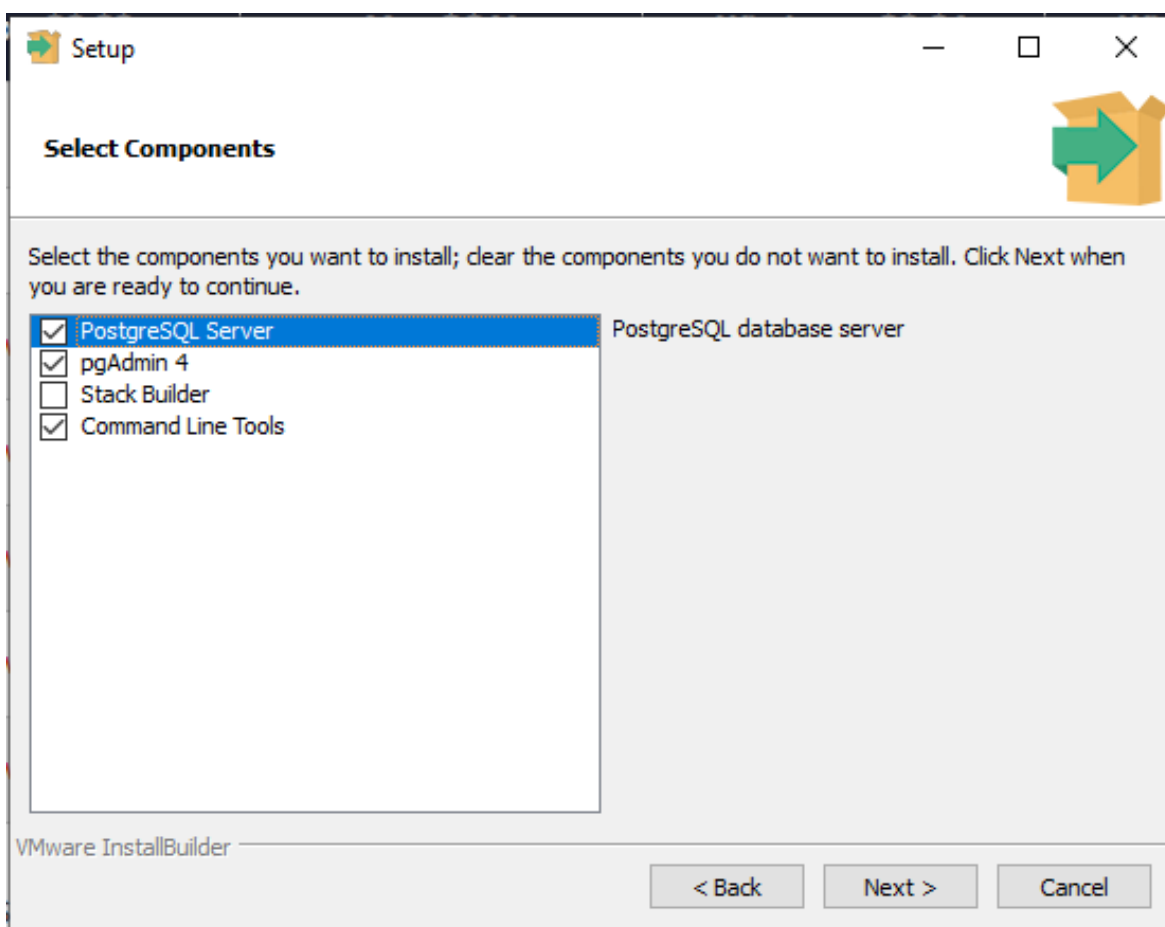
- [1] Uku Saarm. Veebiroomaja rakendus vastavalt meediaagentuuri vajadustele: baka-laureusetöö. Tallinna Ülikool, Tallinn, 2015.
- [2] Andmekaitse ja infoturbe leksikon. <https://akit.cyber.ee/> (15.04.2020)
- [3] Chandi Saini and Vinay Arora. "Information retrieval in web crawling: A survey", 2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), pp 2635-2643.
- [4] Zachary Gold and Mark Latonero. "Robots Welcome: Ethical and Legal Considerations for Web Crawling and Scraping", Washington Journal of Law, Technology & Arts, Vol. 13, Issue 3 (Spring 2018), pp. 275-312.
- [5] C. Castillo (2014). *Effective Web Crawling*. http://chato.cl/papers/crawling_thesis/effective_web_crawling.pdf 08.12.2019
- [6] Junghoo Cho and Hector Garcia-Molina. Effective page refresh policies for web crawlers. ACM Transactions on Database Systems, 28(4), December 2003.
- [7] M. Koster (1995). Robots in the web: threat or treat? ConneXions, 9(4).
- [8] Requests: HTTP for Humans dokumentatsioon. <https://requests.readthedocs.io/en/master/> (15.04.2020)
- [9] Selenium teegi dokumentatsioon. <https://selenium-python.readthedocs.io/> (15.04.2020)
- [10] Teegi Beautiful Soup dokumentatsioon. <https://www.crummy.com/software/BeautifulSoup/bs4/doc/> (14.04.2020)
- [11] Google Vision API pricing. <https://cloud.google.com/vision/pricing> (15.04.2020)
- [12] Node.js raamistiku tutvustus. <https://nodejs.org/en/about/> (18.04.2020)
- [13] What is npm? <https://nodejs.org/en/knowledge/getting-started/npm/what-is-npm/> (18.04.2020)
- [14] Angular'i põhitõed. <https://angular.io/guide/architecture> (18.04.2020)
- [15] André Gardi. How (and why) you should use TypeScript with Node and Express <https://medium.com/javascript-in-plain-english/typescript-with-node-and-express-js-why-when-and-how-eb6bc73edd5d> (18.04.2020)
- [16] Mosh Hamedani. React vs Angular. <https://programmingwithmosh.com/react/react-vs-angular/> (18.04.2020)

Lisad

I. Tarkvaralahenduse paigaldamine, seadistamine ja käivitamine operatsioonisüsteemis Windows 10

a. Andmebaasi haldamise süsteem

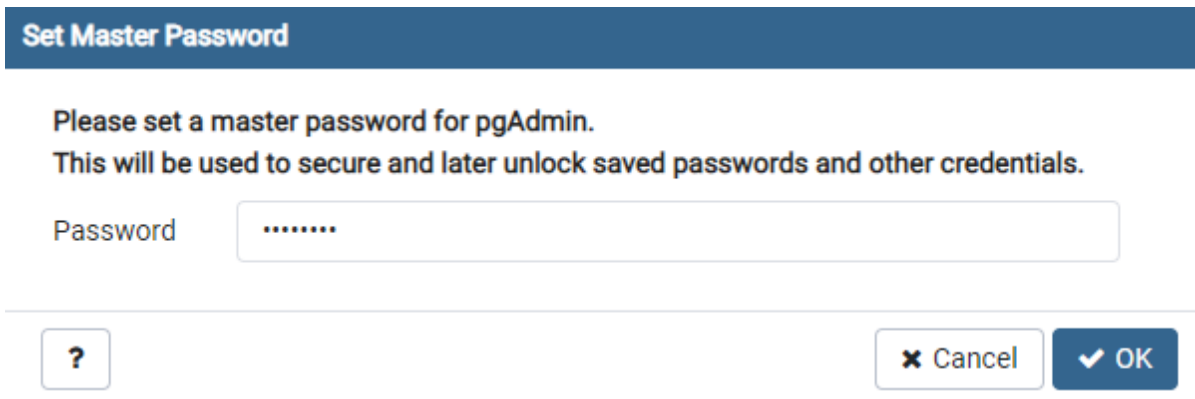
Alustuseks tuleb alla tõmmata ja paigaldada Postgresql version 12.2⁶¹. Paigaldamisel on vajalik komponentide valimisel jätta linnukesed kastidesse PostgreSQL Server, Command Line Tools ja pgAdmin 4, millest viimane on soovituslik, sest lihtsustab uue andmebaasi loomist ja võimaldab edasiste sammude järgimist (vt Joonis 23).



Joonis 23. PostgreSQL komponentide valik.

PgAdmin käivitamisel lülitatakse sisse vajalik teenus ning avaneb uus aken, kus esimesel külalustel tuleb valida üldparool (vt Joonis 24).

⁶¹ <https://www.enterprisedb.com/downloads/postgres-postgresql-downloads>



Joonis 24. PostgreSQL üldparooli valimine.

Seejärel vajutage vasakus servas oleva *Servers* peale, sisestage vajadusel paigaldamise käigus valitud parool ning parema klõpsuga *Databases* peale vajutades valige *Create > Database*. Lisaks valitud andmebaasi nime ja kasutaja meelejätmisele veenduge, et *Definition* all on kodeeringuks *UTF8* ning siis vajuage salvesta. Kui andmebaas on loodud, vajutage sellele parema klõpsuga ja valige *Query Tool...* ning jooksutage ilmuvas tööriistas järgmine käsk asendades *ANDMEBAASI_NIMI* enda andmebaasi nimega: *ALTER DATABASE "ANDMEBAASI_NIMI" SET timezone TO 'Europe/Tallinn';*.

b. Andmekoguja

i. Python'i ja selle teekide paigaldamine

Kõigepealt on vaja tõmmata alla Python versioon 3.8.2⁶². Eelistatavalt 64-bitine variant, sest vastasel juhul võib andmekogujale ligipääsetav muutmälu kogusest puudu jääda. Paigaldamisel on oluline panna linnukene kasti *Add Python 3.8 to PATH*, et hiljem oleks mugavam jooksutada skripti ja lisada vajalikud teegid (vt Joonis 25).

⁶² <https://www.python.org/ftp/python/3.8.2/python-3.8.2-amd64.exe>



Joonis 25. Python'i paigaldamine.

Seejärel tuleb avada Windows'i käsuviip ja paigaldada järgnevad teegid käsuga kujul *pip install Teegi-nimi==versiooninumber*:

- Brotli 1.0.7
- Beautifulsoup4 4.8.1
- Google-api-core 1.16.0
- Google-api-python-client 1.8.0
- Google-auth 1.13.1
- Google-auth-httplib2 0.0.3
- Google-auth-oauthlib 0.4.1
- Lxml 4.5.0
- Psycopg2-binary 2.8.5
- Requests 2.23.0
- Requests-futures 1.0.0
- Selenium 3.141.0

ii. Andmebaasi ühenduse seadistamine

Kui teegid on paigaldatud, siis andmetekoguja *database* kausta *database_operations.py* failis otsige ülesse *connect_to_db* funktsioon ja muutke ära andmebaasi nimi, kasutaja, parool ja port vastavalt sellele, andmebaasi haldamise süsteemi peatükis seadistatud.

iii. Selenium-nimelise teegi sõltuvuste paigaldamine

Selleks, et teek Selenium saaks toimida, installige veebilehitseja Firefox⁶³ või veenduge, et olemasolev oleks vähemalt versioon 60 ning tõmmake alla Geckodriver⁶⁴ versioon 0.26.0. Geckodriver omakorda nõuab, et operatsioonisüsteemil oleks installitud *Microsoft Visual C++ Redistributable for Visual Studio 2015, 2017 and 2019*⁶⁵. Peale seda pakkige Geckodriver'i arhiiv meelepärasesse kohta ja lisage EXE fail Windows'i Path'i järgmiselt:

1. Kirjutage Windows'i sisseehitatud otsingusse *Path*
2. Valige otsingust Edit the system environment variables
3. Klõpsutage nupul *Environment Variables...*
4. *System* -või *User variables* all selekteerige Path ja klõpsutage nupul *Edit...*
5. Klõpsutage nupul *New*
6. Lisage geckodriver.exe asukoht koos failinime ja laiendiga
7. Klõpsutage nupule *OK*

Samuti on vajalik ära muuta andmekoguja peamises kaustas *BrandProductFetcherAbstractClass.py* failis oleva klassi muutuja *geckodriver_dir* väärtust vastavaks sellele, kuhu sai Geckodriver lahti pakitud.

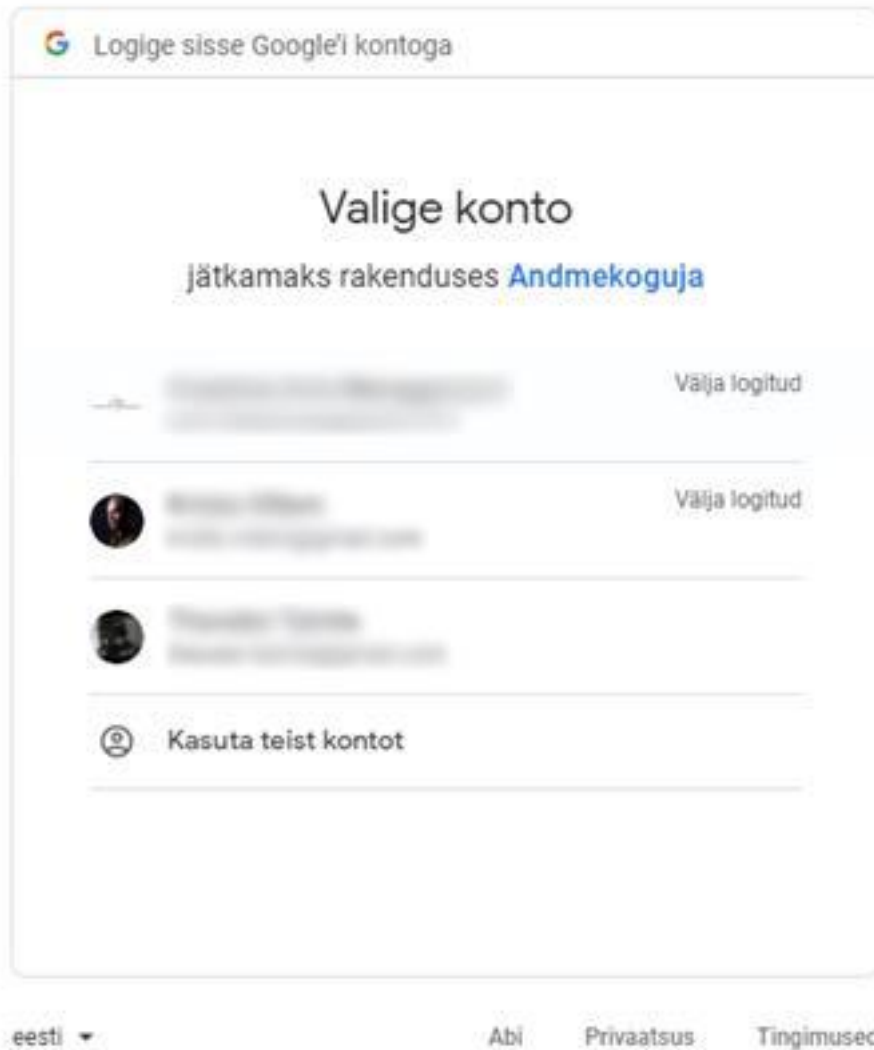
iv. Optilise tekstivastuse lubamine

Main.py faili esmakäivitamisel avaneb brauseris leht, kus palutakse Andmekoguja rakenduses jätkamiseks sisse logida (vt Joonis 26).

⁶³ <https://www.mozilla.org/et/firefox/download/thanks/>

⁶⁴ <https://github.com/mozilla/geckodriver/releases/download/v0.26.0/geckodriver-v0.26.0-win64.zip>

⁶⁵ https://aka.ms/vs/16/release/vc_redist.x64.exe



Joonis 26. Google konto valimine.

Peale konto valimist ja sisselogimist, palutakse kinnitada, et arendaja ning rakendus on kasutaja meelest usaldusväärsed. Seda saab teha klõpsutades nupule *Advanced* ja seejärel lingile, mis on joonisel 27 märgitud kollasega.



This app isn't verified

This app hasn't been verified by Google yet. Only proceed if you know and trust the developer.

If you're the developer, submit a verification request to remove this screen. [Learn more](#)

[Hide Advanced](#)

BACK TO SAFETY

Google hasn't reviewed this app yet and can't confirm it's authentic. Unverified apps may pose a threat to your personal data. [Learn more](#)

[Go to Andmekoguja \(unsafe\)](#)

Joonis 27. Kinnitamata rakenduse usaldamine.

Järgmisel kuval palub rakendus erinevaid õigusi (vt Joonis 28), mis võimaldavad tal teostada optilist tekstituvastust Google Drive'i faile manipuleerides. Rakendus saab õiguste kättesaamisel ligipääsu kõikidele failidele ning seetõttu on soovituslik tekstituvastuse tarvis kasutada tagavara kontot. Kui õigused on antud, siis joonisel 29 on vaade, kus palutakse veelkord kinnitust.

Grant Andmekoguja permission



See, edit, create, and delete
all of your Google Drive files



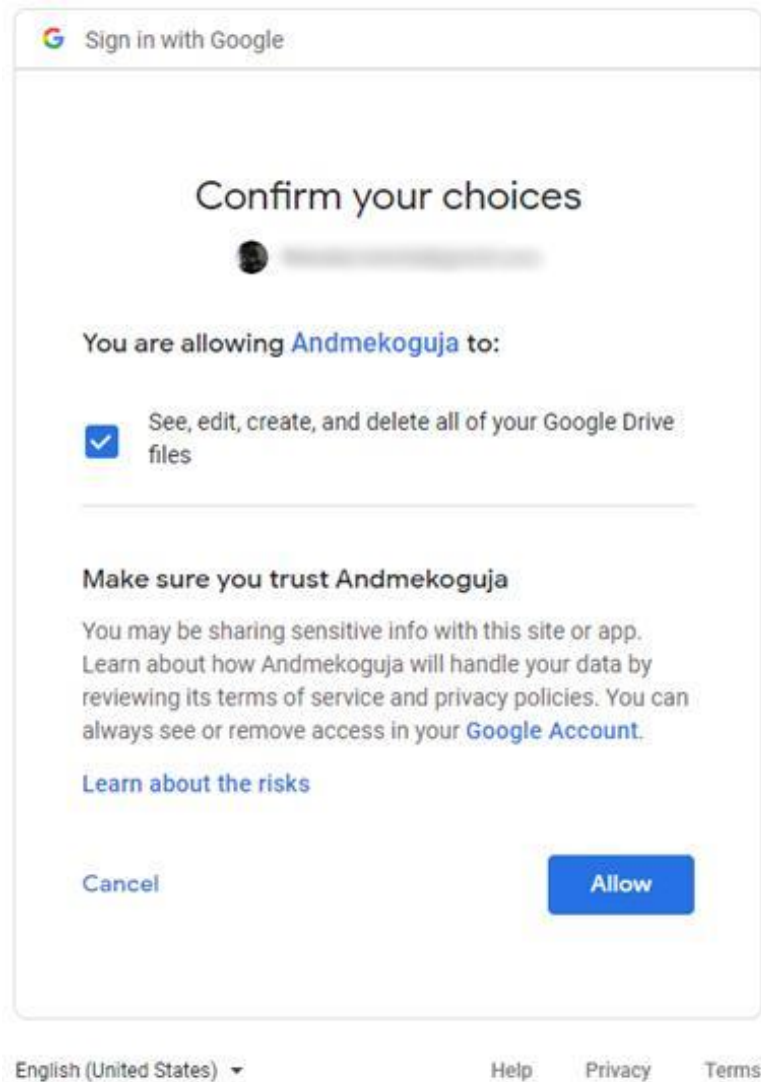
This app wants permission to access everything in your Google Drive. It will be able to do the same things you can do, including:

- See your files
- Upload and download your files
- Delete your files
- See the names and emails of people you share files with
- Share and stop sharing your files with others
- Remove people from your files
- Organize your Drive

There may be private information in your Google Drive, like financial records, medical reports, photos or tax info.

[Deny](#) [Allow](#)

Joonis 28. Google Drive haldamise lubamine.



Joonis 29. Haldamise loa kinnitamine.

c. Veebiserver

i. Java JDK paigaldamine ja seadistamine

Tõmmake alla Java JDK⁶⁶ versioon 14, pakkige lahti meelepärasesse kohta ning lisage Windows'i path'i järgmiselt:

1. Kirjutage Windows'i sisseehitatud otsingusse *Path*
2. Valige otsingust Edit the system environment variables
3. Klõpsutage nupul *Environment Variables...*
4. *System* all klõpsutage nupul *New...*

⁶⁶ https://download.java.net/java/GA/jdk14/076bab302c7b4508975440c56f6cc26a/36/GPL/openjdk-14_windows-x64_bin.zip

5. *System* all selekteerige Path ja klõpsutage nupul *Edit...*
6. Klõpsutage nupul *New...*
7. *Variable name* väärtuseks kirjutage *JAVA_HOME* ning väärtuseks Java JDK lahtipakitud kausta asukoht (Kaust, milles asub kaust nimega *bin*), seejärel klõpsutage nupule *OK*
8. Analoogselt sammudele peatükis “Selenium-nimelise teegi sõltuvuste paigaldamine”, lisage *geckodriver.exe* asukoht koos failinime ja laiendiga.

ii. Andmebaasi ühenduse seadistamine

Veebiserveri *src/main/resources* kataloogis asuva *application.yml* failis tuleb ära muuta *url*, *username* ja *password* vastavalt sellele, mis sai peatükis 10.1 seadistatud. *Url*'i muutuja peab olema kujul *jdbc:postgresql://localhost:PORT/ANDMEBAASI_NIMI*.

iii. Käivitamine

Veebiserveri saab käivitada sisestades juurkataloogis käsku *gradlew run*. See käsk tõmbab alla vajalikud sõltuvused ja käivitab teenuse.

d. Veebiliides

i. Sõltuvuste paigaldamine ja seadistamine

Tõmmake alla ja paigaldage muutmata seadistusega Node.js⁶⁷ versioon 12.16.2 – nii lisatakse vajalik kataloog Path'i automaatselt ja paigaldamisele jääb ka Npm-nimeline pakihaldur. Seejärel navigeerige uues käsuviiba aknas front-end'i juurkataloogi ja sõltuvuste paigaldamiseks sisestage käsk *npm install*.

Soovi korral võib *environments* kaustas *environment.ts* failis ära vahetada *apiUrl* väärtuse aadressi vastu, mis on kommentaaris. Sellisel juhul võetakse andmed mujalt veebiserverist.

ii. Käivitamine

Endiselt juurkataloogist sisestage käsk *npm run ng serve*. Käsu tagajärjel käivitatakse lokaalne veebiserver, millele pääseb ligi brauserist aadressil *localhost:4200*.

⁶⁷ <https://nodejs.org/dist/v12.16.2/node-v12.16.2-x64.msi>

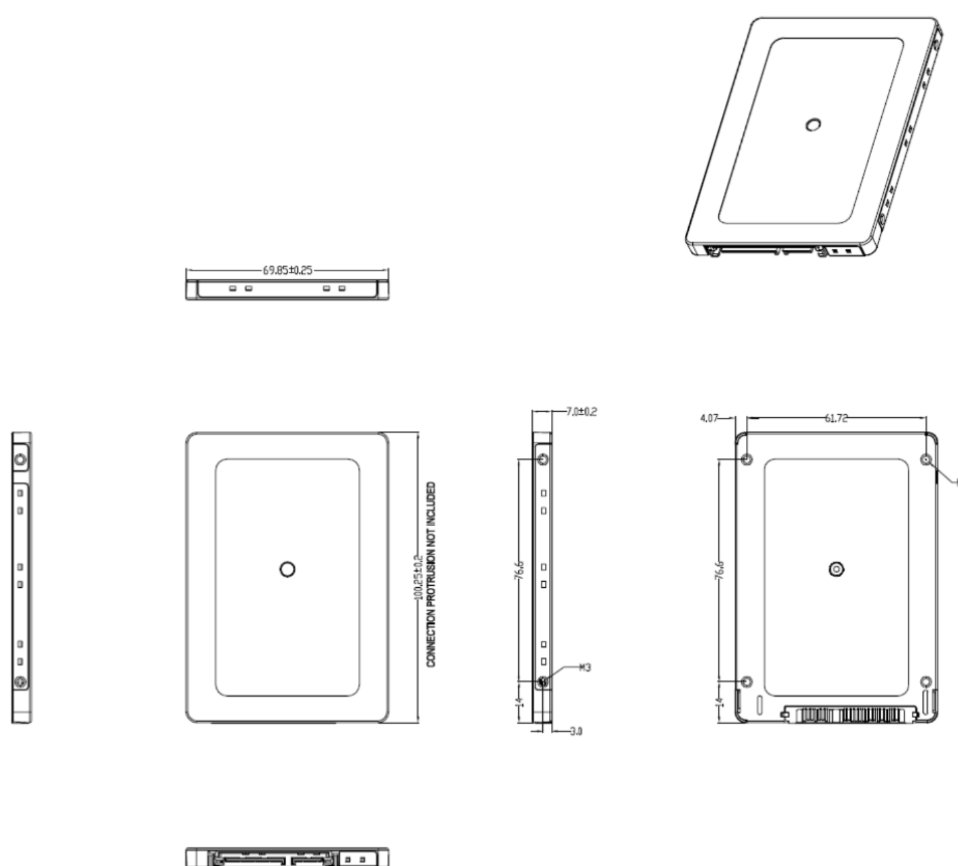
II. Google Drive API ja Tesseract tekstituvastuse võrdlus

Performance

Capacity	ATTO		CDM-QD32T1		AS SSD		4K Random		TBW
	Seq. Read (MB/s)	Seq. Write (MB/s)	Seq. Read (MB/s)	Seq. Write (MB/s)	Seq. Read (MB/s)	Seq. Write (MB/s)	Read (IOPS)	Write (IOPS)	
256GB	Up to 550	Up to 520	Up to 550	Up to 500	Up to 500	Up to 450	65K	75K	200TB
512GB	Up to 550	Up to 520	Up to 550	Up to 500	Up to 500	Up to 450	65K	75K	400TB
1TGB	Up to 550	Up to 520	Up to 550	Up to 500	Up to 500	Up to 450	65K	75K	800TB

* Performance may vary based on SSD capacity, hardware test platform, test software, operating system, and other system variables

Schematics



Joonis 30. ADATA SU750 SSD spetsifikatsioonifaili teine lehekülj⁶⁸.

⁶⁸ http://www.adata.com/upload/downloadfile/Datasheet_SU750_202003.pdf

e Dimensions (L x W x H): 100.45 x 69.85 x 7mm e Certifications: RoHS, CE, FCC, BSMI, VCCI, KC
o Weight: 47.5g e Warranty: 3 years

▲Performance

ATTO CDM-QD32T1

Capacity goq Read Seq.Write Seq.Read

(MB/s) (MB/s) (MB/s)

256GB Upto550 Upto520 Upto550

512GB Upto550 Upto520 Upto550

1TGB Upto550 Upto520 Upto550

* Performance may vary based on SSD capacity, hardware test platform, test software, operating system, and other system variables

Schematics

69.8580.25

100258 02-

CCONNECTION PROTRUSION NOT INCLUDED

Seq. Write
(MB/s)

Up to 500

Up to 500

Up to 500

AS SSD

Seq.Read Seq. Write
(MB/s) (MB/s)

Up to 500 Up to 450

Up to 500 Up to 450

Up to 500 Up to 450

4K Random
Read Write
(Iops) (10PS)

65K 75K

65K 75K

65K 75K

TBW

20078

400TB

800TB

▲

Joonis 31. Tesseract OCR osaline väljund TIFF-failist, mis teisendati Adata SU750 SSD spetsifikatsioonifailist.

Features Ordering Information

- 3D NAND Flash for higher affordability Capacity Model Number EAN Code
- Advanced hardware LDPC ECC technology 256GB ASU750SS-256GT-C 4710273770666
- Supports SLC Caching and DRAM Cache Buffer 512GB ASU750SS-512GT-C 4710273770673
- Free software: SSD Toolbox and Migration Utility 1TB ASU750SS-1TT-C 4710273770680
- Shock, vibration resistant (1500G/0.5ms)

Specifications

- Capacity: 256GB / 512GB / 1TB • Operating Temperature: 0°C-70°C
- NAND Flash: 3D NAND • Storage Temperature: -40°C-85°C
- Interface: SATA 6Gb/s (SATA III) • Shock Resistance: 1500G/0.5ms
- Form Factor: 2.5" • MTBF: 2,000,000 hours
- Dimensions (L x W x H): 100.45 x 69.85 x 7mm • Certifications: RoHS, CE, FCC, BSMI, VCCI, KC
- Weight: 47.5g • Warranty: 3 years

Performance

Capacity

ATTO CDM-QD32T1 AS SSD 4K Random

Seq. Read

TBW (MB/s)

Write

(IOPS)

256GB Up to 550 Up to 520 Up to 550 Up to 500 Up to 500 Up to 450 65K 75K 200TB

512GB Up to 550 Up to 520 Up to 550 Up to 500 Up to 500 Up to 450 65K 75K 400TB

1TGB Up to 550 Up to 520 Up to 550 Up to 500 Up to 500 Up to 450 65K 75K 800TB

* Performance may vary based on SSD capacity, hardware test platform, test software, operating system, and other system variables

Schematics

Seq. Write

Seq. Read

Seq. Write

Seq. Read

Seq. Write

Read

(MB/s)

(MB/s)

(MB/s)

(MB/s)

(MB/s)

(IOPS)

Joonis 32. Google Drive API osaline väljund Adata SU750 SSD spetsifikatsiooni PDF-faili teisendamisel Google dokumendiks ja sealt .txt formaati.

- Shock, vibration resistant (1500G/0.5ms)

Model Number ASU750SS-256GT-C ASU750SS-512GT-C ASU750SS-1TT-C
512GB
4710273770673
1TB
4710273770680
Specifications

- Capacity: 256GB / 512GB / 1TB
- NAND Flash: 3D NAND
- Interface: SATA 6Gb/s (SATA III)
- Form Factor: 2.5"
- Dimensions (L x W x H): 100.45 x 69.85 x 7mm
- Weight: 47.5g
- Operating Temperature: 0°C-70°C
- Storage Temperature: -40°C-85°C
- Shock Resistance: 1500G/0.5ms
- MTBF: 2,000,000 hours
- Certifications: RoHS, CE, FCC, BSMI, VCCI, KC
- Warranty: 3 years

Performance
ATTO
CDM-QD32T1
AS SSD
4K Random
Capacity
TBW
Seq. Read
(MB/s) Up to 550
Seq. Write
(MB/s) Up to 520
Seq. Read
(MB/s) Up to 550
Seq. Write
(MB/s) Up to 500
Seq. Read
(MB/s) Up to 500
Seq. Write
(MB/s) Up to 450
Read (IOPS) 65K
Write (IOPS) 75K
256GB |
200TB
512GB
Up to 550
Up to 520
Up to 550
Up to 500
Up to 500
Up to 450
65K
75K
400TB
1TGB
Up to 550
Up to 520
Up to 550
Up to 500
Up to 500
Up to 450
65K
75K
800TB

* Performance may vary based on SSD capacity, hardware test platform, test software, operating system, and other system variables

Joonis 33. Google Drive API tekstiks teisunduse osaline väljund PDF-faili puhul, kus kogu tekst on pildi kujul.

III. Valminud veebiliides

Ava filtrid		Tühista filtrid		Muutmälud		SSD-kettad	
PARIM		Hind	Jõudlus	Hinna-jõudluse suhe			
Maht: 8/8 GB	1600 MHz	19.9€	gskill				
DDR3	CL11	2.49€/GB	arvutitark				
SO-DIMM	1.35V		Saadavus: KOHAL				
Maht: 8/8 GB	1600 MHz	25.62€	gskill				
DDR3	CL11	3.2€/GB	Müüja: 1a				
SO-DIMM	1.35V		18. mai				
Maht: 8/8 GB	1600 MHz	26.3€	gskill				
DDR3	CL11	3.29€/GB	galador				
SO-DIMM	1.35V		Tarneaeg 3-5 tp				
Maht: 8/16 GB	1600 MHz	52.9€	gskill				
DDR3	CL11	3.31€/GB	galador				
UDIMM	1.5V		Otsas				
Maht: 4/16 GB	2133 MHz	56.9€	kingston				
DDR4	CL14	3.56€/GB	Müüja: 1a				
UDIMM	1.2V		11. mai				
Maht: 8/8 GB	2666 MHz	29.9€	adata				
DDR4	CL19	3.74€/GB	arvutitark				
UDIMM	1.2V		Saadavus: KOHAL				
Maht: 16/64 GB	3200 MHz	239.9€	gskill				
DDR4	CL16	3.75€/GB	arvutitark				
UDIMM	1.35V		Saadavus: KOHAL				
Maht: 16/16 GB	3000 MHz	62.9€	gskill				
DDR4	CL16	3.93€/GB	arvutitark				
UDIMM	1.35V		Saadavus: KOHAL				
Maht: 16/16 GB	2666 MHz	63.9€	gskill				
DDR4	CL19	3.99€/GB	arvutitark				
UDIMM	1.2V		Saadavus: KOHAL				
Maht: 16/64 GB	3200 MHz	259€	gskill				
DDR4	CL16	4.05€/GB	arvutitark				

Joonis 34. Veebiliides muutmälu kategooria nimekirja vaates, filtriaken suletud. Sorteeritud mahu-hinna järgi.

Muutmälud SSD-kettad

Peida kallimad tooted

Peida filtrid Tühista filtrid

PARIM
Hind
Jõudlus
Hinna-jõudluse suhe

Tootja	Maht DIMM kohta	Maht kokku	Mälusid komplektis	Taktisagedus	DDR gen.	Tüüp	Pinge	Latentsus	Veatuvastus	Müüja	Tarne	Hind	Hind/GB
gskill	8GB	8GB	1tk	1600	3	SO-DIMM	1.35V	11	<input type="checkbox"/>	arvutitark	Saadavus: KOHAL	19.9€	2.49€
gskill	16GB	64GB	4tk	3200	4	UDIMM	1.35V	16	<input type="checkbox"/>	arvutitark	Saadavus: KOHAL	239.9€	3.75€
gskill	16GB	16GB	1tk	3000	4	UDIMM	1.35V	16	<input type="checkbox"/>	arvutitark	Saadavus: KOHAL	62.9€	3.93€
gskill	16GB	32GB	2tk	3200	4	UDIMM	1.35V	16	<input type="checkbox"/>	1a	21. mai	135.28€	4.23€
gskill	16GB	32GB	2tk	3000	4	UDIMM	1.35V	16	<input type="checkbox"/>	arvutitark	Saadavus: KOHAL	129.9€	4.06€
gskill	16GB	16GB	1tk	3200	4	UDIMM	1.35V	16	<input type="checkbox"/>	galador	Tarneaeag 5-8 tp	77.2€	4.83€
crucial	16GB	32GB	2tk	3000	4	UDIMM	1.35V	15	<input type="checkbox"/>	galador	Tarneaeag 1-3 tp	166.5€	5.2€
gskill	16GB	32GB	2tk	3600	4	UDIMM	1.35V	16	<input type="checkbox"/>	1a	20. mai	188.99€	5.91€
gskill	16GB	32GB	2tk	2666	4	UDIMM	1.2V	15	<input type="checkbox"/>	1a	20. mai	150€	4.69€
gskill	16GB	32GB	2tk	3000	4	SO-DIMM	1.2V	16	<input type="checkbox"/>	galador	Tarneaeag 5-8 tp	158.9€	4.97€
crucial	16GB	16GB	1tk	2666	4	UDIMM	1.2V	16	<input type="checkbox"/>	arvutitark	Saadavus: KOHAL	70.79€	4.42€
gskill	8GB	8GB	1tk	3000	4	SO-DIMM	1.2V	16	<input type="checkbox"/>	1a	4. juuni	39.89€	4.99€
gskill	8GB	16GB	2tk	2800	4	UDIMM	1.25V	15	<input type="checkbox"/>	1a	3. juuni	79.57€	4.97€
gskill	32GB	256GB	8tk	3200	4	UDIMM	1.35V	16	<input type="checkbox"/>	galador	Tarneaeag 5-8 tp	1380.8€	5.39€
gskill	8GB	32GB	4tk	3000	4	UDIMM	1.35V	16	<input type="checkbox"/>	1a	4. juuni	162.34€	5.07€
crucial	16GB	32GB	2tk	2666	4	UDIMM	1.2V	16	<input type="checkbox"/>	galador	Tarneaeag 1-3 tp	144.7€	4.52€
gskill	8GB	8GB	1tk	2800	4	UDIMM	1.2V	17	<input type="checkbox"/>	1a	4. juuni	35.82€	4.48€
gskill	16GB	32GB	2tk	2400	4	SO-DIMM	1.2V	16	<input type="checkbox"/>	1a	21. mai	132.7€	4.15€
gskill	16GB	32GB	2tk	3600	4	UDIMM	1.35V	18	<input type="checkbox"/>	galador	Tarneaeag 2 nädalat	177.2€	5.54€
gskill	8GB	16GB	2tk	3400	4	UDIMM	1.35V	16	<input type="checkbox"/>	1a	21. mai	94.57€	5.91€
gskill	16GB	32GB	2tk	3466	4	UDIMM	1.35V	16	<input type="checkbox"/>	1a	4. juuni	193.69€	6.05€
gskill	16GB	64GB	4tk	2400	4	UDIMM	1.2V	15	<input type="checkbox"/>	galador	Tarneaeag 6-9 tp	287.9€	4.5€

Joonis 35. Veebiliides muutmälu kategoorias, avatud filtriaknaga tabeli vaates. Tooted sorteeritud hinna-jõudluse suhte järgi.

Muutmälud SSD-kettad

Järgestamine viisil: Jõudluse alus, Lugemine ja kirjutamine, Lugemine ja kirjutamine

Peida kallimad tooted

Peida filtrid Tühista filtrid

tootja müüja maht vastupidavus eluiga kujutegur liides mälu arhitektuur

PARIM	Hind	Jõudlus	Hinna-jõudluse suhe	
Kujutegur: M.2 Mälutüüp: 3D NAND TLC Liides: PCIe 4.0	Maht: 1000 Vastupidavus: 1800 TBW Eluiga: 1.7 mil. tundi	Järj. kirj/lug: 5000/4400 MB/s Suv. kirj/lug: 2930/2930 MB/s	Tootja: adata Müüja: galador Tarneaeg 6-9 tp	265.2€ 0.2652€/GB
Kujutegur: M.2 Mälutüüp: 3D NAND TLC Liides: PCIe 4.0	Maht: 1000 Vastupidavus: 1800 TBW Eluiga: 1.7 mil. tundi	Järj. kirj/lug: 5000/4400 MB/s Suv. kirj/lug: 2930/2930 MB/s	Tootja: adata Müüja: arvutitark Tellitav 3-7 tp	237.92€ 0.2379€/GB
Kujutegur: M.2 Mälutüüp: 3D NAND TLC Liides: PCIe 4.0	Maht: 2000 Vastupidavus: 3600 TBW Eluiga: 1.7 mil. tundi	Järj. kirj/lug: 5000/4400 MB/s Suv. kirj/lug: 2930/2930 MB/s	Tootja: adata Müüja: galador Tarneaeg 6-9 tp	524.5€ 0.2623€/GB
Kujutegur: M.2 Mälutüüp: 3D NAND MLC Liides: PCIe 3.0	Maht: 1024 Vastupidavus: 1200 TBW Eluiga: 1.5 mil. tundi	Järj. kirj/lug: 3500/2700 MB/s Suv. kirj/lug: 1953/1953 MB/s	Tootja: samsung Müüja: arvutitark Tellitav 1-4 tp	360.16€ 0.3517€/GB
Kujutegur: M.2 Mälutüüp: 3D NAND MLC Liides: PCIe 3.0	Maht: 1024 Vastupidavus: 1200 TBW Eluiga: 1.5 mil. tundi	Järj. kirj/lug: 3500/2700 MB/s Suv. kirj/lug: 1953/1953 MB/s	Tootja: samsung Müüja: 1a 29. mai	366.58€ 0.358€/GB
Kujutegur: M.2 Mälutüüp: 3D NAND MLC Liides: PCIe 3.0	Maht: 1024 Vastupidavus: 1200 TBW Eluiga: 1.5 mil. tundi	Järj. kirj/lug: 3500/2700 MB/s Suv. kirj/lug: 1953/1953 MB/s	Tootja: samsung Müüja: galador Tarneaeg 1-3 tp	362.7€ 0.3542€/GB
Kujutegur: M.2 Mälutüüp: 3D NAND MLC Liides: PCIe 3.0	Maht: 1024 Vastupidavus: 1200 TBW Eluiga: 1.5 mil. tundi	Järj. kirj/lug: 3500/2700 MB/s Suv. kirj/lug: 1953/1953 MB/s	Tootja: samsung Müüja: arvutitark Tellitav 3-7 tp	361.28€ 0.3528€/GB
Kujutegur: M.2 Mälutüüp: 3D NAND TLC Liides: PCIe 3.0	Maht: 1000 Vastupidavus: 600 TBW Eluiga: 1.5 mil. tundi	Järj. kirj/lug: 3400/2500 MB/s Suv. kirj/lug: 1953/1953 MB/s	Tootja: samsung Müüja: 1a 11. mai	195€ 0.195€/GB
Kujutegur: M.2 Mälutüüp: 3D NAND TLC Liides: PCIe 3.0	Maht: 1000 Vastupidavus: 600 TBW Eluiga: 1.5 mil. tundi	Järj. kirj/lug: 3400/2500 MB/s Suv. kirj/lug: 1953/1953 MB/s	Tootja: samsung Müüja: galador Tarneaeg 2 nädalat	219.6€ 0.2196€/GB
Kujutegur: M.2 Mälutüüp: 3D NAND TLC Liides: PCIe 3.0	Maht: 1000 Vastupidavus: 600 TBW Eluiga: 1.5 mil. tundi	Järj. kirj/lug: 3400/2500 MB/s Suv. kirj/lug: 1953/1953 MB/s	Tootja: samsung Müüja: arvutitark Tellitav 1-4 tp	205.9€ 0.2059€/GB

Joonis 36. Veebiliides SSD kategoorias, nimekirja vaates avatud filtriga.

Muutmälud SSD-kettad

Järgestamine viisil: Jõudluse alus, Lugemine ja kirjutamine, Lugemine ja kirjutamine

Peida kallimad tooted

Peida filtrid Tühista filtrid

tootja müüja maht vastupidavus eluiga kujutegur liides mälu arhitektuur

PCIe 3.0
 PCIe 4.0
 SATA 6Gb/s
 U.2 PCIe 3.0

Tootja	Kujutegur	Liides	Mälutüüp	Järgestikune lugemine (MB/s)	Järgestikune kirjutamine (MB/s)	Juhuslik lugemine (MB/s)	Juhuslik kirjutamine (MB/s)	Maht	Eluiga (mil. tundi)	Vastupidavus	Müüja	Tarne	Hind	Hind/GB
kingston	M.2	SATA 6Gb/s	3D NAND TLC	520	500	309	98	240	1	100	arvutitark	Tellitav 1-4 tp	46.45€	0.1935€
kingston	M.2	SATA 6Gb/s	3D NAND TLC	520	500	309	98	240	1	100	galador	Tarneaeg 1-3 tp	47.8€	0.1992€
kingston	mSATA	SATA 6Gb/s	3D NAND TLC	520	500	309	98	240	1	100	galador	Tarneaeg 1-3 tp	47.8€	0.1992€
kingston	mSATA	SATA 6Gb/s	3D NAND TLC	520	500	309	98	240	1	100	arvutitark	Tellitav 1-4 tp	46.45€	0.1935€
adata	2.5"	SATA 6Gb/s	3D NAND QLC	520	450	156	254	480	1.5	100	arvutitark	Tellitav 1-3 tp	66.47€	0.1385€
adata	2.5"	SATA 6Gb/s	3D NAND QLC	520	450	156	254	480	1.5	100	galador	Tarneaeg 1-3 tp	66.7€	0.139€

Joonis 37. Veebiliides SSD kategooria tabeli vaates avatud filtriga. Üks filtreerimisparameeter on selekteerimisel, tooted sorteeritud jõudluse järgi.

IV. Litsents

Lihlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Theodor Taimla,

1. annan Tartu Ülikoolile tasuta loa (lihlitsentsi) minu loodud teose

Nutikas otsing parimate pakkumiste leidmiseks muutmälu ja SSD-ketta näitel,

mille juhendajad on Anne Villems ja Aleksei Kelli,

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, alates **08.05.2023** kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Theodor Taimla

08.05.2020