

TARTU ÜLIKOOL  
Arvutiteaduse instituut  
Informaatika õppekava

**Kristofer Lill**

# **Eesti keele kuulamismängud**

**Bakalaureusetöö (9 EAP)**

Juhendaja: Sven Aller

Tartu 2024

# **Eesti keele kuulamismängud**

## **Lühikokkuvõte:**

Käesoleva bakalaureusetöö eesmärk oli luua kuulamismängud, mis kasutavad eesti keele keeleressursse ja sobivad kuulamise harjutamiseks. Tulemusena sai loodud kaks mängu sõnade ja lausete kuulamiseks. Sõnad võeti Eesti Keele Instituudi õpetaja tööriistadest ja laused Eesti Wordnetist. Töö valmis veebisaidina, mis kasutab sõnade ja lausete hääldamiseks Neurokõne rakendusliidest. Töö esimeses pooles antakse ülevaade keeleressurssidest ja keelemängudest koos näidetega. Teises pooles kirjeldatakse andmete töötamise tulemust ja lõplikku mänguprogrammi. Lõpuks analüüsitakse saadud tagasisidet ja edasiarenduse võimalusi.

## **Võtmesõnad:**

Eesti keel, keeleressursid, Neurokõne

**CERCS:** P175 Informaatika, süsteemiteooria

# **Estonian Listening Games**

## **Abstract:**

The aim of this Bachelor's thesis was to create listening games that use Estonian language resources and are suitable for listening practice. Two games for listening to words and sentences were made. The words were taken from the Institute of the Estonian Language Teacher Tools and the sentences from Estonian Wordnet. The work was completed as a website that uses the Neurokõne API for pronouncing words and sentences. The first half of the thesis gives an overview of language resources and language games with examples. The second half describes the result of the data processing and the final game program. Finally, the feedback received and the possibilities for further development are analysed.

## **Keywords:**

Estonian language, language resources, Neurokõne

**CERCS:** P175 Informatics, systems theory

## Sisukord

Sissejuhatus.....	4
1. Keeleressursid .....	5
1.1 Leksikonid.....	5
1.2 Korpused .....	6
1.3 Keeletöötlusvahendid.....	7
2. Olemasolevad keelemängud .....	8
2.1 Keelemängude näited.....	8
2.1.1 Spellcheck .....	8
2.1.2 Sõnaveebi keelemängud.....	9
2.1.3 Strands.....	10
2.1.4 Eesti viipekeeke õppeprogramm.....	11
3. Valminud kuulamismängud .....	13
3.1 Programmide algoritmid .....	13
3.1.1 Eeltöötlus .....	13
3.1.2 Veebiliides .....	16
3.2 Mängimise kirjeldus.....	18
3.3 Tagasiside .....	21
3.4 Edasiarenduse võimalused .....	22
Kokkuvõte.....	24
Viidatud kirjandus.....	25
Lisad.....	27
I. Tagasiside küsimustik.....	27
II. Litsents.....	30

## Sissejuhatus

Keeleliste andmete kasutamine ühiskonna hüvanguks on aastate jooksul üha enam levinud. Selliseid andmeid kogutakse, et keeli paremini uurida ja mõista. Keeleandmed on kasulikud mitmetes valdkondades, näiteks hariduses ja masinõppes. Erinevate keeleressursside ulatus, kvaliteet ja kättesaadavus on varieeruv ning selleks on olulised koostöö- ja standardimisvõimalused. Keeleressursside loomine on eriti oluline väiksemate keelte jaoks, mis peavad kasutama kõiki võimalusi oma keele edendamiseks [1]. Keeltele põhinevad mängud on hea viis keeleõppe levitamiseks ja hõlbustamiseks [2]. Need annavad võimaluse keeltega tutvuda ja keeli harjutada.

Käesoleva rakendusliku bakalaureusetöö eesmärk on luua eesti keele keeleressurssidega veebipõhised keelemängud, mis aitavad harjutada eesti keele kuulamisoskust ja õigekirja. Tulemuseks saadud mängud oleksid kasulikud eesti keele algõppijatele ja keelehuvilistele. Kasutatakse andmed kogutakse Eesti Keele Instituudi õpetaja tööriistadest ja Eesti Wordnetist. Töös kirjeldatakse andmete töötlust ja lõplikku veebirakendust, milles rakendatakse Tartu NLP Neurokõne kõnesünteesi.

Töö peatükkides käsitletakse keeleressursse, olemasolevaid keelemänge ja valminud kuulamismänge. Esimeses peatükis antakse ülevaade keeleressurssidest ja tutvustatakse nende jagunemist leksikonideks, korpusteks ja töötlusvahenditeks. Teises peatükis käsitletakse mängude kui keeleõppe vahendite eeliseid ning tuuakse näiteid eesti ja teiste keelte keelemängudest. Kolmandas peatükis kirjeldatakse loodud kuulamismängude tegemisprotsessi. Selgitatakse eeltöötluste ja veebiliidese lahendusi ning kirjeldatakse mängude samm-sammulist läbimängimist. Lõpus analüüsitakse küsimustikuga saadud tagasisidet ja uuritakse edasiarenduse võimalusi. Kasutatud küsimustik on esitatud lisades.

## 1. Keeleressursid

Andmete kasutamine ühiskonna parendamiseks on tehnoloogia arenguga muutunud levinumaks. Tänapäeva infoajastul koguvad teadlased ja tehnoloogiahuvilised ühtviisi hoolikalt tohtuid andmeressursse. Nende ressursside hulgas on ka inimkeeltega seotud andmed, mille kogumise eesmärgiks on keelte lähedasem uurimine ja mõistmine.

Tavaliselt on keelelised ressursid salvestatud masinloetavas formaadis, mis hõlbustab nende analüüsi ja manipuleerimist erinevate arvutisüsteemide ja tarkvaraprogrammide abil. Need on kasulikud mitmesugustes valdkondades, näiteks loomuliku keele töötlus, masinõpe, tehisintellekt, haridus ja teadusuuringud. Nende mitmekülgset rakendused rõhutavad nende rolli eri valdkondade arengu edendamisel, alates õppematerjalide täiustamisest kuni keelemudelite väljatöötamiseni.

Keeleressursside alaga kaasnevad raskused ja kohustused. Kaalep jt [1] mainivad, et keeleressursside loomine ja nende kasutamine on ajaliselt ja rahaliselt kulukas ettevõtmine, aga siiski väga tähtis. Nad rõhutavad, et see on eriti tähtis väiksemate keelte jaoks, mis vajavad iga võimaluse haaramist keele edenemiseks. Samuti kurdavad nad, et erinevate osapoolte kokkupandud ressursse on raske või mõnikord isegi võimatu ühendada. Selliste põhjuste pärast on vajalik koostöö ja standardiseerimise võimalustega arvestada. Keeleressursside arhiveerimisel ja hooldamisel on oluline koostöö keeleteadlaste, arvutitehnoloogide, leksikograafide ja teiste spetsialistide vahel. Kooskõlastatud meeskonnatöö abil kasutavad erinevad eksperdid oma eriteadmisi ja ainulaadseid vaatenurki, et tagada selliste ressursside kvaliteeti ja asjakohasust. Erinevate keeleressursside ulatus, kvaliteet ja kättesaadavus on varieeruv. Näiteks mingi andmestik võib olla avalikult kättesaadav või hoopis kaitstud ja kättesaadav piiratud arvule isikutele.

Keeleressursse võib jaotada erinevatesse kategooriatesse. Selle töö raames põhineb jaotus Eesti Keeleressursside Keskuse rühmitusel [3]. Keeleressursid jagunevad korpusteks, sõnastikeks ja töötlusvahenditeks. Igäühel nendest on oluline roll keele mõistmisel ja kasutamisel.

### 1.1 Leksikonid

Leksikaalsed ressursid, tavakeeles sõnastikud, on keeleressursside jaoks põhiliseks osaks ja hõlmavad muuhulgas sõnaloendeid, näiteks sagedussõnastikke. Need on peamiselt seotud sõnavaraga. Sõnastikud täidavad üksikisikute igapäevaelus praktilist eesmärki, võimaldades hõlpsasti ligi pääseda kasulikule või huvitavale teabele. Kõige levinumad neist on sõnade mõisted ja tõlked erinevatesse keeltesse.

Digitaalajastul on sõnastikud läbinud mitmeid muutuseid, liikudes paberkandjatelt digitaalse-tesse vormingutesse. Need kaasaegsed versioonid pakuvad palju rohkemat kui ainult mõisteid ja tõlkeid. Kasutajatel on nüüd juurdepääs rohkematele võimalustele, näiteks sõnade häälduse kuulmine ja illustreerivate näidete, sünonüümide ning antonüümide nägemine. See areng suurendab nii keelte ligipääsetavust kui ka kasutatavust, edendades dünaamilisust ja hõlbustades õppimisprotsessi. L'Homme ja Cormier [4] kirjutavad digitaalsete leksikonide headest ja halbadest külgedest. Nad selgitavad kuidas digitaalses formaadis andmeid on lihtsam ja kiirem uuendada võrreldes sõnaraamatute väljapanekutega. Samuti juhivad nad tähelepanu andmehu piirangute praktilisele puudusele. Lisaks toovad nad välja arutelu, et tavakasutajatele võivad igasugused lisa andmed ja võimalused tekitada ülekoormatuse. Tänapäevaks on eesti keele jaoks tehtud kasulik digitaalse sõnastiku keskkond nimega Sõnaveeb<sup>1</sup>, mille eesmärk on kokku koguda erinevad eeltulnud sõnastikud ja kogumid standardiseerimiseks, et hõlbustada kasutusmugavust.

Leksikaalsete ressursside tähtsus ulatub kaugemale kui üksikud kasutajad, hõlmates valdkondi, kus sellised ressursid on osa suurematest rakendussüsteemidest [5]. Nendes valdkondades on leksikaalsed ressursid põhielemendid paljudeks ülesanneteks ja rakendusteks.

## 1.2 Korpused

Korpused on suured andmekogumid, mis koosnevad peamiselt tekstilisest või mõnikord ka audiitiivsest sisust. Hea ülevaate korpustest annab Kaalep [6] oma keeleressursside väljaandes. Ta selgitab kuidas tekstipõhised korpused koosnevad mitmesugustest terviklikest kogumikest, mis esindavad erinevaid valdkondi, žanreid, ajaperioode ja asukohti. Kogumikud omakorda koosnevad mitmetest miljonitest sõnadest. Korpuseid kasutatakse muuhulgas lingvistiliseks analüüsiks, et tuvastada ja märgendada sõnade või lausete vahelisi seoseid. Tulenevad märgendustega korpused on kasulikud mitmeteks eesmärkideks, näiteks õigekirja- ja grammatika-kontrollijateks. Lisaks on need keelelised ressursid olulised uute tekstide genereerimiseks. Märkimisväärsed tekstilised andmed pärinevad erinevatest meediumidest, näiteks raamatutest, ajakirjadest, ajalehtedest ja veebikeskkondadest.

Kõnekorpused on sarnased oma tekstipõhiste vastetega, kuigi neid andmeid salvestatakse teksti asemel helisalvestustena. Sellised helifailid, mis koosnevad peamiselt kõnest, võimaldavad heli ja vastava teksti sisu ühildamist. Nii lahendatakse ülesandeid nagu subtiitrite loomine ehk

---

<sup>1</sup> <https://sonaveeb.ee/>

transkriptsioon ja kõnelemine nutiseadmetega ehk kõnetuvastus. Lisaks sellele annab kõnede manipuleerimine võimaluse genereerida uut kõnekeelt meenutavat heli ehk kõnet sünteesida. Björkenstam kirjutab [7], et üldise kõnekorpusse loomiseks on vaja kaasata kõigi sugude ja vanuserühmade kõnet, milles vaadeldava piirkonna keelt räägitakse. Kõnekorpused jagunevad spontaanseks kõneks, näiteks kahe inimese vaheline dialoog, ja loetud kõneks, näiteks audio-raamat [8].

Teadlased kasutavad korpusid mitmesuguste keeleliste nähtuste uurimiseks. Uurida saab näiteks morfoloogiat, süntaksit, semantikat, pragmaatikat ja diskursust. Korpusete erinevate mustrite analüüsimine annab ülevaate keelekasutusest erinevates kontekstides [7]. See protsess parandab meie arusaamist ideedest ja tähelepanekutest keele kohta.

### 1.3 Keeletöötlusvahendid

Teadlased on välja töötanud mitmesuguseid vahendeid, et hõlbustada keeleliste andmete töötlemist ning suurendada sellise teabe käsitlemise tõhusust ja tulemuslikkust. Keeletöötlusvahendid hõlmavad laiu kasutusvõimalusi, näiteks õigekirjakontroll, sisu analüüs, semantiline märgendamine, info otsimine, keele tuvastamine, kokkuvõtete tegemine, lemmatiseerimine ja tõlkimine. Loomuliku keele töötluste (ingl NLP, *Natural language processing*) põhimõte on arvuti ja inimese vaheline informatsiooni vahetus inimkeeleliselt. Priya jt [9] toovad välja, et NLP rakendamine arvutitel on keeruline, sest inimkeel on hajane ja keerukas. Võib öelda, et arvutite ja inimeste suhtlusviisid on süstemaatiliselt erinevad.

Selle töö rakendusliku osa teostamisel on kasutatud erinevaid eesti keelega seonduvaid vahendeid. Üks selline vahend on EstNLP, mis on mitmekülgne loomuliku keele töötlemise teek (ingl *library*) programmeerimiskeelele Python. See lahendab ülesandeid nagu sõnade, lausete ja lõikude tokeniseerimine ehk arvutile töödeldavaks juppideks jaotamine ning morfoloogiline analüüs ehk sõnadele liikide ja käänete määramine [10]. EstNLP ühendab erinevaid tarkvarasid, nagu Eesti Wordnet, ja teeb need lihtsasti kättesaadavaks ühtse programmeerimisliidese kaudu [11]. Lisaks sellele on käesoleva töö koostamisel kasutatud Tartu NLP Neurokõnet kui kõnesünteesi mudelit [12], mis on võimeline genereerima teksti põhjal kõnet mitmete ettevalmistatud inimsarnaste hääldetega. Selline tarkvara võib aidata kaasa erinevatele juurdepääsetavuse lahendustele. Näiteks saab seda kasutada eestikeelsete uudistesisu kuulamiseks, nagu seda on tehtud Eesti Rahvusringhäälingu veebipõhise uudisteportaali artiklite puhul [13].

## **2. Olemasolevad keelemängud**

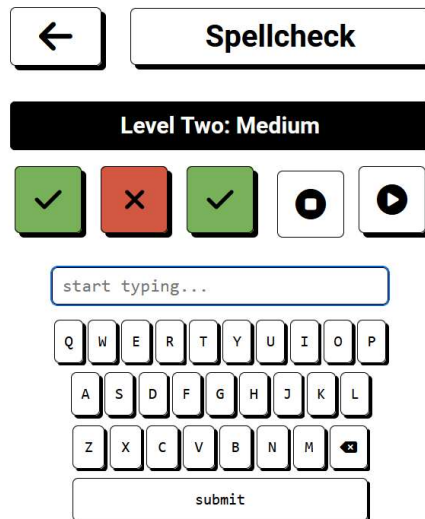
Mängulisuse lisamine igapäevastesse ülesannetesse on teatud viis suurendada keskendumist ja huvi, mis aitab olla tõhusam ja saavutada paremaid tulemusi. Sellest tulenevalt võib kaaluda ka võimalust mängude lõimimist keeleõppe metoodikasse, et kaasa aidata üldisele õppekogemusele. Raheem jt [14] mainivad oma artiklis, et sellised mängud pakuvad tavapärastest õpestiilidest vahelduseks teistsugust lähenemist, mis teevad õpingud nauditavamaks kogemuseks. Seda mõistes võib aru saada, et varieeruvus on kasulik lähenemine teadmiste täiustamiseks. Chen [2] kajastab keelemängude võimekust panna õppijad sügavalt mõtlema. Artiklis tuuakse välja mitmeid eeliseid. Mainitakse, et keelemängud panevad õppijad kesksesse rolli, soodustavad head koostöökeskkonda, ärgitavad loomingulisust, vähendavad õpiärevust ja aitavad kaasa sisulisemalt käsitletavatest teemadest aru saada. Nendest punktidest võib järeldada, et mängude kaasamine õppekeskkonda toob esile kaasahaarava meeole. Seega võivad keelemängud olla head ajaviitmise ja mõtlemapanevad tegevused ka keelehuvilistele või kõigile, kellele meeldib end proovile panna.

### **2.1 Keelemängude näited**

Mängud, mis on käesoleva töö raames nii mõjutajateks kui ka inspiratsiooniks, on head näited keelepõhistest mängudest digitaalses keskkonnas. Nendest mängudest ülevaate saamine aitab paremini mõista teemat tervikuna. Järgmised näited on ligipääsetavad veebilehtedelt.

#### **2.1.1 Spellcheck**

Sõnamängu „Spellcheck“ [15] eesmärk on proovile panna mängija kuulamise ja sõnateadmise oskust. Mäng koosneb ingliskeelsete sõnade kuulamisest ja nende samade sõnade kirja panemisest. Mängida on võimalik kahte erinevat varianti. Joonisel 1 on näha esimest varianti, milles vastusteks olevad sõnad vahetuvad igapäevalt. Päevas on kokku 15 sõna, mis on võrdselt jaotatud kolme raskusastmesse. Teises variandis kasutatakse harjutamiseks kõiki mängus võimalikke sõnu suvalises järjekorras. Selle variandi alguses tuleb valida sama kolme raskusastme vahel. Siinkohal jäävad kõik mängu kestel esinevad sõnad valitud raskusastmesse. Harjutamise variant kestab, kuni on pakutud viis valet vastust. Mõlema mänguvariandi lõpus on näha tulemust, kus esitatakse kõik mängija poolt pakutud õiged ja valed sõnad.



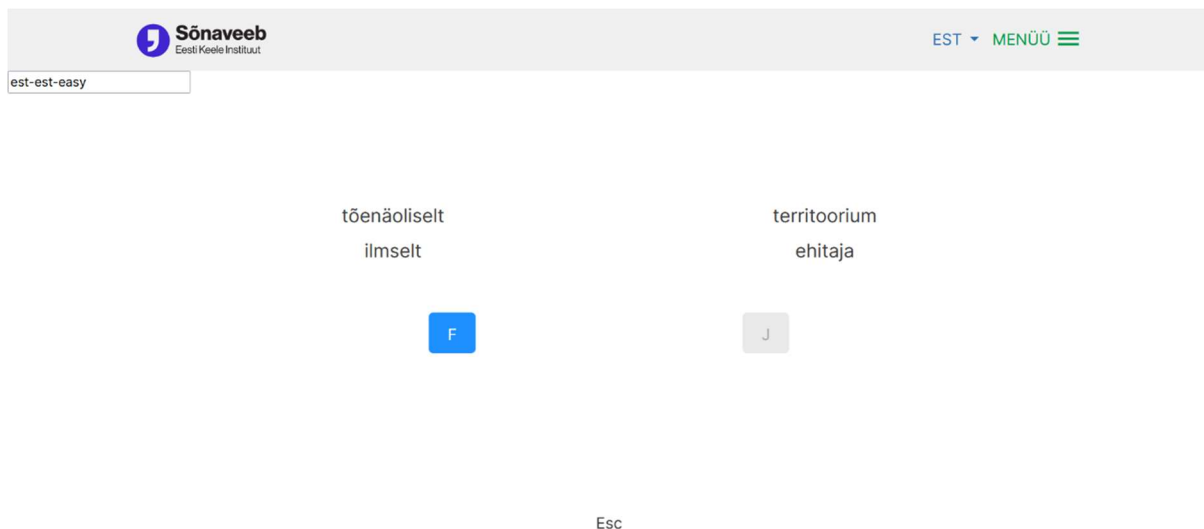
Joonis 1. Veebimängu „Spellcheck“ vaade.

Sõnu hääldab arvuti poolt genereeritud heli koos sõna tähenduse ja etümoloogiaga. Mängija kuulab sõna hääldust, vajutades ruudukujulist helinuppu, ja kirjutab kuuldud sõna välja, kasutades kas füüsilist või ekraanil olevat klaviatuuri.

Harjutamise variant sobib ka selleks, kui on soovi mängida rohkem kui igapäevast 15 vooru. Kuna päevane variant on kõigile ühesugune, siis see võimaldab võrrelda tulemusi teistega, mis võimaldab konkurentsivõimet. Raskusastmed on ka hästi tasakaalustatud. Lihtne tase pakub asjakohaseid sõnu algajatele, keskmine tase keskpärastele ja raske tase pädevatele keeleoskajatele. Kasutajakogemust saab täiustada, näiteks peab mängija iga vooru järel käsitsi vajutama helinuppu, mis eemaldab kursori sisendtekstiväljast. Seda oleks võimalik teha automaatselt vooru alguses, et muuta mäng sujuvamaks.

### 2.1.2 Sõnaveebi keelemängud

Sõnastikuportaal Sõnaveeb pakub enda leheküljel võimalust mõnesid lihtsaid keelemänge mängida. Üheks valikuks on eesti keele sünonüümidel põhinev sõnasarnasuste mäng [16]. Mängus tuleb valida kahe variandi hulgast teineteisele sarnasem sõnapaar (joonis 2). Mäng kestab kokku 50 valikut ehk vooru.



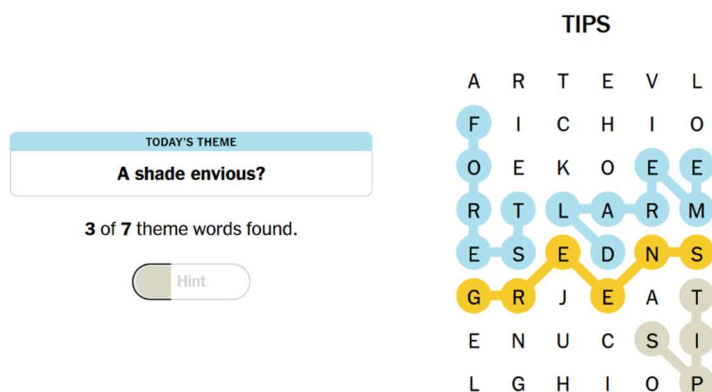
Joonis 2. Sõnaveebi sõnasarnasuse mängu vaade.

Mäng esitab mängijale kaks sõnapaari, mis jäävad vasakule ja paremale. Mängija peab kiiresti reageerima, vajutades kas F klahvi, kui sarnasem sõnade paar on vasakul, või J klahvi, kui sarnasem sõnapaar on paremal.

Mäng mõõdab ka voorude läbimise aega, et mängija saaks enda kiirust mõõta ning end sel viisil proovile panna. Kasutajaliidese komponentide vahel on aga palju tühja ruumi, mis võib jätta mulje, et mäng pole lõplik. Vastuseks olevad sõnad on sageli välimuselt üksteisega väga sarnased (näiteks sõnad „natuke“ ja „natukene“) ning seetõttu tuleb mängijal valida need sõnad, mis näevad välja sarnasemad, selle asemel, et mõelda semantika peale.

### 2.1.3 Strands

The New York Times on Ameerika päevaleht, mille veebilehel pakutakse erinevaid sõnamänge, mis suuresti sarnanevad paberil põhinevate mängudega. Üks nendest mängudest on „Strands“ [17]. See põhineb klassikalisel sõnade otsimise mängul. Eesmärk on leida ingliskeelseid sõnu tähtede ruudustikust (joonis 3). Klassikaliselt tuleb moodustada sõna, tõmmates sirge joone läbi tähtede. Joon võib olla vertikaalne, horisontaalne või diagonaalne. „Strands“ variandis ei pea joon olema sirge. Mängu ruudustik uueneb iga päev ja katab mingit teemat. Möödunud päevade mängi ei ole võimalik tagantjärele lahendada ja seega saab iga päev ainult ühe korra mängida. Mäng on võidetud siis, kui kõik teemaga seotud sõnad on leitud.



Joonis 3. Sõnade otsimise mängu „Strands“ vaade.

Mängu alguses on näha teema sõnastust, mis on meelega kirjutatud ebamääraselt, et panna mängija mõistatama. Mängija saab vihje iga kolme mitte teemaga kattuva sõna leidmisega. Vihje tõstab esile ühe õige sõna tähed. Teemaga seotud sõnad ei kattu üksteisega ja katavad ära terve ruudustiku. Igas ruudustikus on ka nn „spangram“, mis tähendab sõna või fraasi, mis kirjeldab antud teemat ja on märgitud kollasega. See sõna ulatub ruudustiku ühest küljest teiseni. Teemaga seotud sõnad on märgitud sinisega.

Sellel mängul on korralikult kujundatud ja kergesti arusaadav kasutajaliides. Mäng esitab vaba- tahtliku väljakutse võita ilma vihjeid kasutamata. Mõnedel päevadel võib mäng käsitleda teemat, mida mängija ei tunne hästi (näiteks konkreetsed filmipealkirjad). Selle tõttu võimaldab vihjete kasutamine sellist mängu, mis ei ole pingeline ja on sõbralik nendele, kes ei ole teemaga kursis.

#### 2.1.4 Eesti viipekeekele õppeprogramm

On olemas keelemänge, mis on varasemalt tehtud bakalaureusetöö raames. Näiteks Ann Lember on loonud eesti viipekeekele õppeprogrammi [18], mis pakub kolm mängu viipekeekele õppimiseks. Mäng „Sõrmendtähestik“ näitab sõrmendtähestikule vastavaid sõrmendite pilte koos valikvastustega. Mäng „Sõrmendid ja viiped“ näitab samu sõrmendeid ridadena, mis moodustavad sõnu (joonis 4). Mäng „Kas tunnend viipeid?“ näitab sõnale vastavat viipevideot. Viimases kahes mängus peab mängija vastava sõna välja kirjutama.

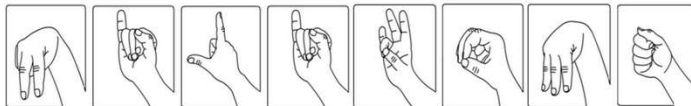
## Sõrmendid ja viiped: Viisakusväljendid

Sõrmentähestik

Sõrmendid ja viiped

Kas tunned viipeid?

Mis sõnale vastab see sõrmendite järjend?



Vastus:

KONTROLLI

Küsimused: 1 / 10

Skoor: 0 %

Joonis 4. Viipekeelee mängu „Sõrmendid ja viiped“ vaade.

Kui mängus „Sõrmendid ja viiped“ on vastus esitatud, siis näidatakse mängijale selle sõna vastavat viipevideot. Sedasi tõuseb raskusaste iga mänguga. Esimene mäng valmistab mängijat ette teiseks ja teine mäng valmistab ette kolmandaks. Lehel tuleb arvamismängude mängimiseks valida teema, milleks on näiteks viisakus, riietus või ametid. Kokku on igas mängus kümme juhuslikult valitud küsimust. Mäng jälgib vastuste õigsust protsendilise skooriga.

Programm toimib hästi viipekeelee õppematerjalina. Raskusaste tõuseb loomulikul viisil, pakudes harjutatavat väikeste koguste kaupa. Näiteks õpetatakse alguses esimesed 8 tähestiku tähte selgeks, siis lisatakse järgmised 8 tähte juurde jne. Õppematerjal piirdub etteantud teemadega ja seega tuleb kasutajal teiste sõnade õppimiseks minna mujale.

### 3. Valminud kuulamismängud

Uurides eesti keele mängude valdkonda, näib, et keele kuulamisel põhinevaid mängu ei ole olemas. Seetõttu püütakse käesolevas töös teha genereeritud kõnel põhinevaid keelemänge. Üheks peamiseks eesmärgiks on edendada erinevate keelevahendite kasutamist, et eesti keele ressurside kasutamisele kaasa aidata. Kuulamismängud rakendatakse veebiliidesena.

Töö käigus valminud kuulamismängud on mõeldud eesti keele harjutamiseks algtasemel. Mängud panevad proovile kasutaja kuulamisoskust eesti keeles. Kokku on loodud kaks mängu, millest üks testib veel kasutaja sõnade kirjutamisoskust ja teine testib valikvastustega lausete analüüsimist. Nende mängude mängimine sobib noortele ja eesti keelt mitte emakeelena rääkivatele kõnelejatele. Neid mängu saab kasutada abivahendiks kuulamisoskuse lihvimiseks, kõnekeele paremaks mõistmiseks ja õigekirja harjutamiseks.

Kirjutamismängu peamiseks inspiratsiooniks on eelmainitud näide „Spellcheck“ [15], milles mängija peab välja kirjutama hääldatud sõna. Valikvastustega mäng nõuab hääldatud lauses oleva sõna valimist etteantud valikutest. Valminud veebiliides sarnaneb ülesehituselt väljatoodud viipekeele õppeprogrammi näitega, mida samuti peeti inspiratsiooniks. Kasutati sarnast lähenemist mängude loetelu esitamiseks ja mänguala kavandamiseks. Protsendilise skoori asemel kasutatakse punaseid ja rohelisi sümboleid, mis sarnaneb mängu „Spellcheck“ lahendusele.

Valminud mängud asuvad veebisaidil, mille nimeks on valitud „Ortograafik“. Nimi tuleneb sõnast „ortograafia“ ja iseloomustab mängude põhimõtet aru saada õigekirjast. Järgnevalt selgitatakse programmi arendusprotsessi, süvenetakse mängija kogemusse, analüüsitakse saadud tagasisidet ja lõpetatakse edasiarenduse võimaluste vaatlusega.

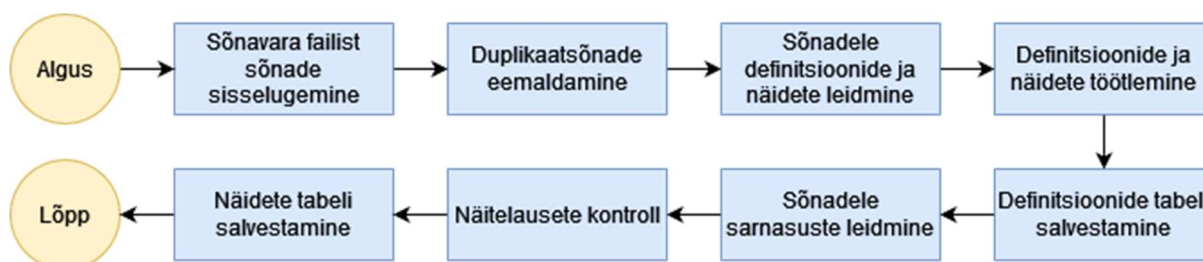
#### 3.1 Programmide algoritmid

Mängu arendusprotsess jaotub kaheks peamiseks etapiks: eeltöötlus ja veebiliidese loomine. Eeltöötlus hõlmab sõnade, nende definitsioonide ja näidislausete kogumist ja töötlemist. Veebiliidese loomine sisaldab kasutajaliidese kavandamist ja rakendamist.

##### 3.1.1 Eeltöötlus

Selle töö raames kasutatakse sõnade kogumiseks Sõnaveebi õpetaja tööriistadest leiduvat sõnavara otsingut [19], mis on kasulik keeleõpetuse seisukohalt. Lisaks on sealsed sõnad liigitatud keeleoskustasemetesse, mis hõlbustab sobiva raskusastme valimist. Sõnad on jaotatud tasemetesse A1, A2, B1, B2 ja C1. Need märgivad algelist, iseseisvat ja vaba keelekasutust. Kokku

on sõnavarast saadaval 10562 sõna. Saadud sõnade lõimimiseks keelemängudega tuleb neid esmalt eeltöötles kasutada definitsioonide ja näidislause leidmiseks. Sellele vastav diagramm on esitatud joonisel 5.



Joonis 5. Eeltöötlusprotsessi diagramm.

Keeletöötlusülesandeid lahendatakse sageli programmeerimiskeele Python abil [20]. Python pakub ulatuslikku teekide raamistikku, on lihtne kasutada ja on kohandatav. Seetõttu on käesolev eeltöötlus läbi viidud keeles Python. Eeltöötlustapp koosneb mitmest sammust. Esimesena loeb programm sisse Sõnaveebist saadud sõnaloendi. Sõnade hulgast eemaldatakse lihtsuse huvides mitmesõnalised fraasid. Seejärel töötleb algoritm sõnade loetelu, kõrvaldades leitud sõnade duplikaadid. Osad sõnad ilmuvad mitmekordselt, sest need on liigitatud mitmesse keeletasemesse või sõnaliiki. Peale fraaside ja dubleeringute eemaldust on alles 9977 sõna.

Programm vaatleb igat sõna üksikhaaval ning üritab leida sõnale definitsiooni ja näitelause. Seda realiseeritakse Eesti Wordneti abil [21], mis on saadaval moodulina (ingl *module*) läbi EstNLTK teegi. Moodul pakub mitmesuguseid funktsioone sõnade töötlemiseks. Eesti Wordnet võtab vastu sõna päringu ning leiab sellele sünonüümihulga (lühidalt sünohulk). Sünonüümihulk tähendab sama mõistet väljendavat sõnade hulka. Sünohulgad on omavahel seotud vastavate semantiliste või leksikaalsete suhete kaudu. Siinkohal leiab programm sünohulgas olevatele sõnadele vastavad definitsioonid ja näitelause. Kui sõnale ei ole olemas definitsiooni või näitelause, siis programm jätab selle sõna vahele.

Leitud laused töödeldakse mooduliga Re [22]. Moodul on loodud keelele Python, et võimaldada regulaaravaldiste (ingl *regular expression*) kasutamist. See moodul tuvastab kasutaja poolt täpsustatud mustreid antud tekstist. Mustreid saab omakorda eemaldada ja asendada. Moodulist Eesti Wordnet saadud laused tihti sisaldavad metaandmeid, näiteks valdkond või allikas. Sellised andmed ei sobi kõnesünteesi lisada. Seega selles töös kasutatakse regulaaravaldisi, et need andmed leida ja eemaldada. Saadud definitsioonis või näites võib olla rohkem kui üks lause. Lisaks kasutatakse Moodulit Re nende lausete tuvastamiseks ja osadeks eraldamiseks. Järele jäetakse igale sõnale pikim definitsioon ja näide. Muude täiustuste hulka kuulub

lühendite laiendamine, et hõlbustada kõnesünteesil hääldamist. Näiteks lühend „jne“ asendatakse sõnadega „ja nii edasi“. Programm eemaldab laused, mis sisaldavad ellipseid ehk on pooliklaused, ja erisümboleid, näiteks „%“ ja „/“, sest need võivad tekitada segadust kõnesünteesis saadud lausete lugemisel.

Peale lausete töötlemist on andmed valmis, et salvestada definitsioonide tabel faili. Programm salvestab andmed CSV-vormingus, mida veebiliides saab sisse lugeda (joonis 6). Eeltöötlus-etapis võimaldab moodul Pandas [23] sisse lugeda faile ning moodul CSV [24] võimaldab töödeldud andmeid vormindada ja salvestada.

SÕNA	DEFINITSIOON	TASE
aabits	raamat kirjaoskuse algete omandamiseks	B1
aasta	ajavahemik, mille vältel Maa teeb tiiru ümber Päikese	A1
aastaaeg	atmosfääri tingimuste järgi jaotatud aasta perioodid	B1
aastakäik	perioodikaväljaande ilmumisaasta järjenumbr alates ilmumise algusest	B2
aastane	üks aasta vana	A2
aastapäev	päev, mil möödub teatud arv aastaid mingist sündmusest, millegi asutamisest, rajamisest ja nii edasi	A2
aastaringelt	läbi aasta toimuvana, terve aasta jooksul	B2
aastatuhat	tuhat aastat, tuhandeaastane periood	B2
aastavahetus	vana aasta lõpp ja uue aasta algus	B1
aatom	millegi üliväike osa või miski üliväikesel kujul	B2
abi	tegevus, millega soodustatakse mingi vajaduse rahuldamist või toetatakse mingite püüdluste või eesmärkide elluviimist	A1
abielluma	abiellu astuma, abielu sõlmima, naist võtma või mehele minema	A2

Joonis 6. Loodud definitsioonidega tabeli algus.

Kirjutamismäng kasutab tabelifaili, mis koosneb sõnadest, definitsioonidest ja keeletasemetest. Kokku on andmeid 2948 sõnale. Suur andmemahu langus on tingitud Eesti Wordnetist, mis ei leia suurele hulga sõnadele täiuslikke ehk kirjaliselt identseid sünohulki või vastavaid definitsioone ja näiteid.

Valikvastustega mängu jaoks leiab programm esialgsest sõnaloendist igale sõnale sarnased vastused, mida mängus kasutatakse valede vastuste valikutena. Sarnaste sõnade leidmiseks kasutatakse mooduli DiffLib [25] SequenceMatcher funktsionaalsust. Selle abil saab tuvastada sõnu, millel on suurim sarnasus antud sõnadega. SequenceMatcher sisuliselt arvutab kahe antud teksti sarnasust, leides pikima kokkulangeva tähtede või sümbolite osajärjestuse ja jättes välja ebaolulised osad nagu tühikud. See ei anna küll parimat sarnasuse tulemust, kuid toimib piisavalt hästi, et leida sõnu, mis tunduvad tavaisikule sarnased. Selline lähenemine sunnib mängijaid kuulama lause konteksti, mida saab kasutada õige sõna tuvastamiseks.

Lõpuks kontrollib programm üle iga näitelause. Lause lemmatiseeritakse, kasutades EstNLTK teegi moodulit Text. Mooduliga teisendatakse lause Text objektiks ja sellele lisandub lemmade ehk sõnade algvormide kiht. Saadud sõna algvormi on võimalik võrrelda esialgse sõnaloendi sõnaga, mis on samuti algvormis. Programm kontrollib, et ühtegi vale vastuse sõna ei leiduks

lauses ning samalaadselt, et õige sõna siiski leiduks. Peale kontrolli salvestab programm näidetega tabeli faili samamoodi, nagu see salvestab definitsioonidega tabeli (joonis 7).

SÕNA	NÄIDE	SARNASED	TASE
aabits	Kas lasteaias aabitsat ka juba õpitakse	abitus, naarits, abitu, abi, ait	B1
aasta	Milline näeb Tallinn välja 2100. aastal	aastane, astma, kaasa, paast, pasta	A1
aastaaeg	Meie kliimas on neli aastaaega	aastane, kaasaeg, ahastama, ajastama, avastama	B1
aastakäik	Ajalehe kuues aastakäik	maastik, allakäik, aastapäev, alasti, külaskäik	B2
aastane	Noorim laps sai eile aastaseks	vastane, iga-aastane, auaste, kastan, saatan	A2
aastapäev	Koor tähistab oma kümnendat aastapäeva	vastlapäev, naistepäev, tänapäev, auaste, esmaspäev	A2
aastaringelt	Ka suusatamist treenitakse aastaringelt	sarnaselt, ööpäevaringselt, ajaliselt, ajutiselt, alaliselt	B2
aastatuhat	Jääajal kattus Põhja-Euroopa korduvalt aastatuhandeteks mandrijäaga	saastatus, statuut, taastumatu, ahastama, ajastama	B2
aastavahetus	Ta loodab aastavahetusel saada uue korteri	saastatus, staatus, vahetus, nädalavahetus, kasvatus	B1
aatom	Vana raamat lagunes ta käes pisikesteks aatomiteks	saatma, anatoomia, alam, auto, daam	B2
abi	Ilma abita ei saanud ta liikuda	abil, abitu, abielu, abinõu, abitus	A1
abielluma	Ema abiellus vanas eas teist korda	abielu, alluma, abielus, paeluma, abiellupaar	A2

Joonis 7. Loodud näidetega tabeli algus.

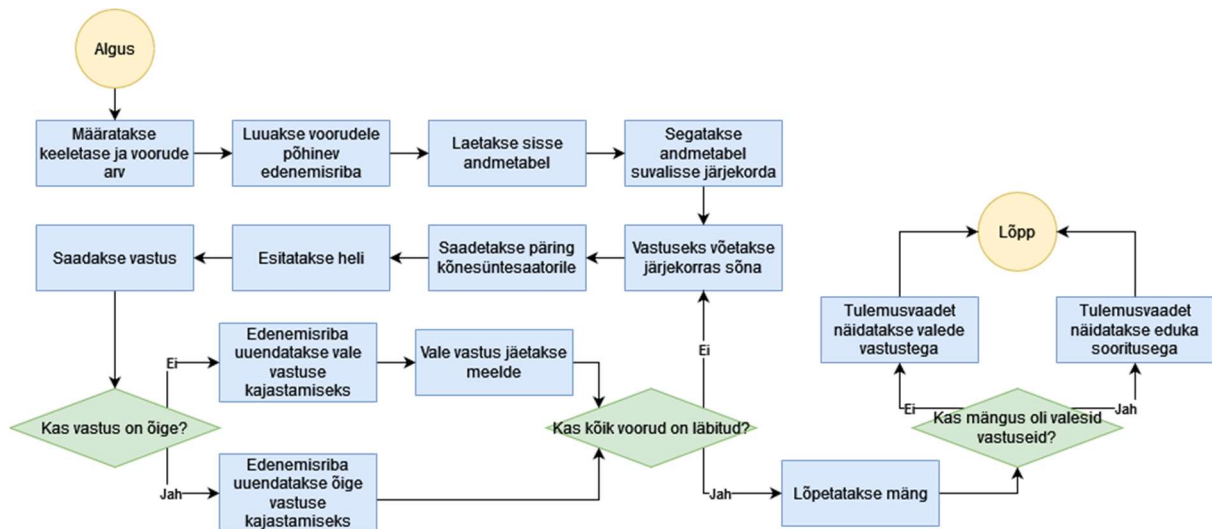
Valikvastustega mäng kasutab tabelifaili, mis koosneb sõnadest, näidetest, sarnastest sõnadest ja keeletasemetest. Kokku on andmeid 2719 sõnale. Nii töödeldud definitsioonid kui ka näitelauseid on kõnesünteesi jaoks loetavad ja terviklikud. Kõnesünteesi ennast ei teostata eeltöötlustapis, sest selle tulemusena tekkivad tuhanded helifailid kasutaksid ära suure koguse salvestusruumi. Käesolevas töös ei ole seda võimalik teha, eelkõige kasutatud veebimajutusteenuse seatud andmemahu piirangute tõttu.

### 3.1.2 Veebiliides

Keelemängud on sageli olemuselt minimalistlikud ja otse veebilehtedel mängitavad, et tagada kasutajatele lihtsat juurdepääsu. Sellist lähenemisviisi kasutatakse käesolevas töös käsitletavate kuulamismängude väljatöötamisel. Mängude kättesaadavaks tegemiseks kasutatakse GitHub veebimajutusteenust GitHub Pages [26]. GitHub on eelkõige ettevõtte, mis pakub teenust failide ja koodi jagamiseks, mis on üldiselt mõeldud koostöö jaoks. Kõik töö raames loodud andme-, ikooni- ja koodifailid hoitakse GitHubis failihoidlas.

Veebiliidese jaoks kasutatakse programmeerimiskeeli HTML ja CSS, mis on veebilehtede arendamisel põhialuseks. Keelt HTML kasutatakse veebilehe struktuurilise raamistiku loomiseks. HTML-kood kirjeldab lehes paiknevate komponentide olemust ja järjestust. Töös märgib see ära näiteks pealkirjad, tekstid ja mängunupud. Seevastu CSS-koodi kasutatakse lehtede esitluse parandamiseks. CSS-kood lisab veebilehele vormingu, värviskeemid, kirjastiilid ja muud stiilile põhinevad komponendid. Näiteks töös paigutatakse komponendid ekraani keskele, lisatakse taustale värv ja asendatakse kirjatüüp Google Fonts [27] pakutud kirjastiiliga „Varela Round“. Stiiliga seotud otsused on tehtud lootusega, et need oleksid köitvad. Töö käigus on kindlaks tehtud, et mängud toimivad mobiilseadmetes nii, et komponendid ei jääks väljapoole ekraani ja ei kattuks üksteistega.

Veebilehele funktsionaalsuse ja mänguloogika lisamiseks kasutatakse programmeerimiskeelt JavaScript. Keel on mõeldud veebilehtede interaktiivsuse lisamiseks. Sellele vastav diagramm on esitatud joonisel 8. Keele JavaScript võimalused loodud CSV-vormingus andmefailide lugemiseks on piiratud, seega selleks kasutatakse JavaScriptile loodud liidest Papa Parse. See on tõhus liides, mis saab hakkama suurte CSV-failidega ja on loodud jõudlust silmas pidades [28].



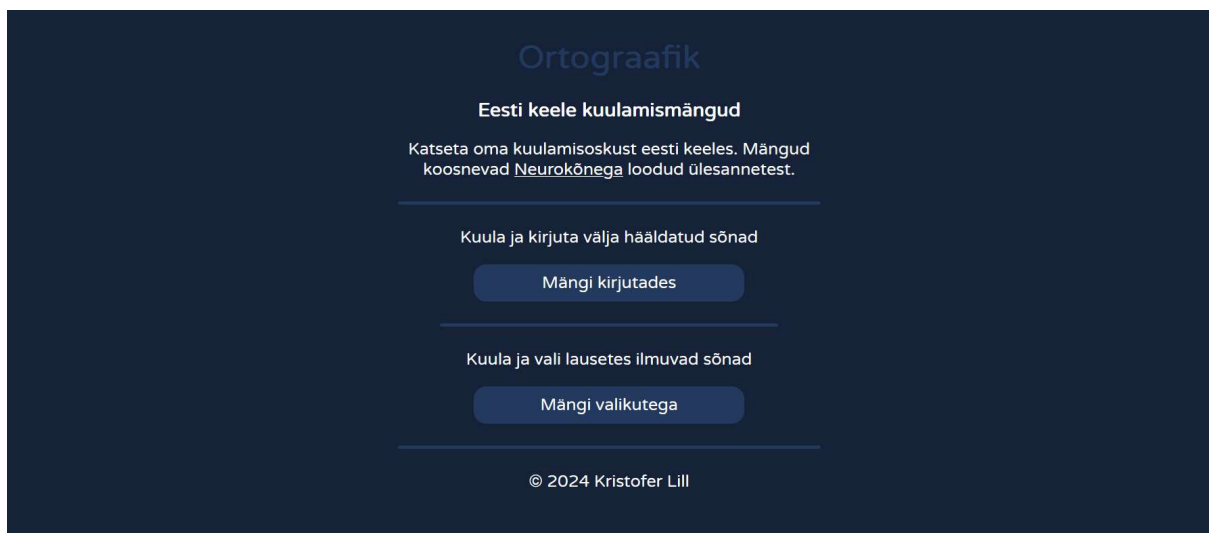
Joonis 8. Kuulamismängude algoritmi diagramm.

Kasutaja valib mängu, satub seadistusvaatesse, soovi korral muudab seadeid ja lõpuks vajutab mängu alustamise nuppu. Kui mäng on alustatud, siis JavaScripti kood peidab ära seadistusvaate ja loob mängule edenemisriba, lisades ringikujulise ikooni igale voorule. Edasi loeb programm Papa Parse abiga andmefaili sisse ning segab saadud andmed ära, kasutades Fisher-Yates algoritmi [29]. See algoritm sobib segamiseks kuna iga permutatsioon on võrdse tõenäosusega ja seega parem võrreldes naiivse lähenemisega, mis võib toota kallutatud tulemust. Programm toob esile mängimiseks vajaminevad komponendid: nupud ja kirjutamismängu korral tekstivälja. Segamini aetud sõnaloendist võetakse vastuseks olev sõna. Valikvastustega mängu korral segatakse õige vastus vastavate valede vastustega kokku ning lisatakse kõik vastuse nupud ekraanile. Järgmisena saadetakse päring Neurokõne rakendusliidesele. Kõnesünteesi loomiseks saab valida kümne hääle vahel, millele on võimalik valida ka hääle tempo. Päringus sisestatakse hääleks „Meelis“, sest peale iga hääle proovimist on see selgeim. Tempo jäetakse vaikewäärtuse ehk ühe peale. Kirjutamismängu korral on päringuks sõna koos selle definitsiooniga. Valikvastustega mängu korral on päringuks näidislause. Kui saabub päringuvastus, siis muutuvad mängu nupud aktiivseks ja mänguvoor algab. Kui seadistuses on automaatne heli kuulamine sisse lülitatud, siis mängib heli kohe vooru alguses, vastasel juhul peab mängija heli kuul-

miseks vastavat nuppu vajutama. Sõltuvalt mängust sisestab mängija oma arvamuse kas kirjutades või valides vastuste hulgest. Programm kontrollib vastust ning uuendab edenemisriba vastavalt vastuse õigsusele. Kui kõik voorud pole veel läbitud, siis võtab programm järgmise sõna ning viib läbi samad sammud. Mäng lõpeb, kui kõik voorud on läbitud. Mängu lõpus peidab programm ära mängu nupud ja toob esile tulemused. Kui kõik vastused on õiged, siis programm teatab seda, vastasel juhul esitab programm nimekirja valedest vastustest ja nende õigetest vastetest. Tulemuste vaates on valik uuesti mängimiseks, mis viib mängija tagasi seadistusvaatesse, ja valik esilehele naasmiseks.

### 3.2 Mängimise kirjeldus

Valminud rakendus on kättesaadav veebiaadressilt <https://ortograafik.github.io/>. Lehele sattunud külastaja kohtub minimaalset kasutajaliidest, mis koosneb lühikesest sissejuhatusesest ja kahest mängima kutsuvast nupust (joonis 9).



Joonis 9. Veebisaidi „Ortograafik“ esilehe vaade.

Kasutajaliidese kujundus on inspireeritud olemasolevatest sõnapõhistest veebimängudest, näiteks Wordle<sup>2</sup> ja Contexto<sup>3</sup>. Eesmärgiks on luua mugav ja mängima kutsuv keskkond koos esteetilise kirjatüübi ja värvidega, samas vältides külastaja ülekoormamist. Lehekülgede ühtne stiil on mõeldud tagada külastajale lihtsat ja otsekohest navigeerimist. Valides esimese kirjutamisele keskenduva alammängu, suunatakse külastaja alamlehele, kus toimub mängu seadistamine (joonis 10).

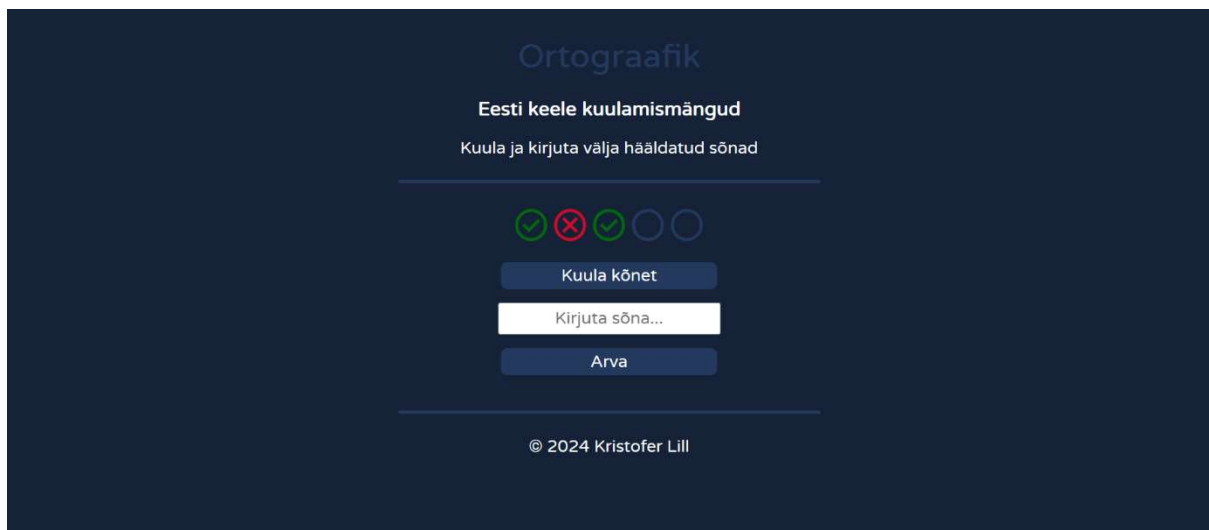
<sup>2</sup> <https://www.nytimes.com/games/wordle/index.html>

<sup>3</sup> <https://contexto.me/>



Joonis 10. Veebisaidi „Ortograafik“ mängu seadistamise vaade.

Seadistada saab sõnade keeletaset (vaikimisi sõnad kõigist keeletasemetest), voorude arvu (vaikimisi 5) ja sünteesitud heli automaatset esitust (vaikimisi sisse lülitatud). Konkreetset keeletasemeid, mille vahel saab valida, on A1, A2, B1, B2 ja C1. Voorude arvu jaoks saab valida kas 5, 10, 15 või 20. Selliste valikute põhimõte on võimaldada kas kiireid või põhjalikumaid katsetusi. Kui seadistused on tehtud, siis saab asuda mängima, vajutades nuppu „Alusta“. Joonisel 11 on näha pooleli olevat kirjutamismängu. Igal hetkel on võimalik pöörduda tagasi avalehele vajutades pealkirja „Ortograafik“.



Joonis 11. Veebisaidi „Ortograafik“ kirjutamismängu vaade.

Kirjutamismängu ajal hääldab kõnesünteesiga loodud heli vastuseks oleva sõna ja selle tähenduse. Tähenduse lisamise eesmärk on aidata mängijal sõnast paremini aru saada. Seejärel palutakse sisestada vastav sõna tekstivälja ning vajutada kas sisestusklahvi või ekraanil olevat

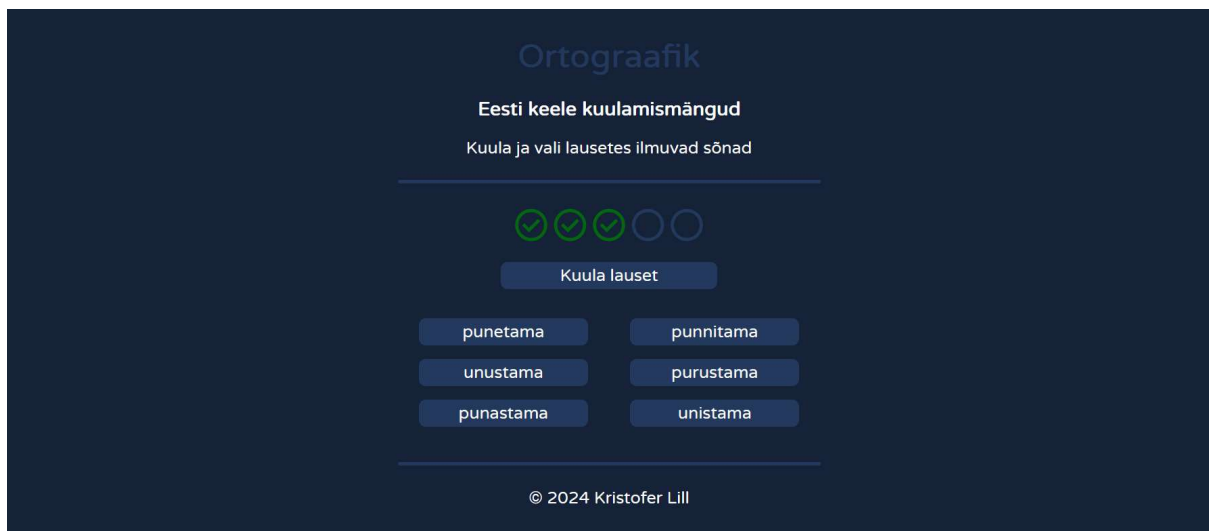
sisestusnuppu „Arva“. Igat sõna on võimalik arvata ainult üks kord. Ringidest koosnev edene-  
misriba näitab voorude kulgemist. Õigete vastuste puhul ilmub ringi sisse roheline linnuke ja  
valede vastuste puhul punane rist. Voor tähendab siin ühe sõna arvamist. Peale voorude läbi-  
mist liigub mäng edasi tulemuste vaate peale. Kui kõik vastused on õiged, siis annab mäng  
seda teada, vastasel juhul tõstab mäng valed vastused esile koos õigete sõnade vastetega. Joonis  
12 näitab tulemust valed vastustega.



Joonis 12. Veebisaidi „Ortograafik“ tulemuste vaade.

Mängijal on võimalus näha oma vigu ja valida, kas mängida sama mängu uuesti või minna  
tagasi pealehele. „Mängi uuesti“ nupp viib mängija tagasi mängu seadistamise vaatele. „Tagasi  
esilehele“ nupp võimaldab navigeerida pealehelt teise mängu alamlehele.

Valikvastustega mäng on üleschituselt sarnane kirjutamismängule. Seadistamise vaade on  
muudatusteta ja pakub samad valikud mängu konfigureerimiseks kui kirjutamismäng. Tule-  
muste vaade on samuti ilma erinevuseta, võimaldades asuda uuesti mängima või pöörduda ta-  
gasi pealehele. Erinevust võib näha mängudes endis, nagu on näidatud joonisel 13, kus on kuju-  
tatud käimasolev valikvastustega mäng.



Joonis 13. Veebisaidi „Ortograafik“ valikvastustega mängu vaade.

Mängijale esitatakse heliga lause. Kui seadistuses on automaatse heli esitus välja lülitatud, siis tuleb vajutada „Kuula lauset“ nuppu. Mängija peab valima lauses esineva sõna kuue valiku hulgast. Kui ei olda arvamuses kindel, siis võib kontrolli mõttes helinuppu uuesti vajutada.

### 3.3 Tagasiside

Veebirakendusele tagasiside saamiseks koostati küsimustik (vt lisa 1). Küsimustikus paluti avaldada arvamust veebirakenduse kujunduse, kasutusmugavuse ja arusaadavuse kohta. Lisaks paluti hinnata mängu raskusastet ja kasutamist meelelahutusena ning keeleõppena. Hindamine toimus viie punkti skaalas. Küsimustik viidi läbi veebipõhiselt [30], et seda oleks lihtne välja saata ja vastajatel oleks mugav seda täita. Kokku oli 9 küsimustikule vastajat. Küsimustikku levitati eestikeelsete kanalite kaudu ja seetõttu eeldatakse analüüsis, et vastajad on eesti keele kõnelejad.

Keskmine hinnang kujundusele oli 4.6. Vastajad kirjeldasid, et veebiliides kujutab selget ja rahulikku keskkonda. Midagi ülemääraselt keerulist ei leitud. Mõned vastajad mainisid, et veebisait on liiga lihtsa kujundusega. Lisaks arvas üks vastaja, et tulemuste vaade võiks olla selgem, sest pealkirjade ja andmete väljad on üksteisele liiga sarnased.

Keskmine hinnang kasutusmugavusele oli 4.8 ja arusaadavusele 4.5. Mainiti, et kõik on mõistetav ja töötab sujuvalt. Näiteks toodi välja selliseid omadusi nagu tekstivälja fookusesse jäämine pärast sõna sisestamist kirjutusmängus ja võimalus kuulata heli korduvalt. Ainus kriitika arusaadavuse kohta oli see, et kirjutamismängus ei ole kohe selge, et tuleb sisestada ainult sõna, mitte kogu tekst. Teisest küljest meeldisid paljudele lühikesed selgitused.

Keskmine hinnang meelelahutuslikkusele oli 3.5 ja raskusastmele 3.9. Arvamused kasutada rakendust meelelahutuse eesmärgil olid väga erinevad. Pooltel vastajatel hakkas mängides kiiresti igav. Esitati ideid, kuidas muuta mängud põnevamaks. Üks idee oli mängimise kiirust mõõta ja mängu lõpus seda näidata. Idee sarnaneb teises peatükis välja toodud Sõnaveebi mängude näitele [16]. Selline täiendus esitaks mängijale suurema väljakutse parima aja saavutamiseks. Teine idee oli tõsta sõnade raskusastet. Konsensus oli, et mängude sõnavara on liiga lihtne. Kuna eeldatakse, et vastajad on eesti keele kõnelejad, siis selline järeldus on oodatav. Soovitati lisada vähetuntuid ja tihti valesti kirjutatavaid sõnu.

Keskmine hinnang keeleõppele oli 4.2. Arvati, et see oleks hea algus noortele ja eesti keelt mitte emakeelena rääkivatele inimestele eesti keele õppimiseks. Pakuti, et see oleks neile hea viis uute sõnade õppimiseks. Samuti väljendati muret, et kõnesünteesi hääldus on kahtlane ja seetõttu on võimalus õppida sõnu valesti hääldama.

Kui probleemid ja ettepanekud olid tagasiside põhjal kogutud, siis uuendati veebiliidest. Kohendati selgituste sõnastust ja lisati täpsustus kirjutamismängus olevate tähenduste kohta. Tulemuste vaates muudeti pealkirjad allajoonituks, et luua kontrasti ja eristada neid vigade nimekirjadest. Kui kirjutamismängus vajutatakse arvamuse tegemiseks sisestusklahvi, siis jääb tekstiväli fookusesse, mis lihtsustab järgmise vastuse kirjutamist. Lisati automaatne tekstivälja fookuseerimine ka olukorraks, kui vajutatakse „Arva“ nuppu.

### **3.4 Edasiarenduse võimalused**

Kuulamismängude edasiseks arendamiseks on palju võimalusi. Valikvastustega mängu üks võimalikke täiustusi oleks lasta mängijatel valida voorus esitatud lausetele mitu leiduvat sõna, et võimaldada lause ja selle ülesehituse sügavamalt mõistmist. Lisades rohkem interaktiivsust, muutuksid mängud keerukamaks ja mitmekesisemaks, mis omakorda köidaks mängijate tähelepanu. Saadud tagasisidele tuginedes võiks mängu raskusastet tõsta, et see sobiks paremini eesti keelt vabalt rääkivale kõnelejale.

Rohkemate mängude lisamine oleks loogiline täiendus. Üks töös teostamata jäänud mäng on selline, kus sarnaselt valikvastustega mängule tuleb esitatud valikutest leida hääldatud lausele sobiv vastus, mis koos algausega oleks loogiline dialoogi tervikosa. Näide sellest oleks öeldud lausele „Kas soovite teed?“ vastuseks valik „Jah, tänan.“. Selline mäng oleks kasulik suhtlusfraaside kasutamise õpetamiseks.

Sobiks kaaluda mängude paremat ühilduvust mobiilseadmetega, et saavutada mobiilsõbralik ja loomulik kujundus. Eelkõige selleks, et tagada laialdast kättesaadavust. Mobiilseadmete kasutus tõuseb Eestis [31] ja seega võib moodustada märkimisväärse osa mängijatest.

Et vähendada projekti sõltuvust väliste ressurssidele, võiks Neurokõne kasutamise üle viia eeltöötlustapile. Niiviisi väheneb tõenäosus, et rakendus töötamast lakkab. Selleks sobiks saadud helifailide salvestamine lokaalselt projekti sisse. Kuna Neurokõne loob helifailid WAV-vormingus, mis on kvaliteedikadudeta salvestusvorming, siis need failid jäävad mahukaks. Et tuhandeid helifaile lokaalselt hoida üleliigse failimahu probleemita, tuleks neid enne teisendada tihendatud failivormingusse. Selleks saab kasutada näiteks OGG-vormingut, mis on märkimisväärselt väiksema mahuga.

## Kokkuvõte

Käesoleva bakalaureusetöö eesmärk oli luua eesti keele keeleressurssidele tuginevad veebimängud, mis võimaldavad harjutada keele kuulamisoskust. Tulemusena loodi veebirakendus, milles on võimalik mängida kahte mängu. Ühes kuulatakse sõnu ja kirjutatakse need välja ning teises kuulatakse lauseid ja valitakse loetelust esinevad sõnad.

Töös vaadeldi olemasolevaid keeleressursse ja keelemänge. Arutati leksikonide, korpuste ja töötlusvahendite olulisust ning kasulikkust. Kaaluti võimalust kasutada mängude mängimist keeleõppeks ja toodi näiteid keelemängudest. Jõuti järeldusele, et keelemängud sobivad hästi keele õppimiseks. Selgus, et puuduvad kuulamisega seonduvad eestikeelsed mängud.

Töös kasutatud andmed koguti Eesti Keele Instituudi õpetaja tööriistadest, millest saadi sõnad, ja Eesti Wordnetist, millest saadi definitsioonid ja näidislauseid. Andmed töödeldi kokku sõnade andmestikuks, milles on 2948 sisestust, ja lausete andmestikuks, milles on 2719 sisestust. Mängimiseks valminud veebiliides kasutab Tartu NLP poolt väljatöötatud Neurokõne kõne-sünteesi, et sõnad ja laused ette lugeda. Valminud veebirakendus on ainulaadne, sest see on teatavasti ainus eestikeelne mäng, mis sisaldab genereeritud kõnet.

Rakendust testiti tagasisideküsimustiku abil, milles kokku oli 9 vastajat. Tagasiside põhjal tehti parandusi veebirakenduse kasutusmugavusele. Enamasti hinnati rakendust positiivselt, märkides selle võimalikku kasulikkust keeleõppes. Tagasisidet arvesse võttes sai eesmärk täidetud. Rakenduse peamiseks puuduseks oli mängude raskusaste, mida soovitati tõsta eesti keele oskajatele. Veebirakenduse edasiarendamiseks on olemas mitmeid võimalusi, mis hõlmaks rohkeimate mängude lisamist ja kasutajaliidese kvaliteedi ning kasutusmugavuse parandamist.

## Viidatud kirjandus

- [1] Kaalep H.-J., Prillop R., Ehasalu E. The role of Internet in creating, financing and integrating language resources. *International Conference on Language Resources and Evaluation*, 1998, lk 1. <https://www.cl.ut.ee/yllitised/lrec1998.pdf> (10.04.2024)
- [2] Chen I.-J. Using Games to Promote Communicative Skills in Language Learning. *The Internet TESL Journal*, 2005, kd 11, nr 2. <http://iteslj.org/Techniques/Chen-Games.html> (04.12.2023)
- [3] Eesti Keeleressursside Keskus. <https://keeleressursid.ee/et/keeleressursid> (10.04.2024)
- [4] L'Homme M.-C., Cormier M. C. Dictionaries and the Digital Revolution: A Focus on Users and Lexical Databases. *International Journal of Lexicography*, 2014, kd 27, nr 4, lk 331–340. <https://doi.org/10.1093/ijl/ecu023> (10.04.2024)
- [5] Muischnek K., Orav H., Kaalep H.-J., Õim H. Eesti keele tehnoloogilised ressursid ja vahendid. Eesti Keele Sihtasutus, 2003. <http://hdl.handle.net/10062/40637> (10.04.2024)
- [6] Kaalep H.-J. Eesti keele ressurside loomine ja kasutamine keeletehnoloogilises arendustöös. *Dissertationes philologiae estonicae Universitatis Tartuensis*, Tartu, 1999, kd 7. [https://www.cl.ut.ee/yllitised/hkaalep\\_dr.pdf](https://www.cl.ut.ee/yllitised/hkaalep_dr.pdf) (10.04.2024)
- [7] Björkenstam K. N. What is a corpus and why are corpora important tools? Stockholm University, 2013. <https://su.diva-portal.org/smash/record.jsf?pid=diva2:686162> (12.05.2024)
- [8] Richey C. Speech Corpora. 2020. [https://web.stanford.edu/dept/linguistics/corpora/material/X\\_Speech\\_Corpora.pdf](https://web.stanford.edu/dept/linguistics/corpora/material/X_Speech_Corpora.pdf) (12.05.2024)
- [9] Priya B., Nandhini J. M., Gnanasekaran T. An Analysis of the Applications of Natural Language Processing in Various Sectors. *Smart Intelligent Computing and Communication Technology*, Advances in Parallel Computing, 2021, kd 38, lk 598–602. <https://ebooks.iospress.nl/doi/10.3233/APC210109> (10.04.2024)
- [10] EstNLTK. <https://github.com/estnltk/estnltk> (29.04.2024)
- [11] Orasmaa S., Petmanson T., Tkachenko A., Laur S., Kaalep H.-J. EstNLTK - NLP Toolkit for Estonian. *The International Conference on Language Resources and Evaluation*, 2016, lk 2460–2466. <https://www.etis.ee/portal/publications/display/24422d21-2ace-440b-a5be-fcebbe343d10> (13.05.2024)
- [12] Tartu NLP Neurokõne. Tartu Ülikool. <https://neurokone.ee/> (29.04.2024)

- [13] Rätsep L. Neurokõne abil saab kuulata ERR-i uudisteportaali artikleid. Tartu Ülikool, 2023. <https://ut.ee/et/sisu/neurokone-abil-saab-kuulata-err-i-uudisteportaali-artikleid> (13.05.2024)
- [14] Raheem F., Ibrahim E. R., Hasan A. A., Hussein B. The Role of Language Games in Developing Linguistic Abilities. *PalArch's Journal of Archaeology of Egypt / Egyptology*, 2020, kd 17, nr 7, lk 5820–5839. <https://archives.palarch.nl/index.php/jae/article/view/2767> (04.12.2023)
- [15] Spellcheck. Answer in Progress. <https://spellcheck.xyz/> (12.05.2024)
- [16] Keelemängud. Eesti Keele Instituut. <https://sonaveeb.ee/games/> (12.05.2024)
- [17] Strands. The New York Times. <https://www.nytimes.com/games/strands> (12.05.2024)
- [18] Lember A. Eesti viipekeele õppeprogramm. <http://prog.keeleressursid.ee/viipekeel/> (12.05.2024)
- [19] Õpetaja tööriistad. Eesti Keele Instituut. <https://sonaveeb.ee/teacher-tools/#/home> (30.04.2024)
- [20] Turing. Which Language Is Useful for NLP and Why? <https://www.turing.com/kb/which-language-is-useful-for-nlp-and-why> (30.04.2024)
- [21] Eesti Wordnet. <https://www.cl.ut.ee/ressursid/teksaurus/> (12.05.2024)
- [22] Mooduli Re dokumentatsioon. <https://docs.python.org/3/library/re.html> (30.04.2024)
- [23] Pandas. <https://pandas.pydata.org/> (30.04.2024)
- [24] Mooduli CSV dokumentatsioon. <https://docs.python.org/3/library/csv.html> (30.04.2024)
- [25] Mooduli DiffLib dokumentatsioon. <https://docs.python.org/3/library/difflib.html> (30.04.2024)
- [26] GitHub Pages. <https://pages.github.com/> (30.04.2024)
- [27] Google Fonts. <https://fonts.google.com/> (30.04.2024)
- [28] Papa Parse. <https://www.papaparse.com/> (30.04.2024)
- [29] Fisher-Yates shuffle. National Institute of Standards and Technology. <https://xlinux.nist.gov/dads/HTML/fisherYatesShuffle.html> (06.05.2024)
- [30] Google Forms. <https://www.google.com/forms/about/> (12.05.2024)
- [31] Kemp S. Digital 2023: Estonia. DataReportal, 2023. <https://datareportal.com/reports/digital-2023-estonia> (30.04.2024)

## Lisad

### I. Tagasiside küsimustik

# Veebirakenduse "Ortograafik" kuulamismängude tagasiside küsimustik

Küsimustik on mõeldud tagasiside saamiseks Kristofer Lille bakalaureusetöös "Eesti keele kuulamismängud" tehtud veebirakenduse jaoks. Veebirakenduse leiab leheküljelt <https://ortograafik.github.io/>.

Veebirakendus koosneb kahest mängust. Kui olete mõlemat mängu proovinud, siis palun vastake järgnevale küsimustele.

Mängude proovimine ja küsimustele vastamine võtab kokku aega umbes 10 minutit.

Küsimustik on anonüümne.

Edenemise salvestamiseks [logige Google'isse sisse](#). [Lisateave](#)

\* Viitab kohustuslikule küsimusele

Milliseid mängu proovisite? \*

☐ Mäng kirjutades

☐ Mäng valikutega

Hinnake veebirakenduse kujundust \*

	1	2	3	4	5	
Halb	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Hea

Miks kujundust selliselt hindate? Kommentaariid kujundusele. \*

Teie vastus

Hinnake veebirakenduse kasutusmugavust \*

	1	2	3	4	5	
Halb	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Hea

Miks kasutusmugavust selliselt hindate? Kommentaarid kasutusmugavusele. \*

Teie vastus

Hinnake veebirakenduse arusaadavust \*

	1	2	3	4	5	
Halb	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Hea

Miks arusaadavust selliselt hindate? Kommentaarid arusaadavusele. \*

Teie vastus

Hinnake veebirakenduse kasutamist meelelahutusena \*

	1	2	3	4	5	
Halb	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Hea

Miks meelelahutusena kasutamist selliselt hindate? Kommentaarid rakenduse kasutamisele meelelahutusena. \*

Teie vastus

Hinnake mängude raskusastet \*

	1	2	3	4	5	
Halb	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Hea

Miks raskusastet selliselt hindate? Kommentaarid raskusastmele. \*

Teie vastus

Hinnake veebirakenduse kasutamist keeleõppena \*

	1	2	3	4	5	
Halb	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Hea

Miks keeleõppena kasutamist selliselt hindate? Kommentaarid rakenduse kasutamisele keeleõppena. \*

Teie vastus

Muud kommentaarid veebirakenduse kohta

Teie vastus

Saada ära

Tühjenda vorm

## II. Litsents

### **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, **Kristofer Lill**,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose  
**„Eesti keele kuulamismängud“**,  
mille juhendaja on Sven Aller,  
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, levi-  
tada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil,  
kuni autoriõiguse kehtivuse lõppemiseni.
3. olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega  
isikuandmete kaitse õigusaktidest tulenevaid õigusi.

*Kristofer Lill*

**15.05.2024**