

TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Janna-Liina Leemets

**Kohanduv veebirakendus eesti keele sõnavara ja
grammatika õppimiseks**

Bakalaureusetöö (9 EAP)

Juhendaja: Sven Aller

Tartu 2020

Kohanduv veebirakendus eesti keele sõnavara ja grammatika õppimiseks

Lühikokkuvõte:

Siinse praktilise bakalaureusetöö eesmärk oli luua veebimäng eesti keele käänete ja sünonüümide omandamiseks, mis kohandub kasutaja vastustele. Programm sobib kasutamiseks esimeses kolmes kooliastmes õpilastele kui ka välismaalastele. Keelemängu sünonüümitestid loodi automaatselt, kasutades kahte korpusi, morfoloogilist analüsaatorit, Eesti Wordnetti, ebasobivate sõnade loetelu ja Word2Vec mudelit. Kohanemise saavutamiseks loodi Leitneri süsteemist lihtsustatud algoritm. Töös kirjeldatakse õpperakendust, testide genereerimisalgoritmi ning adaptiivuse algoritmi.

Võtmesõnad:

keelemäng, e-õpe, keeleressursid, käänded, sünonüümid, masinõpe, adaptiivsus

CERCS: P175

Adaptive Study Application for Learning Estonian Vocabulary and Grammar

Abstract:

The goal of this Bachelor's thesis was to create a web-based study game for learning Estonian cases and synonyms. The application adapts to the answers given by the participant. It is suitable for pupils from elementary to high school and, in addition, is appropriate for people studying Estonian as a non-native language. For the creation of synonym tests the following natural language tools were used: two corpuses, a morphological analyser, Estonian Wordnet, a list of unsuitable words and a Word2Vec machine learning model. To achieve adaptation to the user in the game, a simplification of the Leitner system was built. A description of the application, test generation algorithm and adaptation algorithm are included in the thesis.

Keywords: language game, e-study, language resources, cases, synonyms, machine learning, adaptation

CERCS: P175

Sisukord

1. Sissejuhatus	4
2. Eesti keele õpe	6
2.1 Grammatika õppeprogrammid ja –mängud.....	6
2.2 Sõnavara õppeharjutused.....	9
2.3 Keelemäng grammatika ja sõnavara õppimiseks	11
2.4 Kohanduvus.....	11
2.4.1 Leitneri süsteem.....	12
3. Kasutatavad keeleressurssid ja keele tehnoloogilised vahendid.....	13
3.1 EstNLTK	13
3.2 Korpused	13
3.3 Ebasobivate sõnade loend	14
3.4 Gensim	14
4. Keelemäng.....	15
4.1 Sõnavaratestid	15
4.2 Grammatikatestid	15
4.3 Kasutatud tehnoloogiad.....	16
4.4 Auth0 volitussüsteem	16
4.5 Veebirakenduse ülevaade.....	16
4.6 Õppemängu sünonüümitestide loomine.....	20
4.6.1 Korpuste eeltötlusalgoritm	21
4.6.2 Sünonüümitestide loomine	21
4.7 Keelemängu kasutajale kohanemise algoritm.....	23
4.8 Rakendusele antud tagasiside.....	26
4.9 Probleemid	28
4.10 Edasiarendusvõimalused	28
Kokkuvõte.....	30
Viidatud kirjandus.....	31
Lisad.....	34
I. Küsimustik	34
II. Litsents	39

1. Sissejuhatus

Keel on rahvuse identiteedi ja kultuuri edasikandja ning selle arengutase on üks riigi edukuse tundemärke, mistõttu on keeleõpe väga oluline. Eesti keele emakeelena kõnelejaid on maailmas veidi üle 1 080 000 inimese, kellest umbes 85% asub Eestis ning ülejäänud elavad peamiselt Rootsis, Soomes, Saksamaal, Ameerika Ühendriikides, Kanadas ja Venemaal [1]. Eestis on nende kõnelejate osakaal 68% kõigist elanikest, kelle emakeel on teada [2], ning see arv väheneb püsivalt [3]. Eesti keelt õpetatakse välismaal rohkem kui 30 kõrgkoolis ning õppijaid nendes koolides on aastal 2019 ligikaudu 1000 [4]. Olulised õppijad on ka eestlased ise, kes õpivad eesti keelt emakeele ja teise keele tasemel.

Keeruliseks teeb eesti keele omandamise selle kääneterohkus, mis on probleemiallikas nii eestlastest kui ka välismaalastest õppijate jaoks. Lisaks käänetele on vaja keelt omandades teadlikult harjutada ja kinnistada ka sõnavara. Grammatikat ja sõnavara õpitakse algkoolis, põhikoolis ja gümnaasiumis. Keeleõppe lihtsustamiseks on võimalik kasutada täiendavaid vahendeid nagu e-õpe ja interaktiivsed veebiõpikud. E-õppe üks alamliike on mängupõhine õpe, mis on ajaga muutunud aina populaarsemaks [5] ning loob viisi interaktiivseks ja lõbusaks õppimiseks [6]. Mängudesse saab integreerida mitmeid õpimaterjale: õpikuid, käänete, grammatika ja sõnavara töölehti ning ülesandeid.

Keeleõppe veebimänge on varem loodud ka bakalaureusetöö raames, näiteks käänetemäng [7]. Lisaks leidub mitmeid sarnaseid keeleõppemänge internetis nagu Oahpa! [8], samatähenduslike sõnade harjutus [9] ja vastandtäendusena sõnade ristsõna [10], kuid suurem osa ülejäänud internetist leitavatest rakendustest on kas inglise või muus võõrkeeles, omavad staatilisi kasutajale mitteadapteeruvaid harjutusi või on liiga pealiskaudsed. Need programmid ei ole vastavuses muutunud õpikäsitusega [11], mis rõhutab individuaalset lähenemist igale õpilasele.

Siinne bakalaureusetöö on rakenduslik uurimustöö ja selle eesmärk on luua keeleressursse ning masinõpet kasutades veebimäng, mis võimaldab õppida eesti keelt adaptiivsete ülesannete abiga. Teemavalik tuleneb töö autori soovist õppida keeleressursse paremini kasutama ning huvist veebirakenduste loomise vastu. Loodav rakendus peaks sobima nii algkooli, põhikooli ja gümnaasiumi tasemel õppivatele eesti inimestele kui ka välismaalastele ja võimaldaks neil omandada grammatikat ning sõnavara, täpsemalt käändeid ja sõnade tähendusi sünonüümide kaudu. Õppeprogramm peab meeles, milliste lausete puhul missuguseid vigu õpilane on teinud, et

valida veaohlikke teste tihedamini. Rakenduses kasutatakse EstNLTK morfoloogilist analüsaatorit [12], Eesti Wordnetti [13] ja Word2Vec mudelit [14] ülesannete automaatseks loomiseks. Samuti kasutatakse Anneliis Hallingu valminud käänete omandamise mängu „Õpime käändeid“¹ [7].

Töö on jagatud kolmeks peatükiks. Esimeses kirjeldatakse ja võrreldakse internetis kättesaadavaid õppemänge ning -programme eesti keele õppimiseks ja selgitatakse kasutaja vigadele kohanemise vajalikkust ja ühte selle teostusviisi. Teises peatükis tutvustatakse tehnoloogiaid ja keeleressursse, mida rakenduse koostamisel kasutatakse. Kolmandas peatükis kirjeldatakse valminud õppemängu, adaptiivsuse algoritmi, testide loomise algoritmi, tekkinud probleeme ja võimalusi edasiarendamiseks.

¹ „Õpime käändeid“ mäng: <http://prog.keeleressursid.ee/opimekaandeid/>

2. Eesti keele õpe

Keele sõnavara ja grammatika õppimisele saab ligineda mitmeti. Sõnavara laiendamisele minev rõhk jääb mõnikord grammatika õppimisele alla, kuid tugev grammatikatunnetus ei korva kehva sõnavara ja laia sõnavaraga ning halva grammatikatunnetusega inimene ei saa end õigesti väljendada [15]. Põhjalik keeleõpe peaks sisaldama endas mõlemat, et õpilase areng oleks tasakaalus.

Kristin Saarso [15] on kirjutanud, et tõhus sõnavara omandamine tähendab seotud sõnaühendite kasutamist harjutustes, sest tähendustervik aitab materjali peaaegu 100% meelde jätta, ning õppimisel üksikute sõnade süstematiseerimist ja kategoriseerimist. Selliseks sõnavara omandamiseks on veebimäng hea variant, sest ülesannetes olevad laused tulevad korpusest ning õpilane saab oma sõnavara laiendada konteksti ja sõnavaheliste seoste abil. Sünonüümide tundmine on tähtis osa sõnavarast, mida saab veebimängus harjutada lüngatäitmisülesannetega.

Elle Sõrmus ja Ingrid Krall [16] on välja toonud, et grammatikat saab tõhusalt õppida induktiivselt. Induktiivse ülesande puhul lastakse õppijal iseseisvalt reegleid tuletada näidete ja teksti sisu abil. Sobiv induktiivne grammatikaõppe harjutus on seega näiteks selline, kus lauses küsitakse ühe sõna käänet, sest sellises olukorras saab õpilane lause sisu abiga käände leida ning seeläbi omandada grammatikareegleid alateadlikult.

Eesti keele õppeks leidub mitmeid veebilehti. Tehnoloogiliste vahendite kiire areng on võimaldanud muuta veebirakenduste ja -mängude loomise lihtsaks ning kättesaadavaks ka informaatikas mittepädevatele inimestele. Keeleõppe veebilehti on võimalik luua ühtegi programmeerimiskeelt oskamata näiteks WordPressi, Google Formsi või Wixi abiga, veebimänge aga saab luua Hot Potatoes [17] programmiga või veel lihtsamaid küsitlusi näiteks Kahoot [18] lehel. Veebirakendusi on loodud ka spetsiaalselt eesti keele omandamise jaoks, millest mõningaid kirjeldatakse järgmises peatükis.

2.1 Grammatika õppeprogrammid ja -mängud

Eesti keele grammatika omandamiseks leidub internetis suurema- ja väiksemamahulisi veebiteste ning rakendusi. Siinses töös analüüsitakse mahukamaid grammatikaõppe veebilehti ja -rakendusi, mis on suunatud erinevatele sihtrühmadele või pakuvad erinevaid testitüüpe. Grammatika õppeprogrammides ja -lehtedel on üldlevinud ülesanded, kus tuleb täita lünk õiges käändes sõnaga

või määrata sõna kääne. Levinud on ka testid, kus on ette antud lause, milles küsitav sõna asub, ja siis peab valima selle sõna käände. Leidub ka selliseid, kus peab vaid üksikut sõna käänama.

KÄÄNUK Eesti keele õppemäng "KÄÄNUK" v1.0.1

ANNELI VAATAS YOUTUBE'IST UUE ROSE PINK HEARTI PARFÜÜMI ESITLUST.

ANNELILE MEELDIB VÄGA ERINEVATE PARFÜÜMIDE NUISUTAMINE.

PÄRAST PARFÜÜMIDE ESITLUST VAATAR TA OMA _____, MILLEKS ON "KUIDAS KASUTADA OMA AEGA KÕIGE PAREMINI?"

LEMMIKSAADET
LEMMIKSAATE
LEMMIKSAADE

<< TAGASI AVALEHELE AVA EKI SIHITISEÕNASTIK

Joonis 1. Veebimängu "Käänuk" esimene ülesanne "Kuidas kasutada oma aega kõige paremini?"

Veebimäng Käänuk [19] on eesti keele grammatika õppeks B1- ja kõrgematele tasemetele (joonis 1). Mängu eesmärk on õpetada sihitise käände kasutust. Käänuk koosneb erinevatest elulisi olukordi kirjeldavatest koomiksitest, kus peab täitma lauses oleva lünga ühe sõnaga mitmest variantidest, näiteks ülesanne, kus lause on järgnev:

„Anneli vaatas Youtube’ist uue Rose Pink Hearti parfüümi _____.“

ja valikud on „esitlust“, „esitluse“ ja „esitlus“. Vale vastuse korral antakse uus võimalus sõna valimiseks. Iga ülesande all on võimalik saada abi Eesti Keele Instituudi sihitisesõnastikust.

Käänuk mängu sisu on staatiline, mispärast on puudulik tema kohandumine õpilase vigadele. Õpilase vigade üle arvet ei peeta ning kokkuvõtlikku tagasisidet ei anta. Lisaks võimaldab Käänuk omandada ainult sihitise käänat.

Veebirakendus Oahpa! [8] on õppeprogramm, mis genereerib ülesannete sisu automaatselt ning erinevate raskusastmetega. Oahpa! on avatud lähtekoodiga, mistõttu on see kättesaadaval mitmes keeles, kaasaarvatud ka eesti ja võru keeles. Eestikeelne Oahpa! võimaldab harjutada sõnade käänamist ja pööramist nii eraldiseisvalt kui ka lause kontekstis. Rakenduses saab harjutada kõiki käänendeid peale nimetava, sest sõnad on selles käändes ette antud, näiteks lauseta testides (joonis

2). Peale testi sooritamist näidatakse kasutajale tema tulemus ning esitletakse võimalus näha õigeid vastuseid.

Lausetega testides on ette antud viis küsilauseid, mis on küllaltki lühikesed ja lihtsad ning millele peab vastama sõnaga õiges käändes, näiteks:

„Keda äi nägi? (lehm)“

„Äi nägi _____“.

OAHPA!

MORFA-C MORFA-S LEKSA

Kääne Omastav (kelle? mille?)

Õpik / peatükid E nagu Eesti - all chapters

Uus test

ülekanne

sahtel

sokk

seiklus

hall

Kontrolli vastuseid

Harjuta omastava käände moodustamist.
Kirjuta nimisõnad õiges vormis.

Copyright 2011-2026 The Arctic University of Norway
Kontakt oahpa@uit.no

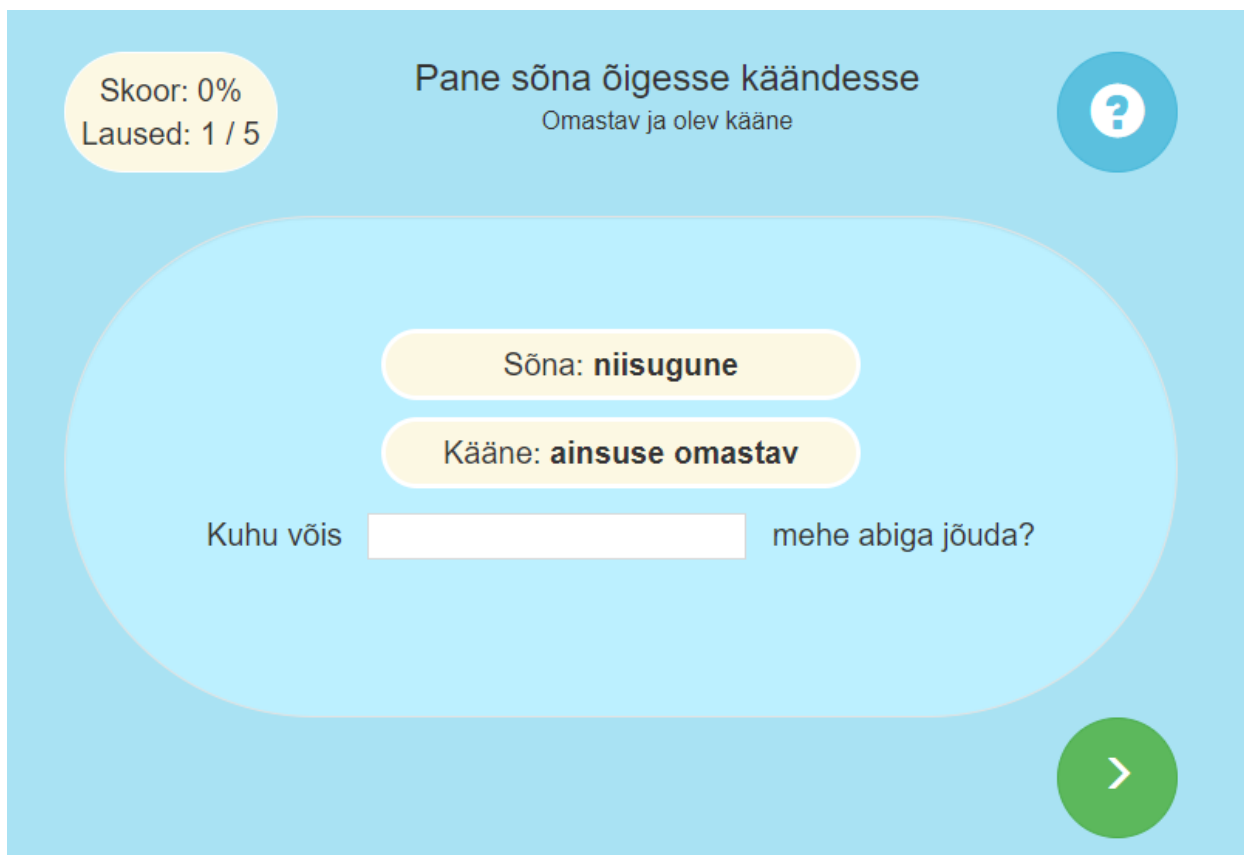
Otselink sellele harjutusele

Joonis 2. Veebirakenduse „Oahpa!“ lausekontekstita ülesande näide

Võrreldes mänguga Käänuk annab Oahpa! kasutajale võimaluse õppida kõiki eesti keele käändeid ning on dünaamilise sisuga. Lisaks annab programm tagasisidet ning näitab vajadusel õigeid vastuseid. Küll aga ei jäta Oahpa! püsivalt meelde õpilase vigu, vähendades nii genereeritud testide kasulikkust.

Veebimäng „Õpime Käändeid“, mille lõi Anneliis Halling oma bakalaureusetöö raames [7], on sisult sarnane Oahpa! rakendusele, kuid pakub lisaks lausetearvu valikut. „Õpime Käändeid“ loob testid, kasutades korpust, Estnltk 1.3 morfoloogilist analüsaatorit ja süntesaatorit,

sagedussõnastikku ning ebasobivate sõnade nimekirja. Testitüüpe on kaks: esimene testidest nõuab kasutajalt õiges käändes sõna sisestamist ning teine korrektse käände valimist valikvastuste hulgast. Esimese puhul peab valima käänetegrupi, mida soovitakse õppida, ja seejärel antakse ülesanded ükshaaval ette (joonis 3), teise puhul on valikvastused kõik 14 eesti keele käänat.



Joonis 3. "Õpime Käändeid" rakenduse "Pane sõna õigesse käändesse" test

„Õpime Käändeid“ pakub sarnast sisu nagu rakendus Oahpa!, kuid annab võimaluse harjutada keelt valikvastustega ülesannetes. „Õpime Käändeid“ on ka selgema kasutajaliidesega, küll aga valitakse testid alati juhuslikult ja kasutajapõhiselt vigu meelde ei jäeta.

2.2 Sõnavara õppeharjutused

Sõnavara õppimiseks ei leidu sarnaseid õpperakendusi või -mänge nagu grammatika omandamiseks, kuid on olemas üksikud harjutuslehed. Kõige põhjalikumad testid kesk- ja kõrgtasemele asuvad Tartu Ülikooli eesti keele (võõrkeelena) osakonna „Eesti keele ja kultuuri kursused“ leheküljel [20]. Materjalid on suunatud eesti keelt võõrkeelena õppijatele, kuid sobib ka emakeelena õppijatele.

Harjutus 2. Nimisõna

Kirjeldus, näpunäited ja õpetus

Valige lausesse sobiv nimisõna.

Sisu

1. Prügi --- suureneb iga aastaga ja tulevikus võib kogu maailm saada üheks suureks prügimäeks.
2. Uurimistöö --- avaldatakse ajakirja järgmises numbris.
3. Sportlane oli viimaste võistluste --- väga rahul.
4. Milliseid --- me sellest uurimusest võiksime teha?
5. --- toime pannud ametnik vabastati ametist.
6. Kes on tema --- selles ametis?
7. --- pealtnägijad pidid kirjutama politseis seletuskirja.
8. Politseiinspektor analüüsis hoolikalt sarnaseid ---.
9. Õnneliku --- läbi sain selle suurepärase töökoha.
10. Telefonisidefirma pakub uudset --- - hommikust äratust.

Kontrolli Alusta uuesti

Joonis 4. Kesktaseme nimisõnade harjutus lehel „Eesti keele ja kultuuri kursused“

Harjutused kesktasemele on nimisõnade, omadussõnade, tegusõnade, asesõnade, määrsõnade, kaassõnade ja sünonüümide tundmise testimiseks, kõrgtasemes homo-, paro-, süno- ja antonüümide, fraseologismide, vanasõnade ning sõnatuletuse peale, kusjuures kõrgtaseme testide

2. Harjutus

Kirjeldus, näpunäited ja õpetus

Igas lauses on üks sõna asendatud antonüümiga. Kirjutage lause ümber, kasutades õiget sõna.

Nt: Käsi peseb jalga. > Käsi peseb kätt.

Sisu

Teisi alt kuulata ei ole ilus. >

Ta teenib oma pühapäevast leiba vabrikutöoga. >

Talle meeldis tugevate meestega selga pista. >

Ta oli oma pere puhas lammas. >

Kontrolli Alusta uuesti

Joonis 5. Kõrgtaseme sünonüümide ja antonüümide harjutus lehel „Eesti keele ja kultuuri kursused“

seas on ka *varia* grupeering, kuhu alla kuuluvad erinevad väiksemad valikvastustega või lüinkharjutused. Harjutusetüpe on palju: lüinkvastused, valikvastused ning küsimuse ja vastuse omavahel vastavusse viimine (joonis 4 ja joonis 5).

Ülesannete sisu on staatiline, kuid seda korvab ülesannete kogus ja põhjalikkus. Kasutajale antakse tagasisidet, kuid õigeid vastuseid vea korral ei näidata, mis muudab mõned raskemad harjutused vähemkogenule lahendamatuks. Kasutajapõhist vigade meeldejätmist pole.

Lihtsamaid harjutusi leidub internetis veel, näiteks samatähenduslike sõnade harjutus [9] ja vastandtähendusega sõnade ristsõna [10], mis on lühikesed ja muutumatu sisuga. Sellised ülesanded on tihtipeale loodud ainult ühe aine või ainetunni raames ja on väga pealiskaudsed. Need testid kasutavad raamistikke, mis ei tööta igas veebilehitsejas või pistikprogrammides nagu eelmainitud ristsõna ja seetõttu on ligipääs nendele oluliselt piiratud.

2.3 Keelemäng grammatika ja sõnavara õppimiseks

Digitaalse keeleõppe materjali peamised tugevused on selle interaktiivsus, kasutajasõbralikkus, motiveerivus, õppimist toetav kohandatavus, kvaliteet, ligipääsetavus, käideldavus ja tehniline universaalsus [21], kuid valdavalt on internetis leiduvatel rakendustel või eelmises alapeatükis kirjeldatud lehekülgedel need vajalikud omadused suuremal või vähemal määral puudulikud. Ükski eelnevatest rakendustest või veebiharjutustest ei ole kasutajale kohanduv ja seega ei vasta õpikäsitusele [11], mis rõhutab individuaalset lähenemist.

Bakalaureusetöö käigus loodav veebimäng on tõsimäng (ingl *serious game*) ehk mäng, mille eesmärk pole meelt lahutada, vaid harida [22]. Mängu looja üritab parandada eelnevate programmide ja harjutuste puudujääke adaptiivsuse ja kasutajasõbralikkuse vallas, olles samas vastavuses ka teiste digitaalse keeleõppe materjalide nõuetega.

2.4 Kohanduvus

Tihedates, kuid lühikestes sessioonides teadmiste omandamine ja ülesannete lahendamine on tõestatud efektiivne õppemeetod [23]. Kindla ajaperioodi vältel kordamine kinnistab õpitud palju paremini, kui näiteks ühe õhtuga korraga õppides. Sellise jaotatud õppimisviisi teeb eriti kasulikuks, kui süstematiseeritult korrata harjutusi, mille puhul on eelnevalt eksitud [24]. Seda lähenemist õppimisele kutsutakse vahedega kordamiseks (ingl *spaced repetition*) ning see on eriti kasulik keeleõppes.

2.4.1 Leitneri süsteem

Vahedega kordamise rakendamiseks leidub algoritm, mille nimi on Leitneri süsteem [25]. Seda kasutatakse enim mälukaartidega õppides, kuid seda saab kohandada siinses töös olevate testide jaoks. Algoritm nõuab ülesannete jaotamist erinevatesse gruppidesse vastavalt nende raskusele õppija jaoks, kus igat gruppi korratakse erineva perioodi tagant. Ülesandele valesti vastates liigub see ühe võrra allapoole tihedamini korratavate ülesannete kategooriasse, õigesti vastates kõrgemasse harvemini nähtavavate kategooriasse. Üks versioon Leitneri süsteemist mälukaartidega töötab näiteks järgnevalt:

1. Kõik kaardid on üheskoos, ühtegi pole läbitud. Kaartide kõrval on N arv karpe, millel on oma arv 1-st N -ni. Õpilane võtab järjest kaarte ja üritab neile õigesti vastata. Kui vastus on õige, pannakse kaart kasti 1. Vastasel juhul läheb kaart paki alla. Nii korratakse, kuni kõik kaardid pakis on läbitud.
2. Järgmisel päeval alustatakse kordamist samast pakist, kui seal on kaarte (kui ei ole, algab süsteem järgmise kasti juures). Õigesti vastatud kaardid liiguvad 1-te ja valed jäävad samasse pakki. Kui pakk läbitud, korratakse sama kastis 1 olevate kaartidega, kuid seekord lähevad õiged vastused kasti 2 ja valed tagasi esimesse pakki.
3. Kolmandal päeval alustatakse kordamist jälle algsest pakist, kui seal on kaarte (vastasel juhul liigub süsteem edasi kasti võrra jne). Korratakse sama protsessi nagu eelmistel päevadel. Seekord aga ei minda järgmise sammuna kast 1 kaartide juurde, vaid kast 2 juurde, kui seal on kaarte. Õigesti vastatud küsimused liiguvad kastist 2 kasti 3.
4. Neljandal päeval alustatakse jälle esimesest pakist. Seejärel vaadatakse läbi karp 1. Viimasena läbitakse karp 3 jne.

Selles näites läbitakse esialgset pakki iga päev, karpi numbriga 1 iga päeva tagant, karpi numbriga 2 iga kahe päeva tagant, karpi 3 iga kolme tagant jne. Kategooriate perioode võib muuta vastavalt soovile.

3. Kasutatavad keeleressurssid ja keeletehnoloogilised vahendid

Bakalaureusetöö praktilises osas kasutatakse testide loomiseks eesti keele jaoks koostatud tehnilisi vahendeid, mis võimaldavad automaatselt tekste analüüsida ning võrrelda. Need ressursid on saadaval Pythoni teekidena, mistõttu on veebirakenduse tagasüsteem loodud selles keeles. Sõnavaratestide loomise põhjaks on tekstikorpused, kust võetakse välja sobivad laused. Grammatikatestidena kasutatakse Anneliis Hallingu bakalaureusetöö tulemusena loodud rakenduse repositooriumist failist „koik_laused.xml“ [26].

3.1 EstNLTK

Eestikeelsete tekstide keeletötlusvahend EstNLTK on avatud lähtekoodiga teek, milles on olemas morfoloogiline analüsaator, süntaktiline analüsaator, lemmatiseerija, *wordnet* jpm [12]. Siinses töös on võtmetähtsusega automaatne teksti lausestamine, sõnestamine, lemmatiseerimine ja sõnaliikide määramine, mida võimaldab teha EstNLTK-s morfoloogiline analüsaator. Morfanalüsaatoriga saab määrata tekstis sõnaliike, sõnavorme ja struktuuri. Lisaks sellele on rakenduses kasutatud ka EstNLTK versioonis 1.4 olemasolevaid funktsioone TEI-formaadis korpuste sisselugemiseks ja Eesti Wordneti liidest sõnatähenduste määramiseks ja sünonüümide valimiseks.

Eesti Wordnet ehk EstWN on andmebaas, kus hoitakse eestikeelseid mõisteid koos nende vaheliste semantiliste ühendustega. Igale sõnale *wordnet*-tüüpi sõnastikus vastab sünonüümihulk ehk sünohulk, mis koosneb ühte mõistet väljendavatest sünonüümidest [13]. Lisaks sünonüümiale hoitakse ka teisi semantilisi suhteid nagu hüpo- ja hüpernüümia, antonüümia, holonüümia jne.

3.2 Korpused

Tekstikorpus on struktureeritud andmekogum, mis on arvutile töödeldavas vormis [27]. Eestikeelsed korpused on kättesaadavad Eesti Keeleressursside Keskuse META-SHARE repositooriumis [28]. Keelemängus on testide genereerimiseks autor kasutanud kahte korpust:

1. eesti ajakirjanduse korpust, mis sisaldab näiteks Eesti Päevalehe, Eesti Ekspressi, Äripäeva, Maalehe ja Postimehe väljaandeid aastavahemikus 1990 kuni 2007. Kokku on korpuses 182 miljonit sõna [29];
2. eesti ilukirjanduse korpust, mis sisaldab ilukirjanduslikke teoseid aastast 1990. Kokku on korpuses 5,6 miljonit sõna [30].

Mõlemad korpused on märgendatud kasutades TEI (*Text Encoding Initiative*) juhtnööre [31] ning on juba lemmatiseeritud. Korpustest on vaja parameetrite järgi sobivad laused välja sorteerida. Antud bakalaureusetöös kasutati korpuste filtreerimiseks Jelena Kallase jt [32] poolt GDEX tööriista jaoks kokkupandud reeglistiku kohandust:

- lause lõpeb kas punkti, küsimärgi või hüüumärgiga;
- lause pikkus jääb 3 ja 10 sõna vahele;
- lauses pole ühtegi sobimatut sõna
- lauses pole ühtegi kaldkriipsu, semikoolonit, kantsulgu, sulgu, alakriipsu, võrdlusmärgi ega täрни;
- lauses on paarisarv jutumärke.

Reeglistik kindlustab, et testidesse jäävad õppetöökõs sobilikud laused, milles pole ühtegi soovimatut sõna ega märki ja mis on paraja pikkusega, et neid analüüsida ja nende põhjal intuiitiivselt järeldusi teha. Võrdlusmärkide reegel eemaldab korpustest kogemata sisse jäänud HTML-elementid. Lausetest eemaldatakse ebavajalikud tühikud ning vale vorminguga jutumärgid asendatakse programmi jaoks sobilikega.

3.3 Ebasobivate sõnade loend

Ebasobivate sõnade loendi on koostatud Kristina Koppel oma doktoritöö jaoks [33]. Loetelus on 912 eesti-, inglisi- ja venekeelset õppetöö konteksti mittesobivat sündsusetut, argikeelset ja mõttetühja sõna või slängi.

3.4 Gensim

Gensim on avatud lähtekoodiga Pythoni teek, mis võimaldab analüüsida tekstide semantilist struktuuri ja teha nende põhjal ennustusi [14]. Gensimi teegis on erinevad masinõppe algoritme kasutavad mudelid nagu Word2Vec, Doc2Vec, FastText jt, millest näiteks esimesed kaks põhinevad neurovõrkudel. Siinses töös on tähtis sõnavara testides leida automaatselt pakutavaid vastusevariante ning Word2Vec on kiire, skaleeruv ja kergesti sätestatav mudel selle eesmärgi täitmiseks. Mudelile saab anda sisendiks lausete loendi ning selle põhjal ennustada, kui sarnased on kaks erinevat sõna vastavalt neid ümbritsevatele sõnadele loendis.

4. Keelemäng

Bakalaureusetöö käigus loodi veebirakendus, mis võimaldab õppida ja süvendada eesti keele tundmist². Selleks pakutakse kasutajale valik kahe testitüübi vahel – grammatika või sõnavara. Nii grammatika- kui ka sõnavaraülesandeid valitakse kasutajale kohanduvalt, arvestades nende eelnevaid sooritusi. Grammatikatestid, mis on Keelemängu toodud Anneliis Hallingu bakalaureusetööst, sobivad tema sõnul kõigile vanuseklassidele, kuigi mõned neist võivad olla esimesele ja teisele kooliastmele liiga keerulised [7]. Sõnavaratestidesse automaatselt valitud sõnadest mitmed ei kuulu põhikooli sõnavarasse [34] ja seega võivad olla liiga rasked õpilasele esimeses kolmes kooliastmes, küll aga sobivad kõik testid juba vanematele kasutajatele.

4.1 Sõnavaratestid

Rakenduse sõnavaratestide funktsionaalsus piirdub ainult sünonüümide ülesannetega. Ülesannete ülesehitus on selline: kasutajale on antud näitelause, küsitav sõna ja valikvastused. Küsitav sõna on alati nimetavas käändes. Valikvastused, mida on neli, on erinevad küsitavale sõnale sarnased sõnad koos õige vastusega. Näiteks üks suvaline sünonüümitest:

Näitelause: „Ta keeras oma vestinööpi ja yabises silmnähtavalt.“

Vali sõnale **vabisema** õige sünonüüm.

Valikvastused on „keerama“, „lõdisema“, „valutama“ ja „taanduma“, millest „lõdisema“ on õige.

4.2 Grammatikatestid

Keelemängu grammatikaülesanded võimaldavad õppida kõiki eesti keele käändeid. Käändetestide ülesehitus on sama, mis sünonüümiülesannetelgi: näitelause konteksti pakkumiseks, küsitav sõna ja valikvastused. Küsitav sõna on käänatud vastavalt sellele, kuidas seda näitelauses on kasutatud, näiteks:

Näitelause: „Ma sulgesin silmad.“

Mis käändes on sõna **silmad**?

Valikvastused on eesti keele neliteist käänat, millest on alati üks õige. Olukordades, kus sõna on mitmes käändes samal kirjakuju, aitab näitelause kindlaks teha, millist tegelikult vastuseks

² Keelemäng asub aadressil <http://167.71.3.184/> ja repositoorium aadressil <https://gitlab.com/janna98/keeleoppebakatoo-ut>

oodatakse. Kui näitelausest ei piisa, saab kasutaja sellest ülesandest teavitada, sest see on liiga segadusttekitav, et jääda õppemängu. Rakendus ei arvesta, kas käändetestis küsitav sõna on ainsuses või mitmuses.

4.3 Kasutatud tehnoloogiad

Keelemängu ees- ja tagarakenduses kasutati erinevaid tehnoloogilisi lahendusi. Kasutajaliides loodi Angular platvormiga, sest töö autoril on sellega varasem kogemus. Angulariga koos kasutati Ng-Zorro [35], mis on vabavaraline raamistik ja pakub võimalust kasutada eelloodud TypeScript ja CSS-i komponente. Ka Ng-Zorroga on töö autoril varasem kokkupuude. Tagarakendus loodi Flask raamistikuga, sest see kasutab programmeerimiskeelt Python ja seega saab rakendust jooksutades lihtsasti välja kutsuda vajalikke Pythoni programme korpuste eeltöötlemiseks, testide loomiseks ja valimiseks. Keelemängu andmebaas on PostgreSQL, sest see on vabavaraline.

4.4 Auth0 volitussüsteem

Auth0 on volitusplatvorm [36], mida saab ühendada erinevate raamistikega ja kasutada mitmes programmeerimiskeeles. Siinse töö autor valis selle platvormi, sest see on tasuta ja sellesse on sisse ehitatud Angulari tugi. Lisaks sellele võimaldab Auth0 autoril vältida uue volitussüsteemi loomist, mis on harilikult vigaderohke protsess.

Seda volitussüsteemi kasutades saavad rakenduse kasutajad registreerida ja sisse logida erinevate sotsiaalmeedia kontodega nagu Facebook ja Twitter või e-posti teenustega nagu Gmail ja Outlook. Siinses töös kasutati ainult Gmaili kontoga või vabalt valitud e-posti aadressi ja parooliga autentimist, kusjuures viimase puhul saadetakse antud aadressile ingliskeelne kinnituskiri. Iga uue kasutaja registreerimisel määrab Auth0 kasutajale unikaalse identifikaatori, mille Keelemäng jätab meelde enda andmebaasi, et tagada ühilduvus volitussüsteemi ja rakendusesiseste kasutajaga seotud protsesside vahel.

4.5 Veebirakenduse ülevaade

Veebirakenduse kasutajaliidese loomisel oli eesmärk minimalistlik disain ja kasutajasõbralikkus, et õppimisprotsess oleks võimalikult müravaba. Sellest johtuvalt on taust heledate toonidega ja tekst ning menüüriba tume, et tekiks tugev kontrast, ning sisu paigutatud vertikaalselt ja horisontaalselt lehe keskele.

Oled Keelemängu avalehel.

Keelemäng on Sinule kohanduv veebirakendus eesti keele õppimiseks. Siin saad kontrollida ja täiendada oma teadmisi käänete ja sünonüümide vallas.

Grammatikatestid on Anneliis Hallingu bakalaureusetööst ja sünonüümitestid on automaatselt genereeritud Keelemängu jaoks.

Keelemängu kasutamiseks on Sul vaja kasutajat.

Logi sisse või loo kasutaja:

Logi sisse

Joonis 6. Keelemängu sisselogimisala

Tere, Jannaliina

Õppimise alustamiseks vali test ja lausete arv:

5 10 15

Grammatika

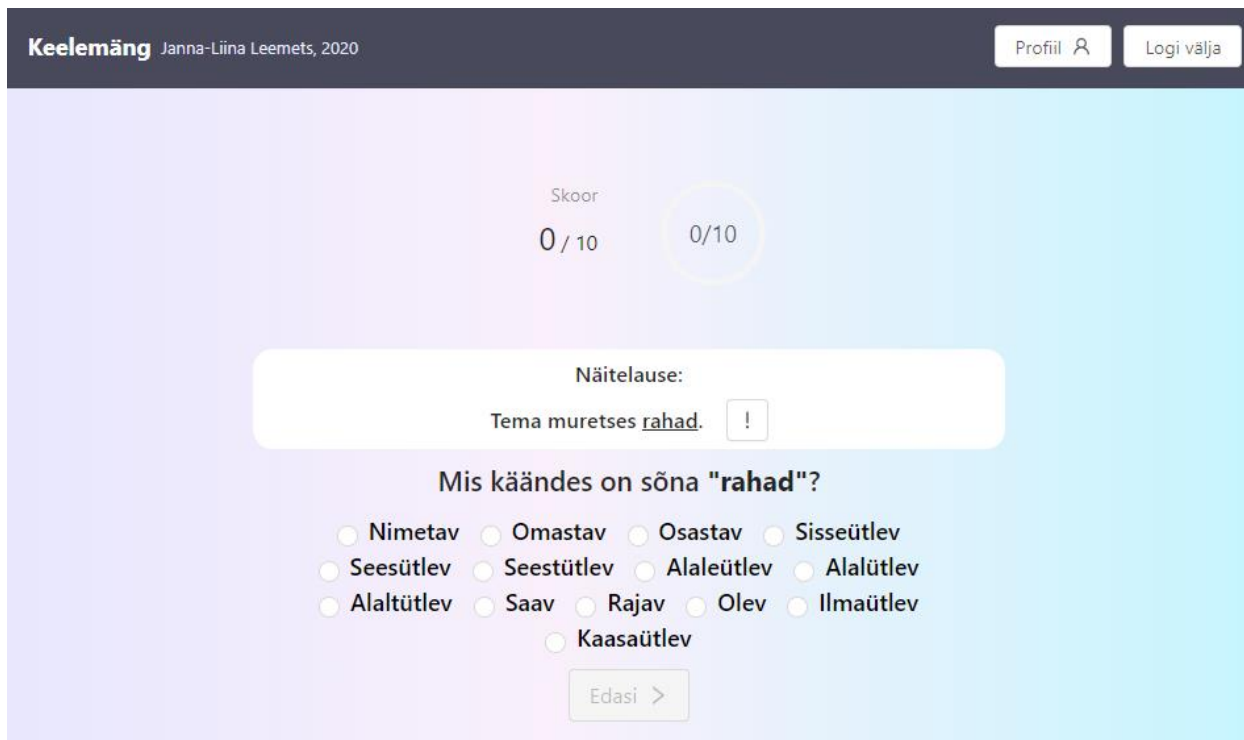
Sõnatähendused

Joonis 7. Keelemängu pealeht

Rakenduse avaleht on sisselogimisala (joonis 6), kus asub rakenduse tutvustus. Kasutaja juhitakse sealt rakenduse kasutamiseks edasi Auth0 volitussüsteemi esmakordseks registreerimiseks või

sisselogimiseks. Järgmise sammuna peale kasutaja autentsuse kinnitamist saadetakse õppija automaatselt tagasi Keelemängu lehele.

Sisselogituna näeb kasutaja pealehte, kus asub valik testitüüpide ja ülesannete arvu vahel (joonis 7). Teste saab lahendada 5, 10 või 15 komplekti kaupa. Autendituna on võimalik igal hetkel näha enda profiili, selleks peab vajutama vastavat nuppu „Profiil“ menüüribal. Väljalogimine viib tagasi joonisel 6 kujutatud avalehele.



Joonis 8. Ühe grammatikatesti leht kümnest

Kummalegi ülesandetüübi nupule vajutades juhitakse kasutaja testimislehele (joonis 8), mis on mõlemal juhul välimuselt samasugune. Menüüribast allpool näidatakse jooksvalt skoori ja läbitud lausete arvu. Iga õige vastuse eest on võimalik saada üks punkt. Seisundiala all asub valges kastis esile tõstetud näitelause, kus küsitav sõna on alla joonitud. Järgmisel real on paksemas ja suuremas kirjas küsimus ja sellele järgnevad valikvastused.

Korpuse põhjal automaatselt genereeritud ülesanded võivad olla ebakorrektsed või vigase näitelausega. Sellisest olukorrast teadaandmiseks on näitelause kõrval hüüumärgiga nupp. Seda vajutades avaneb dialoogiaken, kus küsitakse teavituse saatmiseks kinnitust. Ülesande, mille näitelause või vastus on valeks märgitud, jätab programm meelde ja lisab automaatselt tekstifaili tema identifikaatori. Seejärel loetakse testid uuesti sisse, jättes välja kõik selles failis ülesloetud

harjutused, ja eemaldatakse seda testi refereerivad read andmebaasist. Viimase sammuna kustutatakse see ülesanne lahendamisel olevate testide hulgast ja võimalik saadav skoor selles voorus väheneb ühe võrra. Kui kõik testid hetkekomplektis on sobimatud ja neist on järjest teavitatud, värskendatakse leht ja valitakse uued.

Nupp „Edasi“ on blokeeritud, kuni kasutaja valib ühe vastusevariantidest. Seda tingimust aitab kasutajale selgitada kursoriga sama nupu kohal olles avanev hüpikvihje. Järgmise testi juurde liikudes uueneb õige vastuse korral skoor ja lausetearvu indikaator liigub ühe võrra edasi.

Keelemäng Janna-Liina Leemets, 2020 Profil Logi välja

Tulemused:
Skoor: 3/5

[Mängi uuesti?](#)


[Avalehele](#)

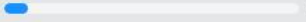
Sõna	Lause	Õige sünonüüm	Sinu vastus
varahommik	Eriti kaugemale ulatus see varahommikul ja hilisõhtul.	varavalge	varavalge ✓
turnima	Restauraatorid turnivad sibulkupli peal kui Moskva tsirkuse artistid.	kippuma	tõusma ✗
nuustik	Issand, kes siis nii kareda nuustikuga hõõrub!	tüügas	tüügas ✓
majesteetlik	Väga majesteetlikud puud olid suure metsa kased.	aristokraatne	aristokraatne ✓
rantsatama	Noatera tungis rantsatades puusse otse mu põse kõrval.	pistma	vajuma ✗


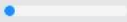
Joonis 9. Tulemuste vaade

Kui kõik ülesanded läbitud, juhitakse kasutaja tulemuste lehele, kus on tabel küsitud sõna, näitelause, kasutaja antud ja õige vastusega. Selguse mõttes asub korrekse vastuse puhul veerus „Sinu vastus“ sõna või käände taga roheline kinnitusmärk, vale puhul punane rist. Kasutajale pakutakse võimalust mängida sama arvu lausetega sama tüüpi testi uuesti või minna avalehele, kus saab tüüpi ja testide arvu muuta (joonis 9).

Jannaliina



Valesti vastamise üldine protsent:
 8% vastustest valed

Valesti vastamise jaotatud protsent:
 8% sünonüümitestide vastustest valed
 8% grammatikatestide vastustest valed

Sõnad, mille sünonüümide valimisel oled enim eksinud:

1 korda	1 korda	1 korda
KIPPUMA ▾	PISTMA ▾	TÕUSETUMA ▾

Käänded, millega oled enim eksinud:

1 korda	1 korda
NIMETAV	OSASTAV

[Tagasi avalehele](#)

Joonis 10. Kasutaja profiili vaade

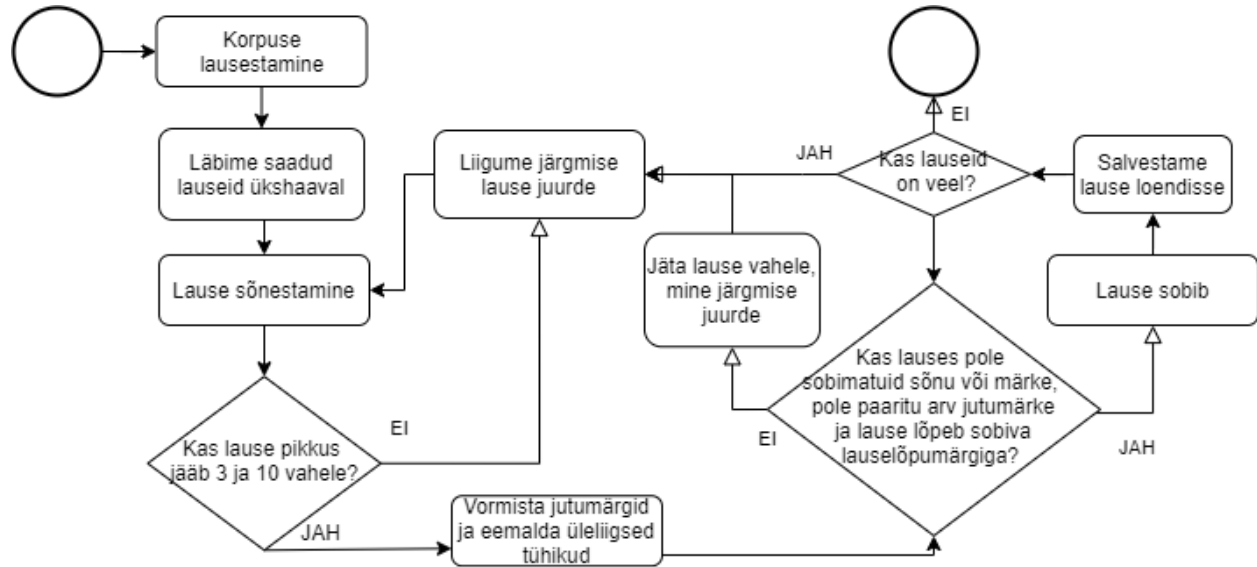
Lehel „Profiil“ kuvatakse kasutajale tema kontoga seotud profiilipilt ja statistika (joonis 10). Esimesena näidatakse üldist valesti vastamise protsenti üle kõikide testide. Protsente illustreerib edenemisnäitur, kus sinine osa kujutab protsenti tervikust. Samasuguse kujundusega on järgmistel ridadel eraldi grammatika- ja sünonüümitestide valesti vastamise protsent. Protsentidele järgneb sõnade statistika, mille puhul on enim eksitud. Esimesena on reas näha sünonüümide esikolmik koos nende vastu tehtud eksimuste arvuga. Vajutades ikoonile sõna kõrval laieneb värvitud kast allapoole, kus on toodud üks näitelause, milles see sõna on esinenud. Samamoodi on järgmine kastiderida esikolmik käänetest, mida kasutaja on kõige rohkem sassi ajanud, kuid need ei avane.

4.6 Õppemängu sünonüümitestide loomine

Sünonüümitestid valiti eelnevas peatükis nimetatud aja- ja ilukirjanduskorpustest. Ajakirjanduskorpuses leidub termineid, mis on keerulised, kuid tõsielulised, ja nendest saab luua Keelemängu raskemaid küsimusi. Ilukirjanduskorpuse sõnavara on mitmekesine ja tagab Keelemängu

sünonüümitestide piisava hulga. Ilukirjanduslikud laused on ka huvitavad, mistõttu köidavad nad tähelepanu ja ei lase mängul igavaks muutuda.

4.6.1 Korpuste eeltötlusalgoritm



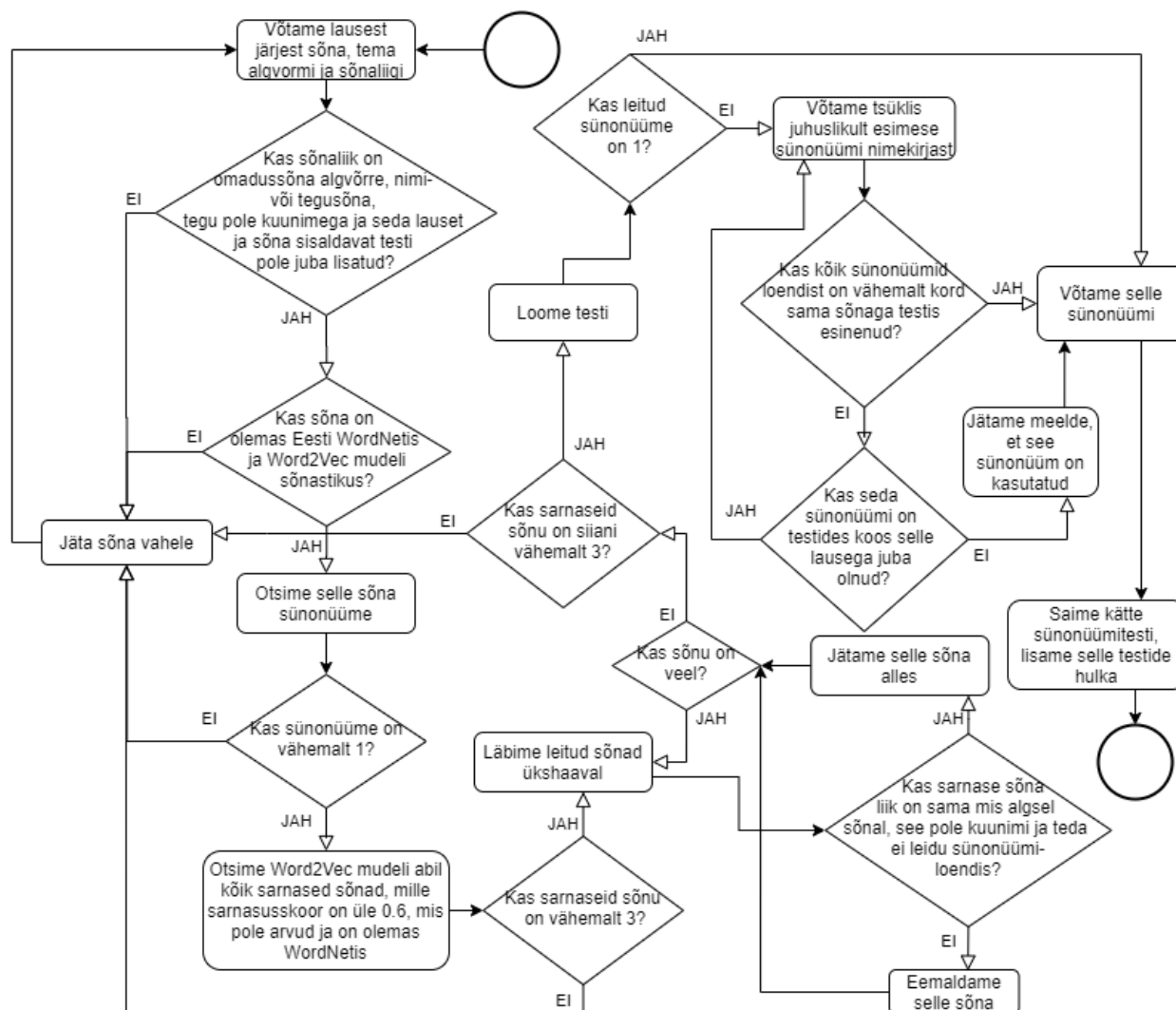
Joonis 11. Korpuste eeltötlusalgoritm

Korpuste eeltötlusalgoritm (joonis 11) esimese sammuna lausestab ja sõnestab korpused EstNLTK morfoloogilise analüsaatoriga ning filtreerib siis välja sobivad laused vastavalt loodud reeglitele. Filtreerimisprotsessi käigus jälgitakse, et ükski lause, milles on ebasobivate sõnade loendisse kuuluv sõna, ei satuks kasutatavate lausete hulka. Kui kõik laused on valitud, antakse nende loend Gensimi Word2Vec mudelile, treenitakse ja salvestatakse see.

4.6.2 Sünonüümitestide loomine

Sünonüümitestide loomisalgoritmi jooksutatakse üks kord eeltötluse käigus enne Keelemängu töölepanekut (joonis 12). Siinkohal on oluline mainida, et korpuste eeltötluse käigus loodud Word2Vec mudel, mida siin algoritmis rakendatakse, märkas ajakirjanduskorpuse olemuse tõttu mitmeid sidemeid kuunimede ja erinevate sõnade vahel ning pakkus neid sarnaste sõnade hulka. Need seosed tulid ilmsiks juba esimestel juhuslikel katsetustel mudeliga teste luua, sest kuunimedega testid moodustasid alati tulemusest suure osa. Küll aga ei sobi kuunimed sünonüümitestidesse, sest vastusevariantidena harilikele omadus-, nimi- ja tegusõnadele on nad ebaadekvaatsed ning isegi sünonüümide leidumisel ei ole need piisavalt tähendusrikkad (näiteks

Eesti Wordnet annab sõnale „pühapäev“ sünonüümiks „P“). Algoritmis kontrollitakse seetõttu tihti kuunimede olemasolu, et vältida nende sattumist testidesse ühelgi kujul.



Joonis 12. Sünonüümitestide loomisalgoritmi ühe sõna töötlemismuster

Algoritm alustab korpuste eeltötlusalgoritmi tulemusena saadud lausete loendi tsüklilise läbimisega. Eesmärk on lause ja kõigi sobivate sõnadega selles luua eraldi testid, kus esimesest saab näitelause ja teisest küsitav sõna. Nagu näha joonisel 12 vaadatakse selleks ükshaaval läbi iga lause kõik sõnad. Esimese sammuna leitakse EstNLTK morfanalüsaatori abil sõna algvorm ja sõnaliik. Algvõrdes omadus-, tegu- ja nimisõnadele, mis pole kuunimed ja mida pole koos hetkel läbitava lausega veel testina olemas, otsib algoritm sünonüümid, kasutades EstNLTK Eesti Wordneti liidest.

EstWN tagastab sõna puhul kõik selle homonüümsed sünohulgad, kui neid leidub. Selliste homonüümide tähenduste automaatne eristamine ja neist hetke lause kontekstis õige valimine aga on eraldi probleem ning jäi siinse töö skoobist välja. Seetõttu jätkab algoritm tööd ainult nende sõnadega, millele EstWN tagastab ühe sünohulga. Sünohulk on ühendatud kõikide selle sõna sünonüümidega ning kui neid sünonüüme on vähemalt üks, siis jätkatakse selle sõnaga tööd. Sellest sõnast saab testis küsitav sõna.

Word2Vec valmismudelisse sisseehitatud sarnasusfunktsioonidega leitakse küsitavale sõnale teisi sarnaseid sõnu, mille sarnasusskoor on vähemalt 0,6 ja mis pole arvud, lühendid jms (ehk kas nad olid olemas EstWN-s). Lävend 0,6 on piisavalt kõrge, et tagada küsitavale sõnale tugevalt sarnaste sõnade valiku, ja niivõrd madal, et neid sõnu oleks vähemalt 3. Tulemusloend läbitakse ükskõik kontrollimaks, kas:

1. sarnase sõna sõnaliik kattub küsitava omaga;
2. sõna ei kuulu sünonüümide hulka;
3. sõna ei ole kuunimi.

Kui peale kontrolli jääb alles ikkagi vähemalt kolm sarnast sõna, siis võib testi hakata kokku panema. Selleks on vaja valida välja õige vastus. Kui sünonüüme on üks, saab sellest õige vastus. Kui neid on üle ühe, valitakse neist esimene suvaline, mida pole üheski testis veel vastusena kasutatud. Kui kõik on juba kord testides esinenud, siis sobib ükskõik milline juhuslik valik. Kui kõik ülesandeks vajalikud osad on olemas, siis salvestatakse loodud test ja liigutakse lauses järgmise sõna juurde. Kui kõik need läbitud, siis korratakse sama järgmise lausega loendist.

Sünonüümitestide algoritm töötab korruga mitmeid tunde, mistõttu kõikide testide failis hoidmine on hädavajalik samm, et Keelemängu tööaega kokku hoida. Kõik ülesanded kirjutati tekstifaili järgmisel kujul: lause unikaalne identifikaator, näidislause, õige vastus, sarnaste sõnade loend, õige vastus. Kõik testid loetakse failist sisse tagarakenduse esimesel käivitamisel ja on kättesaadavad igal hetkel loendi kujul.

4.7 Keelemängu kasutajale kohanemise algoritm

Siinses töös kasutatakse 2.4.1 kirjeldatud Leitneri süsteemi lihtsustust, kus igale läbitud ülesandele antakse hinne (ehk kategooria). Ajaliselt aga kasutaja testimisharjumusi ei jälgita ja hinnetel pole oma päeva või ajahetke, mil need esineda võivad, seega igal hetkel koosneb test samade hinnetega

küsitavatest sõnadest. Hinded salvestatakse andmebaasi kohe peale ülesannete läbimist. Andmebaasis hoitakse järgmiseid andmeid:

- antud küsitava sõna hinnet ehk kui hästi on kasutaja tundnud selle sõna sünonüüme või saanud aru, mis käändes see on;
- iga küsitava sõna kohta tema testide identifikaatoreid, mida õppija on juba läbinud, et programm ei valiks neid teste mitu korda uuesti, vaid leiaks uusi sarnaseid;
- hinde omaniku Auth0 identifikaatorit;
- käänete testi puhul ka küsitava sõna kääned, mille osas täpsemalt kasutaja vea tegi, et saaks selle käände ja küsitava sõnaga teste otsida.

Võimalikud hinded on 1, 2 ja 3, kus hindegaga 1 teste nähakse kõige tihedamini ja hindegaga 3 kõige harvemini. Esimesel korral õigesti vastates antakse testile hindeks 2, valesti vastates 1. Järgmistel kordadel muutub hinne vastavalt korrektsusele nagu Leitneri algoritmis.

Algoritm alustab tööd vastavalt kasutaja poolt valitud testitüübile ja arvule. Andmebaasist päritakse selle kasutaja kõik hinded, küsitavad sõnad ja juba nähtud ülesannete identifikaatorid. Saadud loend järjestatakse juhuslikult, et nende järgi otsitavad testid ei tuleks alati samas järjekorras. Olenevalt valitud ülesandetüübist kutsutakse välja vastav Leitneri süsteemi lihtsustuse funktsioon.

Sõnavaratestide valiku funktsioon, mida on näha joonisel 13, on aluseks ka käändetestide valiku meetodile. Kõigepealt kontrollitakse, kas kasutajal on ühtegi hinnet. Kui ei, siis valitakse tsükliliselt suvalised erinevad testid, kuni soovitud arv on täis. Kui jah, siis arvutatakse välja, mitu ülesannet iga hinne kohta võetakse. Juhul, kui kasutaja tahab teha 5 või 10 ülesannet, siis on hinde kaupa jaotumine selline:

1. sõnu, mille hinne on 1, küsitakse enim ehk 60% testide küsitavatest sõnadest on sellest kategooriast;
2. sõnu, mille hinne on 2, küsitakse vähem ehk 20% teadmiste kinnistamise eesmärgil;
3. sõnu, mille hinne on 3, ei küsita nii vähesel testide arvu puhul üldse;
4. sõnu, mis on täiesti uued (kui neid on), küsitakse sama palju kui hindegaga 2 ehk 20%.

on sõna kindlasti kasutajale selge ning ta võib seda küsimust näha väikese tõenäosusega uuesti ainult 15 lauset valides.

Kui ülesannete arvuline jaotus on leitud, hakkab algoritm hinde kaupa ülesandeid otsima, alustades kõige tihedamini korratava kategooriaga 1. Siin tuleb esile käände- ja sünonüümitestide erinevus: käänete puhul otsitakse kõigepealt teste, kus on sama küsitav sõna ja temaga kaasa käiv andmebaasist pärit kääne. Kui selliseid ei leitud piisavalt, et hindegaga 1 ülesannete jaotus täis saada, otsitakse ainult küsitava sõna järgi, nagu seda tehakse sünonüümitestide puhul. Jaotuse tühjaksjäämisel lisati viimases sammus nii palju veel nägemata ülesandeid, et vajalik arv teste täis saada. Sama protsessi korratakse ka hinnete 2 ja 3 puhul. Kui kasutaja on kõiki teste näinud, valitakse uued suvaliselt, kuid selline olukord juhtub harva, sest sünonüümitestide on ligikaudu 830 ja käändetest 8380. Saadud loend tagastatakse eesrakendusele.

4.8 Rakendusele antud tagasiside

Õppemängu testiti üks kord 10 inimese poolt. Testijatel paluti registreerida konto ning proovida mõlemat testitüüpi kaks korda järjest kas 10 või 15 ülesandega. Seejärel tuli täita küsimustik, mida on näha lisas 1. Küsimustik koosnes neljast alaosast: esimeses küsiti inimese taustaandmete, teises käändetestide, kolmandas sünonüümitestide ja neljandas Keelemängu kohta. Vastanutest kõik rääkisid eesti keelt emakeelena ja olid kas 21-, 22- või 24-aastased. Pärast testimist lahendati leitud probleemid ning muudeti kasutajaliidest ja ülesannete valimisalgoritmi vastavalt soovitudele.

Testimise käigus esines kaks tehnilist viga: kaks kasutajat ei saanud programmi registreerida ja samal kahel kasutajal jooksis programm lõputusse tsüklisse, kui prooviti teavitada sobimatust lausest. Registreerides ei salvestanud Keelemäng nendele Auth0 poolt määratud identifikaatorit oma andmebaasi, kuid kasutajad olid siiski Auth0 haldussüsteemi sisse logitud ja said ligi autentimist nõudvatele rakenduse vaadetele. Kasutaja identifikaatori puudumine tekitas programmis vigu ning kõik katsed tagasüsteemist andmeid pärida või sinna saata ebaõnnestusid. Need vead väljendusid sobimatust lausest teavitades lõpmatu tsükli. Vigane registreerimispäring tagasüsteemi sai parandatud peale tagasiside saamist. Kahjuks ei kajastunud programmi kohanduvus nende kasutajate jaoks, sest nende tulemusi ei saanud andmebaasi salvestada ilma vajaliku identifikaatorita, küll aga said nad anda tagasisidet muus osas.

Grammatikatestides pidas 9 inimest 10-st näitelauseid arusaadavaks. Kõik peale kahe inimese vastasid nende ülesannete jooksul vähemalt kord valesti. Need, kes vastasid valesti, nägid vigaseid

sõnu järgmise vooru jooksul uuesti. Õigesti vastatud küsimusi (hindegaga 2) nähti ka uuesti, kuid mõne inimese jaoks liiga tihti. Mitu inimest kiitis õigesti vastatud sõnade teist korda nägemist, sest see aitab teadmisi kinnistada.

Sünonüümitestides pidas 8 inimest 10-st näitelauseid arusaadavaks, kuigi üks neist arvas, et näitelauseid tegelikult küsimusest paremini aru saada ei aidanud. Üks inimene, kes pidas näitelauseid arusaamatuteks, arvas, et nad siiski on kasulikud ülesande mõistmisel. Kõik peale ühe inimese vastasid nende ülesannete jooksul valesti. Need, kes vastasid valesti, nägid vigaseid sõnu järgmise vooru jooksul uuesti, kuigi ka siin oli ühine arvamus, et õigesti vastatud (hindegaga 2) teste korrati liiga palju.

Vigastest lausetest teavitati kokku 16 korda. Laused, millest anti teada, olid vaatamata korpuste filtreerimisele siiski vigase ülesehituse, arusaamatu näitelause või väära vastusega. Vale vastusega küsimust kohates teadis kasutaja õiget vastust ja märkas programmi eksimust, kui nad ei saanud vastuse eest punkti. Väära vastusega teste näidati sellisel juhul järgmises voorus uuesti, mis andis võimaluse kasutajal sellest autorit teavitada. Sellised valed vastused tulenevad sünonüümitestide puhul sõnatähenduse valesti märkimisest ja grammatikatestide puhul morfoloogilise ühestaja ebakorrektselt tööst [7].

Kõik testijad nõustusid, et adaptiivsus on e-õppes väga kasulik omadus. Idee täitmisel oli Keelemäng pigem edukas, kuid sooviti, et õigesti vastatud ülesanded korduksid vähem ja esimest korda nähtud testid ei kattuks omavahel niivõrd palju, näiteks ühel kasutajal sisaldas 10-st testist 5 küsitavat sõna „alustama“. Kasutajaliidese kujundust peeti üldiselt selgeks ja õppeprotsessi mittehäirivaks. Kiideti ka „Profiil“ vaadet, mis andis sisukat infot kasutaja õppimisprotsessi kohta. Paar kasutajat soovis, et küsimuste raskusaste oleks olnud ühtlasem.

Tagasisidest lähtuvalt muutis autor testide valimisalgoritmi. Varasem algoritm oli ehitatud üles nii, et olukorras, kus ühe hindegaga sõnu ei leitud jaotuse täitmiseks piisavalt, korvati seda puudujääki järgmise astme hindegaga ülesannetega. Selline loogika asendati lõpptulemuses oleva algoritmiga, kus iga hindegaga sõnu on maksimaalselt sellele hindele antud jaotuse jagu ning puudujäägid asendatakse võimalusel kas täiesti uute testidega või, kui kõik testid on nähtud, suvaliste testidega.

4.9 Probleemid

Töö käigus loodud programmi üks eesmärk oli pakkuda sõnavarateste, mis ei jääks pealiskaudseks. Sünonüümitestide loomise käigus aga tekkis probleem sõnatähenduste määramisel, sest EstWN tagastab korraga sõna kõik homonüümid. Automaatne homonüümide tuvastamine väärib eraldi tähelepanu ja jäi bakalaureusetöö mahu tõttu käsitlemata. Genereeritud teste on seetõttu küll piisavalt palju, kuid need on sõnavaliku poolest piiratud.

Korpusest pärit lauseid küll filtreeriti reeglistikuga ja ebasobivate sõnade loendiga, kuid siiski jäi testi alles ebakorrekse struktuuri või sisuga lauseid. Reeglistikku oleks saanud laiendada piisava aja leidmisel, kuid seegi ei aitaks eemaldada ebasobivaid laused täielikult. Töö autori lahendus sellele probleemile on kasutajale antud võimalus sellistest lausetest teada anda.

Word2Vec mudeli pakkumised sarnaste sõnade otsingul jäid mõnikord sobimatuteks. Näiteks pakkus see sõnale „pruun“ valikvastusteks sõnad „rasvane“, „roheline“ ja „roosa“, millest esimene pole loogiline ja teised kaks on segased, sest mitmel juhul saaks kõik kolm värvi näitelauseesse samale kohale sobida. Üks selline näitelause korpustest, mis jõudis testidesse, on näiteks „Küll _____ värvi, aga siiski nahk“.

4.10 Edasiarendusvõimalused

Siinses töös oli põhieesmärk automaatselt testide loomine ja kasutajale kohanduva keelemängu loomine. Leidub erinevaid viise õppemängu kvaliteetsemaks, mängulisemaks ja kohanevamaks muutmiseks. Adaptiivsuse saavutamiseks kasutati lihtsustatud varianti Leitneri süsteemist, kus ei peetud järge aegade üle, millal kasutaja mingi hinde saanud testi sooritas. Algoritmi saaks muuta efektiivsemaks, kui jälgida täpset ajakava testide valikul, näiteks valides hindega 1 sõnadega ülesandeid iga päev, hindega 2 üle päeva jne. See võimaldaks kasutaja valitud testide arvu täita paremini sellel ajahetkel kordamist vajavate ülesannetega.

Sünonüümitestidele saab lisada anto-, paro-, homonüümi jms tundmist testivaid harjutusi. Anneliis Hallingu sõnul [7] saaks näiteks lisada pööramise teste, kus ma-tegevusnimes sõna peaks panema antud pöördesse, või lisada käänetestidele raskusastmeid, kus oleks mitu küsitavat sõna. Olemasolevates käänetestides saaks pakkuda kategooriatena nimisõnade, omadussõnade, arvsõnade ja asesõnade käänamist. Grammatikaülesannetele saab lisada kirjavahemärkide teadmise kontrole. Mängulisuse suurendamiseks saab pidada järge testi sooritamise ajast ja seda piirata.

Genereeritud sünonüümitestid olid väga erinevad oma raskusastmelt. Töö mahu tõttu jäi rakendamata planeeritud raskuse hindamise funktsioon. See võiks baseeruda korpuste sõnadest koostatud sagedussõnastikule. Mida populaarsemad on testi näitelause sõnad sõnastikus, seda lihtsam on see lause tema sõnade laia kasutuse pärast. Sagedussõnastikku kasutades on vaja osata ignoreerida asesõnu, arvsõnu, sidesõnu, nimesid, hüüdsõnu jms, mis ei anna sisukat infot lause keerukuse kohta. Selliste sõnade kogumine manuaalselt jäi väljapoole töö ajakava ja selle töö automatiseerimine nõuab omaette planeerimist ja katsetamist.

Kokkuvõte

Siinse bakalaureusetöö eesmärk oli luua eesti keele õppemäng, mis kohanduks kasutaja tehtud vigadele. Keelemängus on kaks testitüüpi – käände- ja sünonüümitestid vastavalt grammatika ja sõnavara harjutamiseks. See mäng võiks aidata ema- või võõrkeelena eesti keele õppijaid erinevatest vanuseastmetest. Töös uuriti sarnaseid keeleõppemänge ja -programme, mis on internetis kättesaadavad, et nende tugevustest ja vigadest õppida ning neid teadmisi oma rakenduse arendamisel kasutada.

Töö praktilise osana valminud Keelemäng on kasutatav eraldi veebserveril. Rakenduse sünonüümitestide loomisel kasutati eesti keele jaoks loodud keeletehnoloogilisi vahendeid nagu EstNLTK ja Word2Vec masinõppemudel. Ülesanded genereeriti korpuste põhjal, mis võimaldas muuta selle protsessi automaatseks. Korpuseid filtreeriti reeglistiku ja ebasobivate sõnade loendi abil, kuid see ei välistanud vigaste ja sobimatute lausete testidesse sattumist, mistõttu on kasutajaliideses võimalik igast sellisest testist autorit teavitada. Käändetestid on pärit Anneliis Hallingu bakalaureusetööst.

Keelemängu eristab teistest sarnastest programmidest selle kohanemine kasutaja vigadele. Selleks tutvus autor Leitneri süsteemiga, mida kasutatakse harilikult mälukaartidega õppimiseks, kuid mida saab laiendada muude ülesandetüüpide jaoks õppetöö kohandumise saavutamiseks. Leitneri süsteemis läbitakse järjest küsimusi ning iga õige vastuse korral liigutatakse see ülesanne uude kõrgemasse kategooriasse. Valesti vastates liigub küsimus kategooria võrra tagasi. Kohanemise omaduse saavutamiseks loodi rakenduses lihtsustatud ülesannete valimisalgoritm, mis põhineb sellel süsteemil. Igale kasutaja vastatud sõnale antakse vastavalt tulemusele hinne. Algoritm selekteerib kõikidest testidest uusi nende hinnete toetudes. Õigesti vastates saab sõna kõrgema hinde, valesti vastates madalama hinde. Lisaks uute ülesannete valimisele saab õppija salvestatud tulemuste abil näha ka statistikat tehtud vigade kohta.

Rakendust testis 10 inimest, pärast mida parandati testide valimise algoritmi ning Keelemängu kasutajaliidest. Kõik testijad nõustusid, et e-õppes on kasutajale kohanemine väga oluline. Kasutajad arvasid, et mäng täitis adaptiivsuse eesmärki, mistõttu võib öelda, et loodud rakenduse eesmärk sai täidetud, kuigi kindlasti saab kohandumisalgoritmi veelgi paremaks muuta. Selleks võiks anda igale hindele ajaperioodi, mille järel seda valida. Mitu testijat lisas, et Keelemängus võiks ülesandeid raskusastme järgi jaotada, et testid kohanduksid kasutajale veelgi rohkem.

Viidatud kirjandus

- [1] Kilgi A. Eesti keel maailma taustal. Entsüklopeedia Eestist, 24.08.2005. http://www.estonica.org/et/%C3%9Chiskond/Eesti_keel/Eesti_keel_maailma_taustal/ (30.11.2019)
- [2] Tammur A. Kui palju räägitakse Eestis eesti keelt? Statistikablogi, 2017. <https://blog.stat.ee/2017/03/13/kui-palju-raagitakse-estis-est-keelt/> (30.11.2019)
- [3] Haridus- ja teadusministeerium. Eesti keelevaldkonna arengukava 2018-2027. Keel loob väärtust. https://www.hm.ee/sites/default/files/eesti_keelevalkonna_arengukava_eelnou_8.03.2018.pdf (30.11.2019)
- [4] Haridus- ja teadusministeerium. Eesti keele välisõpe. 2019. <https://www.hm.ee/et/tegevused/keeled/eesti-keele-valisope> (30.11.2019)
- [5] Moreno-Ger P., Burgos D., Martínez-Ortiz I., Sierra J. L., Fernández-Manjón B. Educational game design for online education. Computers in Human Behavior, 2008, vol. 24 no. 6, pp. 2530-2540
- [6] Tang, S., Hanneghan, M. State-of-the-Art Model Driven Game Development: A Survey of Technological Solutions for Game-Based Learning. Journal of Interactive Learning Research, 2011, vol. 22, no. 4, pp. 551-605.
- [7] Halling, A. Õpime käändeid. TÜ arvutiteaduse instituudi bakalaureusetöö. <http://prog.keeleressursid.ee/opimekaandeid/> (08.04.2020)
- [8] Oahpa! <http://oahpa.no/eesti/> (01.12.2019)
- [9] Kai testid: sama tähendusega sõnad. <http://kai16.pbworks.com/f/Samatahendussonad.htm> (30.11.2019)
- [10] Vastandtäheendusega sõna ristsõna. http://www.kubbu.com/student/?i=1&a=41216_samat_henduslikud_s_nad (30.11.2019)
- [11] Haridus- ja teadusministeerium. Eesti elukestva õppe strateegia 2020. <https://www.hm.ee/et/eesti-elukestva-oppe-strateegia-2020> (30.11.2019)

- [12] EstNLTK 1.4 dokumentatsioon. <https://estnltk.github.io/estnltk/1.4/> (08.04.2020)
- [13] Eesti Wordnet. <https://www.cl.ut.ee/ressursid/teksaurus/?lang=et> (14.04.2020)
- [14] `models.word2vec` – `Word2vec` embeddings. <https://radimrehurek.com/gensim/models/word2vec.html> (09.04.2020)
- [15] Saarso, K. Kuidas õpetada sõnavara ja koostada sõnavaraülesandeid. Metoodilisi võtteid eesti keele õpetamiseks. Tallinn, 2003, lk 37-40.
- [16] Krall, I., Sõrmus, E. Keeleõpetaja metoodikavihik. Eesti keele grammatika õpetamise võimalusi. Tallinn: TEA. 2000.
- [17] Hot Potatoes. <https://hotpot.uvic.ca/> (29.12.2019)
- [18] Kahoot. <https://kahoot.com/> (29.12.2019)
- [19] Käänuk. <http://k44nuk.ee/> (01.12.2019)
- [20] Tartu Ülikooli eesti keele (võõrkeelena) osakond. Eesti keele sõnavara harjutused kesktasemele. <https://keeleweb2.ut.ee/kursused/eesti-keele-sonavara-harjutused-kesktasemele> (08.04.2020)
- [21] Villem, A., Aluoja, L., Pilt, L., Naulainen, M.-M., Kusmin, M., Rogalevitš, V., Tokko, U. Digitaalse õppematerjali loomise soovitused. 2015. <https://oppevara.hitsa.ee/kvaliteet/#eessona> (29.12.2019)
- [22] Djaouti, D., Alvarez, J., Jessel, J.-P. Classifying Serious Games: the G/P/S model. University of Toulouse, France. http://www.ludoscience.com/files/ressources/classifying_serious_games.pdf (08.04.2020)
- [23] Kang, S. H. K. Spaced Repetition Promotes Efficient and Effective Learning: Policy Implications for Instruction. Dartmouth College. 2016. [https://www.dartmouth.edu/~cogedlab/pubs/Kang\(2016,PIBBS\).pdf](https://www.dartmouth.edu/~cogedlab/pubs/Kang(2016,PIBBS).pdf) (14.04.2020)
- [24] Spaced Repetition for Efficient Learning. 2019. <https://www.gwern.net/Spaced-repetition>
- [25] Bjornstad, S. Spaced Repetition From The Ground Up. <https://controlaltbackspace.org/memory/spaced-repetition-from-the-ground-up/> (15.04.2020)

- [26] Halling, A. Õpime käändeid GitHubi repositoorium. <https://github.com/Anneliis/opimekaandeid/tree/master/laused> (09.04.2020)
- [27] Eesti Keeleressursside Keskus. <https://keeleressursid.ee/et/keeleressursid> (08.04.2020)
- [28] Eesti Keeleressursside Keskuse META-SHARE repositoorium. <https://metashare.ut.ee/repository/search/?q=corpus> (08.04.2020)
- [29] Eesti ajakirjanduse korpus. <https://metashare.ut.ee/repository/browse/eesti-ajakirjanduse-korpus/74b937fc58e311e2a6e4005056b40024c60ddfc6a3054ce196a44cb326d38cf5/> (08.04.2020)
- [30] Eesti ilukirjanduse korpus. <https://metashare.ut.ee/repository/browse/eesti-ilukirjanduse-korpus/aa4851da58e011e2a6e4005056b40024cd765e538c534d399ccd471fce9f4696/> (08.04.2020)
- [31] TEI: Guidelines. <https://tei-c.org/Guidelines/> (08.04.2020)
- [32] Kallas, J., Kilgarriff, A., Koppel, K., Kudritski, E., Langemets, M., Michelfeit, J., Tuulik, M., Viks, Ü. Automatic generation of the Estonian Collocations Dictionary database. 2015. https://elex.link/elex2015/proceedings/eLex_2015_01_Kallas+etal.pdf (09.04.2020)
- [33] Koppel, K. Näitelausete korpuspõhine automaattuvastus eesti keele õppesõnastikele. *Dissertationes linguisticae Universitatis Tartuensis*, 38. Tartu: Tartu Ülikooli Kirjastus. 2020.
- [34] Kitsnik, M., Metsland, H. Põhikooli ainetundide sõnavara. Tallinn. 2011. https://www.meis.ee/sites/default/files/15_PK_ainet_sonavara_KOOND.pdf (09.04.2020)
- [35] Ant Design of Angular. <https://ng.ant.design/docs/introduce/en> (17.04.2020)
- [36] Auth0 Customer Identity Management. <https://auth0.com/b2c-customer-identity-management/> (14.04.2020)

Lisad

I. Küsimustik



Õppemängu Keelemäng testimisküsimustik

Käesolev küsitlus on tagasiside kogumiseks Janna-Liina Leemetsa bakalaureusetöö käigus loodud keeleõppe mängule Keelemäng.

Keelemäng võimaldab kasutajal õppida eesti keele grammatikat käändetestide ja sõnavara sünonüümitestide kujul. Testid on raskusastmelt mõeldud kooliealistele lastele.

Rakenduse üks olulisemaid omadusi on tema kohanduvus kasutajale. Esimesel mängimisel valitakse testid suvaliselt. Kohanduvus tähendab, et programm üritab teile järgmistel mängukordadel enim anda uuesti selliste sõnadega teste, mille puhul olete valesti vastanud. Õigete vastustega teste näidatakse harilikult kaks korda teadmiste kinnistamise eesmärgil, peale mida seda enam ei näidata peaaegu üldse. Iga kord kaasatakse ka kindel arv uusi teste.

Küsitlus on vabatahtlik ja anonüümne, selles osalemise võib igal hetkel katkestada. Saates vastused ära, annab osaleja nõusoleku vastuseid eelmainitud töös kasutada. Küsimustiku tulemustega saab tutvuda 2020. suvel kaitstavas bakalaureustöös.

Kontaktisik: Janna-Liina Leemets

* Required

Vanus *

Your answer _____

Kas eesti keel on Sinu emakeel? *

Jah

Ei

Juhised

Enne mängu juurde asumist palun lugege läbi kogu juhend.
NB! Testid on loodud automaatselt, seega võib seal olla vigu.

Programm on leitav siin: <http://167.71.3.184/> ja nõuab kasutaja registreerimist ja sisselogimist, et teie tulemusi meeles pidada. Registreerida saab Gmaili konto või muu meili ja parooliga, kuid soovitatav on kasutada Gmaili.

Kui oled sisse loginud, proovi vabalt valitud testitüüpi 10 või 15 lausega. Ürita jätta meelde sõnu, mille puhul Sa vastasid valesti. Seejärel vali sama testitüüpi uuesti 10 või 15 lausega ning vaadake, kuidas programm Sinu eelmistele vastustele vastavalt uusi valis. Tee sama ka teise testitüübiga.

Sobimatu lause leidmisel teavita palun autorit, selleks on näitelause kõrval hüüamärgiga nupp.

Seejärel täida küsimustik.

Võid ka uurida lehte "Profiil", kui huvi on, kuid soovitatav on pöörduda siia küsimustikku kohe tagasi, kui veel kõik värskelt meeles on. Kohanduvuse isegi paremaks nägemiseks võid mängida ka kolmandat korda, peale mida kaks korda õigesti vastatud teste enam ei näidata.

Käändetestid

Käändetestid on pärit Anneliis Hallingu bakalaureusetööst ja nende kohta siinses töös pikemalt ei küsita.

Kas näitelauseid olid arusaadavad? *

Jah

Ei

Kas tegid vigu grammatikatestis? *

Jah

Ei

Hinda, kui hästi oskas programm arvestada Sinu eelnevates testides tehtud vigu ja õigeid vastuseid uute testide valikul *

Your answer

Sünonüümitestid

Sünonüümitestid

Kas näitelauseid olid arusaadavad? *

- Ei
 Jah

Kas näitelauseid aitasid küsimusest paremini aru saada? *

- Jah
 Ei

Kas leidsid küsimusi, mille programmi poolt antud õige vastus oli tegelikult vale?
Kui jah, too palun siin näide

Your answer

Kas tegid vigu sünonüümitestides? *

- Jah
 Ei

Hinda, kui hästi oskas programm arvestada Sinu eelnevates testides tehtud vigu
ja õigeid vastuseid uute testide valikul *

Your answer

Keelemäng

Järgnevad küsimused on mängu idee ja teostuse kohta.

Kuidas hindad kohanduvuse idee kasulikkust? *

Your answer

Kui hästi täitis Sinu arvates Keelemäng juhendis selgitatud kohanduvuse eesmärgi? *

Your answer

Kuidas hindad Keelemängu järgnevaid aspekte? *

	Väga hea	Hea	Keskpärase	Ei oska öelda	Halb	Väga halb
Kasutusmugavus	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Programmi töökiirus	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Värvid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ülesannete selgus	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Palun selgita oma valikuid

Your answer

Kui Keelemängus esines tehnilisi tõrkeid, siis palun kirjeldage, mis juhtus

Kui Keelemängus esines tehnilisi tõrkeid, siis palun kirjeldage, mis juhtus

Your answer

Kommentaare rakendusele või autorile:

Your answer

Submit

II. Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, **Janna-Liina Leemets**,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose
„Kohanduv veebirakendus eesti keele sõnavara ja grammatika õppimiseks“,
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, levitada
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.

Janna-Liina Leemets, 08.05.2019