

Tartu Ülikool
Arvutiteaduse instituut
Infotehnoloogia mitteinformaatikutele õppekava

Vilve Roosioks

**EstNLTK kasutamise võimalused eestikeelsete
ülesannete hindamisel eksamite infosüsteemis EIS**

Magistritöö (15 EAP)

Juhendaja: Heili Orav

Tartu 2022

EstNLTK kasutamise võimalused eestikeelsete ülesannete hindamisel eksamite infosüsteemis EIS

Lühikokkuvõte:

Magistritöö eesmärgiks oli analüüsida EstNLTK kasutamise võimalusi eksamite infosüsteemis EIS ja pakkuda välja võimalusi eestikeelsete tekstide hindamisvahendi täiustamiseks.

Loodud lahendused analüüsivad kirjutatud teksti ja annavad hindajale hindamiseks vajalikku lisainfot ning seeläbi lihtsustavad hindamist.

Lahenduste loomisel on kasutatud Tartu Ülikoolis loodud keeletöõtlusteedi EstNLTK ja töö käigus loodud lahendusi on plaanis edasi katsetada ning täiendada.

Võtmesõnad: keeletöõtlus, hindamisvahend, eksamite infosüsteem, EstNLTK

CERCS: P175 Informaatika, süsteemiteooria

Possibilities for Using EstNLTK in Assessing Estonian-language Tasks in The Examination Information System (EIS)

Abstract:

The aim of the Master's thesis was to analyse the possibilities of using EstNLTK in the EIS examination information system and to offer possibilities for improving the assessment tool for Estonian texts.

The solutions created analyse the written text and provide the evaluator with additional information necessary for the evaluation and thereby simplify it.

The language processing library EstNLTK created at the University of Tartu has been used to create solutions and the solutions created during the work are planned to be further tested and supplemented.

Keywords: natural language processing, assessment tool, The Examination Information System

CERCS: P175 Informatics, systems theory, Examination Information System, EstNLTK

Sisukord

Sissejuhatus.....	5
1 Eksamite infosüsteem (EIS).....	7
1.1 Ülevaade erinevatest testiliikidest ja testide korraldamisest.....	8
1.2 Ülevaade ülesandetüüpidest.....	9
1.3 Hindamine EIS-is.....	11
2 Eesti keele oskust mõõtvad ülesanded EISis ja nende hindamine.....	13
2.1 Eesti keele eksami hindamine riigieksamite näitel.....	13
2.1.1 Lugemisosa hindamine.....	13
2.1.2 Kirjutamisosa hindamine.....	14
2.2 Eestikeelsete tekstide hindamine EIS-is.....	15
2.3 Automaatse hindamise võimaluste parendamine EIS-is.....	16
3 EstNLTK kasutamise võimalused teksti hindamisel EIS-is.....	17
3.1 EstNLTK.....	18
3.2 Kirjavigade märgendamine tekstis.....	19
3.3 Teksti analüüs.....	21
3.3.1 Lausete arvu leidmine ja keskmine lausepikkus.....	21
3.3.2 Sõnarikkus.....	21
3.3.3 Pikkade sõnade protsent ja loetavuse indeks (Lix).....	22
3.3.4 Teksti nominaalsus.....	23
3.4 Stopp-sõnade arv tekstis.....	23
3.5 Sõnakorduste leidmine.....	23
3.6 EIS-i hindajavaate täiendamine.....	24
3.7 Mõtted edasiarendustest.....	26
3.8 Analüüs.....	27
Kokkuvõte.....	29
Viidatud kirjandus.....	30
Lisad.....	34
Lisa 1. Ülevaade EIS-is ülesannete koostamiseks kasutatavatest sisuplokkidest.....	34
Lisa 2. EstNLTK-s kasutatavad sõnaliigi tunnused.....	40
Lisa 3. Riigieksami kirjutamisosa hindamisjuhend.....	41

Lisa 4. Hindajavaate prototüüp.....	43
Lisa 5. Vigaste sõnade arv tekstis ja nende märgendamine	44
Lisa 6. Teksti analüüs.....	46
Lisa 7. Stoppsõnade ja sõnakorduste leidmine tekstist	48
Litsents.....	50

Sissejuhatus

Omaloominguliste kirjalike tekstide, nagu näiteks eesti keele riigieksamite kirjutamisosa, hindamine on hindajate jaoks alati suur töö, kuna lisaks sisule tuleb hinnata ka õigekeelsust, teksti ülesehitust, stiili [1].

Eesti keele riigieksami hindamisjuhendi [1] põhjal hinnatakse lugemisosa ühe hindaja poolt, kirjutamisosa hindavad aga kaks teineteisest sõltumatut hindajat ja kui kahe hindaja tulemuste erinevus on suurem kui 30%, läheb kirjand kolmandale hindamisele ning alles pärast kolmandat hindamist kujuneb lõpptulemuseks kahe lähema tulemuse keskmine.

Praegu toimuvad eesti keele riigieksamid ja põhikooli lõpueksamid paberil ning tekst kirjutatakse käsitsi. Riiklikus testiarenduses liigutakse järjest enam ainult elektroonilise testimise suunas. Eesti haridusvaldkonna arengukavas 2035 [2] on seatud õppe tulemuslikkuse suurendamise ja õppija arengu pideva toetamise eesmärgiks arendada ja kasutada õppetöös digilahendusi kui haridusuuenduste tööriistu, mis võimaldavad õpet, sh õppimist toetavat hindamist, mitmekesistada ja personaliseerida; tõsta õppeprotsessis osalejate teadlikkust infoühiskonna võimalustest ja ohtudest; luua süsteemne lähenemine uute lahenduste kasutuselevõtuks. Samuti on eesmärgina välja toodud eestikeelse hariduse ja eesti keele õppe kättesaadavuse parandamine digilahenduste abil.

Hetkel on koolides toimunud juba eesti keele e-riigieksami esimene katsetus. Kui sooritajad sisestavad teksti arvutis ja hindajad peavad seda ka arvuti vahendusel hindama, siis aitab nende tööaega kokku hoida see, kui arvutiprogramm teksti eelneva analüüsi ära teeb.

Eesti keele e-riigieksamit on planeeritud läbi viia Eksamite Infosüsteemis EIS¹. Eksamite Infosüsteem (EIS) on keskkond e-ülesannete ja elektrooniliste testide (e-testide) koostamiseks, e-testide läbiviimiseks, lõpu- ja riigieksamite tulemuste säilitamiseks, eesti keele tasemeeksamite ja kodakondsuseksamite korraldamiseks. 2013. aastast alustati järkjärgulist üleminekut elektroonilistele taseme- ja eksamitöödele ning neid viiakse läbi EIS-is [3].

Infosüsteemi arendamist alustati aastatel 2008–2013 läbi viidud ÕKVA² programmi kaudu [4].

Kuigi suurem osa EIS-is kasutatavaid ülesandetüüpe on arvutihinnatavad, siis keeletestides on siiski oma roll ka loomingulistel tekstidel, mida hindab inimene. Hindajal tuleb hinnata nii grammatilist kui sisulist korrektsust ning õigekeelsust.

EstNLTK (NLTK ehk Natural Language ToolKit) on peamiselt Pythoni programmeerimiskeeles kirjutatud kogumik teke eestikeelsete tekstide tötluseks [5]. Selles kogumikus on olemas rida vahendeid teksti analüüsimiseks.

¹ <https://eis.ekk.edu.ee/eis>

² ESF-i programm „Õppe kvaliteedi parandamine õppeasutuste sise- ja õpitulemuste välishindamissüsteemi arendamise kaudu“

Haridus ja Teadusministeeriumi keeletehnoloogia teadus- ja arendustegevuse programmi „Eesti keeletehnoloogia 2018-2027“ üks eesmärke on olemasolevate keeletehnoloogiliste lahenduste kasutuselevõtmine võimalikult paljudes valdkondades [6]. Antud töös võetaksegi tähelepanu alla EstNLTK vahendite kasutamisevõimalused EIS-is kirjalike tekstide hindamisel, sest sarnaselt EstNLTK-le on ka EIS programmeeritud Pythonis.

Magistritöö eesmärgiks on leida vastust küsimusele, mil moel on võimalik kasutada EstNLTK võimalusi õpilaste tekstide hindajate töö lihtsustamiseks.

Selleks analüüsitakse eesti keele riigieksami hindamisjuhendit, EIS-is olevaid ülesandetüüpe ja pakutakse välja võimalikke lahendusi eestikeelsete tekstide hindamisel EstNLTK võimalusi kasutada.

Töö esimeses osas antakse ülevaade eksamite infosüsteemis EIS olevatest peamistest testi- ja ülesandetüüpidest ning sellest, kuidas toimub nende hindamine.

Töö teine osa keskendub eesti keele oskust mõõtvatele ülesannetele EIS-is ja nende hindamisele ning riigieksami hindamisjuhendi õigekeelsuse osale. Analüüsitakse olemasolevaid eestikeelsete tekstide hindamise võimalusi ning pakutakse välja võimalused automaatse hindamise parendamiseks EIS-is.

Kolmandas osas antakse ülevaade EstNLTK põhifunktsioonidest ja pakutakse välja võimalusi nende rakendamiseks EISis. Kuna töö eesmärgiks on leida võimalusi lihtsustada hindajate tööd, siis luuakse siin ka lahendus hindamise vaate täiendamiseks.

Lisadena on välja toodud ülevaade EIS-i ülesannete sisuplokkidest ja nende hindamisvõimalustest, sõnaliikide tuvastamise tabel, riigieksami kirjutamisosa hindamisjuhend viide töö käigus valminud täiendatud hindajavaate prototüübile ja EstNLTK kasutusvõimalustel põhinevad lahendused hindaja töö lihtsustamiseks.

1 Eksamite infosüsteem (EIS)

Eksamite infosüsteem (EIS) on keskkond e-ülesannete ja e-testide koostamiseks, e-testide läbiviimiseks, lõpu- ja riigieksamite tulemuste säilitamiseks, eesti keele tasemeeksamite ja kodakondsuseksamite korraldamiseks [7]. EIS pakub hetkel teste ja testimisvõimalusi alusharidusele ning üldharidusele. Lisaks kasutavad süsteemi ka kõrgkoolid erinevate alus- ja üldharidust puudutavate projektide läbiviimiseks.

EIS-i kasutajatel on kolm erineva funktsionaalsusega vaadet:

- Avalik vaade (ei vaja kasutaja autentimist). Selles vaates on võimalik lahendada ligikaudu 2000 erinevat ülesannet ja kontrollida lahenduse õigsust, kuid tulemusi ei salvestata. Lisaks on seal võimalik näha eksamite ja tasemetööde isikustamata statistikat nende testide puhul, kus tulemused on avalikustatud. Samuti saab selles vaates kontrollida riigieksamite tunnistuste kehtivust.
- Avalik vaade (vajab autentimist). Vastavalt kasutaja rollile saab siin teha erinevaid tegevusi. Õpilased saavad lahendada neile suunatud teste ja iseseisvaid töid ning vaadata sooritatud testide tulemusi. Õpetajad saavad kasutada valmis teste, Innove³/Harno⁴ poolt tellitud ülesannetest ise teste kokku panna ja õpilastele lahendamiseks suunata ning ülesannete koostamise moodulis ise ülesandeid luua. Koolijuhid näevad kooli testide tulemusi, saavad oma kooli piires isikutele rolle anda. Avalikus vaates saavad ka hindajad hinnata ülesandeid, mille hindajateks nad on määratud.
- Sisevaade (vajab tugevat autentimist, Harno töötajad ja lepingulise partnerid, kellel on lepingust tingitult see ligipääs vajalik). Siin on vastavalt kasutaja rollile EIS-is võimalik ülesandeid luua, nendest teste kokku panna ja neid keskselt koolidele korraldada, luua testidele tagasisidet, sisestada pabereksamite tulemusi ja luua tunnistusi. Eelloetletud tegevused on määratud kasutajate rollidega süsteemis.

Eksamite infosüsteemi kasutajaliideses on võimalik sisestada ülesandeid ja luua nendest teste ning neid korraldada. Infosüsteemil on olemas avalik vaade, kus on võimalik lahendada avalikuks kasutamiseks mõeldud ülesandeid, vaadata eksamite kohta avaldatud üldist statistikat ja otsida isikule välja antud eksamitunnistusi. Kõik ülejäänud tegevused nõuavad kasutajalt autentimist. Autentida on võimalik ID-kaardi, mobiili-ID, SmartID, HarIDga. Teatud juhtudel on võimalik autentida ka kasutajanime (isikukood) ja parooliga. Viimast võimalust saavad kasutada ainult testi sooritajad.

³ Haridus- ja Noorteamet eelkäija, mille üks tegevus oli ka elektrooniliste hindamisvahendite arendamine.

⁴ Haridus- ja Noorteamet

1.1 Ülevaade erinevatest testiliikidest ja testide korraldamisest

EIS-is on võimalik luua väga erineva funktsionaalsuse ja korraldusega teste. Igal testi on olemas kindel ülesehitus ehk struktuur, kuhu saab lisada ülesandeid.

Testidel võib olla üks või mitu testiosa, mida saab läbi viia soovi korral ka erinevatel päevadel. Testiosasid võib omakorda jagada alatestideks, mille lahendamisele saab seada piiranguid ja tingimusi (näiteks lahendamise piirae, ühesuunalisus⁵). Testi lisatud ülesanded moodustavad ülesandekomplekti, neid võib ühes testis olla ka mitu – nii tekivad ühe testi erinevad variandid, mis võivad olla nii testi korraldaja poolt määratud kui ka sooritaja poolt valitavad.

Testide korraldamise all mõeldakse seda, kuidas jõuab valmis test sooritajani. Riiklikud tasemetööd ja eksamid korraldatakse Harno poolt spetsiaalse korraldusliidese kaudu [7]. Lisaks on testid, mida saavad korraldada õpetajad ja koolide poolt määratud isikud ka ise EIS-i avalikus vaates.

Enamkasutatavad testiliigid e-testide puhul on:

- tasemetööd,
- diagnostilised testid (d-testid),
- lähtetasemetestid.

Harno testide ja hindamise osakonna veebilehel [8] kirjeldatakse riiklike tasemetööd, kui õppeaine või ainevaldkonna–alastest ülesannetest koosnevaid hindamisvahendeid, mida korraldatakse iga-aastaselt Haridus- ja Noorteameti poolt ning mille abil saadakse riikliku õppekava eesmärkidest lähtudes tagasisidet õpilaste õpitulemustest erinevates ainevaldkondades põhikooli esimeses, teises ja kolmandas kooliastmes ning gümnaasiumis.

D-testid on loodud õpetamist toetavaks hindamiseks. Need pakuvad õpetajatele tuge uudsete hindamismeetodite kasutamisel nii kujundava õppeprotsessi juhtimisel kui ka õpilaste teemapõhiste teadmiste ja oskuste testimisel [9]. D-testid annavad sooritajale täpsemat tagasisidet selle kohta, kui hästi ta mingi teemaploki on omandanud ja sageli ka soovitusi selle kohta, mida peaks teadmiste parandamiseks edasi tegema. Õpetaja saab tagasisidena teada, kuidas on tema õpilased antud teema omandanud – kes on valmis uue teemaga edasi liikuma, kellele on vaja anda veel aega teemaga tegeleda. Sõltuvalt testi ülesehitusest võivad d-testid olla ka sooritusega kohanduvad ehk hargnevad, see tähendab, et vastavalt sooritaja poolt antud vastusele kuvatakse järgmine ülesanne. Sellel juhul ei pruugi kõik sooritajaid samu ülesandeid lahendada.

Lähtetasemetestide loomisega alustati 2021. aastal, põhjuseks 2020.a. suhteliselt pikk distantsõppeperiood koroonapiirangute tõttu ja HTM⁶-i poolt tellitud uuringust „Eriolukorrast

⁵ Testi lahendamise viis, kus ülesanded tuleb lahendada kindlas, koostaja poolt määratud, järjekorras ning juba lahendatud ülesandele enam tagasi minna ei saa.

⁶ Haridus- ja Teadusministeerium

tingitud distantsõppe kogemused ja mõju Eesti üldharidussüsteemile“ [10] selgunud vajadus tagada kvaliteetne digiõppevara, mis sisaldaks ka õpilastele enesekontrolliteste. Lähtetaseme testide eesmärgiks on õppeaasta alguses selgitada, kuivõrd on õpilastel meeles eelmisel aastal õpitu ning kaardistada õppeaine need osad, mis vajavad kas osade õpilaste või terve klassi puhul üle kordamist [11].

EIS-i kasutatakse ka riigieksamite ja põhikooli lõpueksamite korraldamiseks. Keskkonnas toimub eksamitele registreerumine, eksamite soorituskohtade ja läbiviijate määramine, protokollide täitmine ja kinnitamine ning tulemuste sisestamine ning koondtulemuste arvutamine. Riigi- ja põhikooli eksamite tulemuste sisestamiseks luuakse samuti vastava liigiga test, millesse lisatakse lihtsustatud kujul ülesanded (sisuliselt võimalusega sisestada eksamitöölt ülesande eest saadud punktid).

1.2 Ülevaade ülesandetüüpidest

Ülesanded pannakse EIS-is kokku sisuplokkidest⁷. Ühes ülesandes saab koostaja kasutada erinevaid sisuplokke ja kuvada neid ka üksteise sees. Sisuplokke eristatakse funktsionaalsuse põhjal:

- Ülesande sisend (9 sisuplokki) - sisuplokkid, mille abil saab kuvada ja luua ülesande lahendamiseks olulist infot (tekst, valemid, multimeediafail, pildid, juhuarv) ja muuta ülesande väljanägemist (HTML koodi lisamise võimalus), aga lahendajal ei ole võimalust sisestada vastust ning puudub ka hindamismatriks⁸.
- Vastuse tüüp ehk interaktsiooni sisuplokkid (38) - võimaldavad kasutada ülesandes erinevaid vastuse tüüpe. Antavaid vastuseid on võimalik võrrelda hindamismatriksiga. 28 sisuploki puhul on võimalik sõltuvalt hindamismatriksi seadistusest kasutada täielikku arvutihinnatavust ja neist omakorda 11 sisuplokki saabki kasutada ainult arvutihinnatavana. Ülejäänud sisuplokkid on kas ainult inimhinnatavad või on võimalus kasutada hübriidhindamist⁹.
- Ülesande väljund (1 sisuplokk) – on mõeldud ülesande koostajale kasutamiseks, võimaldab rakendada täiendavaid valemid erinevate sisuplokkide vastuste hindamiseks ja ei ole lahendajale nähtav.

Hetkel on võimalik kasutada 39 hinnatavat sisuplokki ja 9 kirjeldavat/ abistavat sisuplokki. Sisuplokkide kirjeldusi on võimalik lugeda lisast 1.

⁷ Sisuplokk on kindla funktsionaalsusega komponent, mis on seotud ühe ülesandetüübiga [12].

⁸ Vastuse hindamisel kasutatav tabel, kus iga rida vastab ühele võimalikule vastusele ning näitab, mitu hindepalli selle vastuse eest lahendajale antakse [12].

⁹ Ülesanded, mille puhul hindab arvuti juhul, kui antud vastus kattub 100% hindamismatriksis toodud vastusega. Juhul kui antud vastus hindamismatriksis olevaga ei kattu, läheb ülesanne hindajale hindamiseks [12].

Ülesande koostamise vaates (joonis 1) kuvatakse koostajale vasakul sisuplokkide valik, kust ta saab hiirega klõpsates vajaliku sisuploki ülesandesse lisada. Sisuplokkide järjestus on võimalik hiljem hiirega lohistades muuta.

The screenshot shows the 'Ülesande koostamine' (Question editing) interface. At the top, there are navigation tabs: 'Üldandmed', 'Ülesande sisu', 'Aspektid', 'Koostamine', 'Tekstide versioonid', 'Koostamise ajalugu', and 'Kasutamise ajalugu'. Below these, there are sub-tabs: 'alus', 'AK', and 'Lahendamise'. The main area is divided into two columns. The left column, titled 'Lisa ülesandele sisuplokk' (Add content block to question), contains a table of content blocks and a section for question types. The right column, titled 'Sisuplokid' (Content blocks), shows a list of added blocks with actions like 'Kopeeri', 'Kustuta', 'Lisa mäрге', 'Vaata', and 'Muuda'. The first block is 'alus (1) Alustekst' and the second is 'AK (2) Avatud vastusega küsimus'. Below the second block, there is a text input field and a 'Klõpsa lüngal ja kirjuta sobiv vastus.' (Click the gap and write a suitable answer.) instruction.

Joonis 1. Ülesande koostamise vaade [13].

Vastuse tüüpi sisuplokkide puhul on enamikel juhtudel võimalik lisada hindamismaatriks (joonis 2).

The screenshot shows the 'Vastuse hindamine' (Question grading) interface. On the left, there are various input fields for configuring the question: 'Küsimuse ID' (K02), 'Selgitus', 'Küsimuse miinimumpunktid' (0), 'Küsimuse maksimumpunktid', 'Õige vastuse vaikumisi pallid', 'Vaikumisi punktid vastuse eest, mida maatriksis pole', 'Vastuste arv, mida ületades antakse miinimumpallid', 'Õigete vastuste vähim lubatud arv', 'Vastuse õigsuse määramiseks kasutatav küsimus', 'Kasuta vastuseid testi edasistes ülesannetes algseisuna', 'Kasuta testi varasemate ülesannete vastuseid algseisuna', and 'Ära arvesta vastust tulemustes'. There are also checkboxes for 'Arvutihinnatav' and 'Hübriidhinnatav'. On the right, the 'Hindamismaatriks' (Grading matrix) section shows a table with columns: 'Jrk', 'Vastus', 'Tingimus', 'Punktid', 'Õige', and 'Tabamuste loendur'. The first row contains the value '1' in the 'Jrk' column, 'A (Esimene)' in the 'Vastus' column, '0,5' in the 'Punktid' column, and a checked box in the 'Õige' column. Below the table is a '+ Lisa' button.

Joonis 2. Valikvastustega küsimuse hindamismaatriks [13].

Arvutihinnatavate ülesannete puhul tuleb hindamismaatriks kindlasti täita, sest selle põhjal hinnatakse antud vastust. Hindamismaatriksid on erinevad sisuplokkide lõikes, aga reeglina märgitakse hindamismaatriksis lahendajalt oodatavad vastused ning nende eest antavad punktid. Lisaks õigele vastusele saab lisada hindamisele veel lisakriteeriume ja tingimusi – näiteks kasutada vastuse hindamisel eelmise sisuploki vastust või kasutada antud vastust mõne järgmise sisuploki hindamisel.

1.3 Hindamine EIS-is

Elektrooniliste testide puhul on üheks eelduseks, et testi läbiviimine arvutis aitab vähendada inimeste tööd. Ülesannete, mille puhul vastus on üheselt määratav ja seda on võimalik kirjeldada hindamismaatriksi abil, on arvutihinnatavus võimalik. Paraku loomingulisemate ülesannete puhul, kus vastus eeldab vastaja pikemat kirjalikku selgitust, ei ole praeguste vahenditega võimalik arvuti abil ülesannet hinnata. Peamiselt puudutab see EIS-i ülesandetüüpi avatud vastusega küsimus, kus sooritaja peab vastusena sisestama pikema, enamasti omaloomingulise teksti.

Avatud vastusega küsimustele antud vastuseid hindavad EIS-is hindajad kasutades selleks spetsiaalselt hindajale mõeldud EIS-i kasutajaliidest. Hinnatavad tööd on hindaja jaoks anonüümsed, hindaja näeb sisestatud vastust ja ülesannet, mida lahendati. Kui testi koostaja on hindajale ülesande juurde lisanud juhendmaterjali, siis on hindajal võimalus vaadata ka seda. Enamasti hindab hindaja järjest kõikide sooritajate ühte ülesannet ja siis saab hindamiseks järgmise. Juhul kui sooritaja on jätnud vastusevälja tühjaks (ei ole küsimusele vastanud), siis hindab EIS sellise vastuse automaatselt 0 punktiga ning seda vastust hindajale hindamiseks ei saadeta.

Hindamiseks on hindajal võimalus sisestada kastikesse hindepunktide arv (sisestada saab punkte ülesande hindamismaatriksi koostaja poolt ettenähtud vahemikus). Ülesande hindamisel saab kasutada ka hindamisaspekte¹⁰ - näiteks keeleülesannete puhul võib pikema teksti hindamisel olla vajalik hinnata erinevalt sisu vastavust ülesandele, teksti sidusust, sõnastust, õigekeelsust ja vormistust (joonis 3).

The screenshot displays the EIS evaluation interface. On the left, a question titled 'Ülesanne' is shown with a text input field and a 'Kirjuta ülevaate tekst siia.' button. Below the question, a table titled 'Ülesande vastuse hindamine hindamisaspektide kaupa' lists evaluation aspects. The table has columns for 'Aspekt', 'Toorpunktid', 'Max', and 'Kaal'. The 'Sisu ja ülesande täitmine' aspect is selected, showing a score of 4 out of 4. On the right, a 'Hindamisaspektid' panel shows a list of criteria for the selected aspect, including 'Ülesanne on täitmata.', 'Suurem osa andmetest on analüüsitud, üldistused/järeldused puuduvad või on valed.', and 'Andmeid on analüüsitud, kuid üldistused/järeldused on ebapäsed või puuduvad, on esitatud ebaolulisi detaile.'

Aspekt	Toorpunktid	Max	Kaal
Sisu ja ülesande täitmine	4	4	1

Aspekt	Toorpunktid	Max	Kaal
0p	Ülesanne on täitmata. / Tekst ei vasta ülesandele		
1p	Ülesanne on suures osas täitmata. Suurem osa andmetest on analüüsitud, üldistused/järeldused puuduvad või on valed. Peamiselt on esitatud ebaolulisi detaile või on andmed tekstist otse maha kirjutatud.		
2p	Ülesanne on täidetud osaliselt. Andmeid on analüüsitud, kuid üldistused/järeldused on ebapäsed või puuduvad, on esitatud ebaolulisi detaile.		
3p	Ülesanne on suures osas täidetud. Andmete analüüs on üldjoontes asjakohane, kuid mõned üldistused puuduvad või on esitatud ebaolulisi detaile.		
4p	Ülesanne on täielikult täidetud. Andmete analüüs on asjakohane ja täpne, üldistused/järeldused on asjakohased.		

Joonis 3. Hindaja vaade ülesande hindamisel [14].

¹⁰ Sisuploki eest tervikuna antavad punktid jagunevad erinevate, ülesande koostaja poolt kirjeldatud tingimuste ehk aspektide vahel. Hindaja saab iga tingimust hinnata ülesande koostaja määratud vahemikus.

Hindajaid saab määrata Harno, aga saab määrata ka kool - sõltub testi ja hindamise korraldusest.

Avatud vastusega lünga ülesandetüübi puhul saab ülesande koostaja hindamismaatriksis kirjeldada vastused, mis õigeks loetakse. Kui selline ülesanne on märgitud süsteemis arvutihinnatavaks, siis iga väiksema kõrvalekalle hindamismaatriksis kirjeldatud vastusest tähendab automaatselt vale vastust. Juhul kui antud vastuseid analüüsides on siiski vajadus ka mõned hindamismaatriksis mitte kirjeldatud vastused õigeks lugeda, saab selleks kasutada vastuste standardiseerimist ehk arvutihinnatavaks märgitud avatud vastustega lünkade hindamist inimese poolt [15]. Selleks on EIS-is olemas vastuste analüüsi plokk, kus testi omanik saab lisada hindamismaatriksisse sooritajate poolt antud vastused, mida lisaks esialgselt maatriksis kirjeldatule tuleks samuti õigeks lugeda. Joonisel 4 toodud näite puhul on algselt hindamismaatriksisse lisatud vastus “päev“, aga sooritajad on avatud vastusega lünka kirjutanud ka muid versioone, mis sõltuvalt ülesande püstitusest võib samuti õigeks lugeda ja maatriksisse lisada.

Antud vastused

Leiti 61 kirjet

1 2 → Kuvan korraga: 20

	Jrk	Vastus	Selgitus	Esinemiste arv	Sagedus	Õige
<input type="checkbox"/>	1	päev		223	55,335%	Õige
<input type="checkbox"/>	2			40	9,926%	Vastamata
<input type="checkbox"/>	3	päeval		21	5,211%	Õige
<input type="checkbox"/>	4	2		17	4,218%	Vale
<input type="checkbox"/>	5	päeva		15	3,722%	Õige
<input type="checkbox"/>	6	paev		9	2,233%	Õige
<input type="checkbox"/>	7	kaks		8	1,985%	Vale
<input type="checkbox"/>	8	kaks / 2		6	1,489%	Vale
<input type="checkbox"/>	9	kaks/2		4	0,993%	Vale
<input type="checkbox"/>	10	paeval		3	0,744%	Õige

Joonis 4. Vastuste standardiseerimine [14]

Tulemuste uuesti arvutamisel loetakse kõik sellised vastused automaatselt õigeks ja sama ülesande kasutamisel mõnes uues testis jäävad selliselt lisatud vastused ülesande hindamismaatriksisse ja neid kasutatakse vastuse õigsuse kontrollimisel.

2 Eesti keele oskust mõõtvad ülesanded EISis ja nende hindamine

Eestikeelset teksti tuleb hinnata eesti keele ja eesti keel teise keelena õppeainete tasemetöodes ja testides, kus sooritaja vastuseks on avatud küsimuse vastusena sisestatud pikem kirjalik tekst. Praegu kontrollitakse õpilaste teadmisi eesti keeles kirjalike e-tasemetöödega, mis sisaldavad lugemis-, kuulamis- ja kirjutamisülesandeid ning ülesandeid keeleteadmiste kohta I ja II kooliastmes [16]. Kuulamis-, lugemis- ja keeleteadmiste ülesanded on nii palju kui võimalik arvutihinnataval kujul, kirjutamisosa ülesandeid hinnatakse hindajate poolt kolmest aspektist lähtuvalt: sisu ja ülesehitus, lausestus ja sõnastus ning õigekeelsus ja vormistus [16].

Eesti keel teise keelena tasemetööd on samuti koostatud I ja II kooliastme õpitulemuste põhjal. Nii I kui II kooliastmes mõõdetakse A1 keeletaseme osaoskuseid – kuulamis-, lugemis-, kirjutamis- ja rääkimisoskust ning testis on nii kirjalik kui suuline osa. Kirjutamis- ja rääkimisülesandeid hindab hindaja [17].

Kirjutamisülesannete sisule lisaks peab hindaja hindama ka teksti õigekeelsust.

2.1 Eesti keele eksami hindamine riigeksamite näitel

Riigeksamid on gümnaasiumi lõpus toimuvad eksamid, mis tuleb kooli lõpetamiseks sooritada kõigil abiturientidel. Vabatahtlikult võivad riigeksameid sooritada ka kutseõppeasutuste õpilased ja juba varasemalt keskhariduse omandanud inimesed, kellel on riigeksami tulemust tarvis õpingute jätkamiseks kõrgkoolis [18].

Praegu toimub eesti keele riigeksam kirjalikus vormis ja paberil [19], aga Harno testide ja hindamise osakond juba tegeleb nende eksamite digitaliseerimisega. Kui mõelda kirjandi kirjutamisele, siis kindlasti on tänapäeva sooritajale mugavam ja harjumuspärasem kirjutada teksti arvutis.

Eesti keele eksamid on kaheosalised - lugemisosa, kus mõõdetakse tekstimõistmise oskust ja kirjutamise osa, kus mõõdetakse tekstiloome oskust [19]. Lugemisülesanne koosneb kuni 1000-sõnalise alustekstist, mille põhjal peab sooritaja lahendama ülesandeid.

Eksami teises osas peab sooritaja kirjutama umbes 400-sõnalise sidusa teksti [19].

Eesti keele riigeksami eristuskirja kohaselt hindavad riigeksamitöid haridus- ja teadusministri käskkirjaga kinnitatud hindajad ning hindamine toimub hindamisjuhendi põhjal [19]. Eksami lugemisosa hindab üks hindaja, kirjutamisosa aga teineteisest sõltumatult kaks erinevat hindajat.

2.1.1 Lugemisosa hindamine

Lugemisosa vastuste sisu hinnatakse vastavalt iga eksami puhul koostatud hindamisskaalale [19]. Vastuste eest saadavate punktide arv sõltub küsimuse keerukusest ja vastuse sisust. Hinnatakse ka iga vastuse õigekeelsust, mille eest on võimalik saada maksimaalselt 2 punkti.

Õigekeelsust hinnatakse vastavalt juuresolevale tabelile (tabel 1).

Tabel 1. Lugemisülesannete õigekeelsus hindamine [19].

Punktid	Hindamiskriteeriumid
2	Vastuses ei ole õigekeelsusvigu.
1	Vastuses on 1–2 õigekeelsusviga.
0	Vastuses on 3 või rohkem õigekeelsusviga.

Juhul kui vastuse sisu hinnatakse 0 punktiga või vastus puudub, siis õigekeelsust eraldi ei hinnata [19].

2.1.2 Kirjutamisosa hindamine

Kirjandis hinnatakse sisu, õigekeelsust, stiili/sõnastust ja vormi/ülesehitust skaalal 0–5. Hinnatavate aspektide osatähtsus lõpptulemuses on erinev (sisupunktid korrutatakse teguriga 5, õigekeelsuspunktid teguriga 4, stiilipunktid teguriga 2 ja vormipunktid teguriga 1). Sisu eest on võimalik saada maksimaalselt 25 punkti, õigekeelsuse eest 20 punkti, stiili eest 10 punkti ja vormi eest 5 punkti. Kogu kirjand hinnatakse 0 punktiga, kui see ei põhine eksamiülesandel või on lühem kui 250 sõna. [19]

Sisu.

Kirjutamisosa pikkus peab olema umbes 400 sõna [19]. Kirjutamisosas võib kasutada ka lugemisosa tekste ja neid refereerida või tsiteerida, aga lugemisosa alustekste ei tohi ulatuslikult ümber kirjutada – sellisel juhul alandatakse sisu eest antavaid punkte.

Õigekeelsus.

Eesti keele riigieksami eristuskirja kohaselt [19] arvestatakse kirjandis üheks veaks:

- 1) ühes sõnas esinevad eri tüüpi õigekeelsusvead;
- 2) samas sõnas või morfoloogilises vormis korduvad vead;
- 3) sama lausetarindi puhul korduvalt ära jäetud või üleaarune kirjavahemärk või kahepoolne vahemärk (jutumärgid, sulud);
- 4) kõrvallause, lauselühendi, kiilu, sh ütte ja hüüundi, järellisandi ja -täiendi ning lisanditaoliste määruste puhul korduvalt ära jäetud koma(d) või mõttekriips(ud);
- 5) sama lausetüübi puhul korduvalt ära jäetud lõpumärk.

Stiil

Stiili puhul hinnatakse teksti ladusust, täpsust, isikupära ja sõnavaralist rikkust [19]. Kirjandi hindamisest annab ülevaate lisa 3.

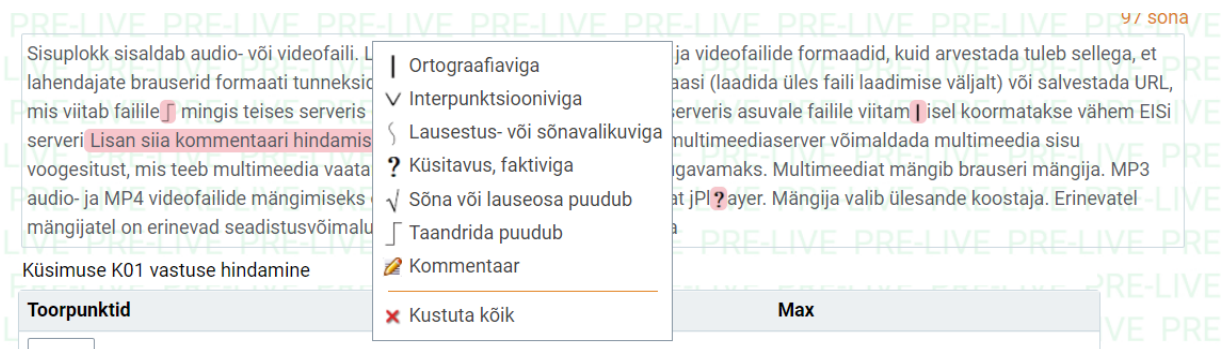
2.2 Eestikeelsete tekstide hindamine EIS-is

Eestikeelsete tekstide hindamiseks on hetkel olemas avatud vastusega lünga puhul võimalus, et hindaja saab tekstis vigu märkida. See võimalus tuleb aktiveerida ülesannet luues sisuplokis.

Vead, mida hindaja saab tekstis märkida (joonis 5):

- ortograafiaviga;
- interpunktsiooniviga;
- lausestus- või sõnavalikuviga;
- küsitavus, faktiviga;
- sõna või lauseosa puudub;
- taandrida puudub.

Lisaks on hindajal võimalus lisada teksti kommentaare.



Joonis 5. Vigade märkimine tekstis [14].

Ülesande koostamise saab avatud vastusega küsimuse sisuplokis sisse lülitada sõnade arvu kuvamise vastuses. Sellisel juhul on ülesande lahendajal võimalus näha, mitu sõna tema kirjutatud tekstis on (oluline juhul, kui maksimumpunktide saamiseks on vajalik teatud arvu sõnadega tekst) ja ka hindajale kuvatakse hiljem sisestatud teksti sõnade arv. Teksti sisu, ülesehituse, lausestuse ja sõnastuse jms hindamise jaoks on hindajale hindamisjuhendis kirjeldatud erinevad aspektid ning punktide arv, mis kirjeldusele vastab. Vastavalt sellele saab hindaja sisestada „Toorpunktid“ lahtrisse oma hinnangu.

2

Kirjutamisülesanne 2 Tulemus: 16,5p 20-st, 82,5%

Ülesande ID: 33049

Kirjuta kirjand vastuse aknasse. Ära unusta pealkirja. Märgi see poolpaksus kirjas ja joonda keskele. Reavahetus alustab uut lõiku, seepärast ei ole tekstis vaja kasutada taandridu.

402 sõna

Vihakõne ja selle problemaatika

Tänapäeva riikides esineb sõnavabadus. Samuti ka Eesti Vabariigis. Selle tulemusena paljud inimesed arvavad, et nad võivad öelda välja igal ajal ja ükskõik mida ning ükskõik, millise retoorika ja sõnastusega. Igauks, kes vihakõnet peab, põhjendab tihti peale just põhiseadusest pärit sõnavabadusega. Kuid igal ajal on mingisugunegi piir. Kust kohast tõmmata piir vihakõnele? Kui suureks see Eestis on eskaleerunud?

Hindamisjuhend Öige vastus Lahendaja vaade Küsimused ✕

Hindamisaspektid

Aspekt	Max pallid	Hindamisjuhend
Sisu	4	0p Ülesanne on täitmata. / Tekst ei vasta ülesandele. Ülesanne on suures osas täitmata. Probleemid on avatud pinnapealselt. Tekst on valdavalt jutustav/kirjeldav.
		1p Probleemikäsitus on lünklik. Analüüs on puudulik, jutustamine domineerib analüüsi ja argumentatsiooni üle / esineb küsitavusi ja faktivigu. Ülesanne on täidetud osaliselt. Teema on avatud üldsõnaliselt. Tekst on kohati jutustav/kirjeldav. Ülesanne on täidetud osaliselt.
		2p Probleem(id) on avatud üldsõnaliselt. Tekst on kohati iutustav/kirjeldav.

Ülesande vastuse hindamine hindamisaspektide kaupa

Aspekt	Toorpunktid	Max	Kaal
▼ Sisu	4	4	1
▼ Alustekstide kasutamine	3	4	1
▼ Ülesehitus ja sidusus	4	4	1
▼ Lausestus ja sõnastus	3	4	1
▼ Õigekeelsus ja vormistus	4	4	1
Toorpunktid kokku			
18	Koeffitsient	Hindepallid	
	1	18	

Joonis 6. Hindaja vaade tekstile ja hindamisaspektidele [14].

Kui sooritaja jätab vastusekasti tühjaks, siis hindab EIS selle automaatselt 0 punktiga ja hindajale seda ülesannet hindamiseks ei suuna.

2.3 Automaatse hindamise võimaluste parendamine EIS-is

Eeloleva põhjal oleks automaatset hindamist võimalik parandada 0 punktiga hinnatavate ülesannete osas. Ülesande koostajal oleks võimalus:

- määrata ülesande sisuplokis teksti pikkus sõnades, millest väiksema sõnade arvu korral hinnatakse tekst automaatselt 0 punktiga;
- märkida õigekeelsusvigade arv, millest rohkemate vigade korral hinnatakse õigekeelsus automaatselt 0 punktiga.
- märkida alustekstide kasutamise puhul protsent, millest rohkem ei tohi kirjutamise osas olla kattuvusi alustekstidega, vastasel juhul hinnatakse tekst automaatselt 0 punktiga.

Omavahel seotud sisuplokkide või aspektide hindamise puhul saab arvestada eelmise sisuploki/ aspekti tulemust järgmise hindamisel (näiteks kui vastuse sisu hinnatakse 0 punktiga, siis ei tule hindajale enam õigekeelsuse hindamise võimalust).

3 EstNLTK kasutamise võimalused teksti hindamisel EIS-is

Kairi Tamuri, Ethel Roosna jt kirjutavad “Õpetajate Lehe” artiklis [20]: “Elektroonilise hindamise üks oluline eelis on võimalus säästa õpetaja aega: andmed on ülevaatlikud, arhiveeritavad ning lubavad jälgida õpilase õpitulemuste muutumist ja arengut õppekava eesmärkide kontekstis. Samuti pakub e-hindamine kiiret ja mugavat võimalust anda tagasisidet kõigile osalistele – õpilasele, lapsevanemale, õpetajale, koolile ja riigile.”

Gümnaasiumi eesti keele e-riigieksami arendamisega alustati 2015.a. ning esimene suurem katsetus toimus 2017.a. kevadel [21]. 2022.a. kevadel katsetati esimest e-eksami versiooni. E-riigieksami töörühma seisukoht on, et e-eksam jääb sisult samasuguseks, kui praegune pabereksam, muudetakse vaid lahendamise keskkonda: eksamitöö on endiselt kaheosaline, kuid see sooritatakse arvutis [22]. See tähendab, et eksamiülesannete tüübiks on valdavalt avatud vastusega lünk või avatud vastusega küsimus, kus sooritaja saab vastusekasti sisestada vaba teksti. Kõik sellised ülesanded on inimese poolt hinnatavad. Küsitledes 2022. kevadel toimunud e-eksami katsetööde hindajaid, siis pikkade tekstide puhul kasutati hindamisel näiteks varianti, kus kopeeriti sooritaja tekst EIS-ist mõnda spellerit omavasse tekstitötlusprogrammi ning saadi nii kiiremini ülevaade kirjavigadest. EIS on praegu võimeline kokku lugema sõnade arvu tekstis ja hindajal on võimalus ise tekstis vigu märkida. Samas aitaks EstNLTK¹¹ kasutamine anda hindajale teksti analüüsi kaudu märksa rohkem infot, mis teda hindamisel abistaks.

EIS-is õpilaste pikemaid tekste hinnanud õpetajatelt saadud tagasiside põhjal võiksid olla hindamist abistavateks järgmised näitajad:

- sõnade arv tekstis (seda hetkel juba kuvatakse);
- vigaselt kirjutatud sõnade märgendamine tekstis;
- lausete arv tekstis;
- lausete keskmine pikkus (sõnades);
- pikkade (üle 7-tähealiste) sõnade protsent tekstis;
- sõnarikkus;
- nominaalsus (nimisõnade %);
- sõnakordused;
- stopp-sõnade %.

Kuigi riigieksameid veel elektrooniliselt ei korraldata, siis on nende hindamisjuhendid hindajate jaoks kõige konkreetsemad ja sellepärast võetakse EstNLTK võimaluste kasutuselevõtmise analüüsil aluseks just need.

¹¹ Allolevate näidete puhul kasutatakse teksti analüüsimisel EstNLTK versiooni 1.6.7.

3.1 EstNLTK

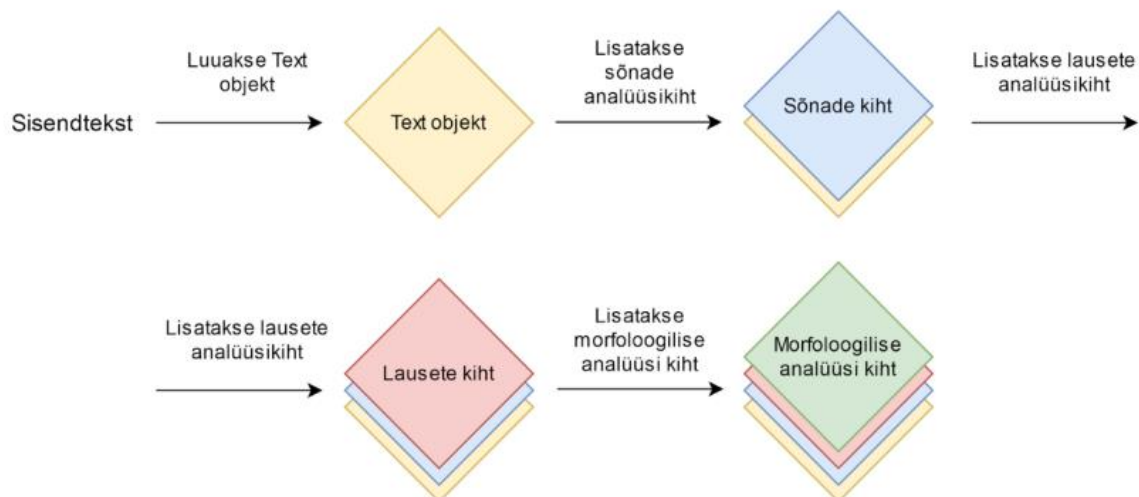
EstNLTK (*Estonian Natural Language Processing Toolkit*) on Pythoni programmeerimiskeeles kirjutatud teek eesti keele töötlemiseks, sisaldades põhilisi loomuliku keele töötamise võimalusi nagu sõnestamist, morfoloogilist ja süntaktilist analüüsi, lemmatiseerimist ja nimeüksuste tuvastust (NER - *named entity recognition*), osalausestamist (*clause segmentation*), ajaväljendite ja verbiahelate tuvastamist [23].

EstNLTK abil on võimalik jagada tekst väiksemateks osadeks – lõikudeks, lauseteks ja sõnadeks, mida saab siis omakorda täpsemalt analüüsida.

EstNLTK kasutamiseks tuleb teek installeerida keskkonda, milles seda kasutada soovitakse¹². Olulisim klass EstNLTK-s on analüüsitava teksti, teksti metaandmeid ja analüüsi käigus loodavaid märgenduskihte sisaldav *Text*, mis tuleb kasutamiseks programmi importida [24]. Analüüsitava tekstile märgenduskihtide lisamiseks tuleb see esmalt muuta *Text* klassi objektiks [24].

Tag_layer() meetodi abil saab märkida *Text*-objektile 5 kihti: laused (*sentences*), sõned (*tokens*), *compolund_tokens* (*liitsõned*), sõnad (*words*), morfoloogiline analüüs (*morph_analysis*). Morfoloogilise analüüsi kiht sisaldab omakorda järgmisi atribuute: *text* ja *normalized_text* olid juba sõnade kihil, ent juurde on lisatud sõna algvorm ehk lemma (lemma), tüvi (*root*), tüveosised (*root_tokens*), käände-/pöördelõpp (*ending*), gi/ki liide (*clitic*), sõna morfoloogiline vorm (*form*), sõnaliik (*partofspeech*) [25].

Eelkirjeldatu annab hästi edasi Karl Erik Karindi bakalaureusetöös esitatud joonis 7 [26].



Joonis 7. EstNLTK kihiline struktuur [26].

¹² Selles töös on koodinäidiste loomisel kasutatud Jupyter Notebook graafilist rakendust Anaconda virtuaalkeskonnas.

Teet Kalmuse magistritöö [27] kohaselt võib EstNLTK põhifunktsioonidena välja tuua:

1. **teksti segmenteerimine** ehk jagamine väiksemateks osadeks – lõikudeks, lauseteks ja sõnadeks.
2. **morfoloogiline analüüs**, mille abil saab leida tekstisõne kohta tema lemma ehk algvormi, sõnaliigi, sõnalõpud, sõnavormid ja järelliited. Morfoloogilise analüüs tehakse vaikimisi koos ühestamise¹³, oletamise¹⁴ ja pärisnimeanalüüsi pakkumisega¹⁵[25].
3. **õigekirja kontroll ja morfoloogiline süntees**, viimane on morfoloogilise analüüsi pöördprotsess, mis võimaldab genereerida etteantud lemmale ja vormitunnustele vastava sõnavormi.
4. **nimeüksuste tuvastamine** (*named entity recognition* – NER), mille abil on võimalik tekstist leida inimeste, organisatsioonide ja asukohtade nimed.
5. **ajaväljendite tuvastamine**, mis võimaldab tekstist otsida ajaväljendeid.
6. **osalausete leidmine** ehk pikkade ja keeruka ülesehitusega lausete tükeldamine väiksemateks osadeks (osalauseteks) [28].
7. **verbiahelate tuvastamine** leiab verbidega seotud ühe- ja mitmesõnalised üksused.
8. **grammatikate kirjutamise töövahend** reeglipõhise info eraldamise jaoks. Reeglite koostamisel saab nõ ehitusklotsidena kasutada sõnaliike, lemmasid ja regulaaravaldisi (regex), kombineerides neid omavahel.

Loetletud põhifunktsioonidest leiavad allolevate näidete puhul kasutamist teksti segmenteerimine, morfoloogiline analüüs ja õigekirja kontroll.

3.2 Kirjavigade märgendamine tekstis

Kirjavigade märgendamiseks saab EstNLTK-s kasutada märgendajat *SpellCheckRetagger*. Märgendajaid saab lisada *estnlk.taggers* kaudu [29]. Enne märgendaja rakendamist tuleb sisendtekstile lisada vajalikud sõltuvuskihid – vigaste sõnade otsimiseks kiht *words*. *SpellCheckRetagger* lisab valesti kirjutatud sõnadele õige vormi [24]. *Words* kihi *normalized_form* annab selle märgendaja rakendamisel õigesti kirjutatud sõnade puhul väärtuseks *None*. Seega saab vigaste sõnade leidmiseks otsida neid sõnu, mille *normalized_form* kiht seda ei ole. Teine võimalus on kasutada märgendajas lippu

¹³ Kui ühe sõnavormi analüüsiks on mitu võimalust, siis valitakse välja antud kontekstis kõige tõenäolisem tõlgendus [25].

¹⁴ Kui analüsaator ei suuda oma sõnastiku põhjal sõna analüüsida, siis pakutakse eeldatav analüüs [25].

¹⁵ Suurtähega algavate nimisõnade puhul üritab analüsaator konteksti põhjal ennustada, kas tegemist on nimega [25].

add_spellcheck. See lisab sõnakihile atribuudi *spelling*, mille väärtus *False* näitab, et tegemist ei ole õigesti kirjutatud sõnaga (joonis 8).

```
from estnltk.taggers import SpellCheckRetagger

spelling_tagger=SpellCheckRetagger(add_spellcheck=True)
spelling_tagger.retag(tekst)
```

Joonis 8. Klassi *SpellCheckRetagger* importimine ja rakendamine *Text*-objektile.

Tekstis kirjavigade märgendamiseks tuleb need sealt esmalt üles leida. Eesti keele eksami tekstiloomel hindamisel arvestatakse üheks veaks samas sõnas või morfoloogilises vormis korduvad vead [19], sellepärast on mõistlik vigade loendamisel kasutada hulka, sest sinna ei ole võimalik korduvaid elemente salvestada (joonis 9).

```
veasonad=set()
for lause in tekst.sentences:
    for sona in lause.words:
        if sona.spelling[0] == False:
            veasonad.add(sona.text)
```

Joonis 9. Vigaselt kirjutatud sõnade leidmine ja lisamine hulka „veasonad“.

Vigaste sõnade märgenduskihi loomiseks luuakse klass *VigasteSonadeTagger*, mille väljundiks on märgenduskiht, mille saab vigaste sõnade kuvamiseks tekstile lisada (joonis 10).

```
from estnltk import Text, Layer
from estnltk.taggers import Tagger

class VigasteSonadeTagger(Tagger):
    """Märgistab vigased sõnad."""
    conf_param = ['vead']

    def __init__(self, veasonaloend, output_layer='veasonad3', input_morph_analysis_layer='morph_analysis'):

        self.input_layers = [input_morph_analysis_layer]
        self.output_layer = output_layer
        self.output_attributes = ['vigased']

        self.vead = veasonaloend

    def _make_layer(self, text, layers, status):
        layer = Layer(name=self.output_layer, attributes=self.output_attributes, text_object=text)
        for span in layers['morph_analysis']:
            for annotation in span.annotations:
                if annotation['normalized_text'] in self.vead:
                    layer.add_annotation(span.base_span, token_type='VIGA')
                    break

        return layer
```

Joonis 10. Märgenduskihi *VigasteSonadeTagger* loomine [28].

Joonisel 11 on toodud uue märgendaja loomine, mis võtab aluseks hulka „veasonad“ kogutud kirjavigadega sõnad ning nende põhjal uue kihi tekitamine. Meetodiga *tag* saab kihi tekstile lisada ning funktsiooniga *display()* tekstile kuvada.

```
vigadeMargendaja=VigasteSonadeTagger(veasonad)
vigadeMargendaja.tag(tekst)
tekst.veasonakiht.display()
```

Joonis 11. Uue märgendaja loomine, selle rakendamine tekstile ning vigaste sõnade kuvamine.

Unikaalsete vigaste sõnade arvu saab leida vigaste sõnade hulga pikkusest. Lisas 5 on üks võimalik lahendus vigaste sõnade leidmiseks ja märgendamiseks tekstis.

3.3 Teksti analüüs

Teksti analüüsil toetuti hindajate poolt soovitud näitajate leidmisele ja kuvamisele. Allpool toodud alapeatükkides kirjeldatu koodinäite leiab lisas 6.

3.3.1 Lausete arvu leidmine ja keskmine lausepikkus

Lausete arvu leidmiseks ja arvutustehete tegemiseks tuleb importida *math* teek. Sisendtekstile tuleb meetodi *tag_layer* abil lisada vajalikud sõltuvuskihid. Peale seda on tegelikult võimalik lausete arv juba lausete kihi *sentences* kaudu leida, aga EstNLTK loeb sõnadeks ka kirjavahemärke, seega tuleb keskmise lausepikkuse (sõnades) arvutamisel need eelnevalt lausest eemaldada (joonis 12). Kirjavahemärgid on EstNLTK-s tähistatud sõnaliigiga 'Z' (lisa 2 tabeli põhjal kirjavahemärgi tähis).

```
sonedearv=0
for lause in tekst.sentences:
    for sona in lause.words:
        if sona.partofspeech[0]!='Z':
            sonedearv+=1
keskminelausepikkus=sonedearv/len(tekst.sentences)
lausetearv=len(tekst.sentences)
```

Joonis 12. Teksti keskmise lausepikkuse ja lausete arvu leidmine.

Keskmise lausepikkuse leidmiseks jagatakse kogu teksti sõnade arv lausete arvuga.

3.3.2 Sõnarikkus

Sõnarikkuse leidmiseks soovitatakse artiklis „Sõnavara loomulik rikkus haritud keeleoskaja tekstides“ [30] kasutada Uberi indeksit (1), mille puhul kehtib seos, et mida suurem on indeksi väärtus, seda rikkalikum on sõnavara.

$$U=(\log N)^2/(\log N-\log V), \quad (1)$$

kus N on sõnede (*tokens*) arv ja V eri sõnade arv [30].

Kuna hindajate jaoks võib olla Uber'i indeksi põhjal teksti sõnarikkuse hindamine tundmatu, siis kasutatakse ka lihtsamat varianti, kus võetakse unikaalsete sõnade protsent kõikide sõnade arvust [31].

$$\text{Sõnarikkus} = V/N * 100 \quad (2)$$

Nii sõnarikkuse lihtsama versiooni kui Uber'i indeksi leidmiseks tuleb esmalt eemaldada tekstist kirjavahemärgid, et teada saada sõnade arv N. Erinevate sõnade arvu leidmiseks tuleb sõnad esmalt lemmatiseerida ja seejärel lisada sõnade lemmad hulka, mille pikkus võrdub unikaalsete sõnade arvuga V.

3.3.3 Pikkade sõnade protsent ja loetavuse indeks (Lix)

Pikkade sõnade arvu P leidmiseks loetakse tekstis kokku sõnad, mis on 7 ja rohkem tähemärki pikad. Antud suurust saab Helin Puksandi järgi [32] kasutada teksti raskusastme mõõtmiseks ka eestikeelsete tekstide puhul edukalt kasutatud loetavuse indeksi (Lix) arvutamisel, milleks summeeritakse lausete keskmine pikkus ja pikkade sõnade osatähtsus ning ümardatakse tulemus täisarvuks. Lix indeksi tõlgendamisel saab kasutada 1968. aastal rootsi keelele välja töötatud tabelit (tabel 2):

Tabel 2. Teksti raskusastme tõlgendamine Rootsis [32].

LIX	Teksti raskusaste	Lausepikkus (sõnades)	Pikki sõnu (%)
20	Väga lihtne	9	11
30	Lihtne	13	17
40	Keskmine	16	24
50	Raske	19	31
60	Väga raske	22	38

Teksti hindajatele ei pruugi Lix-indeks olla tuttav, seega kuvatakse neile teksti analüüsil EIS-is ka pikkade sõnade protsent (joonis 13).

```
Lix=round(keskminelausepikkus+pikkadesonadeosakaal,1)
pikkadesonadeosakaal=round(pikadsonad/sonedearv*100,2)
```

Joonis 13. Pikkade sõnade protsendi ja Lix-indeksi leidmine.

Pikkade sõnade protsendi leidmiseks jagatakse pikkade sõnade arv (pikadsonad) teksti sõnade arvuga (sonedearv) ning viiakse tulemus protsendi kujule.

3.3.4 Teksti nominaalsus

Teksti nominaalsuse ehk nimisõnavormide osatähtsuse [33] määramiseks leitakse tekstis leiduvate nimisõnade protsent teksti sõnade arvust.

$$\text{NOM} = N_{\text{nim}} / N * 100 \quad (3)$$

Nimisõnade (N_{nim}) leidmiseks analüüsitakse teksti sõnaliikide tasandil ja loendatakse sõnad, mille liigiks on „S“ (Lisa 2 tabeli põhjal nimisõna ehk substantiivi tähis EstNLTK-s).

3.4 Stopp-sõnade arv tekstis

Allkivi ja Kippari [34] sõnul on stopp-sõnad sõnad, mis ütlevad teksti sisu kohta vähe. Nendeks on reeglina sidesõnad (aga, et, sest), asesõnad (see, tema, oma), asemäärsõnad (nii, siis, seal) ja grammatilise tähendusega tegusõnad (olema, saama).

Lisas 7 toodud programmis kasutatakse stopp-sõnade leidmiseks võrdlusmaterjalina Kristel Uiboaia lemmatiseeritud stopp-sõnade loendit [35], millele lisatakse teksti analüüsiks vajalik märgenduskiht. Edasi võrreldakse tekstis olevate sõnade lemmasid stopp-sõnade loendis olevatega ja juhul kui sõna leidub stopp-sõnade hulgas, siis suurendatakse vastavat loendurit. Lisaks tekstis leiduvate stopp-sõnade arvule leitakse ka nende protsent teksti sõnade arvust.

3.5 Sõnakorduste leidmine

Sõnakordused on üks stiilivigadest, mida kirjandi puhul hindamisel arvestatakse [36]. Siin otsustati korduvaid sõnu otsida lauses ja kahes üksteisele järgnevas lauses. Analüüsitava sõnade hulgast jäeti välja stopp-sõnad, kasutades selleks eelmises alapeatükis kirjeldatud programmis loodud testis sisalduvate stopp-sõnade hulka (stopp). Lauses esinevatest sõnadest moodustatakse sõnastik (lauseskordused), kus võtmeks on sõna ja väärtuseks sõna esinemiste arv lauses. Nendest on võimalik leida lauses korduvad sõnad (esinemiste arv suurem kui 1). Kahe järgneva lause omavaheliseks võrdluseks luuakse uus sõnastik (kordused) nii, et võtmeks on lause number tekstis ja väärtuseks lause sõnadest eelnevalt moodustatud sõnastik (lauseskordused) (joonis 14).

```
#sõnastik lausete ja nendes sisalduvate sõnade jaoks
kordused={}
lausenr=0

for lause in tekst.sentences: #tekst vaadatakse lausehaaval läbi
    lausenr+=1 #lause numbrid algavad 1-st
#luuakse sõnastiks lauses korduvate sõnade jaoks
    lauseskordused={}
#lause järjekorranumbriga luuakse tühi sõnastik
    kordused[lausenr]={}
    for sona in lause.words:
        if sona.partofspeech[0]!='Z': #kirjavahemärkide eemaldamiseks
            if sona.lemma[0] not in stopp: #stopp-sõnade eemaldamiseks
                #print(sona.lemma[0])
                if sona.lemma[0] not in lauseskordused:
```

```

#iga sõna lisatakse sõnastikku 1 kord
    lauseskordused[sona.lemma[0]]=1

    else:
#kui sõna juba on sõnastikus, siis suurendatakse tema esinemisarvu
    lauseskordused[sona.lemma[0]]+=1

#lauses olevate sõnade ja nende arvu sõnastik lisatakse lausete sõnastikku
    kordused[lausenr]=lauseskordused

```

Joonis 14. Lausetest ja sealt olevate sõnade esinemiste arvust sõnastiku loomine.

Võrreldes nüüd kahte üksteisele järgnevat lauset on võimalik leida sõnad, mis leiduvad mõlemas lauses (joonis 15).

```

#kuna lauseid nummerdati sõnastikus alates 1-st, siis muutuja nr väärtustatakse
#1-ga
nr=1
lausetearv=len(tekst.sentences) #lausete arv tekstis

kordublauses=set()
korduvad=set() #kogub korduvad sõnad märgendamise jaoks
while nr<lausetearv:#vaadatakse järjest läbi kõik välimise sõnastiku kirjed

    uus=nr+1 #järgnevate lausete võrdlemiseks
    #nummerdatud laused vaadatakse sõnastikus läbi

    for voti in kordused[nr]:
#kui sõna on lauses korduvalt, siis lisatakse sõna korduvate hulka
        if kordused[nr][voti]>1:
            kordublauses.add(voti)
            print("lauses "+str(nr)+" kordub: "+voti, kordused[nr][voti])

#kui sama sõna on järgmises lauses, siis lisatakse samuti korduvate hulka

        if voti in kordused[uus]:
            korduvad.add(voti)
            print("lausetes "+str(nr)+" ja "+str(uus)+" kordub: "+voti, kordused[nr][voti]+kordused[uus][voti])

    uus=0
    nr+=1

```

Joonis 15. Korduste leidmine lausetes ja kahes teineteisele järgnevas lauses.

Ülaloleva näite puhul väljastatakse kordused lauses ja teineteisele järgnevas lauses tekstina.

3.6 EIS-i hindajavaate täiendamine

Tekstianalüüsi kuvamiseks hindajale plaanitakse täiendada olemasolevat hindaja vaadet avatud vastusega küsimuse hindamisel sakiga „Tekstianalüüs“, mille avamisel kuvatakse testi analüüsil arvutatud väärtused, mis võiksid aidata hindajal teksti kiiremini ja objektiivsemalt hinnata.

Töö koostamise käigus loodi planeeritavast hindajavaatest ka Figma¹⁶ keskkonnas prototüüp (lisa 4).

Hindajale kuvatakse teksti analüüsimiseks tehtud arvutuste tulemused (joonis 16). Kirjavigadega sõnu ja sõnakorduseid võib tekstis märkeruudu valimisel nähtavaks muuta (joonis 17).

The screenshot shows a web application interface for text analysis. On the left, there is a document titled "Kirjutamisülesanne 2" with a preview of text. On the right, a sidebar titled "Tekstianalüüs" displays various statistics:

Nimetus	Näitaja	Märgi tekstis	Hindaja tegevus
Lauseste arv tekstis	9		
Keskmine lausepikkus sõnades	21		
Vigaste sõnade arv tekstis	5	<input type="checkbox"/>	Valesti vigaseks märgitud sõnad: 0
EstNLTk abil leitud vigaste sõnade arv	5		
Hindaja poolt märgitud vigaste sõnade arv	0		Vigaseks märkimata jäänud sõnad: 0
Pikkade (7 ja rohkem tähte) sõnade %	41,4		
Texti jälgitavus (Lix indeks)	62,1		
Nimisõnade % (teksti nominaalsus)	35,5		
Sõnarikkus	67,2		
Sõnarikkust iseloomustav Liberi indeks	29,8		
Sõnakordused lauses		<input type="checkbox"/>	
Sõnakordused järjestikustes lausetes		<input type="checkbox"/>	
Stopp-sõnade % tekstis	45,2		

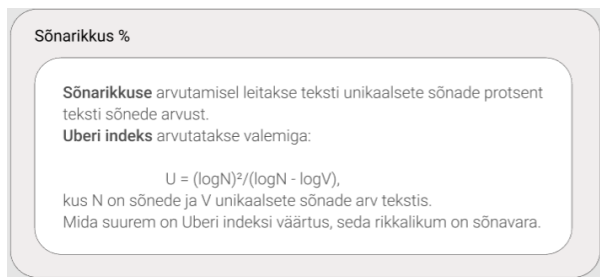
Joonis 16. Tekstianalüüsi sakk, arvutatud väärtuste ja valikute leht (vt lisa 4).

Vigaste sõnade analüsaatori puhul on tegemist automaatse analüüsiga, seega siin võib esineda vigu. Sellepärast on hindajal võimalus automaatselt märgistatud sõnadele lisaks ka ise vigaseid sõnu märkida ning sõnadelt, mis on keskkonna poolt valesti vigaseks märgitud, ka märgit maha võtta. Seda tulemust kajastatakse hindajale ka teksti analüüsi saki lehel (joonis 16).

This screenshot is identical to the previous one, but with the checkboxes in the "Märgi tekstis" column selected, indicating that the tool is showing spelling corrections for the detected errors.

Joonis 17. Järjestikustes lausetes esinevate sõnakorduste kuvamine (vt lisa 4).

¹⁶ <https://www.figma.com/>



Joonis 18. Sõnarikkuse infonupu alt avanev selgitus (vt lisa 4).

Teksti analüüsil kasutatud arvutuste selgitused on lisatud infonupukese alla (joonis 18). Nupul klõpsates avaneb tekstikast infoga (selgitus, arvutuskäik jms). Kastikese sulgemiseks tuleb sellel klõpsata.

3.7 Mõtted edasiarendustest

Magistritöös käsitletud arenduste rakendamisel EIS-is planeeritakse järgmise etapina lahendust ülesannete hindajatega testida, nendelt tagasisidet küsida ja hindamisvõimalusi vastavalt nende ettepanekutele täiendada.

Eesti keele e-riigieksamile mõeldes leiaks hindamisel kindlasti kasutust HTM-i keeletehnoloogia teadus- ja arendustegevuse programmi „Eesti keeletehnoloogia 2018 – 2027“ [6] raames loodav grammatikakontrollija, mis leiab lausest:

- komavead,
- ühildumis- ja rektsioonivead (ta võis tegema),
- sõnad, mille vorm ei sobi lausesse (muutis elu praemaks).

Harno eesti keele valdkonna spetsialistidelt saadud ettepanekute hulgas on lisaks kirjavigade märgendamisele ka suure ja väikese algustähe õigekiri, sõnade kokku- ja lahkukirjutamisel tehtud vigade leidmine ja hindajale kuvamine.

Karl Erik Karindi on oma bakalaureusetöös „Lõputöö teksti analüsaator“ [26] kirjeldanud ja loonud tsitaadituvastaja. Sellest võiks olla abi riigieksami esimese osa, lugemisülesande, hindamise automatiseerimisel nende ülesannete puhul, kus lahendaja peab vastusena loetud teksti tsiteerima. Samuti annaks selle abil kontrollida, kas kirjutamisülesandes ei ole liigselt alusteksti tsiteeritud.

EIS-is saab EstNLTK vahendeid kasutada ka avatud vastusega lünga, kuhu vastaja peab enamasti vastuseks kirjutama sõna, hindamismaatriksis. Praegu lisatakse hindamismaatriksisse õige sõna ja EIS kontrollib sisestatud sõna vastavust hindamismaatriksis olevale, 100% vastavuse korral annab vastuse eest automaatselt punktid. Kui ülesandes ei kontrollita õigekirja, siis loetakse hiljem vastuste standardiseerimise käigus ülesande koostaja poolt õigeks ka sõnad, mille kirja pilt on vigane, aga mis sisult on vastusena aktsepteeritavad. EstNLTK-d koos

ühendamise ja oletamisega rakendades oleks võimalik automaatselt õigeks lugeda tunduvalt rohkem vastuseid, ilma et hindamismatriksisse peaks standardiseerimisel lisama kirjavigadega vastuseid.

Sõnaveebi õpetaja tööriist [37] pakub võimalust teksti kirjutaja sõnavara põhjal keeletaset hinnata. Harno testide ja hindamise osakonnas koostatakse teste ka õppeainele eesti keel teise keelena ja eesti keele tasemeeksameid, kus määratakse sooritajate keeletase. Siin oleks võimalus ka EIS-is analoogset lahendust kasutades saada esmane hinnang ülesande lahendaja keeletasemele.

Lisaks hindajate töö lihtsustamisele võimaldab teksti analüüs anda põhjalikumat tagasisidet teksti kohta ka kirjutajale. Selleks oleks vaja lisada sooritajale vastuse vaatamise vaatesse võimalus näha tekstianalüüsi ja hindaja täpsustusi.

3.8 Analüüs

EstNLTK vahendite abil on võimalik eestikeelset teksti analüüsida ja kuvada hindajatele lisainformatsiooni, mis aitaks neil teha hindamisotsuseid. Käesolevas töös käsitletud vahendite valik põhineb hindajatelt saadud tagasisidel ja töö koostajale teada olevatel EstNLTK võimalustel. Loodud lahendusi ei jõutud töö raames hindajatega testida.

Kirjavigade märgendamisel leitakse praegu ainult sõnad, mis *SpellCheckRetagger* atribuuti *spelling* kasutades osutuvad vigaseks. Samas töö käigus analüüsitud tekstinäites esineb vigu, kus sõna on küll õigesti kirjutatud, aga antud lause kontekstis vale. Näiteks allolevas lauses (joonis 19) on märgitud kollasega kirjavigade märgendajaga leitud vigased sõnad, aga lause tervikuna ei ole korrektne. Ebakõlad lauses on siin näites allajoonitud.

Selliste inimestega ei või kunagi kindel olla kas nad räägivad tõtt või mitte aga tõeks on see tulnud Eestist väljastpool näiteks Ameerika Osariikides, kus vähemusinimesi kiusatatakse taga nende naha värvuse poolest, see ei piiranud kui politsei väideltavalt kägistas ära värvusega mehe.

Joonis 19. Sõnade õigekiri ja lause kontekst.

Praegu on lahenduseks see, kui hindaja saab ise täiendavalt märkida, kui sõna ei ole korrektne.

Lisaks on võimalus, et tekstis märgitakse valeks sõna, mida tuleks lugeda õigeks. Joonisel 20 on kirjavigade märgendaja märkinud vigaseks sõna „EKRE“, mis antud kontekstis tuleb õigeks lugeda (sõna „Naiteks“ on õigesti märgitud).

Naiteks, EKRE pooldajad vihkavad pagulasi, sest esiteks paljud pagulased ei taha rääkida eesti keeles ja teiseks neil on teistsugune arvamus usu suhtes.

Joonis 20. Vigaseks märgendatud sõnad lauses.

Probleemiks on ka sõnad, mis jäävad analüüsil mitmeseks. Näiteks sõna „kui“ võib olla nii side- kui määrsõna, sõna „teisele“ nii järgarv- kui asesõna. Käesolevas töös on mitmeseks jäänud sõnade puhul kasutatud esimesena pakutud varianti.

Kirjavigade märgendaja ei leia praegu tekstist ka kokku- ja lahkukirjutamise vigu ega suure algustähe vigu lauses.

Kindlasti on oluline peale töös käsitletud EstNLTK võimaluste rakendamist EIS-is neid ka hindajatega analüüsida ning vastavalt hindajate ettepanekutele täiendada.

Kokkuvõte

Käesoleva magistritöö eesmärgiks oli uurida, milliseid EstNLTK võimalusi saab kasutada EIS-is eestikeelsete vaba vastusega ülesannete hindamisel ning vastavalt Harno e-testide hindajate ettepanekutele pakkuda välja lahendused hindamist hõlbustavaks tekstianalüüsiks.

Püstitatud eesmärgi täitmiseks anti ülevaade eksamite infosüsteemi EIS ülesannete ja testide koostamisega seotud funktsionaalsusest ning eesti keele eksami hindamisest riigieksamite näitel. Selleks uuriti eesti keele riigieksami eristuskirja ja hindamisjuhendit, anti ülevaade erinevatest hindamisvõimalustest EIS-is ning tehti parandusettepanekud ülesannete automaatse hindamise osas.

Automaatset hindamist oleks võimalik parandada 0 punktiga hinnatavate ülesannete osas:

- Ülesande koostaja saab EIS-ile ette anda teksti pikkuse sõnades, millest väiksema sõnade arvu korral hinnatakse tekst automaatselt 0 punktiga ja õigekeelsusvigade arvu, millest rohkemate vigade korral hinnatakse õigekeelsus automaatselt 0 punktiga.
- Alustekstide kasutamise puhul saab ülesande koostaja määrata protsendi, millest rohkem ei tohi kirjutamise osas olla kattuvusi alustekstidega, vastasel juhul hinnatakse tekst automaatselt 0 punktiga.
- Omavahel seotud sisuplokkide hindamise puhul saab arvestada eelmise sisuploki tulemust järgmise hindamisel.

Töö käigus loodi koodinäidised tekstis:

- kirjavigade märgendamiseks;
- lausete arvu, keskmise lausepikkuse, sõnarikkuse, pikkade sõnade protsendi, loetavuse ja nominaalsuse leidmiseks;
- stopp-sõnade arvu ja sõnakorduste leidmiseks.

Lisaks koostati prototüüp hindaja vaate täiendamisest EIS-is.

Loodud abivahend ei loo veel võimalust õpilaste omaloomingulisi teksti täielikult arvuti abil hinnata, aga hindaja tööd pikemate eestikeelsete tekstide hindamisel peaks see siiski lihtsustama.

Edaspidi plaanitakse testide katsetamisel loodud võimalusi katsetada ja hindajate tagasiside põhjal hindamisvahendit täiustada ning võtta kasutusele ka selle töö fookusest välja jäänud EstNLTKi tekstianalüüsiks mõeldud vahendeid.

Viidatud kirjandus

1. Eesti keele RE 2021 põhiksam kirjutamisosa hindamisjuhend. (2021). https://innovesa.sharepoint.com/:b:r/sites/SA_Innove_testide_keskus/Shared%20Documents/Riigiksamid/Eesti%20keel/2021/eesti%20keele%20RE%202021%20p%C3%B5hiksam_kirjutamisosa%20hindamisjuhend.pdf?csf=1&web=1&e=S3vfQT (vaadatud 11.04.2022)
2. Haridusvaldkonna arengukava 2021-2035 (2020). Haridus- ja teadusministeerium. https://www.hm.ee/sites/default/files/eesti_haridusvaldkonna_arengukava_2035_seisuga_2020.03.27.pdf (vaadatud 10.05.2022)
3. Vetik, L. (2016). Kuidas olla valmis elektroonilisteks tasemetöödeks? Koolielu. Haridusportaal. <https://koolielu.ee/uudiskiri/readnews/505640/kuidas-olla-valmis-elektroonilisteks-tasemetoodeks> (vaadatud 11.04.2022)
4. Voolaid, H. (2017). Õppeasutuste järelevalvest 2001–2017. *Ülevaade haridussüsteemi välishindamisest 2016/2017. õppeaastal, lk 47-52.* https://www.hm.ee/sites/default/files/ulevaade_haridussusteemi_valishindamisest_2016-2017_0a.pdf (vaadatud 12.04.2022)
5. EstNLTK <https://estnlk.github.io/> (vaadatud 11.04.2022)
6. Haridus Ja Teadusministeeriumi keeletehnoloogia teadus- ja arendustegevuse programm „Eesti keeletehnoloogia 2018-2027“. (2022). Keeletehnoloogia. <https://www.keeletehnoloogia.ee/et> (vaadatud 11.05.2022)
7. EIS-i kasutusjuhendid (avalik vaade) (2022). <https://projektid.edu.ee/pages/viewpage.action?pageId=41683038> (vaadatud 11.05.2022)
8. Tasemetööd. (2022). Harno testide ja hindamise osakonna veebileht. <https://projektid.edu.ee/pages/viewpage.action?pageId=68650948> (vaadatud 11.05.2022)
9. E-kogud ja d-testid (2021). Harno testide ja hindamise osakonna veebileht. <https://projektid.edu.ee/display/THO/E-kogud+ja+d-testid> (vaadatud 11.05.2022)
10. Eriolukorrast tingitud distantsõppe kogemused ja mõju Eesti üldharidussüsteemile. (2020). Tallinna Ülikool. HTM-i poolt tellitud uuringu „Valdkondliku teadus- ja arendustegevuse tugevdamine“ vaheraport. https://www.hm.ee/sites/default/files/tlu_raport_distantsope_yldharidus_2810.pdf
11. Lähtetaseme testid õppeaasta alguses (2022). Harno testide ja hindamise osakonna veebileht. <https://projektid.edu.ee/pages/viewpage.action?pageId=88477563>
12. EIS kasutusjuhendid (asutusesiseseks kasutamiseks). (2022). <https://projektid.edu.ee/display/EK/EISi+kasutusjuhendid> (vaadatud 10.05.2022)
13. Eksamite infosüsteemi testikeskkond (asutusesiseseks kasutamiseks) (2022). <https://test.ekk.edu.ee/ekk/> (vaadatud 10.05.2022)

14. Eksamite infosüsteemi pre-live keskkond (asutusesiseseks kasutamiseks) (2022). <https://prelive.ekk.edu.ee/ekk/> (vaadatud 10.05.2022)
15. EISI kasutusjuhend e-testide spetsialistidele (asutusesiseseks kasutamiseks) (2020). <https://projektid.edu.ee/display/EK/EISI+kasutusjuhend+e-testide+spetsialistidele>
16. Eesti keele e-tasemetööd I-II kooliaste (2021). Harno testide ja hindamisosakonna veebileht. <https://projektid.edu.ee/pages/viewpage.action?pageId=88480282> (vaadatud 26.04.2022)
17. Eesti keel teise keelena (2021). Harno testide ja hindamisosakonna veebileht. <https://projektid.edu.ee/display/THO/Eesti+keel+teise+keelena> (vaadatud 26.04.2022)
18. Riigieksamid. (2022). Haridus- ja Teadusministeeriumi veebileht. <https://www.hm.ee/et/tegevused/valishindamine/riigieksamid> (vaadatud 26.04.2022)
19. Eesti keele riigieksami eristuskiri. (2022). Haridus- ja Noorteamet. [eesti keele RE eristuskiri 2020.pdf](https://www.hm.ee/et/tegevused/valishindamine/riigieksamid) (vaadatud 26.04.2022)
20. Tamuri, K.; Roosna, E.; Mürsepp, T.; Mõisja, K. (2019). Eesti keele kui teise keele õpitulemuste e-hindamisest. *Õpetajate Leht*, 20. detsember 2019. <https://dea.digar.ee/?a=is&oid=opetajateleht20191220&type=staticpdf> (vaadatud 26.04.2022)
21. Puksand, H.; Uusen, A. E-hindamine tuleb nagunii! *Õpetajate Leht* 15.11.2019. <https://dea.digar.ee/?a=is&oid=opetajateleht20191115&type=staticpdf> (vaadatud 26.04.2022)
22. Puksand, H. (2017). Eesti keele riigieksam liigub e-eksami poole. *Oma Keel* 1/2017. https://www.emakeeleselts.ee/omakeel/2017_1/OK-1-2017_06.pdf (vaadatud 26.04.2022)
23. Orasmaa, S.; Petmanson, T.; Tkachenko, A.; Laur, S. ja Kaalep, H-J. (2016). EstNLTK—NLP Toolkit for Estonian. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). <https://aclanthology.org/L16-1390.pdf> (vaadatud 11.04.2022)
24. EstNLTK baastarkused sisukorraga. Eesti keele töötlus Pythonis. (2020). Tartu Ülikool. https://github.com/d009/EstNLP/blob/master/estnltk_baastarkused_sisukorraga.ipynb (vaadatud 12.05.2022)
25. Praktikum 2. Eesti keele töötlus Pythonis. Koostajad Orasmaa, S., Eiche, S., Särg, D. (2021). Tartu Ülikool. <https://github.com/d009/EstNLP/tree/master/Praktikum2> (vaadatud 10.05.2022)
26. Karindi, K., E. (2020). Lõputöö teksti analüsaator. Bakalaureusetöö. Tartu Ülikool arvutiteaduse instituut. https://comserv.cs.ut.ee/ati_thesis/datasheet.php?id=69686&year=2020 (vaadatud 11.05.2022)
27. Kalmus, T. (2019). Metoodika kasvatusväärtuste automaatseks eraldamiseks online-meediatekstidest. Magistritöö. Tartu Ülikool arvutiteaduse instituut. https://comserv.cs.ut.ee/home/files/Teet_Kalmus_magistritoo.pdf?study=ATILoputoo

- [&reference=04D105DFE7880BCA344DF6DE8337E42883D4028B](#) (vaadatud 26.04.2022)
28. Praktikum 9. Eesti keele töötlus Pythonis. Koostajad Orasmaa, S., Eiche, S., Särg, D. (2021). Tartu Ülikool <https://github.com/d009/EstNLP/tree/master/Praktikum9> (vaadatud 12.05.2022)
29. Text segmentation: Spellcheck and normalization suggestions for words. `estnlTK/B_03_segmentation_words_spelling_normalization.ipynb` (2020). [https://github.com/estnlTK/estnlTK/blob/version 1.6/tutorials/nlp_pipeline/B_03_segme ntation_words_spelling_normalization.ipynb](https://github.com/estnlTK/estnlTK/blob/version%201.6/tutorials/nlp_pipeline/B_03_segmentation_words_spelling_normalization.ipynb) (vaadatud 11.05.2022)
30. Pajupuu, H.; Kerge, K.; Alp, P. Sõnavara loomulik rikkus haritud keeleoskaja tekstides (2009). Eesti Rakenduslingvistika Ühingu aastaraamat. https://www.researchgate.net/publication/270016020_Sonavara_loomulik_rikkus_haritud_keeleskaja_tekstides (vaadatud: 03.05.2022)
31. Hout, R.; Vermeer, A. (2007). Comparing measures of lexical richness. In: H. Daller, J. Milton & J. Treffers-Daller (eds.), *Modelling and assessing vocabulary knowledge* (93-116). Cambridge: Cambridge University Press. https://www.researchgate.net/publication/254801850_Comparing_measures_of_lexical_richness (vaadatud: 03.05.2022)
32. Puksand, H. (2004). Loetavusindeksi Lix sobivusest eestikeelsele tekstile. Teksti ja taustad III. Lingvistiline tekstianalüüs. *Tartu Ülikooli eesti keele õppetooli toimetised* 28. lk 108-119. <https://docplayer.ee/180914672-Tekstid-ja-taustad-iii.html> (vaadatud: 03.05.2022)
33. Kerge, K. (2012). Kas õpikuteksti mõõdetavad tunnused on kõnekad? *Oma Keel* 2/2021 lk 23-34. https://www.emakeeleselts.ee/omakeel/2012_2/OK_2012-2_03.pdf (vaadatud: 03.05.2022)
34. Allkivi-Metsoja, K.; Kippar, J. (puudub). Kvantitatiivne sisuanalüüs: sõnasagedusloendid ja võtmesõnad. *Keel ja tehnoloogia*. Tallinna Ülikool. http://www.tlu.ee/~jaagup/dh/keeletehnoloogia/06_Kvantitatiivne_sisuanalyys_sonasedus_votmesonad.pdf (vaadatud: 09.05.2022)
35. Uihoaed, K. (2018). Eesti keele stoppsõnad. Tartu Ülikooli raamatukogu. <http://datadoi.ee/bitstream/handle/33/78/estonian-stopwords-lemmas.txt?sequence=2&isAllowed=y> (vaadatud: 09.05.2022)
36. Eesti keele riigieksam 2017. Kirjutamisosa hindamisjuhend (2017). https://www.innove.ee/wp-content/uploads/2018/10/Eesti-keele-RE-2017_Kirjutamisosa-hindamisjuhend.pdf
37. Kallas, J.; Koppel, K.; Pool, R.; Üksik, T.; Tsepelina, K. (2022). Eesti keele kui teise keele õpetaja tööriistad. Eesti Keele Instituut. DOI: 10.15155/3-00-0000-0000-0000-08C04L; <https://sonaveeb.ee/teacher-tools> (09.05.2022)

38. POS tag descriptions. (2018) EstNLTK tutorials.
https://github.com/estnltk/estnltk/blob/develop_1.6/tutorials/nlp_pipeline/A_02_morphology_tables.ipynb#Table:-POS-tag-descriptions

Lisad

Lisa 1. Ülevaade EIS-is ülesannete koostamiseks kasutatavatest sisuplokkidest.

Ülesande sisend

Nimetus	Kirjeldus	Hindamisvõimalused
Alustekst	Ülesande sisuteksti sisestamiseks. Võimalus kasutada kireva teksti redaktorit, lisada tabeleid, pilte.	Puuduvad
Multimeedia	Sisuplokki saab sisestada audio- või videofaili. Multimeedia mängimiseks saab ülesande koostaja valida kas kasutatakse brauseri mängijat või jPlayerit	Puuduvad
Matemaatiline tekst	Matemaatikatehete sisestamiseks saab kasutada lihtsat matemaatikaredaktorit. Sisestamiseks ilmub nupuriba enamlevinud sümbolite sisestamiseks.	Puuduvad
Matemaatiline tekst (MathType)	Matemaatikatehete sisestamine Wirise moodulit kasutades.	Puuduvad
Pilt	Sisuplokki saab üles laadida 1 või mitu pildifaili, mida saab ülesande sisuplokkide kujunduses kasutada. Soovitatav on kasutada .png, .jpg või .gif formaati.	Puuduvad
Muu fail	Sisuplokki saab salvestada faili, mida ülesande lahendaja saab failina ka EIS-ist alla laadida.	Puuduvad
HTML päis	Võimalus HTML-is lisada ülesandele täiendavaid kujundusvõimalusi.	Puuduvad
Google Charts	Võimalus Google Charts vahendi abil kujundada erinevaid diagramme.	Puuduvad
Juhuarv	Võimalus anda ette vahemik, milles EIS genereerib juhuslikke arve (täisarvud või reaalarvud). Juhuarve saab ID abil kasutada ülesannetes.	Puuduvad

Ülesande/ vastuse tüüp

Nimetus	Kirjeldus	Hindamisvõimalused
Alade värvimine	Pildil saab määrata piirkonnad, mida lahendaja saab eri värvidega märkida. Ülesande koostaja määrab piirkonnad ja värvide tähendused.	Iga piirkonna jaoks luuakse hindamismatriks, kus määratakse mitu punkti mingi värvi kasutamine selles piirkonnas annab. Sisuplokk on täielikult arvutihinnatav.
Avatud vastusega küsimus	Lahendaja saab kuvatud kastikesse kirjutada suvalise pikkusega vabas vormis vastuse. Koostaja saab määrata, kui suur on	Võimalus kuvada sõnade arv vastuses. Võimalus kasutada maski.

	vastuse sisestamise aken ja kas lahendaja saab kasutada kireva teksti võimalust.	Tekst tüüpi vaba teksti vastuste puhul ei ole mõistlik hindamismaatriksit täita. Kui vastuse tüübiks on valitud täis- või reaalarv, siis saab hindamismaatriksis anda ka õige vastuse ning ülesanne on arvutihinnatav.
Avatud vastusega lünk	Lahendajale kuvatakse tekst, milles ta peab täitma avatud lüngad (lünkasid võib olla mitu).	Igal lüngal on eraldi hindamismaatriks. Lünkade puhul eeldatakse lühemat vastust, kui avatud vastusega küsimuse puhul. Olenevalt vastuse pikkusest on võimalik lisada iga lünga hindamismaatriksisse õige vastus ja muuta ülesanne täielikult arvutihinnatavaks.
Desmos	Võimaldab ülesande sisuplokis kasutada Desmose graafikakalkulaatorit.	Ülesanne on ainult inimhinnatav.
Faili salvestamine	Lahendajale antakse võimalus suvalist tüüpi faili üles laadida.	Ülesanne on ainult inimhinnatav.
GeoGebra	Võimaldab ülesande sisuplokis kasutada GeoGebra graafikakalkulaatorit.	Ülesanne on ainult inimhinnatav.
Joonistamine	Võimalik kasutada arvutihinnatavat joont või mitte-arvutihinnatavat joont. Saab joonistada ka kindlaid kujundeid. Joonistamise põhjaks tuleb süsteemi laadida pilt. Pildile saab lisada ka teksti.	Arvutihinnatava joone puhul on olemas hindamismaatriks ja sisuplokk on täielikult arvutihinnatav. Mitte-arvutihinnataval joonel hindamismaatriksit ei ole ja sisuplokk on täielikult inimhinnatav.
Järjestamine	Lahendaja peab õigesse järjekorda panema koostaja poolt sisestatud valikud.	Hindamismaatriksis saab seadistada, kuidas järjestust hinnatakse. Sisuplokk on täielikult arvutihinnatav
Järjestamine pildil	Taustapildile kantakse piirkonnad, millel lahendaja peab kindlas järjekorras klõpsama.	Hindamismaatriksis saab seadistada, kuidas järjestust hinnatakse. Sisuplokk on täielikult arvutihinnatav
Kõne salvestamine	Lahendaja saab vastusena salvestada heli. Sõltuvalt sisuploki seadistustest saab ka salvestust kuulata ja uuesti salvestada.	Ülesanne on ainult inimhinnatav.

Liugur	Sirglõik, millel on võimalik kuvada numbrilised vahemikud ja millel saab liigutada liugurit.	Hindamismaatriksisse saab sisestada liuguri õige väärtuse. Sisuplokk on täielikult arvutihinnatav
Lühivastusega küsimus	Lahendajale kuvatakse üks või mitu sisestusvälja, millesse saab sisestada vastuse.	Üks hindamismaatriks kehtib kõikide väljade puhul. Kui vastusena oodatakse konkreetset sõna või arvu, siis saab sisuploki muuta ka täielikult arvutihinnatavaks. Muidu inimhinnatav.
Matemaatilise teksti sisestamine	Lahendajale kuvatakse üks või mitu sisestusvälja, milles klõpsates avaneb matemaatiliste sümbolite sisestamiseks mõeldud nupuriba.	Üks hindamismaatriks kehtib kõikidele sisestusväljadele. Juhul kui on 1 väli ja vastusena oodatakse konkreetset arvu või avaldist, siis võimalik muuta arvutihinnatavaks. Muidu inimhinnatav.
Matemaatilise teksti sisestamine (MathType)	Lahendajale kuvatakse üks või mitu sisestusvälja, mille kohale on Wirise matemaatilise teksti sisestamise nupuriba.	Üks hindamismaatriks kehtib kõikidele sisestusväljadele. Ülesanne on ainult inimhinnatav.
Mitme valikuga tabel	Lahendajale kuvatakse raadionuppude või märkeruutudega tabel, kus ta saab valikuid teha.	Hindamismaatriksit saab seadistada kas ühesugusena kõikidele küsimustele või iga küsimuse puhul eraldi. Sisuplokk on täielikult arvutihinnatav.
Märkimine pildil	Lahendajale kuvatakse taustapilt ja ta peab sellel märkima ühe või mitu kohta. Märkekohale ilmub väike ring.	Hindamismaatriksis määratakse piirkonnad, kuhu sattunud märkimisi hinnatakse. Piirkondi lahendajale ei näidata. Sisuplokk on täielikult arvutihinnatav.
Märkimine pildil II	Põhimõttelt samasugune ülesanne, kui „Märkimine pildil“, aga siin on võimalik eraldi hinnata ka märkimisi väljaspool piiratud ala.	Hindamismaatriksis määratakse piirkonnad, hinnata saab nii määratud piirkondadesse kui ka sellest väljaspool tehtud märkimisi. Piirkondi lahendajale ei näidata. Sisuplokk on täielikult arvutihinnatav.
Pangaga lünk	Lahendajale kuvatakse lünktekst ja valikute pank, milles olevad valikud tuleb lünkadesse lohistada.	Hindamismaatriksis antakse vastus paarina valiku ID ja sellele vastava lünga ID. Igal

		lüngal peab olema ülesande piires unikaalne ID. Sisuplokk on täielikult arvutihinnatav.
Pildi avamine	Ülesandes on peidetud pilt, mille peal on ruudustik. Iga ruudukese kohta on küsimus, millele õigesti vastates avaneb see ruut ja pildiosa ilmub nähtavale.	Hindamismaatriks puudub ja sisuplokk punkte ei anna.
Pildil oleva kujundi valik	Taustapildil on märgitud piirkonnad (võivad olla lahendaja jaoks nähtamatud), mida lahendaja saab valida.	Hindamismaatriksis antakse vastus piirkonna koodiga. Sisuplokk on täielikult arvutihinnatav.
Piltide lohistamine	Lahendajale kuvatakse suure taustapildi kõrval pisemad pildid, mis tuleb hiirega lohistada taustapildil õigesse kohta (ülesande koostaja poolt määratud piirkonnad). Pilte võib taustapildil lohistada ka mujale (nn määramata piirkonda)	Hindamismaatriksis näidatakse piirkonnad taustapildil ja hindepallide arv, mis lahendaja saab, kui ta pildi vastavasse piirkonda lohistab. Sisuplokk on täielikult arvutihinnatav.
Piltide lohistamine II	Analoogne sisuplokk eelnevaga võrreldes selle vahega, et hindamismaatriksis saab kasutada ka määramata ala (ehk siis märgitud alast väljaspoole jäänud piltide puhul seda hindamismaatriksis arvestada).	Hindamismaatriksis näidatakse piirkonnad taustapildil ja hindepallide arv, mis lahendaja saab, kui ta pildi vastavasse piirkonda lohistab. Sisuplokk on täielikult arvutihinnatav.
Piltide lohistamine kujunditele	Sisuplokk sarnaneb „Piltide lohistamise“ plokile aga siin saab pilte lohistada ainult määratud piirkondadesse.	Hindamismaatriksis seotakse ID abil piirkonna ja lohistatava pildi paar. Sisuplokk on täielikult arvutihinnatav.
Ristsõna	Lahendajale kuvatakse ristsõna, mida saab täita.	Hindamismaatriks on iga sõna kohta. Õigesti kirjutatud sõna annab määratud punktid. Sisuplokk on täielikult arvutihinnatav.
Seostamine	Lahendajale kuvatakse sõnade või ka pikemate tekstide hulk, kus tuleb hulgas olevad objektid paari kaupa tabeli lahtritesse lohistada (ühel real on üks paar).	Hindamismaatriksis näidatakse paar seostatavate valikute koodide paarina. Sisuplokk on täielikult arvutihinnatav.
Seostamine pildil	Lahendajale kuvatakse taustapilt, millel on näha piirkonnad. Lahendaja ülesanne	Hindamismaatriksis antakse paaridena seosed 2 piirkonna

	hiirega ühest piirkonnast teiseni joont tõmmates need omavahel seostada.	vahel ja nende eest antavad punktid. Sisuplokk on täielikult arvutihinnatav.
Sobitamine	Lahendajale kuvatakse 2 valikute hulka ja ta peab nende vahel looma paare.	Hindamismaatriksis näidatakse ühele hulgale vastav vaste teises hulgas. Sisuplokk on täielikult arvutihinnatav.
Sobitamine kolme hulgaga	Lahendajale kuvatakse 3 valikute hulka ja ta peab looma paare 1. ja 2. hulga ning 2. ja 3. hulga vahel.	Hindamismaatriksis võetakse aluseks 2. hulk ja kirjeldatakse selle seoses vastavalt 1. ja 3. hulga elementidega. Seoseid hulkade vahel hinnatakse eraldi. Sisuplokk on täielikult arvutihinnatav.
Sobitamine kolme hulgaga kolmikute hindamisega	Lahendajale kuvatakse 3 valikute hulka ja ta peab looma seose kõigi 3 hulga vahel.	Hindamismaatriksis kirjeldatakse seos kõigi 3 hulga vahel ja hinnatakse kolmiku õigsust. Sisuplokk on täielikult arvutihinnatav.
Teekonna märkimine	Lahendajale kuvatakse taustapilt, millele on lisatud ruudustik. Ruudul klõpsates tekib jälg ja jäljerida viib õige ruuduni. Jälg peab olema jätkuv ehk ruudukesi ei tohi vahele jätta.	Hindamismaatriksisse lisatakse õige teekonna moodustavate ruutude selgitused. Sisuplokk on täielikult arvutihinnatav.
Tekstide lohistamine	Lahendajale kuvatakse suure taustapildi kõrval tekstid, mis tuleb hiirega lohistada taustapildil õigesse kohta (ülesande koostaja poolt määratud piirkonnad). Pilte võib taustapildil lohistada ka mujale (nn määramata piirkonda)	Hindamismaatriksis seotakse tekst ja piirkond, kuhu tekst tuleb lohistada. Sisuplokk on täielikult arvutihinnatav.
Tekstide lohistamine kujunditele	Sarnaneb „Tekstide lohistamise“ sisuplokile, aga tekste saab lohistada ainult määratud piirkondadesse.	Hindamismaatriksis seotakse ID abil piirkonna ja lohistatava teksti paar. Sisuplokk on täielikult arvutihinnatav.
Tekstide seostamine kujunditega	Lahendaja saab taustapildil määratud piirkonnad (kujundid) seostada lisatud tekstivalikutega.	Hindamismaatriksis seotakse tekst määratud piirkonna ID-ga. Sisuplokk on täielikult arvutihinnatav.
Tekstiosa valik	Lahendajale kuvatakse tekst, milles osad või kõik sõnad on valitavad (kas sõnal	Hindamismaatriksis on märgitavatel tekstiosadel oma ID

	klõpsamise või sõna ette lisatud märkeruuduga, võimaluse valib ülesande koostaja).	ja õiged valikud saab nende abil maatriksisse lisada. Sisuplokk on täielikult arvutihinnatav.
Tekstiosa värvimine	Lahendaja saab värviga märkida tekstis erinevaid osi. Värv valimiseks tuleb sellel klõpsata ja järgmisena klõpsata tekstiosal, mida selle värviga soovitakse märgistada.	Hindamismaatriksis seatakse tekstiosade ID vastavusse värviga, millega lahendaja peab selle tekstiosa märgistama. Sisuplokk on täielikult arvutihinnatav.
Valikvastusega küsimus	Lahendajale antakse vastusevalikud kas raadionuppude (1 valik) või märkeruutudena (mitu valikut).	Hindamismaatriksisse lisatakse õige (õigete) valiku (valikute) ID. Sisuplokk on täielikult arvutihinnatav.
Valikvastusega lünk	Lahendajale kuvatakse tekstis lüngad, milles avanevad rippmenüüs valikud.	Igal sisestatud lüngal on oma hindamismaatriks, kus valikute ID järgi määratakse selle lünga õige vastus. Sisuplokk on täielikult arvutihinnatav.
Võrguülesanne	Lahendajale kuvatakse taustapildil koostaja poolt lisatud piirkonnad, mida ta peab õiges järjekorras klõpsates üksteisega seostama, et tekiks vastuste võrk.	Hindamismaatriksisse lisatakse piirkonnad vastavalt sellele, mis järjestuses tuleb neid valida. Sisuplokk on täielikult arvutihinnatav.

Ülesande väljund

Nimetus	Kirjeldus	Hindamisvõimalused
Arvutatud väärtus	Sisuplokki lahendajale ei kuvata, seda saab kasutada ülesande koostaja, et erinevate valemite abil arvutada teiste küsimuste vastuste, saadud punktide ja tabamuste loendurite alusel „vastus“.	Hindamismaatriksis saab anda punkte vastavalt valemiga välja arvutatud väärtusele. Sisuplokk on täielikult arvutihinnatav.

Lisa 2. EstNLTK-s kasutatavad sõnaliigi tunnused.

Tabel 3. POS tag descriptions [38]

POS tag	Description	Example
A	omadussõna - algvõrre (adjektiiv - positiiv), nii käänduvad kui käändumatud	kallis või eht
C	omadussõna - keskvõrre (adjektiiv - komparatiiv)	laiem
D	määrsõna (adverb)	kõrvuti
G	genitiivatribuut (käändumatu omadussõna)	balti
H	pärinimi	Edgar
I	hüüdsõna (interjektsioon)	tere
J	sidesõna (konjunktsioon)	ja
K	kaassõna (pre/postpositsioon)	kaudu
N	põhiarvsõna (kardinaalnumeraal)	kaks
O	järgarvsõna (ordinaalnumeraal)	teine
P	asesõna (pronoomen)	see
S	nimisõna (substantiiv)	asi
U	omadussõna - ülivõrre (adjektiiv - superlatiiv)	pikim
V	teigusõna (verb)	lugema
X	verbi juurde kuuluv sõna, millel eraldi sõnaliigi tähistus puudub	plehku
Y	lühend	USA
Z	lausemärk	-, /, ...

Lisa 3. Riigieksami kirjutamisosa hindamisjuhend.

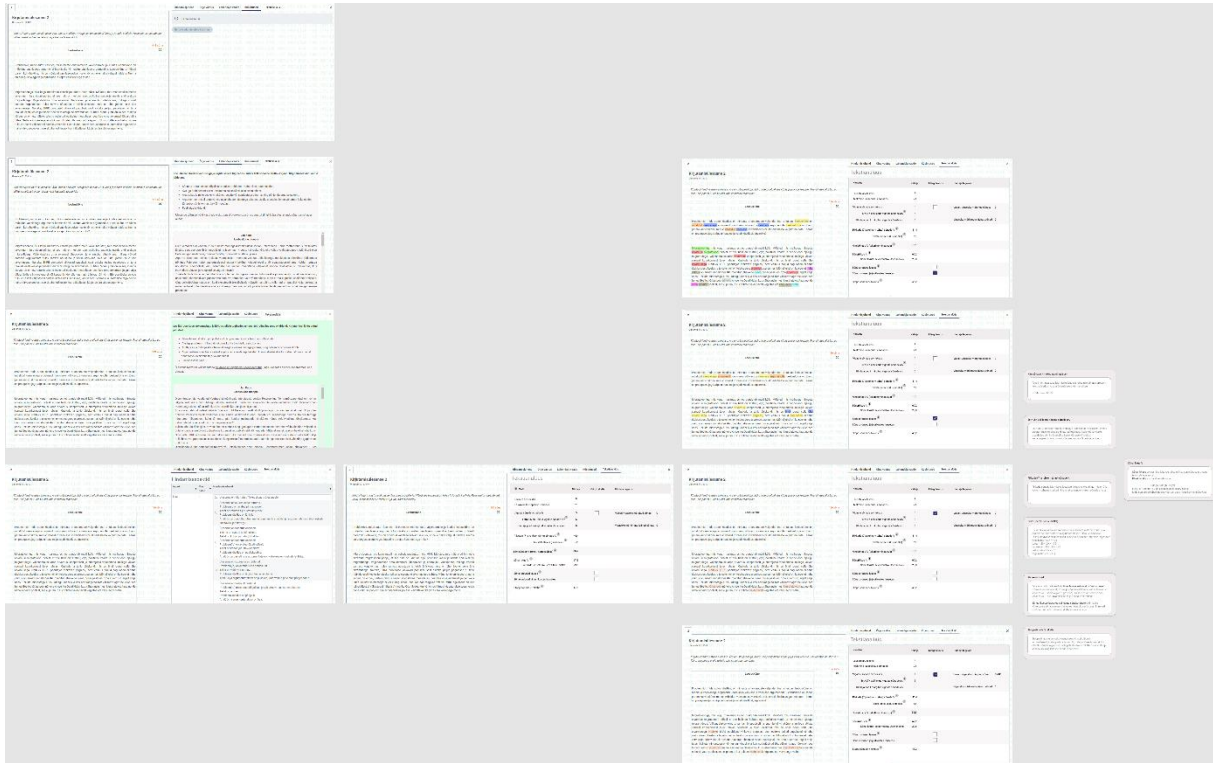
Tabel 4. Kirjandi hindamisskaala [13].

PUNK TID	SISU	ÕIGEKEELS US	STIIL	VORM
5	<ul style="list-style-type: none"> • Tekst põhineb eksamiülesandel. • Tekst on arutlev kirjand. • Probleemikäsitus on põhjalik. • Tekst on üldistusjõuline. • Analüüs ja argumentatsioon on väga hea. • Argumente toetavad näited on asjakohased. 	<ul style="list-style-type: none"> • Tekstis on kuni 2 viga 	<ul style="list-style-type: none"> • Teksti sõnastus on selge, ladus, täpne ja isikupärane. • Sõnavara on rikkalik. 	<ul style="list-style-type: none"> • Tekst on ülesehituselt sisulisvormiline tervik. • Teksti liigendamine lõikudeks on loogiline ja täpne. • Tekst on sidus.
4	<ul style="list-style-type: none"> • Tekst põhineb eksamiülesandel. • Tekst on arutlev kirjand. • Probleemikäsitus on põhjalik. • Analüüs ja argumentatsioon on hea. • Argumente toetavad näited on asjakohased. 	<ul style="list-style-type: none"> • Tekstis on 3-4 viga 	<ul style="list-style-type: none"> • Teksti sõnastus on selge ja ladus. • Sõnavara on mitmekülgne. 	<ul style="list-style-type: none"> • Tekst on ülesehituselt tervik. • Teksti liigendamine lõikudeks on üldiselt loogiline. • Tekst on sidus.
3	<ul style="list-style-type: none"> • Tekst põhineb eksamiülesandel. • Tekst on arutlev kirjand. • Probleemikäsitus on piisav. • Analüüs ja argumentatsioon on piisav, kuid väiteid ei ole alati põhjendatud. • Argumente toetavad näited on valdavalt asjakohased. 	<ul style="list-style-type: none"> • Tekstis on 5-6 viga 	<ul style="list-style-type: none"> • Teksti sõnastus on üldiselt ladus, kuid kohati ühekülgne/konarlik. • Sõnavara on piisav. 	<ul style="list-style-type: none"> • Tekst on ülesehituselt tervik. • Teksti liigendamisel lõikudeks esineb vigu (nt teksti lõikude proportsioonid ei ole alati põhjendatud jms). • Tekst on üldiselt sidus.
2	<ul style="list-style-type: none"> • Tekst põhineb eksamiülesandel, kuid esineb üksikuid kõrvalekaldeid. • Tekst on arutlevkirjand, kuid sellel on kohati jutustava/kirjeldava kirjandi tunnuseid. • Probleemikäsitus on pealiskaudne/ tekst on üldsõnaline. • Analüüs on pinnaline ja argumentatsioon väheveenev. 	<ul style="list-style-type: none"> • Tekstis on 7-8 viga 	<ul style="list-style-type: none"> • Teksti sõnastus on konarlik, kuid arusaadav. • Sõnavara on ühekülgne. 	<ul style="list-style-type: none"> • Teksti ülesehituses on puudujääke (nt alguse venitatus, välja arendamata või liiga pikk lõpetus, mitmekordne algus või lõpetus jms). • Teksti liigendamisel lõikudeks esineb vigu (nt teksti lõikude proportsioonid ei ole põhjendatud jms).

	<ul style="list-style-type: none"> • Esineb üksikuid faktivigu. • Argumente toetavaid näiteid on vähe / osa näiteid ei ole asjakohased / tsitaatide, refereeringute ja/või näidetega on liialdatud. 			<ul style="list-style-type: none"> • Tekst ei ole kohati sidus.
1	<ul style="list-style-type: none"> • Tekst põhineb eksamiülesandel üldjoontes. • Tekst on valdavalt jutustav/kirjeldav kirjand. • Probleemikäsitlus on lünklik. • Analüüs on puudulik, jutustamine domineerib analüüsi ja argumentatsiooni üle. • Esineb küsitavusi ja faktivigu. • Näited ei ole käsitletava probleemiga piisavalt seotud/näited puuduvad. 	<ul style="list-style-type: none"> • Tekstis on 9-10 viga 	<ul style="list-style-type: none"> • Teksti sõnastus on algeline ning kohati raskesti mõistetav. • Sõnavara on piiratud. 	<ul style="list-style-type: none"> • Teksti ülesehitus on ebaloogiline (nt teksti algus ja lõpetus on käsitletava probleemiga nõrgalt seotud, teksti mõni osa ei täida sellele pandud ülesandeid jms). • Teksti liigendamisel lõikudeks esineb vigu (nt teksti lõikude proportsioonid ei ole põhjendatud jms). • Suurem osa tekstist ei ole sidus.
0	<ul style="list-style-type: none"> • Tekst vastab eksamiülesandele osaliselt, st suurem osa mõttearendusest ei põhine ülesandel. 	<ul style="list-style-type: none"> • Tekstis on 11 või enam viga. 	<ul style="list-style-type: none"> • Tekst on algeline ja raskesti mõistetav. • Stiil ei sobi käsitletava probleemiga. 	<ul style="list-style-type: none"> • Tekst ei ole ülesehituselt tervik (nt teksti mõni osa (sissejuhatus, arendus, lõpetus) puudub). • Teksti liigendus on ebaloogiline. • Tekst ei ole sidus.

Lisa 4. Hindajavaate prototüüp

https://www.figma.com/proto/SLWfHjP8XNPPTaPth5VyA8/eesti_keeke_hindamine?node-id=4%3A32&scaling=min-zoom&page-id=0%3A1&starting-point-node-id=4%3A32



Joonis 18. Hindaja vaate prototüüp Figma's

Lisa 5. Vigaste sõnade arv tekstis ja nende märgendamine

```
from estnltk import Text
#muutuja vastus on õpilase sisestatud tekst
vastus="Probleemid, mida autor käsitles, oli inimeste eriarvamuste väljenda
mine ja kuidas leebuskõne on mõeldud vaenlasega argumendi loomiseks või vas
tus vaenlase argumendile. Leebuskõne on tõesti parem kui vihakõne, mis on m
õeldud vaenlase solvamiseks või, et neile halba valgust näidata. Teema on p
raeguse ajaga väga aktuaalne ja vajab käsitlust, aga miks? Migratsiooniga,
mis kogu maailmas esineb puudutab meid kõiki. Mõisted, mis määravad inimest
e arvamusi migratsiooni põhjal, et see halb on: kultuur, ego, poliitiline v
aade ja nende endi ajalugu migrantidega. Väljendatakse oma arvamusi ühepool
selt ja enamjuhul vihakõnena, millega võivad samast kogukonnast tulev rahva
s nõustuda ja tekib ühiskond, mis on ühel joonel selle ühe arvamusega. Nait
eks, EKRE pooldajad vihkavad pagulasi, sest esiteks paljud pagulased ei tah
a rääkida eesti keeles ja teiseks neil on teistsugune arvamus usu suhtes. M
õnel juhul on ka see, et naha värvus on erinev, näiteks tekstis mainitud eb
akaine mees bussi peal kes oma arvamust jäigalt välja ütles. Selliste inime
stega ei või kunagi kindel olla kas nad räägivad tõtt või mitte aga tõeks o
n see tulnud Eestist väljastpool näiteks Ameerika Osariikides, kus vähemusi
nimesi kiusatatakse taga nende naha värvuse poolest, see ei piiranud kui po
litsei väideltavalt kägistas ära värvusega mehe."

#Sisestatud tekst muudetakse Text-objektiks
tekst=Text(vastus).tag_layer()

from estnltk.taggers import SpellCheckRetagger

spelling_tagger=SpellCheckRetagger(add_spellcheck=True)
spelling_tagger.retag(tekst)
veasonad=set()
for lause in tekst.sentences:
    for sona in lause.words:
        if sona.spelling[0] == False:

            veasonad.add(sona.text)

from estnltk import Text, Layer
from estnltk.taggers import Tagger

class VigasteSonadeTagger(Tagger):
    """Märgistab vigased sõnad."""
    conf_param = ['vead']
    def __init__(self, veasonaloend, output_layer='veasonakiht', input_morp
h_analysis_layer='morph_analysis'):
        self.input_layers = [input_morph_analysis_layer]
        self.output_layer = output_layer
        self.output_attributes = ['vigased']
        self.vead = veasonaloend

    def _make_layer(self, text, layers, status):
        layer = Layer(name=self.output_layer, attributes=self.output_attri
butes, text_object=text)
        for span in layers['morph_analysis']:
            for annotation in span.annotations:
                if annotation['normalized_text'] in self.vead:
                    layer.add_annotation(span.base_span, token_type='VIGA')
                    break

        return layer
vigadeMargendaja=VigasteSonadeTagger(veasonad)
```

```
vigadeMargendaja.tag(tekst)
tekst.veasonakiht.display()
```

Probleemid, mida autor käsitles, oli inimeste eriarvamuste väljendamine ja kuidas leebuskõne on mõeldud vaenlasega argumendi loomiseks või vastus vaenlase argumendile. Leebuskõne on tõesti parem kui vihakõne, mis on mõeldud vaenlase solvamiseks või, et neile halba valgust näidata. Teema on praeguse ajaga väga aktuaalne ja vajab käsitlust, aga miks? Migratsiooniga, mis kogu maailmas esineb puudutab meid kõiki. Mõisted, mis määravad inimeste arvamusi migratsiooni põhjal, et see halb on: kultuur, ego, poliitiline vaade ja nende endi ajalugu migrantidega. Väljendatakse oma arvamusi ühepoolset ja enamjuhul vihakõnena, millega võivad samast kogukonnast tulev rahvas nõustuda ja tekib ühiskond, mis on ühel joonel selle ühe arvamusega. Näiteks, EKRE pooldajad vihkavad pagulasi, sest esiteks paljud pagulased ei taha rääkida eesti keeles ja teiseks neil on teistsugune arvamus usu suhtes. Mõnel juhul on ka see, et naha värvus on erinev, näiteks tekstis mainitud ebakaine mees bussi peal kes oma arvamust jäigalt välja ütles. Selliste inimestega ei või kunagi kindel olla kas nad räägivad tõtt või mitte aga tõeks on see tulnud Eestist väljastpool näiteks Ameerika Osariikides, kus vähemusinimesi kiusatatakse taga nende naha värvuse poolest, see ei piiranud kui politsei väideltavalt kägistas ära värvusega mehe.

```
print("Vigaste sõnade arv: ",len(veasonad))
Vigaste sõnade arv: 5
```

Lisa 6. Teksti analüüs

```
from estnltk import Text

vastus="Probleemid, mida autor käsitles, oli inimeste eriarvamuste väljenda
mine ja kuidas leebuskõne on mõeldud \
vaenlasega argumendi loomiseks või vastus vaenlase argumendile. Leebuskõne
on tõesti parem kui vihakõne, mis on \
mõeldud vaenlase solvamiseks või, et neile halba valgust näidata. Teema on
praeguse ajaga väga aktuaalne ja vajab \
käsitlust, aga miks? \
Migratsiooniga, mis kogu maailmas esineb puudutab meid kõiki. Mõisted, mis
määravad inimeste \
arvamusi migratsiooni põhjal, et see halb on: kultuur, ego, poliitiline vaa
de ja nende endi ajalugu migrantidega. \
Väljendatakse oma arvamusi ühepoolset ja enamjuhul vihakõnena, millega või
vad samast kogukonnast tulev rahvas \
nõustuda ja tekib ühiskond, mis on ühel joonel selle ühe arvamusega. Naitek
s, EKRE pooldajad vihkavad pagulasi, \
sest esiteks paljud pagulased ei taha rääkida eesti keeles ja teiseks neil
on teistsugune arvamus usu suhtes. \
Mõnel juhul on ka see, et naha värvus on erinev, näiteks tekstis mainitud e
bakaine mees bussi peal kes oma arvamust \
jäigalt välja ütles. Selliste inimestega ei või kunagi kindel olla kas nad
räägivad tõtt või mitte aga tõeks on see \
tulnud Eestist väljastpool näiteks Ameerika Osariikides, kus vähemusinimesi
kiusatakse taga nende naha värvuse poolest, \
see ei piiranud kui politsei väideltavalt kägistas ära värvusega mehe."

tekst=Text(vastus).tag_layer()
#Tekst muudetakse Text-objektiks ja lisatakse märgenduskiht.
sonedearv=0      #Muutuja sõnede loendamiseks
sonad=set()     #Hulk unikaalsete sõnade kogumiseks
keskminelausepikkus=0
pikadsonad=0
nimisonadeary=0

for lause in tekst.sentences:      #Tekst vaadatakse läbi lause kaupa
    for sona in lause.words:      #Lause vaadatakse läbi sõna kaupa
        if sona.partofspeech[0]!='Z':
            #Kirjavahemärkide eemaldamiseks kontrollitakse, et sõnaliik ei oleks 'Z'
            sonedearv+=1          #sõnede loendamine
            sonad.add(sona.lemma[0])
            #sõnade hulka lisatakse sõna lemma, hulka kasutatakse seetõttu, et seal
            #ei saa olla korduvaid elemente
            if len(sona.text) >=7:
                #pikkade sõnade leidmiseks vaadatakse sõna pikkust
                pikadsonad+=1

            if sona.partofspeech[0]=='S':
                #nimisõnade arvu leidmiseks suurendatakse muutujat juhul,
                #kui sõna sõnaliigiks on 'S'

                nimisonadeary+=1

keskminelausepikkus=sonedearv/len(tekst.sentences)
pikadesonadeosakaal=round(pikadsonad/sonedearv*100,2)
print('Lause arvu tekstis on '+str(len(tekst.sentences)))
print('Keskmine lausepikkus on '+str(round(keskminelausepikkus))+ ' sõna')
```

```
print('Pikkade sõnade %: '+str(pikkadesonadeosakaal))
print('Teksti nominaalsus NOM on: '+str(round(nimisonade arv/sonde arv*100,2
))+ '%')
print('Pikkade sõnade arv: '+str(pikkadesonad))
print('Nimisõnade arv: '+str(nimisonade arv))
sonde arv=len(sonad)
```

```
#Uberi indeksi arvutamine
```

```
import math
```

```
U=(math.pow(math.log10(sonde arv),2)/(math.log10(sonde arv)-math.log10(sona
de arv)))
```

```
print('Uberi indeks: '+str(round(U,2)))
```

```
#Sõnarikkuse arvutamine
```

```
sonarikkus=round(sonde arv/sonde arv*100,2)
```

```
print('Sõnarikkus: '+str(sonarikkus)+'%')
```

```
#Lix indeksi arvutamine
```

```
Lix=round(keskminelause pikkus+pikkadesonadeosakaal,1)
```

```
print('Teksti jälgitavust iseloomustav Lix-indeks: '+str(Lix))
```

Lisa 7. Stoppsõnade ja sõnakorduste leidmine tekstist

```
from estnltk import Text
vastus="Probleemid, mida autor käsitles, oli inimeste eriarvamuste väljenda
mine ja kuidas leebuskõne on mõeldud \
vaenlasega argumendi loomiseks või vastus vaenlase argumendile. Leebuskõne
on tõesti parem kui vihakõne, mis on \
mõeldud vaenlase solvamiseks või, et neile halba valgust näidata. Teema on
praeguse ajaga väga aktuaalne ja vajab \
käsitlust, aga miks? \
Migratsiooniga, mis kogu maailmas esineb puudutab meid kõiki. Mõisted, mis
määravad inimeste \
arvamusi migratsiooni põhjal, et see halb on: kultuur, ego, poliitiline vaa
de ja nende endi ajalugu migrantidega. \
Väljendatakse oma arvamusi ühepoolset ja enamjuhul vihakõnena, millega või
vad samast kogukonnast tulev rahvas \
nõustuda ja tekib ühiskond, mis on ühel joonel selle ühe arvamusega. Näitek
s, EKRE pooldajad vihkavad pagulasi, \
sest esiteks paljud pagulased ei taha rääkida eesti keeles ja teiseks neil
on teistsugune arvamused usu suhtes. \
Mõnel juhul on ka see, et naha värvus on erinev, näiteks tekstis mainitud e
bakaine mees bussi peal kes oma arvamust \
jäigalt välja ütles. Selliste inimestega ei või kunagi kindel olla kas nad
räägivad tõtt või mitte aga tõeks on see \
tulnud Eestist väljastpool näiteks Ameerika Osariikides, kus vähemusiinimesi
kiusatakse taga nende naha värvuse poolest, \
see ei piiranud kui politsei väideltavalt kägistas ära värvusega mehe."

tekst=Text(vastus).tag_layer()

#Võrdlusfail on alla laetud sellelt lehelt https://datadoi.ee/bitstream/han
dle/33/78/estonian-stopwords-lemmas.txt
with open('estonian-stopwords-lemmas.txt', 'r', encoding='utf-8') as stopp:
    # Faili sisu lugemine ja märgenduskihi lisamine
    stoppsonad = Text(stopp.read()).tag_layer()
#stoppsonad
stoppsonadearv=0 #Muutuja tekstis sisalduvate stopp-sõnade loendamiseks
sonedearv=0      #Muutuja sõnade arvu leidmiseks

for lause in tekst.sentences:
    for sona in lause.words:
#Tekstist kirjavahemärkide eraldamine

        if sona.partofspeech[0]!='Z':
            sonedearv+=1
            for el in stoppsonad.words:
#sõna võrreldakse stopp-sõnade loetelus olevate sõnadega
                if sona.lemma[0]==el.text:
                    stoppsonadearv+=1
#kui sõna on stopp-sõnade hulgas, siis suurendatakse loendurit

print('Tekstis olevate stopp-sõnade arv: '+str(stoppsonadearv))
print('Stopp-sõnade % tekstis: '+str(round(stoppsonadearv/sonedearv*100,2))
+'%')

#sõnastik lausete ja nendes sisalduvate sõnade jaoks
kordused={}
lausenr=0

for lause in tekst.sentences: #tekst vaadatakse lausehaaval läbi
```

```

lausenr+=1                                #lause numbrid algavad 1-st

#luuakse sõnastiks lauses korduvate sõnade jaoks
lauseskordused={}
#lause järjekorranumbriga luuakse tühi sõnastik
kordused[lausenr]={}
for sona in lause.words:

#kirjavahemärkide eemaldamiseks
if sona.partofspeech[0]!='Z':
    if sona.lemma[0] not in stopp: #stopp-sõnade eemaldamiseks

#iga sõna lisatakse sõnastikku 1 kord
    if sona.lemma[0] not in lauseskordused:
        lauseskordused[sona.lemma[0]]=1

    else:
#kui sõna juba on sõnastikus, siis suurendatakse tema esinemisarvu
        lauseskordused[sona.lemma[0]]+=1

#lauses olevate sõnade ja nende arvu sõnastik lisatakse lausete sõnasti
kku
    kordused[lausenr]=lauseskordused

#kuna lauseid nummerdati sõnastikus alates 1-st, siis muutuja nr väärtu
statakse 1-ga

nr=1
lausetearv=len(tekst.sentences) #lausete arv tekstis

kordublauses=set()
korduvad=set() #kogub korduvad sõnad märgendamise jaoks

#vaadatakse järjest läbi kõik välimise sõnastiku kirjed
while nr<lausetearv:

    uus=nr+1                                #järgnevate lausete võrdlemiseks

#nummerdatud laused vaadatakse sõnastikus läbi
    for voti in kordused[nr]:

#kui sõna on lauses korduvalt, siis lisatakse sõna korduvate hulka
        if kordused[nr][voti]>1:
            kordublauses.add(voti)
            print("lauses "+str(nr)+" kordub: "+voti, kordused[nr][voti
])
#kui sama sõna on järgmises lauses, siis lisatakse samuti korduvate hul
ka
        if voti in kordused[uus]:
            korduvad.add(voti)
            print("lausetes "+str(nr)+" ja "+str(uus)+" kordub: "+voti,
kordused[nr][voti]+kordused[uus][voti])

    uus=0
    nr+=1

```

Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Vilve Roosioks,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „EstNLTK kasutamise võimalused eestikeelsete ülesannete hindamisel eksamite infosüsteemis EIS“, mille juhendaja on Heili Orav, 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.

Vilve Roosioks

17.05.2022