

TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Karoli Kapp

Luuletuste automaatne genereerimine

Bakalaureusetöö (9 EAP)

Juhendaja: Sven Aller, MSc

Tartu 2024

Luuletuste automaatne genereerimine

Lühikokkuvõte:

Bakalaureusetöö käsitleb luuletuste automaatset genereerimist. Töö käigus luuakse luuletuste automaatne generaator, mis genereerib kasutaja esitatud teksti alusel luuletuse. Luuletus peab järgima kasutaja valitud poeetika reegleid nagu riimi, värsimõõtu ja stroofilist struktuuri. Lisaks sisaldab töö ülevaadet sarnastest olemasolevatest rakendustest. Samuti kirjeldatakse töös valitud poeetika reegleid ning programmi algoritmilist ülesehitust. Kokkuvõttes esitatakse ülevaade rakenduse tulemustest, pakutakse välja lahendusi tarkvara parandamiseks ning võimalikke edasiarendusvõimalusi luuletuste automaatse genereerimise teemal.

Võtmesõnad:

luuletuste automaatne genereerimine, loomuliku keele töötlus, eest keel

CERCS: P175 - Informaatika, süsteemiteooria.

Poetry automatic generation

Abstract:

This Bachelor's thesis is on the topic of automatic poetry generation. The aim of the thesis is to create a poetry generator which conducts a poem based on a text provided by the user. The poem must follow certain poetic elements such as rhyme, metrical rhythm and strophic structure, according to the user's choice. In addition, the paper includes an overview of similar existing applications and describes poetics as a larger subject. An overview of chosen poetic rules is presented, and the derivation of algorithms based on these rules is described. To conclude, an overview of the application's weaknesses and strengths is provided, as well as solutions for improving the software, along with a description of potential future development opportunities on this topic.

Keywords:

poetry automatic generation, natural language processing, estonian language

CERCS: P175 - Informatics, system theory.

Sisukord

Sissejuhatus	4
1. Luule ja poeetika	6
1.1 Riim	8
1.2 Meetrika	9
1.2.1 Silbilisrõhuline värsisüsteem	9
1.3 Stroofika	12
1.4 Varasemad tööd	13
1.4.1 aipoemgenerator.org	14
1.4.2 editpad.org	14
1.4.3 poemgenerator-ai.com	14
1.4.4 ChatGPT-3.5	15
2. Tehnilised vahendid	16
2.1 Python	16
2.2 EstNLTK teekide kogum	16
2.3 Google Colab	17
3. Luuletuste generaator Värsivigur	19
3.1 Algoritmiline ülesehitus	19
3.1.1 Stroofika	19
3.1.2 Riim	19
3.1.3 Värsimõõt ja süntaks	20
3.2 Teksti ettevalmistus	24
3.3 Kasutajaliides	25
4. Tulemused	27
4.1 Tagasiside küsimustiku tulemuste analüüs	28
4.1.1 Luuletuste poeetika	28
4.1.2 Luuletuste sisu	29
4.1.3 Üldine mulje ja kasutajakogemus	29
4.2 Diskussioon ja edasiarendusvõimalused	30
4.3 Võrdlus sarnaste rakendustega	32
Kokkuvõte	33
Viidatud kirjandus	34

Sissejuhatus

Ilukirjandusliku teksti, spetsiifilisemat luuleteksti genereerimise vastu on huvi tundnud mitmed arvuti- ja kirjandusteadlased juba aastakümneid (Oliveira, 2017). Inimestel on alati olnud huvi testida arvutite võimekust ning avastada, kas ja kuidas need suudavad teha tegevusi, mis meenutavad või isegi ületavad inimintellekti. On loogiline eeldada, et arvutid võiksid samuti proovida luua tekste, mis vastavad inimese loomingulistele võimetele. Luule, olles üks esimesi säilinud kõrgkultuurilisi tekstiliike, on seega keeleline väljund, mida arvuti peaks inimintellekti saavutamiseks suutma genereerida.

Luule ja teiste kirjanduslike tekstide automaatse genereerimisega on tegeletud aastakümneid, kuid endiselt kahtlevad kirjanikud, näiteks Meelis Friedenthal artiklis “Kas robotite loodud kirjandus vallutab maailma?” (2023) ja keeleteadlased Aedmaa & Kaljumäe artiklis “Skandaalne ChatGPT – kuidas panna see eesti keele heaks tööle?” (2023) selle tekstiloomes võimekuses. Siiki pakub see rohkelt kõneainet ning näitab, kui palju see neile ja laiemale lugejaskonnale huvi pakub. Artiklid nagu “Üks tont käib klassis ringi – tehisintellekt!” (Juurak, 2023) ja “ChatGPT eetilisest kasutamisest koolides” (Laas, 2023) viitavad samuti, et selliseid tekstiloomes võimekusega programme kasutatakse usinalt nii tööülesannete täitmisel kui ka hariduslikel eesmärkidel.

Luuletuse generaatoreid leidub internetiavarustes laialt, mis annab samuti au nende populaarsusele ja meelelahutuslikule väljundile. Lõputöö kirjutamise ajal ei leidu autorile teadaolevalt ühtegi rakendust, mis eestikeelset luulet selle ainueesmärgiga automaatselt genereeriks.

Töö eesmärk on uurida luuletuste automaatse genereerimise võimalusi. Uurimuse käigus plaanitakse luua tarkvaralahendus, mis suudaks konstrueerida kasutaja esitatud tekstil põhineva ning kasutaja valitud kindlaid poeetilisi reegleid järgiv eestikeelne luuletus. Poeetilistest parameetritest on siinses uurimistöös keskendutud 1) luuletuse riimiskeemile ehk täpsemalt lõppriimi paiknevusele salmis ning 2) värsimõõdule ehk silpide ja rõhkude korrapärasele järgnevusele, mida tarkvara luuletuse koostamisel aluseks võtab. Genereeritud luuletuse kvaliteedi ning seega tarkvaralahenduse edu määravateks tunnusteks on peaosas kasutaja valitud poeetiliste reeglite järgimine, kuid võrdselt püütakse teada saada, kuidas nende parameetrite kasutamise tagajärjel algne sisendtekst luule sisus kajastub ning poeetika reeglite rakendamine luuletuse sisu üldiselt muudab.

Uurimuse laiem objekt on püüda poeetika reegleid süstemaatiliselt uurida keeletehnoloogilisest vaatepunktist. Algoritmiline alus praktiliseks tööks tuletatakse teoreetilisest taustast. Uurimistöö reeglipõhine lähenemine on tehisintellekti kõrval, mida tänapäeval luule genereerimisega peaosas seostatakse, vanem, kuid üks võimalik lähenemisviis, mida eesti keeles veel teadaolevalt katsetatud pole. Kokkuvõtlikult võiks tarkvara olla algpunktiks samal või sarnasel teemal kirjutatud eestikeelsetele lahendustele ning üks paljudest viisidest leida lahendus tänapäeval tüüpiliselt tehisintellekti kasutavale probleemile.

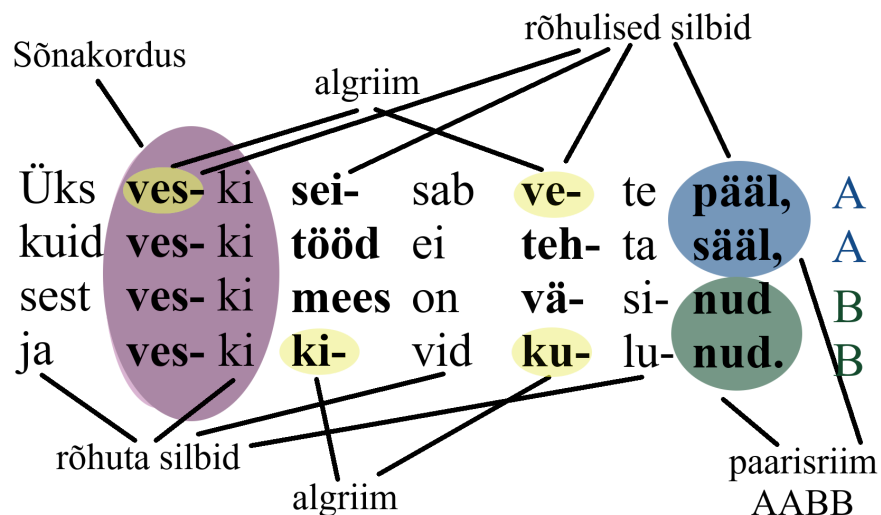
Bakalaureusetöö koosneb neljast peatükist. Esimeses antakse teoreetiline ülevaade luulest, luule poeetikast, hõlmates seejuures töö mõistmiseks oluliste põhimõistete avamist. Samuti tutvustatakse uurimistöö temaga seotud varasemaid rakendusi ja uuringuid. Teises peatükis keskendutakse programmi loomisel kasutatud tehniliste vahendite tausta selgitamisele ning põhjendatakse tehnoloogiate valikut. Kolmandas peatükis kirjeldatakse valminud tarkvaralahenduse algoritmilist ülesehitust, kasutaja valitud sisendteksti analüüsimise protsessi ning rakenduse kasutajaliidest. Töö viimane peatükk keskendub valminud tarkvara tagasiside küsimustiku tulemuste analüüsile, diskussioonile ning esitab teema edasiarendusvõimalusi. Töö lisadest leiab lingi tarkvaralahendusele ning ülevaate tagasiside küsimustikust.

1. Luule ja poeetika

Mõiste “luule” kannab endas palju tähendusi, laiemalt märgib see kogu loomingut üldse (Krull, 2011), kitsamalt märgib see poeesiat ehk keelelist loomingut, mis on üles märgitud värss- ehk seotud kõnes (Merilai et al., 2011). Luuletus on lüürika¹ ehk kirjanduse põhiliigi vorm, mille alla kuuluvad sonetid, haikud, tankad, eelegiad jpt (Merilai, 2011). Luuletuse kui vormi määratlemine on märksa keerulisem, kui pealtnäha paistab, eriti tänapäeval, kus erinevate vormi- ja žanrimängude tulemusena piirid hägustuvad. Lihtsustamise huvides on käesolevas töös võetud kasutusele Hasso Krulli artiklis “Mis on luule?” (2011) ilmunud luuletuse kõige üldtuntum mõiste: luuletus on rütmilist korrapära (mida loovad värsimõõt, riim, skandeeriv esitus) omav lühike tekst, mis kujutab minategelase tundemaailma. Selle definitsiooni kohaselt, ning toetudes “Poeetika gümnaasiumiõpikule” (Merilai et al., 2011), on iga luuletekst mõõdetav selle 1) sisus ja 2) poeetilises väljendusviisis. Selles uurimistöös valminud tarkvaralahenduse fookuses on luuletuse poeetiline mõõde ehk kujundlik keelekasutus, sisuline seejuures allub sellele väljendusviisile, mistõttu sisu “tundelisust” ei saa genereeritud luuletustes garanteerida.

“Poeetika gümnaasiumiõpikus” on eesti kirjanduse professor Arne Merilai (2011) esitanud poetikale kaks tähendusvälja: 1) poeetika kui “luulekunsti või laiemalt ilukirjanduslikkuse õpetuse” (2011:9) ning 2) poeetika kui kunstilise mõjuga sõnateose vormilised omadused, mis on uurimisobjektiks luulekunstiõpetuses ning mis alluvad poeetikaanalüüsile. Siinse uurimistöo raames mõistetakse poeetika all peamiselt selle teist tähendusvälja ehk sõnateose poeetilisi vormiomadusi. Poeetiliste omaduste all peetakse õpikule tuginedes silmas tekstile endale (tavaliselt luulele) omistatud vormilisi tunnuseid, mis häälvivad tavapärasest kõnekeelest. Õpikus on nendena loendatud rütmiline kõne (värsimõõt), keeleline heakõla (epiteedid, anagramm, foneemi valik), tekstiosade kordused (algriim, lõppriim, silbi- või sõnakordus) ning spetsiifiliselt klassikalisemate luuletuste puhul teksti korrapärane eraldatus (salmideks) jne. Neid kujundlikke keelekasutuse võtteid on äärmiselt palju.

¹ Definitsioon sõnale “lüürika” <https://sonaveeb.ee/search/unif/dlall/dsall/1%C3%BC%C3%BCrika/1>



Joonis 1. Hando Runneli “Üks veski seisab vete pääl” (2018) esimese salmi poeetilist analüüsi kujutav joonis, loodud “Poeetika gümnaasiumiõpiku” eeskujul (Merilai jt, 2011:36). Joonisel 1 on esitatud üks viis mõnede poeetiliste võtete analüüsiks. Joonisel 1 kujutatud salm on esmalt poeetilise analüüsi tarbeks eraldatud sõnadeks ja silpideks. Skandeeriva esituse abil on leitud sõnarõhud ja nende tõusud, mis on märgitud rasvaselt. Riimumise analüüsiks on silbid jaotatud eraldi häälikuteks, et neid omavahel salmis ja värsireas paiknevuse alusel võrrelda. Selle uurimistöö praktilises osas rakendatud poeetilised võtted valiti tuginedes Hasso Krulli välja pakutud klassikalisele luuletuse definitsioonile. Spetsiifilisi valikuid põhjendatakse neid tutvustavates alapeatükkides.

Luule (laiemas kui kitsamas tähenduses) olemust ja eesmärki on üritanud põhjendada mitmed juba sadu, kui mitte tuhandeid aastaid (Merilai, 2011), alates Aristotelese “Luulekunstist: poeetika” (2001) ja Martin Opitz’i “Raamat saksa luulekunstist” (2016) ning lõpetades Jürgen Rooste luuletusega “Milleks on vaja luulet” (2005). Vastajaid küsimusele otsitakse tüüpilist kirjanike seast ning ilmselt on ka küsijad nad ise või inimesed, kes juba luulet hingeliselt hindavad. Tuhandeid vaatamisi kogunud TEDTalk konverentside sarja kõnes “*Why people need poetry*” (eesti keeles “Miks inimesed vajavad luulet”) püüab kirjanik ja luulearmastaja Stephen Burt (2014) leida vastust samale küsimusele just “luulevõõrate” inimeste perspektiivist. Tema sõnul võiks luule olla inimestele elus teejuhiks ning õpetada meile, kuidas elus erinevates olukordades toime tulla, aga ka kuidas teised inimesed on meile võõrastes olukordades käitunud ja ennast neis tundnud. Kokkuvõtlikult on luule peidetud kujul empaatiaõpetus.

Siinses töös on fookuses poeetiline aspekt ehk kujundlik keelekasutus, mis Arne Merilai sõnul (2011) võimendab luuletuse sisus väljenduvaid vaimseid väärtusi. Luuletuse mõistmiseks tuleb tema arvates osata seda kujundlikku keelt dešifreerida, sest vaid nii saab avada selle teksti täieliku sisu. Semiootik Mihhail Lotman on ajakirjas Vikerkaar (1986) selgitanud, kuidas poeetika on pelgalt üks mõiste kirjeldamaks nähtust, mis leidub kõigis märgisüsteemides, olgu selleks maalikunst, geneetika või programmeerimine - sisu mõistmiseks peame tundma süsteemi, milles see on kirja pandud, sest vaid siis avaldab see oma tõelise sisu.

1.1 Riim

Jaak Põldmäe selgitab “Eesti värsiõpetuses” (2002) lahti riimi defineerimise keerukuse, kuid jõuab neutraalse definitsioonini ning nimetab riimiks “mingites kindlates värsipositsioonides fikseeritavat vastavust mingite rõhurühmade või nende osade foneemikoostiste vahel” (Põldmäe, 2002:256). Sisuliselt hõlmab riim mingite sõnaosade või häälikute ehk foneemide kokkulangevust, milles on arvestatud ka rõhulise kokkulangevusega. Eesti kirjanduse professor Arne Merilai kirjutab “Poeetika gümnaasiumiõpikus” (2011), et riim on rütmilise kõne kõrval teine oluline luuletuse kõlavuse kujundaja. Ta selgitab, et riim ühendab nii salmid kui kogu luuletuse ühtsesse kõlatervikusse ning selle abil saab tõsta esile luuletuses uusi tähenduslikke väljendusi. Õpikus on riimi kategoriseeritud vastavalt tema asukohale 1) sõnas, 2) värsireas või 3) salmis ning ka kokkulangevate silpide arvule toetudes, näiteks on meesriim ühesilbiline kokkulangevus, naisriim kaheasilbiline jne. Merilai nendib, et nendest erinevatest jaotustest hoolimata mõistetakse üldiselt riimi all vaikimisi lõppriimi. See on ka üks põhjuseid, miks selles uurimistöös on valitud üheks poeetiliseks parameetriks lõppriim.

Lõppriimi all võib Merilai sõnul (2011) mõista nii värsiridade lõpus paiknevat riimi kui ka sõna lõppriimi, kuid üldiselt hõlmab see mõlemat korraga. Lõppriimiga on Merilai arvates hõlbus eristada salme ja värsse, aga ka neid üksteisele mõtteliselt lähendada, samuti aitab see luule rütmilisust esile tõsta. Viimasega Jaak Põldmäe “Eesti värsiõpetuses” päris ei nõustu, öeldes, et see on liialt rõhutatud.

Järgnev lõik ja selle näited pärinevad “Poeetika gümnaasiumiõpikust” (Merilai et al., 2011). Sõna lõppriim moodustub tavaliselt 1) rõhusilbi täishäälikutest (**ta** : **sa**, **taat** : **saad**, **tee** : **veskisse**) ning esineb mõnikord ka koos 2) sellele järgneva kaashäälikuühendiga (**rind** : **sind**, **müts** : **kets**) või 3) hõlmab veel rõhusilbile järgnevaid häälikuid (**valu** : **tal**, **lõngast** :

rõngast). Salmis paiknemise alusel jaguneb lõppriim väga paljudeks vormideks, millest siinses uurimistöös on kasutatud kolme tüüpilisemat:

- 1) süliriim (abba, abbba...),
- 2) paarisriim (aabb, aa, bb,...) ja
- 3) ristriim (abab).

Joonisel 1 on kujutatud muuhulgas riimi analüüsi rakendamist Hando Runneli “Üks veski seisab vete pääl” esimesele salmile. Joonisel on näha värsiridade lõpus paiknevat paarisriimilise skeemiga lõppriimi. Skeemi esimeses ja teises reas kattuvad sõnade pikk täishäälik ning sellele järgnev kaashäälik (A), kolmandas ja neljandas reas omadussõna *-nud* lõpud. Joonisele on lisatud ka teiste siin töös käsitlemata jäänud riimide analüüs, näiteks sõnakordus, mis mõnedes genereeritud luuletustes siiski esines.

1.2 Meetrika

Meetrika² on “värsiõpetuse osa, mis käsitleb rütmi, värsimõõtu ja värsisüsteemi; luuleteose rütmistruktuur”. Luuletuse värsimõõd tekib sarnaste keeleosakeste, näiteks silpide ja sõnade, süsteemipärase korrastamise tulemusena (Merilai et al., 2011). Värsimõõdu analüüsimiseks tuleb vaadelda kõiki värsimõõdu loomise viise, nende hulgas silbirõhku, silpide arvu ning nende koosmõju, samuti sõna- ja lauserütmi (Merilai et al., 2011).

Merilai (2011) nendib, et tänapäeva eesti luules on peamiseks lähtepunktist rõhuasetus, mis on tihti kooskõlas silbiarvu korrapärasusega. Samuti leidub tema sõnul tänapäeval rohkelt vabavärsilist ehk korrapäratu värsisüsteemiga luuletusi. Peamiselt eristatakse kolme suuremat värsisüsteemi, mis on süstemaatilisuse ja korrapära alusel kahanevalt järjestatud : 1) silbilisrõhuline, 2) rõhuline ja 3) vabavärs. Siinse uurimistöö praktilises osas on valitud keskenduda värsisüsteemidest silbilis-rõhulisele oma süstemaatilisema olemuse tõttu teiste variantide kõrval. See lihtsustab oluliselt selle algoritmistamist.

1.2.1 Silbilisrõhuline värsisüsteem

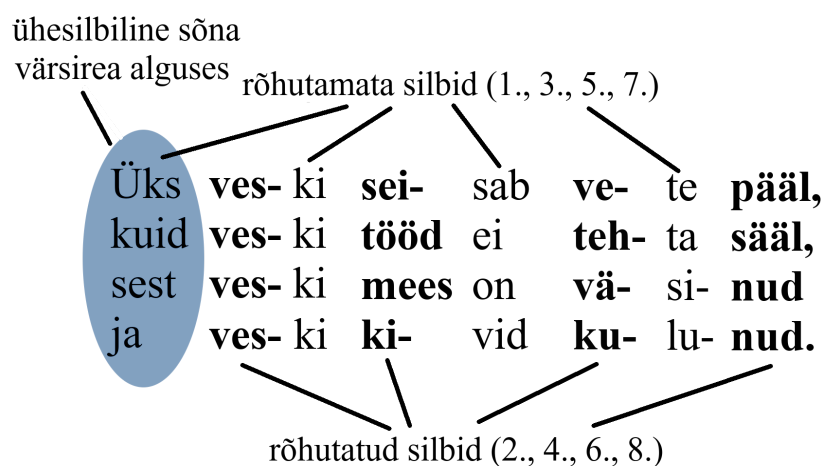
Arne Merilai (2011) sõnul moodustab silbilisrõhulise värsisüsteemi värsijalad ehk erilised kõnetaktid, mis koosnevad ühest rõhusilbist ehk värsitõusust ning selle juurde kuuluvatest rõhuta silpidest ehk värsilangustest. Põldmäe definitsioon on sarnane (2002), kuid ta kasutab värsijala tõusude märkimiseks nime “itkus” ning languste märkimiseks “mitteitkus” või ka sõna “intervall”. Erinevates allikates, k.a “Eesti värsiõpetuses” (Põldmäe, 2002) ja “Poeetika

² Definitsioon sõnale “meetrika” <https://sonaveeb.ee/search/unif/dlall/dsall/meetrika/1>

gümnaasiumiõpikus” (Merilai et al., 2011), on nende põhjal skeemide koostamiseks kasutatud mitmeid erinevaid märke (seda ka ühe allika siseselt), näiteks on iktus ja intervall märgitud vastavalt “+” ja “-” märkidega, iktus ja mitteiktus vastavalt “x’” ja “x” ning arsis ja teesis “—” ja “U”. Sisuliselt langevad need luuletust ette lugedes omavahel kokku, ehkki mõnikord kasutatakse üht või teist väljendamiseks pikka ja lühikest silpi, mis silbilisrõhulises süsteemis langevad vastavalt rõhulise ja rõhuta silpidega kokku.

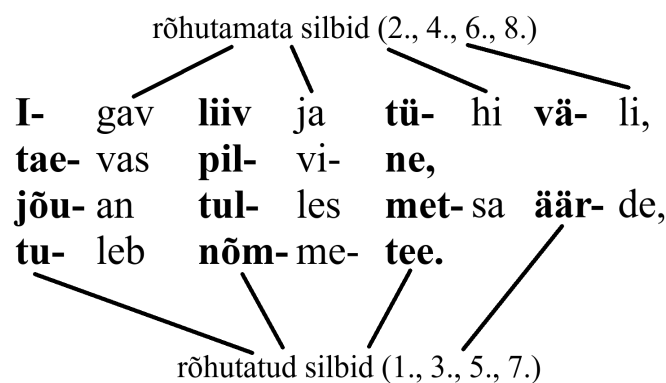
Arne Merilai esitatud (2011) värsimõõdu korrastusalustest on silbilisrõhulises süsteemis määrav värsis esinevate silpide arvu ning rõhkude üheaegne korrastatus. Eesti kirjandusteadlane Jaak Põldmäe kinnitab seda oma teoses “Eesti värsiõpetus” (2002:85), öeldes: “pearõhusilpide arvu muutusele värsis kaasneb muutus värsi silbi arvus”. Sellest saab järeldada, et kindlas värsimõõdus ning kindlate rõhkude arvuga (ning seega ka rõhuta silpide arvuga) värsirea pikkus on silpides alati sama. Samas pole ebatavaline leida luuletusi, mille viimane sõna katkestab rütmi või jätab selle lõpetamata, st rõhusilbile liitub ülejäänud värsireast erinev arv rõhuta silpe või ei liitu ühtegi. Selline ebakorrapära on täiesti lubatud ning ei muuda luuletuse värsimõõtu. Selles uurimistöös ei pruugi riimis sõnade rütmid samuti kattuda valitud värsimõõduga, mistõttu silpide arv ja värsimõõt üksteist ei mõjuta.

Joonisel 2 on esitatud üks salm Hando Runneli luuletusest “Üks veski seisab vete pääl” silbilisrõhulises värsimõõdus jamb, mis moodustub ühest rõhuta ja sellele järgnevast ühest rõhulisest silbist. Joonisel rasvasena märgitud silbid on rõhutõusud. Sinisega on ringitatud ühesilbisõna värsirea alguses, mida jambis nimetatakse alarõhutatud oleku tõttu anakruusiks.



Joonis 2. Hando Runneli “Üks veski seisab vete pääl” (2018) esimese salmi värsimõõdu analüüsi kujutav joonis loodud “Poeetika gümnaasiumiõpiku” eeskujul (Merilai et al., 2011:36).

Jamb on üks mitmest silbilisrõhulisest värsimõõdust. Selles uurimistöös on valitud keskenduda trohheusele, mida nii Merilai “Poeetika gümnaasiumiõpik” (2011) kui ka Põldmäe “Eesti värsiõpetus” (2002) defineerivad värsijalana, mille moodustavad üks rõhuline ning sellele järgnev üks rõhuta silp (“x’x” või “+–” või “⊂–”). Programmis kasutati rõhkude märkimiseks täisarve 0 ja 1, sest samasugune märkimisviis oli kasutusel ka EstNLTK morfoloogilises analüsaatoris Vabamorf. Töös valiti peamiseks värsimõõduks trohheus, kuna see on eesti kõnekeeles kõige loomulikum, sest sõnarõhk langeb eesti keeles esimesele silbile (EKI, 2007). Lisaks on trohheus ajalooliselt nii Põldmäe kui Merilai sõnul eesti rahvaluuletraditsiooni tõttu kõige tüüpilisem eestikeelse luule värsimõõt. Trohheuses kirjutatud luulet on kujutatud joonisel 3.



Joonis 3. Juhan Liivi “Nõmm”³ (Liiv, 1903) esimese salmi värsimõõdu analüüsi kujutav joonis, loodud “Poeetika gümnaasiumiõpiku” eeskujul (Merilai et al., 2011:36).

Siiski tuleb joonise 2 ja joonise 3 vahel märgata märkimisväärsed sarnasusi. Nimelt on trohheus kahesilbilise silbilisrõhulise süsteemi aluseks, jamb pelgalt lisab sellele valmis värsimõõdule ette anakruusi ehk ühesilbilise alarõhutatud sõna. Sellega nihkuvad trohheusest jambi tegemiseks kõik värsijala piirid ühe võrra edasi (paremale). Sellist võtet kirjeldavad mõlemad poeetika uurijad ka “Poeetika gümnaasiumiõpikus” (Merilai et al., 2011) ning “Eesti värsiõpetuses” (Põldmäe, 2002).

Selles uurimistöös on värsimõõduks mitte ainult trohheus, vaid ka jamb, millest viimane saadakse suvalise ühesilbilise luulelise “tüüpsõna” lisamisel trohheuses valmistatud rea ette, näiteks lisatakse sõna “Oh” või “Ah”. Nende sõnade valikul lähtuti eesti luulet aastatel 1917–1940 koondava sarja “Sõnarine 2” (Muru, 1992) luuletustes kasutatud värsirida

³ Luuletus ilmus A. Vinkel koostatud Juhan Liivi luulekogusid koondavas raamatus “Sinuga ja sinuta” (1989)

alustavatest ühesilbisõnadest, näiteks oli Marie Under tuntud sõna “Ah!” kasutamise poolest oma värssides. Valitud sõnadeks osutusid “Ah”, “Oh”, “Ka”, “Oi”, “Oo”.

1.3 Stroofika

Stroofid ehk salmid⁴ on luuleteoses sarnase ülesehitusega värsiridade üksused. Stroofika⁵ ehk salmiõpetus on “stroofivorme käsitlev värsiõpetuse osa”. Arne Merilai sõnul (2011) võivad luuletuse salmid olla ühesuguse ülesehitusega või oma korrapäralt erineda, kuid ei tohi olla värsiridade arvult ega ehituselt täiesti erinevad - sellisel juhul on tegemist näivsalvidega. Merilai järeldeb, et salme ei erista ainult näiline kirjpilt ehk taandrida, suurtähed või salmivahe, vaid ka teised poeetilised aspektid nagu ühtlane riim, värsimõõt, värsiridade arv ning sisulisest küljest tähenduslik tervik. Selle uurimistöö tarkvaralahenduses on salvide koostamisel lähtutud peaaesjalikult riimitervikust ning läbivast värsimõõdust, millele lisandub ka taandrida salvide vahel.

Alljärgnev lõik põhineb Jaak Põldmäe raamatule “Eesti värsiõpetus” (2002). Salme on võimalik nimetada selles esineva värsiridade arvu järgi, näiteks kaksikvärss ehk distihhon, kolmikvärss ehk tertsett, nelikvärss ehk katraän, viisikvärss ehk kvint jne. Luuletusi saab žanriliselt jaotada temaatika ehk sisu alusel (näiteks ballaad, poeem, ood) või poeetika ehk vormitunnuste (peamiselt salvide ja värsimõõdu) korrastatuse alusel kinnisvormideks. Selliste kinnisvormide alla kuuluvad näiteks

- sonett (14-realine, salmid jaotuvad värsiridade arvu alusel traditsioonilisel 4+4+4+2 (inglise sonett) või 4+4+3+3 (itaalia sonett), aga ka arhailiselt 8+6 vms);
- tanka (31-silbiline luulevorm, silbid jaotuvad tüüpiliselt (5-7-5)+(7-7));
- haiku (3-värsiline luuletus, silbid alati 5-7-5).

Arne Merilai nendib “Poeetika gümnaasiumiõpikus” (2011), et nendel kinnisvormidel on ajalooliselt tihti lisaks vormilistele tunnustele ka sisulised tunnused, tuntakse ju haikut jaapani traditsioonis loodusluulena, samas kui tankal on zen-budistlik traditsioonitaust. Kirjandusteadlane Rebekka Lotman on doktoritöös “Eesti sonett” (2019) täheldanud, et tänapäeval on sellised sisulised piirangud oma olulisust kaotanud, samas kui vorm on jäänud samaks. Tema sõnul on vastavalt keelelisele eripärale ja kirjakultuuri traditsioonidele alget vormi muudetud. Näiteks sonetivorm, mis tema töös vaatluse alla kuulus, on itaalia traditsioonis üksteistsilbik (kindla värsimõõduta 11 silbist koosnev värsirida), samas kui

⁴ Definitsioon sõnale “stroof” <https://sonaveeb.ee/search/unif/dlall/dsall/stroof/1>

⁵ Definitsioon sõnale “stroofika” <https://sonaveeb.ee/search/unif/dlall/dsall/stroofika/1>

inglise-, saksa- ja eesti keeles klassikaliselt viisikjamb (5 rõhulisest silbist koosnev rida, kus vahelduvad rõhuta ja rõhutatud silp). Siinses uurimistöös pole vaatluse all ükski täpne kinnisvorm, kuid kõik võimalused poeetilisest aspektist on selleks olemas. Uurimistöö tagasiside küsimustikuks valmistatud katsetes määrati genereeritavate luuletuste silpide arvud neid kinnisvorme silmas pidades 7, 10 ja 11.

1.4 Varasemad tööd

Portugali Coimbra Ülikooli teadlane Hugo Gonalo Oliveira koostas 2017. aastal kokkuvõtliku ülevaate luule genereerimise ajaloost (Oliveira, 2017). Esimesed tõsisemad katsetused algasid 1960ndatel aastatel, mil Queneau üritas suurest hulgast luuletustest ridade ümbertõstmise kaudu luua uusi luuletusi. Oliveira mainib ka hilisemaid katsetusi, näiteks 1981. aastal prantsuse teadlaste loodud programmi *rimbaudelaires*, mis kasutas kirjaniku Arthur Rimbaud luuletuste struktuuri ning täitis selle sõnavaraga Baudelaire'i luuletustest. Artikli autor nendib, et tõsisem trend luule genereerimisel tõusis 21. sajandi alguses, mil esmakordselt prooviti selleks kasutada ka tehisintellekti. Oliveira meelest muutis see genereeritud teksti olemust, kuna püüti esmakordselt panna rõhku esteetikale ja teksti poeetikale. Esimesed tehisintellekti kasutavad lahendused loodi hispaania ja portugali keeles, artikli autor olles portugali keelse lahenduse autor (Oliveira, 2017).

Valdkonna ingliskeelsete uurimuste hulgast leiab vägagi sarnase probleemiasetusega Imperial College London ja University College Dublin koostöös ilmunud artikli “Full-FACE Poetry Generation” (Colton et al., 2012). Viimane üritab korpuspõhise lähenemise abil luua “päevaluulet”, mis kasutab algtekstina nii brittide ajalehte “The British Guardian” kui ka kuulsate autorite üksikuid tekste. Nagu ka selles uurimistöös, asetab artikkel oma fookuse mitte üksnes luuletuse sisule ja loogilisusele, vaid ka poeetilistele aspektidele, nende hulgas meetrikale ja riimile, aga ka üldisemalt keelele, näiteks esteetika ja tunde tabamisele, kordustele ja väljenditele ehk keelekujunditele laiemalt.

Keelelisest aspektist eesti keelega sarnasemate probleemidega silmitsi seisvad soome teadlased (Hämäläinen et al., 2019) on samuti kasutanud luule genereerimiseks FACE korpuspõhist tehnoloogiat (inspireeritud Coltoni jt tööst). Uurimustöö autorite Mika Hämäläineni ja Khalid Alnajjari probleemiasetus tuli keelest tulenevalt morfoloogial, mille poolest soome keel on sarnaselt eesti keelele morfoloogiliselt väga rikas.

Mitteformaalseid lahendusi luule genereerimiseks arvuti abiga on ohtralt. Üks kõige lihtsam viis on kasutada otsingumootorit ja proovida erinevaid luuletuste generaatoreid, mis on

kasutajatele kergesti kättesaadavad. Guugeldades “luule generaator” või “luuletuste genereerimine”, ei leidu otsingutulemustes ühtegi eestikeelsete luuletuste generaatorit. Guugeldades “*poem generator*” või “*poetry generation*” või muid analoogseid kāske, saab leida mitmeid ingliskeelset luulet genereerivaid veebilehekülgi. Lõviosa võtab enda alla tehisintellekti kasutavad lahendused (või seda väitvad, sest kood pole avalik). Need rakendused küsivad sisendiks 1) üksikut sõne või lauset vabas vormis (näiteks luuletuse teemat); 2) teha valik erinevate parameetrite vahel, millest kõige tavalisem on luuletuse tüüp (sonett, haiku jne) ja pikkus või 3) segu nendest erinevatest valikutest. Esimeste otsingutulemuste seas on näiteks *aipoemgenerator.org*⁶, *editpad.org*⁷ ja *poemgenerator-ai.com*⁸.

1.4.1 aipoemgenerator.org

Selles luuletuse generaatoris oli vaid üks tekstiväli, mis küsis kasutajalt soovitud luuletuse teemat (“*Generate a poem about...*”). Iga genereeritud luuletus oli juhusliku ridade ja salmide arvuga. Iga luuletuse salm oli paarisriimiga ning igas salmis oli alati 4 rida.

1.4.2 editpad.org

Editpad.org veebirakenduses saab luuletuse genereerimisel valida luuletuse teema (*poem topic*), pikkuse ehk lühikese, keskmise või pika (*Length: short, medium, long*) ning luuletuse tüübi (*Type: sonnet, ode, Ballad*) ehk erinevate kinnisvormi vahel. Automaatselt oli mitmetes genereeritud luuletustes ka värsirea lõpusõnad riimuvad, tüüpiliselt üle ühe ehk ristiimis. Generaator on üpris mitmekesine ehkki selles leidusid mõned loogikavead, näiteks oli paigutatud kinnisvormide alla *Free Verse* ehk vabavärss, mis on värsimõõt ning võib olla ükskõik millise kinnisvormi osa. Siiski teadis programm, et sonett peab olema 14-realine ning haikus silpide arvud 5-7-5. Kõigis genereeritud luuletustes ta neid tingimusi ei järginud, kuid tegi seda enamikel juhtudest.

1.4.3 poemgenerator-ai.com

AI Poem Generator on teine mitmekesisemate parameetritega luulegeneraator. Rakenduses palutakse kasutajal valida kuue kinnisvormi vahel ehkki mõned neist, näiteks armastusluule (ingl love poem) on rohkem teema valik. Seejärel saab kasutaja vabalt kirjeldada luuletuse temaatikat ning valida erinevate keelte (inglise, prantsuse, vene, itaalia, saksa, hispaania, portugali) vahel. Generaatorit katsetades selgus, et vastavalt genereeritava luule keelele,

⁶ <https://aipoemgenerator.org/>

⁷ <https://www.editpad.org/tool/poem-generator>

⁸ <http://poemgenerator-ai.com/>

suudab programm muuta ka vormi, näiteks ingliskeelne sonett järgib inglise soneti värsiridade jaotust ning itaaliakeelne itaalia sonetivormi traditsioone.

1.4.4 ChatGPT-3.5

Tehisintellekti tekstilooma võimekuse üle on meedias peale ChatGPT-3.5 (OpenAI) avalikkusele tutvustamist olnud palju arutelusid. Katsetades ChatGPT-ga luuletuse genereerimist, on tänu selle vabalt valitavale viibale (ingl *prompt*) võimalik paluda tarkvaral teha täpselt endale sobiv luuletus. Viibale “Tee luuletus, mille teema on hobused ning mis riimub” tagastab ChatGPT 12-realise luuletuse hobustest. Selles ei leidu silmaga nähtavaid eesti keele ortograafia vigu ning luuletus pole äratuntavalt arvuti loodud. Esimese korraga luuletus ei riimu, kuid selgitades programmile pikemalt, mida riimumine tähendab, saab peale mõneminutulist katsetamist soovitud luuletuse. Kahjuks tulevad sellisel viisil sisse ka ortograafiavead, kuna tehisintellekt ei muuda luuletust täielikult, vaid püüab muuta sõnalõppe, tekitades käände- ja pöördedorme, mida eesti keeles ei eksisteeri. Kindlas värsimõõdus loodud luuletuse tegemisega ChatGPT lõpuni hakkama ei saanud, isegi mitte selgitatud reeglite alusel.

Poeetilisi parameetreid kui ka sisu arvesse võtvaid luulegeneraatoreid eestikeelsel kujul ei eksisteeri, seega panustab uurimistöö nii eesti keeletehnoloogia kui ka kirjanduse ja digihumanitaaria valdkonda.

2. Tehnilised vahendid

Siinses peatükis esitatakse lühike ülevaade tarkvaralahenduses kasutatud tehnoloogilistest vahenditest. Rakenduses on kasutatud programmeerimiskeelt Python, EstNLTK teekide kogumit (EstNLTK) ja programmeerimiskeskonna Google Colab (Google Colab) veebirakendust.

2.1 Python

Rakenduse loomiseks kasutati programmeerimiskeele Python 3.10.12 versiooni. Programmeerimiskeele Python kasuks otsustati selle eesti keele töötlusteegi EstNLTK olemasolu tõttu, mis võimaldas eestikeelset teksti hõlpsamini analüüsida ning sünteesida.

2.2 EstNLTK teekide kogum

EstNLTK (ingl *Estonian natural language toolkit*) on programmeerimiskeele Python baasil loodud teek eesti keele töötluseks. EstNLTK dokumentatsioon on Jupyter Notebooki kujul. Samuti on Tartu Ülikooli kursuse “Eesti keele töötlus Pythonis” aine raames loodud põhjalik teegi võimalusi tutvustav õppematerjal⁹. See peatükk tugineb sellele õppematerjalile.

Eesti Keeletehnoloogia Riikliku Programmi rahastatud ning Tartu Ülikooli teadlaste eestvedamisel arendatud loomuliku keele töötluse (ingl NLP ehk *natural language processing*) teegi funktsionaalsuste hulka kuuluvad sõnestamine ja lausestamine, morfoloogiline analüüs, ajaväljendite ja nimeolemite tuvastus jpm. Teek on ainus samalaadne eesti keele töötluseks, seega polnud ühtegi teist alternatiivi, mida oleks saanud rakenduse loomiseks kasutada. Selle uurimistöö valmimiseks on tarvilikud järgnevad EstNLTK funktsioonid:

- 1) sõnestamine ehk sõnapiiride tuvastamine;
- 2) lausestamine ehk lausepiiride tuvastamine;
- 3) osalausestamine ehk liitlause osalauseteks jagamine ning grupeerimine olulisuse alusel;
- 4) süntaksanalüüs ehk lauseosade (sõnade) funktsiooni määramine;
- 5) sõnade silbitamine ning sõnarõhkude määramine neis silpides.

Funktsionaalsuste kasutamiseks tuleb esmalt EstNLTK importida PyPA (*Python Packaging Authority*) repositooriumist. EstNLTK põhiline klass on *Text*, mis luuakse soovitud sõnetüüpi

⁹ Kursuse materjalid üleval Github versioonihalduse rakenduse veebilehel <https://github.com/d009/EstNLP>

tekstist. *Text*-objektile rakendatakse seejärel teksti analüüsimiseks vajalikud märgenduskihid meetodiga *tag_layer()*. Selles uurimuses on tarvis poeetilise analüüsi teostamiseks kasutaja valitud sisendteksti lausetasandist häälikuni analüüsida. Mõned kihtidest on *tag_layer()* meetodis automaatselt märgendatud (*text.words* ehk sõnad, *text.sentences* ehk laused jne), teised on vaja märkida vastavalt vajadusele tervele tekstile või väiksemale üksusele, näiteks lausele süntaksanalüüsi kiht või sõnale silbitamiskiht.

```
!pip install estnltk==1.7.2
import estnltk
from estnltk import Text
from estnltk.vabamorf.morf import syllabify_word

tekst = "Siia tuleb kopeerida soovitud sisendtekst. Selleks sobib
näitaks artikkel või proosatekst."
text = Text(tekst).tag_layer('maltparser_syntax')
for lause in text.sentences:
    süntaks = lause.maltparser_syntax
    for sõna in lause.words:
        silbitatult = syllabify_word(sõna)
```

Joonis 4. EstNLTK *text*-objekti loomine ning sellel erinevate märgenduskihtide rakendamine.

Joonisel 4. kujutatakse *text*-objekti loomist, sellele *tag_layer()* meetodi rakendamist ning mõnede märgenduskihtide kasutamist. Uurimistöö peatükis “Teksti ettevalmistus” on pikem kirjeldus selles uurimuses rakendatud märgenduskihtidest ja nende väljundite kasutamisest.

2.3 Google Colab

Colab (lühend väljendist *Google Colaboratory* ehk Google’i koostöölabor) on Google’i väljatöötatud veebirakendus, mis võimaldab jooksutada Jupyter Notebooki tarkvara kaugserveril (Google Colaboratory). Jupyter Notebook on tarkvara (Jupyter Notebook latest), mis pakub kiiret ja interaktiivset keskkonda koodi prototüüpimiseks ja selgitamiseks, andmete uurimiseks ja visualiseerimiseks ning ideede jagamiseks teistega. Jupyter Notebooki teenus on kättesaadav Google Colabi veebirakendusena, kuid töötab ka kasutaja arvutis lokaalselt ilma Google Colabita.

Jupyter Notebook on sisuliselt märkmik, milles on lihtne jooksutada koodi nii nagu teistes programmeerimiskeskondades. Lisaks suudab märkmik talletada ka teksti kujul andmeid ning visualiseerida neid, näiteks saab sellega hõlpsasti teha programmeerimist nõudvaid õppematerjale. Seda võimaldavad rakenduses lisatavad tekstilahtrid (ingl *textbox*), milles

saab teksti vormistada pealkirjadeks, paragraafideks ja loeteludeks. Colabi koodilahtrites saab samuti teha vajalikke muudatusi, näiteks lisada erinevatele andmetüüpidele interaktiivseid lahtrid, liugureid ja mitme valikuga rippmenüüd. Koodi ennast on võimalik hõlpsasti kasutaja eest varjule panna, kuid igal märkmiku kasutajal on võimalik neid ka avada ning oma soovi järgi muuta või koodi uurida.

Bakalaureusetöös eelistati Jupyter Notebooki kasutamist Colabi kaudu, sest selle abil on võimalik saada tasuta ligi Google'i poolt pakutud arvutusressurssidele, mis võimaldab kiiremat töötlust võrreldes lokaalselt jooksvat serveriga. Selle keskkonna kaudu saab ka hõlpsasti jagada enda märkmikku laialdaseks kasutamiseks ilma ülesseadmise vaevata. Valikul mängis rolli ka sisseehitatud mugav ja loogiline lahtrite ja selgitava teksti lisamise võimalused, tänu millele saavad seda kasutada ka koodi lugemisoskusega kasutajad.

Alternatiivselt saaks programmi üles panna veebirakendusena, kus serverisse jäävad vajalikud teegid. Nii saaks kasutajatele näidata ainult eesliidest (ingl *front-end*).

3. Luuletuste generaator Värsivigur

Bakalaureusetöoks valminud tarkvaralahendus on üleval Google Colab keskkonnas lingil https://colab.research.google.com/drive/1c7W2GKW_Kpp5NyGQOVJVeQA1SU5Cq81G?usp=sharing. Selle peatüki esimeses alapeatükis kirjeldatakse programmi algoritmi, teises sisendteksti ettevalmistust luuletuse genereerimiseks ning viimases esitatakse ülevaade rakenduse kasutajaliidest.

3.1 Algoritmiline ülesehitus

Kuigi poeetika reegleid on süstemaatiliselt uuritud ja neist kirjutatud, pole nende rakendamist ega analüüsimist algoritmiliselt kirja pandud, st nende määramine ja analüüsimise järjekord on tüüpiliselt tunnetuslik protsess. Töö autor ei leidnud ühtegi käsitlust, mis kirjeldaks samm-sammult poeetilise analüüsi rakendamist ega luuletuse loomist soovitud poeetilisi reegleid järgides.

3.1.1 Stroofika

Materjalide puudumise tõttu otsustati lähtuda teistest loogilistest alustest ehk alustati üldisemast tasandist ning liiguti detailisemale, selles kontekstis stroofikalt (lauseid ja sõnu hõlmavast) meetrikale (silpe ja nende rõhke hõlmavale). Stroofika käitus luuletuse raami ehk struktuurina. Selles töös on üks konstantne stroofiline parameeter värsiridade arv salmis (4), teist kahte olulisemat ehk salmide arvu (1–4) ning silpide arvu värsireas (5–20) saab kasutaja vabalt valida. Värsiridade arv (4) valiti töösse klassikalise neljarealisele riimiskeemide süsteemi (AABB, ABBA, ABAB) alusel.

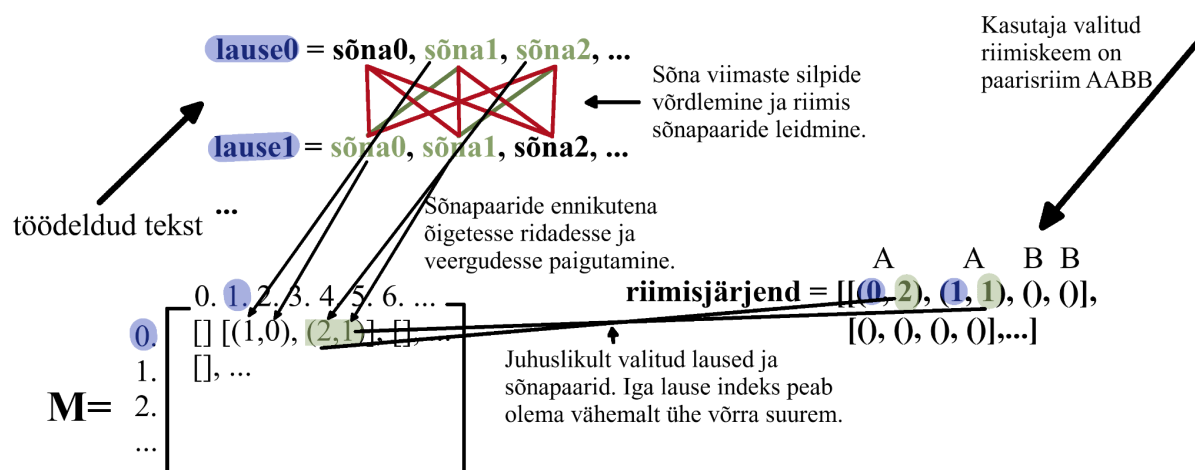
3.1.2 Riim

Samaaegselt stroofikaga tuleb luuletuse koostamiseks paika panna ka riimis sõnad ning valida nende skeem. Rakenduses sooviti vältida riimis sõnade otsimist väljaspool teksti. Samas mõisteti, et ülejäänud värsirea eelneval paika panemisel, seda nii süntaktilisest aspektist kui ka värsimõõdu osas, ei pruugi enam lausest ülejäänud sõnade hulgast riimis sõna leiduda. Seetõttu otsustati tegeleda riimiga eelisjärjekorras.

Esmalt leiti kõikide teksti lausete kõikide sõnapaaride võrdlemisel omavahel riimuvad sõnad. Esialgu määrati lõppriimi tunnuseks tagasihoidlikult sõnade viimase silbi vokaali kattumine, hilisemal katsetamisel lisati sellele ka vokaalile järgnevate häälikute kattuvus nagu Merilai (2011) ja Põldmäe (2002) olid riimi kirjeldanud. Kõik riimis sõnapaariid salvestati indeksi kujul ennikutena kogu teksti riimis sõnu sisaldavasse maatriksisse. Maatriksi moodustasid

järjendid, mille asukoht reas määras ennikus esimesel kohal oleva sõna lause indeksi ning asukoht veerus teisel kohal oleva sõna lause indeksi.

Teksti kõiki riimis sõnade paare sisaldavast maatriksist valiti juhuslikult välja soovitud riimiskeemi järgivad sõnad. Samaaegselt salvestati ka laused, kust riimis sõnad pärit olid, millega hakati hiljem värsirida koostama. Salmi riimis sõnade valimisel arvestati teksti algse lausejärjekorraga. Kogu luuletuses seda teha ei saanud, kuna riimis sõnu sisaldavaid lauseid polnud igas tekstis piisavalt. Nii oluks salmide arv mõne tekstiga piiratud.



Joonis 5. Lausest riimis sõnade valimine.

Joonis 5 kujutab töödeldud tekstist riimis sõnade leidmist ja valimist. Joonisel on tähisega “M” kujutatud kõiki riimis sõnu sisaldav maatriks. Rohelise värviga on märgitud sõnad, mis riimi kontrolli läbisid. Nendest valitakse juhuslikult välja üks paar, mis lisatakse soovitud riimiskeemi reegleid järgides *riimisjärjendisse*. Sinisega on märgitud lausete indeksid.

3.1.3 Värsimõõt ja süntaks

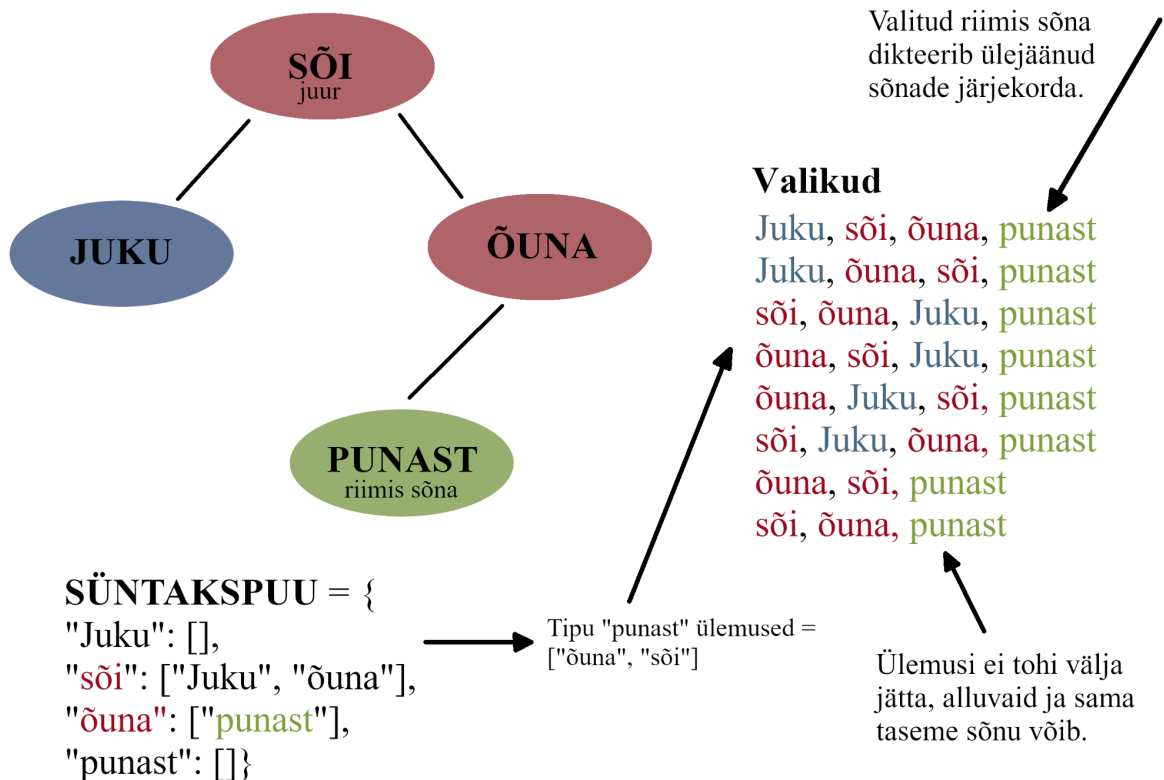
Kui kõikide salmide riimis sõnad ning neid sisaldavad osalaused olid valitud, hakati koostama ülejäänud värsiridu rida rea haaval. Kasutajalt küsitud värsimõõd ei muutnud esialgu sõnade valikuprotsessi, kuna trohheusest saab hõlpsasti värsirea ette ühesilbilise sõna lisamisega teha jambi. Trohheilise värsimõõdu algoritmi tuletamine oli töö kõige keerulisem osa, sest ükski kirjandusteaduslik allikas ei selgitanud täpselt selle silpide arvu ja sõnarõhkude alusel määramist. Selleks on ka hea põhjendus: sõnarõhud ei toimi alati ettenägelikult, st vajadusel saab neid mõningate reeglite eiramisega alarõhutada või ülerõhutada, kuid see on “tunnetuslik” protsess, mida luuletuse lugeja või koostaja saab oma peas või ettelugemise kaudu katsetada. Uurimistöö autor oli sunnitud lähtuma juba

kirjapandud reeglitest ning neid eesti kõnekeele reeglitega sünteesides tuletama soovitud algoritmi.

Eesti keeles on tüüpiliselt sõnarõhk esimesel silbil, sõna kõik teised silbid on rõhuta. Võõrsõnadel võib rõhusilp langeda ka teistele silpidele. Jaak Põldmäe, kes on eesti ühe põhjalikuma värsikäsitluse “Eesti värsiõpetus” (2002) autor, vihjab oma teose sissejuhatuses, et trohheust soosivad eesti keeles kahesilbilised sõnad (“+–”), samas kui kolmesilbilised sõnad soosivad daktüli (“+––”). Seda väidet järgides, tulnuks tekstist peaaegu kõik sõnad välja visata või kasutada sõnu väga hüppeliselt, mis jätnuks ka sisusse suured lüngad. Samas oli teada, et eesti keeles on kolmesilbilistel sõnadel kolmas silp kaasarõhuline (EKI, 2007), mis tähendab, et seda saab muuta vastavalt vajadusele rõhuliseks (“+–+”). Trohheilise rütmi saavutamiseks peaks seega kolmesilbisõnale järgnema rõhuta silbiga algav sõna, milleks on juba mainitud võõrsõnad ning ühesilbilised sõnad, mida on samuti lubatud alarõhutatada. Kahesilbisõnade ja kolmesilbisõnadele rakendatud loogikat laiendati vastavalt kõikidele paaris- ning paaritusilbilistele sõnadele, välja arvatud juba nimetatud ühesilbilised sõnad.

Selline trohheuse algoritm on hüpoteetiline, kuna ükski varasem allikas pole seda kasutanud. Näiteks võib kolmesilbilist sõna, millele järgneb alarõhutatud ühesilbisõna, lugeda hoopis daktüüliliselt (“+––+”). Samas võib oletada, et luuletuse lugeja on mõjutatud värsireas ja salmis juba eksisteerinud trohheilisisest rütmist ning loeb selle siiski tunnetuslikult “õigesti”.

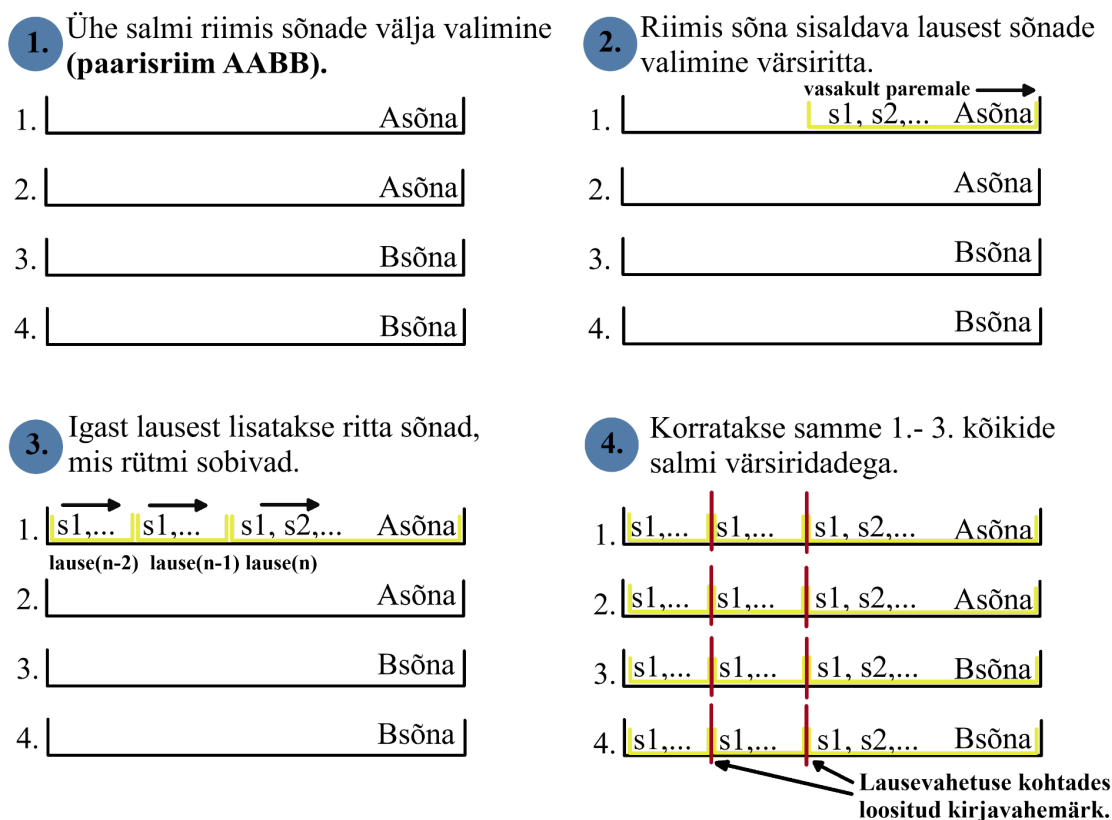
Sellise algoritmi rakendamise tülikus seisneb vajalike sõnade leidmises ühest lausest. Isegi kui selliseid sõnu lauses leidub, ei saa neid ilma põhjuseta ära jätta ega ümber tõsta, ilma et kannataks lause sisu ja üldine süntaktiline ülesehitus. Lause süntaktilist ülesehitust selgitab joonis 6. Lausest “Juku sõi punast õuna” on loodud süntakspuu, mis näitab sõnade hierarhiat lauses. Tipu lisamiseks värsiritta peab vähemalt osaliselt algse lause süntaks säilima, näiteks ei saa lausest kasutada sõna “punane” ilma sõnata “õuna” ning “õuna” ilma sõnata “sõi”. See tähendab, et iga sõna kasutamiseks peavad olema tuletatud lauses esindatud ka tema ülemused. Teisisõnu on tegemist laiuti läbimisega (ingl *breadth-first*). Lisaks tuli enne värsiritta lisamist kontrollida, kas sõna sobis trohheuse rütmi ning mahtus järele jäänud silpide arvu sisse.



Joonis 6. Lause süntakspuu alusel lausete moodustamine.

Selline algoritm muutis mitme kõrvallausega põimlausetest sõnade valimise väga keeruliseks. Riimis sõna võis puu juurest olla väga kaugel, mistõttu oli tal palju tippe kuni juurtipuni. Sellist lahendust katsetati, kuid leiti kiiresti, et nende ülemuste järjekorra vahetamine oli värsimõõdu sobitamiseks vajalik, kuid selle tagajärjel sai kannatada lause süntaktiline ülesehitus ja lause sõnum. Probleemi lahendamiseks segmenteeriti tekst osalauseteks, mille süntakspuud polnud nii sügavad, ja käsitleti neid edaspidi kui eraldiseisvaid lauseid.

Need osalaused jäid ühe värsirea koostamiseks väikeseks. Värsiritta vajalike lisasõnade leidmiseks kasutati tekstist teisi lauseid, mis jäid juba valitud (riimis sõna sisaldavast) lausest ettepoole. Lauseid vaadati sõna haaval laiuti läbimisega, kuni värsireas oli soovitud arv silpe. Paraku oli ka sellisel lausete ja sõnade valimisel oma negatiivsed küljed. Erinevate lausete kokkupuutepiiridele oli tarvis panna kirjavahemärk, sest lauses alustatud mõttekäik ajutiselt värsireas katkes. Samuti tekitas see olukorra, kus ei saanud kontrollida kokkupuutepiiridel värsimõõdu jätkumist. Värsimõõdu tagamiseks lubati sõnu ka värsireas järjestikuliselt korrata (sõnakorduse poeetiline võte), et iga sisemine tuletatud lause algaks rõhulise silbiga ning lõppeks rõhuta silbiga, tagades nii trohheilise rütmi säilimise.



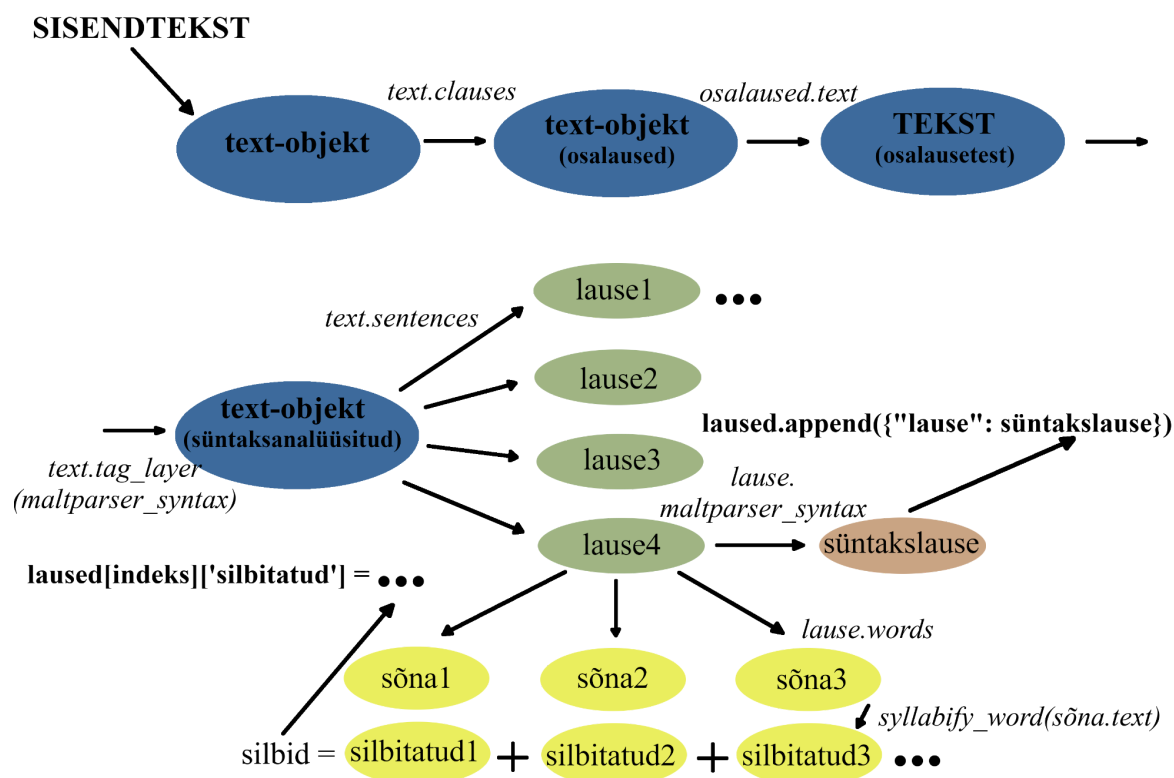
Joonis 7. Ühe salmi kokkupanemise skeem.

Joonisel 7 on esitatud algoritmilise järjekorra abstraktsioon. Joonise 7 esimesel (vasakul ülal) illustratsioonil on näidatud ühe salmi algne seis peale riimis sõnade välja valimist kasutaja valitud riimiskeemi alusel. Teisel illustratsioonil (paremal ülal) on kujutatud järgmise tööfaasi lõppu, kus riimis sõna ette on valitud sõnade järjend esimesest lausest (sellest, mis riimis sõna sisaldas). Kolmanda faasi (vasakul all) eel võtab algoritm uute sõnade lisamiseks uue lause (andmestruktuuris indeksiga -1) ning kordab sõnade valimist nii, et iga järjend algab rõhulise silbiga ning lõppeb rõhuta silbiga. Samuti kontrollitakse etappides 2. ja 3. iga sõna lisamisel, kas vajalik silpide arv on saavutatud. Joonise neljandal illustratsioonil (paremal all) on kujutatud salmi loomise töötsükli lõppu ehk iga salmi reaga on läbitud kõik joonise eelnevad etapid. Punase joonega on märgitud lausevahetuse kohad, milles tekib mõtteline katkestus. Selle mõttelise katkestuse tähistamiseks lisatakse ritta juhuslikult universaalne kirjavahemärk, näiteks hüüumärk, kolm punkti vms. Neljanda etapi lõpus salvestatakse salm maatriksina järjendisse ning alustatakse töötsükli uuesti, kuni on saavutatud kasutaja soovitud salmide arv.

3.2 Teksti ettevalmistus

Teoorias selgus, et poetiliste parameetrite tagamiseks luuletuses tuleb esialgne tekst segmenteerida lauseteks, sõnadeks, silpideks ja häälikuteks. Neid üksuseid on vaja mitmes luuletuse kokkupanemise osas, mistõttu otsustati luua suur teksti kõiki lauseid koondav järjend (nimega *laused*). See sisaldab endas 1) sõnastikke võtmega 'lause', mille väärtusteks on laused süntaksandmetega, ning 2) ühe lause kaupa grupeeritud järjendeid, mis kätkevad kõiki selle lause sõnu silbitatult.

Vajalike keeleüksuste kättesaamiseks tehti kasutaja valitud tekstist (sõnest) EstNLTK Text klassi objekt. *Text*-objektide märgendamine on ajamahukas töö, mistõttu peeti oluliseks need kihid lisada võimalikult suurele osale algtekstist korraga.



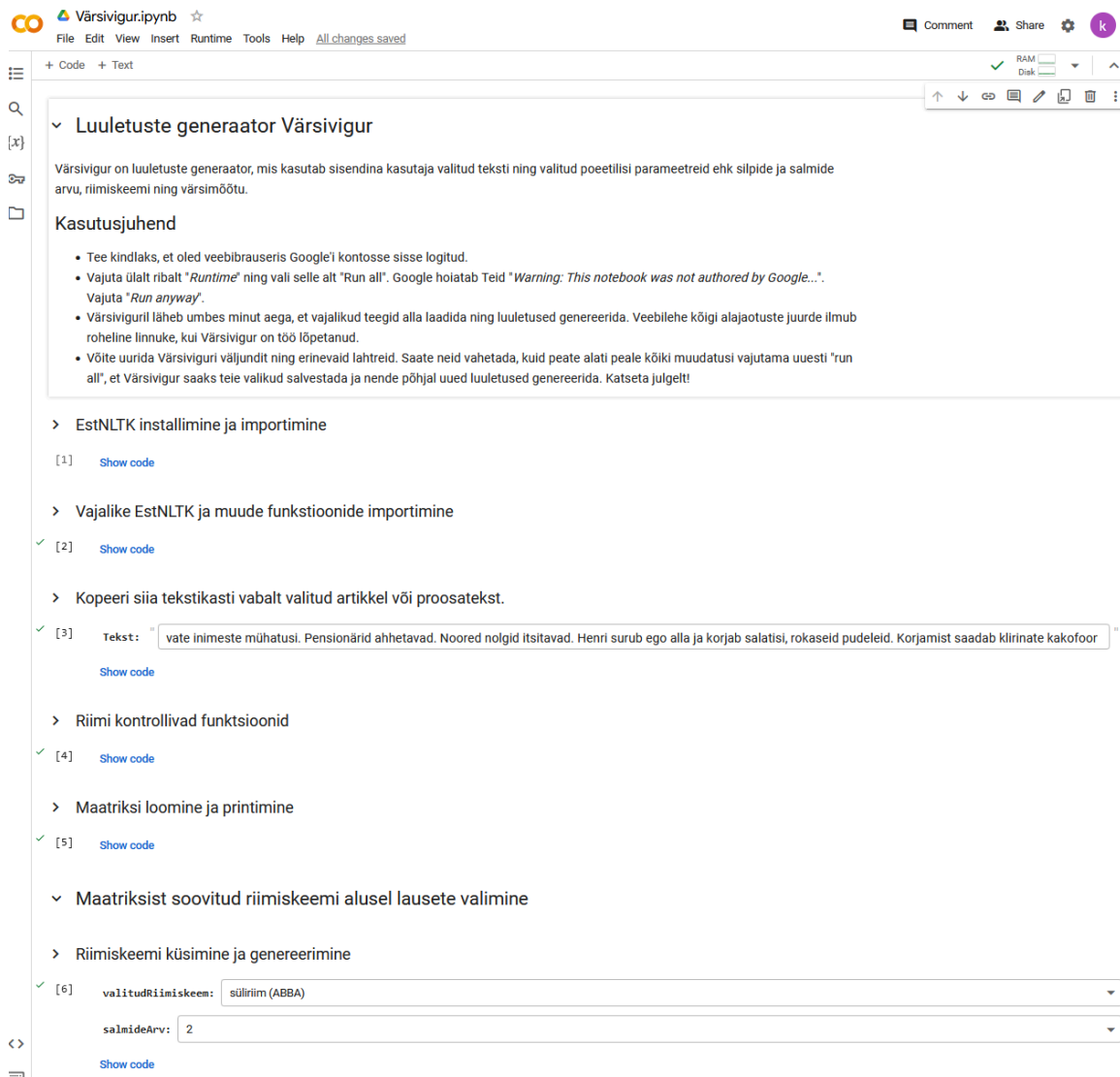
Joonis 8. Sisendteksti segmenteerimine ja töötlus.

Joonisel 8 on kujutatud tarkvaralahenduse tekstitöötamise abstrahheeritud ja lihtsustatud skeem. Tekstitöötlus jaguneb mõtteliselt kaheks osaks: sisend tekstist osalausestatud teksti loomine (ülemine rida) ning osalausest teksti segmenteerimine ja edasine analüüs (skeemi alumine osa).

Spetsiifiliste andmestruktuuride valikul sai kohati määravaks ka juba EstNLTK-sse sisseehitatud funktsioonide väljundid, sest nii välditi andmestruktuuride ümbertegemise käigus tekkivaid võimalikke vigu. Lisaks olid andmestruktuurid sobilikud, kuna kasutasid sõnastikke, mis seeläbi vähendasid indeksitele tuginevat otsimist. Viimast oli andmestruktuuride rohke pesastatuse tõttu keeruline jälgida ka lõplikus versioonis.

3.3 Kasutajaliides

Värsiviguri kasutajaliides disainiti veebirakenduse Google Colab Forms (Google Colab Forms) võimalustega. Kasutajalt soovitud sisendid valmistati ette Formsi lahtritega, selleks et ka programmeerimisega mitte kokku puutunud huvilised saaksid kasutada programmi hõlpsasti. Ühtlasi tõstis see valminud rakenduse üldist interaktiivsust ning muutis kasutajaliidese intuitiivsemaks.



Joonis 9. Värsiviguri kasutajaliides.

Joonisel 9 on kujutatud tarkvaralahenduse vaade. Kasutaja eest on kood ära peidetud. Mugavamaks arendustööks on kood ja funktsioonid nende eesmärgi järgi grupeeritud, näiteks “Riimis sõnade leidmisega seotud funktsioonid”, mida saab vajadusel avada ning oma soovi järgi virtuaalses sessioonis muuta. See on kasulik just neile, kes tahaksid koodi isiklike eesmärkide nimel uurida ja katsetada selle täiendamist. Samuti on võimalik teha koodist sarnaselt Google Drive’i (Google Drive) võimalustele koopia, mida saab jäädavalt muuta ja salvestada (mitte ainult ühe sessiooni siseselt).

4. Tulemused

Selle uurimistöö käigus valminud tarkvara testimiseks kasutati tagasiside küsimustikku. Tarkvaralahendus ei valminud kindlale sihtrühmale, seega otsustati kasutada mugavusvalimit. Vastajatelt nende isikuga seotud informatsiooni ei kogutud ning seda vastustega ei kõrvutatud. Küsimustiku eesmärgiks oli saada esmast tagasisidet Värsviguri üldise toimimise ning selle genereeritud luuletuste kohta.

Küsimustik jagunes neljaks osaks:

- 1) sissejuhatus ja kasutusjuhend;
- 2) esimene katse ehk uudisteartikli alusel genereeritud kolme luuletuse põhjal küsimustele vastamine;
- 3) teine katse ehk proosateksti alusel genereeritud kolme luuletuse põhjal küsimustele vastamine ning
- 4) üldisemad küsimused Värsviguri kasutamiskogemuse ja luuletuste kohta.

Katsed olid Värsviguris spetsiaalselt ettevalmistatud ühe katsena. Katsete vahepeal pidi vahetama ainult sisendiks valitud teksti: esimeses uudisteartikkel, teises proosatekst. Ühe katse siseselt olid fikseeritud luuletuste värsimõõt, riimiskeem ja salmide arv. Katse luuletused erinesid üksteisest silpide arvu poolest: esimeses luuletuses oli 7 silpi, teises 10 silpi ning kolmandas 11 silpi. Arvud valiti luuletuste kinnisvormide (sonett, haiku) silpide arvu järgi. Katsete vahepeal lubati vastajatel värsimõõtu, riimiskeemi ning salmide arvu muuta, kuid see polnud kohustuslik. Mõlemas katses koguti vastajatelt luuletuste genereerimiseks kasutatud teksti link ning valminud luuletused. Katse käigus sooviti teada saada:

- mida vastanud arvavad luuletuste poetikast ja sisust;
- kas ja kuidas mõjutab luuletuste genereerimiseks kasutatud teksti tüüp luuletuse sisu ja poetikat;
- milline oli kasutajate üldine kogemus Värsviguri kasutamisel.

Küsimustiku loomiseks kasutati Google Forms veebirakendust (Google Forms). Küsimustik on piltidena leitav töö lisadest.

4.1 Tagasiside küsimustiku tulemuste analüüs

Küsimustik saadeti välja 10 inimesele, kellest 6 küsimustikule vastasid. Küsimustiku vastused olid piisavad esmase tagasiside kogumiseks ning edasiarendusvõimaluste täiendamiseks.

4.1.1 Luuletuste poeetika

Luuletuste poeetika tagasisides paluti vastajatel sisestada katses valitud muutujad (värsimõõt, riimiskeem) ning neid vabas vormis kirjeldada. Küsimustiku värsimõõduga seotud küsimustes paluti vastanutel analüüsida üldist luuletuse rütmilisust. Täpset värsimõõdu kontrolli kasutajatel teha ei palutud, kuna valimisse ei valitud ühtegi poeetika reeglite tundjat. Seetõttu ei tulnud küsimustikust välja, kas valitud värsimõõd tõepoolest saavutati. Täpse värsimõõdu küsimise asemel oli palutud vastajatel otsustada, kas luuletus on rütmiline, ning seda rütmi kirjeldada. Vastanud arvasid, et luuletused olid rütmilised. Kolm vastajat kuuest leidsid, et mõnedes luuletustes oli rütm katkendlikum kui teistes. Nende vastajate genereeritud luuletusi teiste kolme vastaja omadega võrreldes selgus, et nende luuletustes oli väga palju kirjavahemärke ehk lausevahetusi ühe värsirea sees. Sellest tuli ka rütmi katkendlikkus. Nendest kolmest vastajast kaks lisasid, et rütmilisust segavad ühes reas järjestikulised sõnade kordused. Üks neist lisas, et sõnakordused riivasid silma eriti esimese katse esimeses luuletuses ehk uudisteartikli põhjal genereeritud 7-silbilise värsireaga luuletuse puhul.

Riimiskeemi vastavuse hindamisega kasutajatel probleeme polnud, kuigi selle kirjeldamisel jäädi pigem napisõnaliseks. Kõik vastajad nõustusid, et nende valitud riimiskeemi saavutasid kõik genereeritud luuletused. Ühe vastaja luuletuses oli ka täisriimi ehk riimused ühe salmi kõik ridade lõpud. Kaks vastanut pidasid imelikuks sama riimis sõna kordumist erinevates salmides. Nende genereeritud luuletusi ja tekste kõrvutades selgus, et genereerimiseks valitud sisendtekstid olid teiste vastanute omadest märkimisväärselt lühemad, seetõttu oli tõenäoliselt Värsiviguri riimis sõnade valik piiratum.

Erinevate tekstide põhjal loodud luuletuste poeetikaga seotud vastuste võrdluses tõdeti, et kasutajad eelistasid proosatekstil põhinevaid luuletusi uudisteartikli alusel valminutele. Põhjendusi selleks on mitmeid. Vastanud pidasid proosatekstil põhinevate luuletuste riimis sõnu “kuidagi ilusamateks ja paremateks” ning “mitmekesisemaks”. Ilmselt on proosateksti sõnavara juba kujundlikumat laadi kui uudises ning sõnade mitmekesisuse tõttu leidub selles rohkem riimuvaid sõnu, mis välistab nende kordamise tõenäosust. Ka värsimõõdus tundusid

vastused vähem kahtlevad, kuid seda võib põhjendada ka katsete järjekorraga - vastajad juba teadsid, mida katsest oodata ning võrdlesid teise katse luuletusi esimesega.

4.1.2 Luuletuste sisu

Luuletuste sisu puudutavate küsimuste vastused olid väga varieeruvad, kuid sisukad ja põhjalikud. Enamik vastanutest nentisid, et luuletused andsid esialgse teksti sisu edasi ainult märksõnades. Meeldivamaks hinnati neid luuletusi, milles kasutati algses tekstis olulisemaid märksõnu. Näiteks toob üks vastaja ravimtaimede artikli põhjal loodud luuletustes esile, et pikemad luuletused (võib eeldada, et 10- ja 11-silbilise värsireaga) kasutasid rohkem artikli sisu edasiandvaid sõnu, selles näites täpsemalt sõnu “palderjan” ja “ravim”. Selliseid sõnu tuleks uurimuse autori arvates luuletuste genereerimisel rohkem valida. Samas kirjeldasid kaks vastanut kuuest, et luuletuste sisu kvaliteet oli kõigis ühe katse luuletustes ühtlane ehk silpide arv seda ei mõjutanud. Üks vastaja arvas, et luuletustel polnud otsest sisu ega seost algse tekstiga üldse ning nentis samuti, et need read, kus leidis vähem kordusi, olid paremad.

Üldiselt kehtis ka sisus tendents eelistada proosatekstil loodud luuletusi uudisteartikli luuletustele. Põhjusi on selleks autori arvates mitmeid. Ilmselt ei suutnud uudisteartiklist genereeritud luuletused algse teksti sõnumit edasi anda, kuna tekstid olid oma loomult natuke faktikesksemad ning nende oluliste detailide edastamiseks polnud arenduses eraldi tähelepanu pööratud. Proosatekst jäi samuti tekstikaugeks, kuna ei suutnud selle keerulist narratiivi edastada, kuid selle üldmulje oli parem, sest kasutas “huvitavamaid väljendeid”. Tõenäoliselt tuleb siin mängu proosateksti kujundlikum keelekasutus, milles juba leidub igapäevakeelest harvemini kasutatavaid sõnu.

4.1.3 Üldine mulje ja kasutajakogemus

Vastajatel paluti anda üldine hinnang Värsiviguri kasutamise kogemusele ning selle genereeritud luuletustele 10 palli skaalal. Värsiviguri üldine hinnang kuue vastaja põhjal oli 7,6p/10p. Ettepanekutesse anti tagasisidet peamiselt Värsiviguri üldise toimimise kohta. Kaks vastanut kuuest ütlesid, et pidid katsetama erinevate tekstidega, kuna esimestel katsetustel oli tekst liiga pikk ning programm “hangus ära”. Samuti lisasid nad, et rohkemate salvide arvu korral kasvas programmi tööaeg märgatavalt, kuid jäi “normi piiridesse”. Uurimuse autor arvab, et optimaalse sisendteksti suuruse leidmiseks tuleks tulevikus teha põhjalikum seda uuriv küsitlus.

Genereeritud luuletuste keskmine üldhinnang oli 5,3p/10p. Luuletuste osas oli üldine tagasiside väga positiivne ja toetav. Vastanud arvasid, et luuletused olid hea struktuuri ja

ülesehitusega ning neid oli huvitav genereerida. Samas lisasid ka siin mitmed, et neid jäid sõnakordused häirima ning see muutis genereeritud luuletused ettearvatavaks. Kokkuvõttlikult ütlesid mitmed, et genereeritud luuletused tunduvad nagu luuletused, kuid sisu ja täpsemaid sõnu vaadeldes kaob esmane efekt ära.

4.2 Diskussioon ja edasiarendusvõimalused

Autori katsetustes ning hinnangul töötas värsimõõdu algoritm ettenähtult, st hüpoteetiline sõnade järjestuse korrapära silpide ja rõhkude arvu alusel saavutati. Samas nõustub autor tagasiside küsitlusele vastanutega: süntaktilise poole klapitamise ning lausete vahetamise tõttu mõjus rütm siiski liiga katkendlikult. Töö autor alahindas lause tähendusliku terviku olulisust värsimõõdu loomisprotsessis. Tulevikus tuleks kindlasti sellist värsimõõdu algoritmi veel katsetada variandis, kus ritta sobivaid sõnu oleks rohkem nii, et need moodustaks värsirea siseselt ühe mõttekäigu. Selleks sobiks näiteks sarnase ülesehitusega wordneti abil sünonüümide leidmine (näiteks paaritusilbilisele sõnale leitakse paarissilbiline vaste). See annaks luuletuse sisule huvitava uue mõtme, mis võib väljavahetatud sõnade tõttu mõjuda kohati ebaloomiliselt ja sürreaalselt, kuid on siiski veel katsetamata lahenduskäik.

Riimi osas oli tulemused väga head. Lõppriim toimus edukalt. Selline algoritmiline ülesehitus, mis riimis sõnade leidmise eelisjärjekorras ära tegi, toimis hästi ning seda võiks sarnastes rakendustes edaspidigi kasutada. Riimile saaks tulevikus täiendavalt lisada ka algriimi, näiteks värsiridade algusesse või ühe värsirea siseselt nagu tüüpiliselt eesti vanasõnades ja rahvaluules.

Luuletuste sisu polnud töö eesmärk, küll aga on see oluline luuletuse osa ning kindlasti saaks sellele tulevikus rohkem tähelepanu pöörata. Uurimuse autor nõustub küsimustikus esitatud sisu kriitikaga ning et sõnade valik mängib luuletuse sisus suurt rolli. Sisu parandamiseks saaks Värsivigurit täiendada põhjalikuma teksti töötlusega, näiteks leida algsest tekstist üles olulised või korduvad erilised sõnad ja väljendid, mida luuletused kindlasti sisaldama peaksid. Suurematest tekstidest võiks proovida teha ka kontsentreeritud kokkuvõtteid, milles kirjeldavaid vahe- ja täitelauseid vähem oleks ning seega ei satuks neid nii palju ka luuletusse. Nende muutuste implementeerimine ei mahtunud töö skoopi.

Mõned genereeritud luuletused sisaldasid juhuslikult ka töösse valimata jäänud poetikavõtteid. Sõnade järjestikuline kordus oli üks esialgu planeerimata poeetiline võte, mis värsiritta soovitud värsimõõdu saavutamiseks lisati. See võte osutus tagasiside küsimustiku alusel segavaks, kuid mõjus autori arvates mõnel juhul kurjakuulutatavalt ja loitsivalt. Samas

nõustub autor, et kordust esines kohati liiga palju ning see mõjus genereeritud luuletustes pigem segavalt. Samas mõjus kordus teinekord sarnaselt Hando Runneli “Üks veski seisab vete pääl” (2018) salmi ülesehitusele, milles sõna “veski” igas reas kordus. Näiteks Värske Rõhu proosarubriigist valitud Krismar Rosina tekstist “Kerjused ei vali” (2024) genereeritud luuletuses:

“Ah tiksub tiksub langeb viska selga

Ah tiksub langeb viska ette viina

Oh kostub Taamal Taamal. surutud on seina

Ah riisus riisus... vali vali, minna minna... Mul on aega

Ka tiksub tiksub langeb viska aia

Ah tiksub ette langeb Naabrinaine

Oh Tiksub Tiksub! pudel pudel. Teine Teine, pudel teine

Oo vali vali. minna minna, Mul on lasteaia”.

Selles näites mõjub kordus hästi ka pelgalt sõnavaliku tõttu, sest sõna “tiksub” viitab kellale, mis ise on rütmilise omadusega ning selle jälgendamine luuletuses on kaval poeetiline võte.

Samuti esines mitmes luuletuses siirdeid ehk “mõtte jätkumist järgmises värsireas või salmis” (Merilai et al., 2011:51). Seda põhjustas näiteks sidesõna “ja” või “ning” või sama verbi leidumine kahes järjestikus värsireas, näiteks “Oo longib longib. Ootab foori // Ka Ootab surnumatja”. See on samuti üks näide kordusest, mis silma ei riiva, vaid täiendab luuletust. Paraku jäävad need korduste juhtumid pigem haruldasteks. Tulevikus võiks kordust vähendada, kuid iga luuletuse mõnes värsireas sellega samuti siiski katsetada.

Tööd saaks täiendada nende ja paljude teiste poeetiliste parameetrite lisamise ja täpsustamistamisega. Sellesse töösse valitud poetikast saaks arendada välja kõik silbilisrõhulised värsimõõdud (trohheus, jamb, daktül, amfibrahh, anapest), lisada erinevaid riime (algriim, keskriim, lõppriimi täiendada) ning luua võimalus genereerida erinevaid kinnisvorme (sonett, haiku, tanka) jne.

4.3 Võrdlus sarnaste rakendustega

Võrdluses töö peatükis “Varasemad tööd” esitatud rakendustega, on Värshivigur ainulaadne, kuid jääb mõnes aspektis varasematele lahendustele alla. Tarkvara on esimene teadaolev eestikeelset luulet genereeriv lahendus. Suurimat konkurentsi lahendusele pakub ChatGPT-3.5, mille genereeritud luuletuste üldmulje on Värshiviguri loodud luuletustest sisu, loogilise lauseehituse ja voolavuse poolest etem. Samas oli Värshiviguri peamiseks eesmärgiks luua luuletusi kindlaid poeetilisi parameetreid järgides, mille tegemisega ChatGPT-3.5 hästi hakkama ei saanud või pidi talle neid pikemalt selgitama.

Värshiviguri genereeritud luuletusi on teiste lahenduste luuletustega natuke keerulisem võrrelda, eriti poeetilisest aspektist, kuna nende keele erinevus muudab poeetiliste parameetrite võrdlemise raskendatuks. Enamus rakendustest suutis riimida, kuid ei suutnud tagastada spetsiifilist riimiskeemi nagu Värshivigur. Värshiviguri osas ükski rakendus seda eraldi määrata ei võimaldanud, seega on nende võrdlus asjatu. Samas tuleb nentida, et üldiselt olid need luuletused vähem katkendlikud ning seega mõjusid kohati korrapärasema rütmiga kui Värshiviguriga genereeritud luuletused. Stroofikas on mõned rakendused tänu valitavatele kinnisvormidele paremad, näiteks editpad.org ja poemgenerator-ai.com. Samas pole need kinnisvormid alati õigete poeetiliste tunnustega, mistõttu ei saa neid võrdluses iga kinnisvormiga arvestada.

Kokkuvõte

Uurimistöös “Luuletuste automaatne genereerimine” loodi tarkvaralahendus, mis suudab genereerida kasutaja esitatud teksti põhjal ja valitud poeetilisi parameetreid järgiv luuletus. Keelemudelite kiire arengu tõttu on luuleteksti genereerimise teema aina aktuaalsemaks muutunud, kuid eestikeelseid poeetikat hõlmavaid lahendusi probleemile töö tegemise ajal polnud.

Lõputöös uuriti põhjalikult rakendusse valitud poeetilisi parameetreid, esitati ülevaade olemasolevatest sarnastest lahendustest ning uurimustest.

Uurimistööst järeldus, et valitud algoritmilise ülesehitusega on võimalik genereerida valitud poeetikat järgiv luuletus. Kasutades teksti segmenteerimiseks ja töötlemiseks Pythoni EstNLTK teekide kogumi võimalusi, saavutati kasutaja valitud värsimõõt, ehkki selle käigus pidi luuletuse sisus kompromisse tegema. Peamised negatiivsed mõjutused oli 1) osade sõnade väljajätmine, mis rikkus teksti terviklikkust ning 2) rütmi säilitamiseks sõnade kordamine, mis mõjus kohati ebanaturaalselt. Lõppriim on luuletuses kasutaja valitud riimiskeemi järgides võimalik saavutada erinevate tekstitüüpide, näiteks uudisteartikli ja proosateksti puhul. Luuletuse esialgse teksti sisu edasiandmine oli juhuslik, kuid sõltus osaliselt valitud silpide- ja salmide arvust ehk luuletuse kogupikkusest kui ka riimis sõnaks juhusliku valimi teel sattunud sõna olulisusest algses tekstis. Luuletuse üldine mulje olenes suures osas sinna valitud sõnade erilisusest ning kujundlikkusest, mistõttu oli uurimuse käigus läbiviidud tagasiside küsimustikus eelistatud uudisteartiklist kujundlikuma keelekasutusega proosateksti põhjal genereeritud luuletused.

Poeetika ja luuletuste genereerimine on valdkonnad, milles on veel palju arenguruumi. Tulevikus saaks rakendust edasi arendada parandades peaosas genereeritud luuletuste sisu ning rütmilisust, lisaks võiks anda kasutajale võimalusi suurema hulga poeetiliste parameetrite vahel valimiseks, näiteks lisada stroofilisi kinnisvorme või keelekujundeid.

Viidatud kirjandus

Aedmaa, E. Kaljumäe, H. (17.02.2023). Skandaalne ChatGPT - kuidas panna see eesti keele heaks tööle? *ERR portaal*. Vaadatud 04.04.2024

<https://novaator.err.ee/1608889298/skandaalne-chatgpt-kuidas-panna-see-eesti-keele-heaks-toole>

Aristoteles. (2001). *Luulekunstist: poeetika*. Tartu: Tartu Ülikooli klassikalise filoloogia õppetool. Vaadatud 17.03.2024 <https://antiik.ut.ee/tolked/aristoteles/index.html>

Burt, S. (2014). *Why people need poetry*. Vaadatud 10.05.2024 <https://www.youtube.com/watch?v=08ZWROqoTZo>

Eesti keele käsiraamat: Ortograafia, Häälikud ja foneemid, Rõhk. (2007). *Eesti Keele Instituut*. Vaadatud 10.05.2024

<https://www.eki.ee/books/ekk09/index.php?p=2&p1=5&id=21>

EstNLTK dokumentatsioon. Vaadatud 10.05.2024 <https://estnltk.github.io/estnltk>

Friedenthal, M. (30.04.2023). Kas tehisintellekti loodud kirjandus vallutab maailma? *ERR portaal*. Vaadatud 10.05.2024

<https://www.err.ee/1608960913/meelis-friedenthal-kas-robotite-loodud-kirjandus-vallutab-maailma>

Google Colab koduleht. Vaadatud 14.05.2024 <https://colab.google/>

Google Colab Forms. Vaadatud 14.05.2024

<https://colab.research.google.com/notebooks/forms.ipynb>

Google Drive. Vaadatud 14.05.2024 <https://www.google.com/drive/>

Google Forms veebirakendus. Vaadatud 14.05.2024 <https://www.google.com/forms/about/>

Jakobson, R. (2012). Lingvistika ja poeetika. *Akadeemia*. Vaadatud 10.05.2024 <https://lepo.it.da.ut.ee/~silvi11/Jakobson%20Lingvistika%20ja%20poeetika.pdf>

Juurak, R. (20.01.2023). Üks tont käib klassis ringi - tehisintellekt!. *Õpetajate Leht*. Vaadatud 12.04.2024 <https://opleht.ee/2023/01/uks-tont-kaib-klassis-ringi-tehisintellekt/>

Jupyter Notebook Latest. Vaadatud 14.05.2024

<https://jupyter-notebook.readthedocs.io/en/latest/>

Klementi, J. (27.02.2023). Arvuti kirjutab raamatuid, kuid keegi ei tea kes on autor. *ERR: raadiouudised*. Vaadatud 04.05.2024

<https://www.err.ee/1608898823/arvuti-kirjutab-raamatuid-kuid-keegi-ei-tea-kes-on-autor>

Krull, H. (2011, Juuni). Mis on luule?. *Vikerkaar*. Vaadatud 14.03.2024

<https://www.vikerkaar.ee/archives/12667>

Laas, O. (22.03.2023). ChatGPT eetilise kasutamisest koolides. *ERR portaal*. Vaadatud 10.04.2024

<https://www.err.ee/1608922886/oliver-laas-chatgpt-eetilise-kasutamisest-koolides>

Lotman, M. (1986). Poeetika lühikursus. *Vikerkaar*, 1, 68–71.

Lotman, R. (2019). *Eesti Sonett*. Tartu: Tartu Ülikooli Kirjastus.

Merilai, A., Annus, E., Saro, A. (2011). *Poeetika gümnaasiumiõpik*. Tartu: Tartu Ülikooli Kirjastus.

Muru, K. (1992). *Sõnarine 2*. Tallinn: kirjastus “Eesti Raamat”.

Oliveira, H. G. (2017, September). A survey on intelligent poetry generation: Languages, features, techniques, reutilisation and evaluation. In *Proceedings of the 10th international conference on natural language generation*, 11-20.

OpenAI (2022). ChatGPT (3.5) <https://chat.openai.com>

Põldmäe, J. (2002). *Eesti värsiõpetus*. Tallinn: Eesti Keele Sihtasutus.

Python 3.10.12. <https://www.python.org/doc/>

Rooste, J. (2005). *Ilusaks inimeseks: saladokumente igavikuarhiivi peavalitsusest : [luuletused]*. Tallinn: Verb kirjastus.

Rosin, K. (18.03.2024). Kerjused ei vali. *Värske Rõhk*. Vaadatud 14.05.2024
<https://va.ee/kerjused-ei-vali/>

Runnel, H. (2018). *Jaanipäevaks kõrgeks kasvab rohi*. Tartu: Ilmamaa.

Liiv, J., Vinkel, A. (1989). *Sinuga ja sinuta*. Tallinn: Eesti Raamat.

Väljataga, M. (2013). Mis on luule?. *Keel ja Kirjandus*. Vaadatud 11.04.2024
<https://keeljakirjandus.ee/ee/archives/24627>

Lisad

I. Tarkvara Värsivigur

link:

https://colab.research.google.com/drive/1c7W2GKW_Kpp5NyGQOVJVeQA1SU5Cq81G?usp=sharing

II. Tagasiside küsimustik

Tagasiside küsimustik luuletuste generaatorile Värsivigur

Tere!

Mina olen Tartu Ülikooli informaatika eriala tudeng Karoli Kapp. Minu bakalaureusetöö eesmärgiks on luua luuletuste generaator nimega Värsivigur, mis suudaks kasutaja (Teie) etteantud teksti kasutades väljastada kindlaid poeetika reegleid järgiv luuletus. Selle küsimustiku eesmärgiks on saada tagasisidet valminud tarkvaralahendusele, mille käigus tuleb proovida Värsiviguriga luuletusi genereerida.

Küsimustik koosneb kolmest osast. Esimeses kahes on palutud teil viia läbi Värsiviguri failis ettevalmistatud katsed ning võrrelda omavahel kolme luuletust. Küsimustiku viimases osas on mõned üldisemad küsimused teie kasutajakogemuse kohta. Küsimustikule vastamiseks läheb umbes 15 - 20 min.

karolikas.kapp@gmail.com [Vaheta kontot](#)



Pole jagatud

Kasutusjuhend

- Tee kindlaks, et oled veebibrauseris Google'i kontosse sisse logitud.
- Ava Värsivigur veebiaadressil https://colab.research.google.com/drive/1c7W2GKW_Kpp5NyGQOVJVeQA1SU5Cq81G?usp=sharing
- Vajuta ülalt ribalt "Runtime" ning vali selle alt "Run all". Google hoiatab Teid "Warning: This notebook was not authored by Google...". Vajuta "Run anyway".
- Värsiviguril läheb umbes minut aega, et vajalikud teegid alla laadida ning luuletused genereerida. Veebilehe kõigi alajaotuste juurde ilmub roheline linnuke, kui Värsivigur on töö lõpetanud.
- Võite uurida Värsiviguri väljundit ning erinevaid lahtreid. Saate neid vahetada, kuid peate alati peale kõiki muudatusi vajutama uuesti "run all", et Värsivigur saaks teie valikud salvestada ja nende põhjal uued luuletused genereerida. Katseta julgelt!

Järgmine

Tühjenda vorm

Tagasiside küsimustik luuletuste generaatorile Värsivigur

karolikas.kapp@gmail.com [Vaheta kontot](#)



Pole jagatud

* Viitab kohustuslikule küsimusele

Esimene katse: uudisteartiklist loodud luuletuste võrdlus

Järgnevas katses genereeritakse luuletused Teie valitud artiklile tuginedes.

1. Liigu Värsiviguri Colabis jaotuseni pealkirjaga "Katse: **uudisteartiklist**/proosatekstist loodud luuletused".
2. Leia endale meeldivast meediaväljaandest (näiteks ERR portaalist) kopeeritud **artikkel**, kleebi see tekstilahtrisse "Tekst". Kui soovid, vaheta ka mõni teine muutuja, näiteks värsimõõt või salmide arv endale sobivaks (suuremate salmide arvu puhul võtab luuletuse genereerimine pisut kauem).
3. Vali ülevalt ribalt "Runtime" ning selle alt "run all". Oota kuni kolm erinevat luuletust sinu ees on.
4. Vasta järgnevatele küsimustele.

NB! Kui mingil põhjusel Värsivigur ära hangub või viskab veakoodi, saad rakenduse taaskäivitada vajutades ülevalt ribalt "Runtime" alt "Restart session and run all".

Kopeeri siia tekstikasti kasutatud artikli link. *

Teie vastus

Kopeeri siia tekstikasti genereeritud luuletused. *

Teie vastus

Millise riimiskeemi valisite? *

- ☐ Süliiriim ABBA (riimuvad 1. ja 4. rida ning 2. ja 3. rida)
- ☐ Paarisriim AABB (riimuvad 1. ja 2. rida ning 3. ja 4. rida)
- ☐ Ristriim ABAB (riimuvad 1. ja 3. rida ning 2. ja 4. rida)

Kas genereeritud luuletused tagastasid soovitud riimide paiknevuse? Kuidas kirjeldaksite üldiselt Värsiviguri loodud luuletuste riimi ja kas mõnes luuletuses on riim parem kui teistes? *

Teie vastus

Millise värsimõõdu valisite? *

- ☐ Trohheus
- ☐ Jamb

Kas genereeritud luuletused tunduvad teile rütmilisena? Kuidas kirjeldaksid seda rütmi (pidev või katkev, korrapärane või korrapäratu, kiire või aeglane jne)? Kirjelda lühidalt. *

Teie vastus

Kuidas kirjeldaksite luuletuse sisu? Kas luuletuste sisu on esialgse tekstiga sarnane? Kas mõni luuletus annab sisu paremini edasi kui teine? *

Teie vastus

Tagasi

Järgmine

Tühjenda vorm

Tagasiside küsimustik luuletuste generaatorile Värsivigur

karolikas.kapp@gmail.com [Vaheta kontot](#)



Pole jagatud

* Viitab kohustuslikule küsimusele

Teine katse: proosatekstist loodud luuletuste võrdlus

Järgnevas katses genereeritakse luuletused Teie valitud **proosatekstile** tuginedes.

1. Liigu Värsiviguri Colabis jaotuseni pealkirjaga "Katse: uudisteartiklist/**proosatekstist** loodud luuletused".
2. Leia endale meeldivast meediaväljaandest (näiteks siit: <https://va.ee/rubriigid/proosa/>) kopeeritud **proosatekst**, kleebi see tekstilahtrisse "Tekst". Kui soovid, vaheta ka mõni teine muutuja, näiteks värsimõõt või salvide arv endale sobivaks.
3. Vali ülevalt ribalt "Runtime" ning selle alt "run all". Oota kuni kolm erinevat luuletus sinu ees on.
4. Vasta järgnevatele küsimustele.

NB! Kui mingil põhjusel Värsivigur ära hangub või viskab veakoodi, saad rakenduse taaskäivitada vajutades ülevalt ribalt "Runtime" alt "Restart session and run all".

Kopeeri siia tekstikasti kasutatud artikli link. *

Teie vastus

Kopeeri siia tekstikasti genereeritud luuletused. *

Teie vastus

Millise riimiskeemi valisite? *

- ☐ Süliriim ABBA (riimuvad 1. ja 4. rida ning 2. ja 3. rida)
- ☐ Paarisriim AABB (riimuvad 1. ja 2. rida ning 3. ja 4. rida)
- ☐ Ristriim ABAB (riimuvad 1. ja 3. rida ning 2. ja 4. rida)

Tagasiside küsimustik luuletuste generaatorile Värsivigur

karolikas.kapp@gmail.com [Vaheta kontot](#)



Pole jagatud

* Viitab kohustuslikule küsimusele

Teine katse: proosatekstist loodud luuletuste võrdlus

Järgnevas katses genereeritakse luuletused Teie valitud **proosatekstile** tuginedes.

1. Liigu Värsiviguri Colabis jaotuseni pealkirjaga "Katse: uudisteartiklist/**proosatekstist** loodud luuletused".
2. Leia endale meeldivast meediaväljaandest (näiteks siit: <https://va.ee/rubriigid/proosa/>) kopeeritud **proosatekst**, kleebi see tekstilahtrisse "Tekst". Kui soovid, vaheta ka mõni teine muutuja, näiteks värsimõõt või salmide arv endale sobivaks.
3. Vali ülevalt ribalt "Runtime" ning selle alt "run all". Oota kuni kolm erinevat luuletus sinu ees on.
4. Vasta järgnevatele küsimustele.

NB! Kui mingil põhjusel Värsivigur ära hangub või viskab veakoodi, saad rakenduse taaskäivitada vajutades ülevalt ribalt "Runtime" alt "Restart session and run all".

Kopeeri siia tekstikasti kasutatud artikli link. *

Teie vastus

Kopeeri siia tekstikasti genereeritud luuletused. *

Teie vastus

Millise riimiskeemi valisite? *

- ☐ Süliirim ABBA (riimuvad 1. ja 4. rida ning 2. ja 3. rida)
- ☐ Paarisriim AABB (riimuvad 1. ja 2. rida ning 3. ja 4. rida)
- ☐ Ristriim ABAB (riimuvad 1. ja 3. rida ning 2. ja 4. rida)

Kas genereeritud luuletused tagastasid soovitud riimide paiknevuse? Kuidas kirjeldaksite üldiselt Värsiviguri loodud luuletuste riimi ja kas mõnes luuletuses on riim parem kui teistes? *

Teie vastus

Millise värsimõõdu valisite? *

☐ Trohheus

☐ Jamb

Kas genereeritud luuletused tunduvad teile rütmilisena? Kuidas kirjeldaksid seda rütmi (pidev või katkev, korrapärane või korrapäratu, kiire või aeglane jne)? Kirjelda lühidalt. *

Teie vastus

Kuidas kirjeldaksite luuletuse sisu? Kas luuletuste sisu on esialgse tekstiga sarnane? Kas mõni luuletus annab sisu paremini edasi kui teine? *

Teie vastus

Tagasi

Järgmine

Tühjenda vorm

Tagasiside küsimustik luuletuste generaatorile Värsvigur

karolikas.kapp@gmail.com [Vaheta kontot](#)



Pole jagatud

* Viitab kohustuslikule küsimusele

Küsimused Värsviguri kasutajakogemuse kohta

Kuidas hindaksite oma üldist kogemust kasutades luulegeneraatorit? *

	1	2	3	4	5	6	7	8	9	10	
Ei saanud Värsvigurit tööle	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Värsvigur töötas laitmatult

Kuidas hindaksid üldiselt genereeritud luuletusi? *

	1	2	3	4	5	6	7	8	9	10	
Väga halb	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Suurepärane

Lisakommentaari ja ettepanekud Värsviguri kohta. *

Teie vastus

[Tagasi](#)

[Saada ära](#)

[Tühjenda vorm](#)

Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Karoli Kapp,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose **Luuletuse automaatne genereerimine**, mille juhendaja on Sven Aller, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Karoli Kapp

15.05.2024