

TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Marvin Vihman

Ristsõnade automaatne genereerija

Bakalaureusetöö (9 EAP)

Juhendaja Sven Aller

Tartu 2022

Ristsõnade automaatne genereerija

Lühikokkuvõte:

Töö eesmärk on luua eestikeelsete ristsõnade genereerija, mis kasutab eesti keele keele-ressursse (analüsaatorid, morfoloogiline süntesaator, sõnastikud jne) ja kogub muu vajaliku info veebist. Tulemuseks saadud ristsõna on nii paberile trükitav kui ka lahendatav ja kontrollitav otse veebilehel.

Võtmesõnad:

Keeletehnoloogia, informaatika, ristsõnade genereerija

CERCS: P175 Informaatika

Automatic Crosswords Generator

Abstract:

The purpose of this thesis is to create an Estonian crosswords generator, which uses the Estonian language resources (analyzers, morphological synthesizer, dictionaries and the like) and collects the other needed information from the web. The resulting crossword would be printable so that it can be solved in paper format and also it would be verifiable directly on the webpage.

Keywords:

Language technology, computer science, crosswords generator

CERCS: P175 Informatics

Sisukord

Sissejuhatus.....	5
1. Ristsõnade taustalugu.....	7
1.1 Ristsõnade kasulikkus	7
1.2 Erinevad ristsõnade tüübid	8
1.2.1 Krüptilised ristsõnad	9
1.2.2 Blokeeritud ruudustikuga ristsõnad	10
1.2.3 Võrega ristsõnad	11
1.2.4 Temaatilised ristsõnad	11
1.2.5 Koodsõna ristsõnad.....	12
1.2.6 Sõna otsimise ristsõnad.....	13
1.3 Ristsõnade genereerimise viisid	14
2. Ristsõnade automaatse genereeriija rakendus.....	16
2.1 Kasutatud teegid.....	16
2.1.1 EstNLTK.....	17
2.1.2 Pandas	17
2.1.3 Flask.....	17
2.1.4 Re	18
2.2 Andmete eeltöötlus.....	18
2.3 Veebirakenduse algoritmide loogika.....	19
2.3.1 Ristsõna algoritm	19
2.3.2 Veebirakenduse alus	20
2.4 Kasutusjuhend	20
2.5 Testimine ja tulemused.....	22
2.6 Edasiarendusvõimalused	22
Kokkuvõte.....	24

Viidatud kirjandus.....	25
Lisad.....	28
I. Terminid.....	28
II. Litsents.....	29

Sissejuhatus

Ristsõnu on meile teada-tuntud kujule loodud juba aastakümneid (Novaator, 2013). See sai alguse välismaal ning aja möödudes on jõutud selleni, et ristsõnu enam ei valmistata põhiliselt käsitsi, vaid nende loomiseks kasutatakse programmide abi. Põhjuseks on arvuti võime koostada rohkem variante kindlatest sõna paaridest, seda ka kiiremini, kui inimene selleks võimeline on. Kui eestikeelseid ristsõnu koostatama hakati, tehti seda kas manuaalselt või kasutati võõrkeelseid tarkvarasid. Kuid neid programme, mis automaatselt genereerisid eesti keelseid ristsõnu, vabalt kättesaadaval ei leidunud.

Programmid, mis tegelevad ristsõnadega, on rohkem suunatud võõrkeelsete sõnadele. Lisaks rakendused, mis on suunatud ristsõnade genereerimisele, on vähemuses võrreldes nendega, mis tegelevad ristsõnadele lahenduste pakkumisega. Veebist leiab erinevates programmeerimiskeeltes kirjutatud programme, kuid need töötavad põhimõttel, et antakse ette lahendussõna ja vihje paar, mis tähendab ikkagi, et sõnade ja vihjete eeltöötlus on vaja teha enne, kui ristsõna saab hakata koostama. Õppetöös on ristsõnad kasulikud, kuid neid tehakse rohkem võõrkeele tundides või teatud õppevaldkondades.

Lõputöö eesmärk on luua automaatselt eestikeelseid ristsõnu genereeriv veebirakendus. Rakenduse põhimõte on see, et seda on võimalik lahendada nii arvutis kui ka paberkujul lahendamiseks välja printida. Töö raames loodi terviklik veebirakendus, millega on võimalik genereerida ristsõnu, neid veebis lahendada ning ka välja printida. Programmi baaskoodi üritati kirjutada universaalselt, et tulevikus oleks lihtsam edasiarendusi luua ning täiendada olemasolevat veebirakendust. Töö tulemusena saadi rakendus, mille algoritmide tulemust hinnati genereeritavate ristsõnade põhjal. Nende tulemuste põhjal on võimalik järgnevat edasiarendusvõimalusi teostada ning programmi rohkem optimeerida.

Käesolev töö koosneb sissejuhatusest, kokkuvõttest ning kahest sisupeatükist, mis omakorda sisaldavad alapeatükke. Esimeses peatükis antakse ülevaade ristsõnade ajaloost, ajast kui neid tehti meile tänapäeval teada-tuntaval kujul (Novaator, 2013), lisaks antakse ülevaade teadustöödest, mis käsitlevad ristsõnade kasulikkusest õppetöös. Seejärel on kirjeldatud erinevad ristsõnade tüübid ning seejärel ristsõnade genereerimise viisid. Teine sisupeatükk tutvustab loodud rakendust. Sinna kuulub lühike kokkuvõte kasutatud moodulitest, seejärel kirjeldatakse andmete eeltöötlust. Peale seda käsitletakse rakenduse tööks vajalikke algoritme, millele järgneb kasutusjuhend veebirakenduse kasutamisest. Seejärel kirjeldatakse testimist ja

selle tulemusi ning lõpuks tuuakse välja erinevad edasiarendusvõimalused, mida käesoleva projekti puhul saab edendada.

1. Ristsõnade taustalugu

Käesolev peatükk tutvustab ristsõnade loomise ajalugu. Lisaks tuuakse välja erinevad ristsõnade tüübid ning ristsõnade genereerimise viisid.

Ristsõnasid on tehtud teada-tuntud sõna kujul juba alates aastast 1913, 21. detsembril jõudis ajakirjandusse üks esimesi ristsõnamõistatusi (Novaator, 2013). Väljaanne, kus ristsõna ilmus, oli The New York World (Wikipedia, 2022) ning mõistatuse autor selleaegne ajakirjanik Arthur Wynne. Kuigi ajakirjanduses oli ka varasemalt ilmunud ristsõnu, seda 1890. aasta 14. septembril Itaalia väljaandes Il Secolo Illustrato della Domenica, siis Wynne'i ristsõna oli lähedasem sellele vormile, millisena me seda tänapäeval tunneme. 1920ndaiks oli ristsõnahullus jõudnud sinnamaani, et raamatukogudes tekkis probleeme sõnastike ja entsüklopeediate puuduse tõttu: tudengid ja teised lugejad, kes vajasisid neid oma igapäevategemiste jaoks, ei saanud seda teha (*ibid.*).

1924. aastal ilmus Guinnessi andmete järgi USA-s ristsõnade esimene kogumik (Bellis, s.a.). Raamat loodi Dick Simon ja Lincoln Schuster koostöös, väljaanne Simon & Schuster¹ all, ning kogumiku nimetus oli „Cross Word Book²“. Väljaanne jätkab tänapäevalgi ristsõnade tootmist (Bellis, s.a.).

Alles hiljuti, aastal 1997, tekkis esimene arvutitarkvaraprogramm Crossword Weaver, mida kasutati abivahendina ristsõnade loomisel arvutis. Programm loodi ja patenteeriti Variety Games Inc. poolt (Bellis, s.a.).

1.1 Ristsõnade kasulikkus

Ristsõnu on kasutatud nii meelelahutuseks kui ka õppetöös. Seda põhjusel, et ristsõnade lahendamine arendab inimese mälu, ehk teravdab ajutööd mõtteprotsesside osas, kirjutavad Abuelo jt (2016). Selle tõestamiseks on testitud erinevate valdkondade tudengeid, nt Indias esimese aasta veterinaaria tudengite peal tegid katse Abuelo, Castillo ja May (Abuelo, Castillo, & May, 2016) ning farmakoloogia tudengeid testisid Gaikwad ja Tankhiwale (Gaikwad &

¹ Väljaande Simon & Schuster veebileht: <https://www.simonandschuster.com/>.

² Tänapäeval on John M. Samsoni poolt täiendatud kogumik müügis lehel <https://www.simonandschuster.com/books/Simon-and-Schuster-Crossword-Puzzle-Book-248/John-M-Samson/9780743283137>.

Tankhiwale, 2012), kus mõlemal juhul kasutati sarnast protsessi. Abuelo jt (2016) tõdesid, et tavaliselt on suunatud tudengeid õppima terminoloogiat pähe õppimise meetodil, mida tudengid ise ei eelista. Autorid ütlevad, et teiste valdkondade jaoks on varem tehtud juba uuringuid, mis tõestavad ristsõnade kasutamise efektiivsust õppimise vahendina, kuid uuringut veterinaaria valdkonnaga seotud terminoloogia puhul pole tehtud. Seega tehti test, kus kasutati 42 esimese aasta veterinaaria bakalaureuse tudengit, nad jagati juhuslikult 2 gruppi ning tudengeid testiti varem õpitud teadmiste põhjal. Seejärel esimene grupp sai 30 sõna ja definitsiooniga paberi, teine grupp sai sama, kuid lisaks kuus spetsiaalselt loodud ristsõna, mis hõlmasid õpitud termineid. 50 minuti möödudes tehti vast õpitu põhjal test. Tulemusi võrreldi ning need näitasid, et ristsõnade abil õppinud tudengid said parema tulemuse. Abuelo jt (2016) tegid ka hiljem eraldi küsimustiku, millega saadi teada tudengitelt, et ristsõnade kaasamine õpingutesse on teretulnud.

Gaikwad ja Tankhiwale puhul viidi läbi uuring, millega sooviti (Gaikwad & Tankhiwale, 2012):

- hinnata ristsõnade kasulikkust iseõppevahendina farmakoloogias;
- hinnata õpilaste arusaamu ristsõnadest kui iseõppevahendist.

Uuring viidi läbi India meditsiini õppekava (II MBBS) tudengite peal. Nad jagati samuti kahte gruppi, mõlemas 35 liiget. Mõlemale grupile tehti nii enne kui ka pärast test, milles hinnati tööõigsust. Viimaseks testiks kordasid esimese grupi liikmed tööks nii tavaviisil, milleks on õpikuga kordamine ehk meeldetuletamine, kuid lisaks anti ka ristsõnu, mille abil sai korrata. Teine grupp kordas ainult tavaviisil ehk õpikuga. Mõlemad grupid said ühe tunni aega testi tegemiseks. Testi tulemuseks saadi, et ristsõnu saanud grupi tudengid said testi tulemusteks enne ja pärast kordamist vastavalt 6.09 ja 12.87 ning teise grupi tulemused olid 6.03 ja 9.74. Selle tulemuse põhjal oli esimesel grupil tulemuste areng 33.90% ja teisel grupil 18.55%, mis on märgatav vahe. Gaikwadi ja Tankhiwale (2012) vastati 85.71% juhul, et ristsõnade baasil õppimine on kasulikum tavapärasest meeldejätmise ja kordamise meetodist. Seega saab väita, et ristsõnad on kasulikud nii meelelahutuses meeldejätmise võime arendamiseks kui ka õppetöös erinevates valdkondades.

1.2 Erinevad ristsõnade tüübid

Ristsõnasid on erinevaid liike. Neid saab defineerida nii ruudustiku kui ka ristsõna stiili, kuju, suuruse ja raskusastme järgi, samuti ka keelte, vihjete ning teemade alusel (Weyer, 2022).

Lisaks sõna tüüpi ristsõnadele on veel numbritel ja abstraktsetel sümbolitel põhinevad mõistatused, mida nimetatakse sudokuks. Abstraktsete sümbolitega sudokudes on numbrid, tähed ja sõnad asendatud erinevate sümbolitega (Fragar, 2018).

Kultuurides, kus keeles grammatiliselt kirjutatakse vasakult paremale, on tulemuse sõnad ning fraasid paigutatud ruudustikku vasakult paremale ja ülevalt alla. Sõnade ja fraaside eraldamiseks üksteisest kasutatakse varjutatud ruute (The360reporters, 2022).

1.2.1 Krüptilised ristsõnad

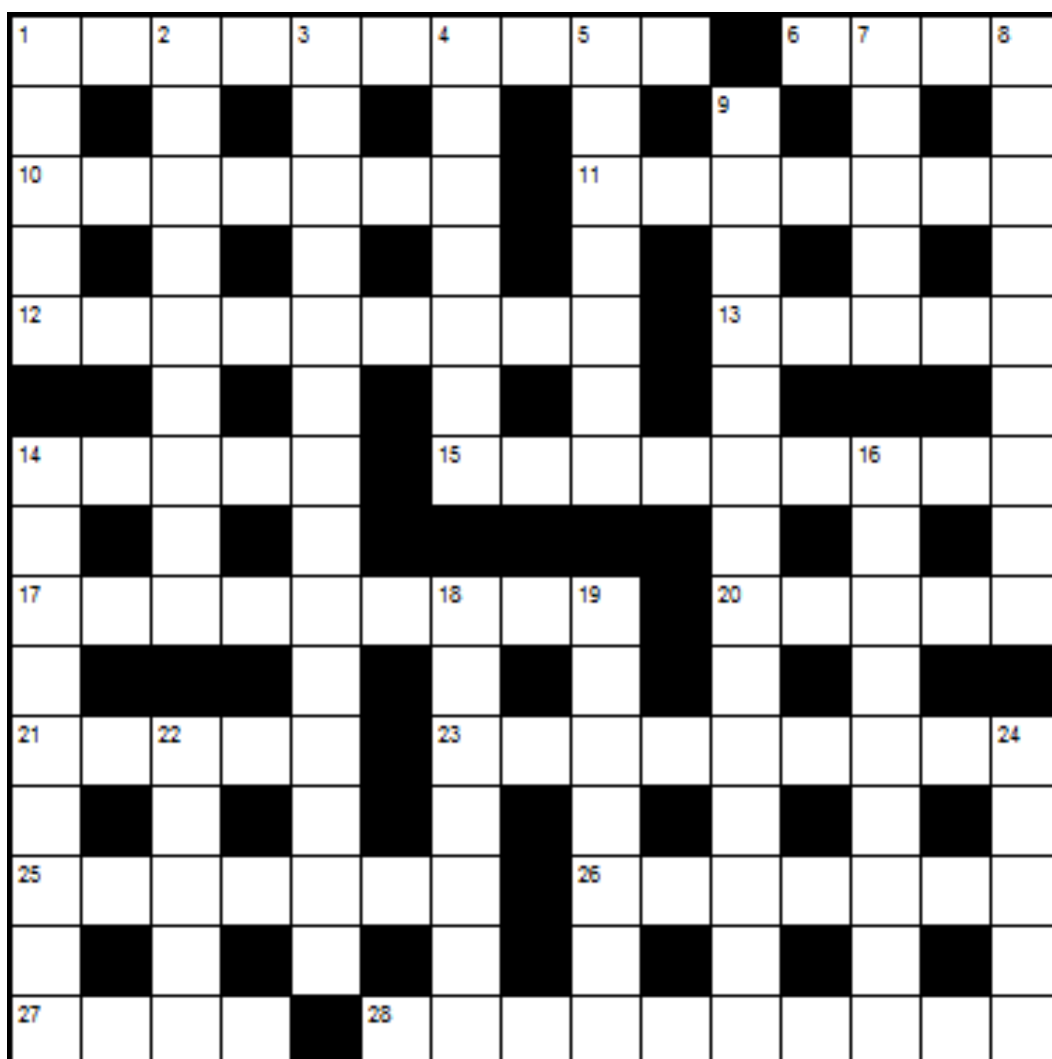
Krüptilise stiili sõnapuslesid on tavaliselt kasutatud põgenemistubades³ ja MIT-i mõistatuste jahis⁴ (Fragar, 2018). Vastused seda tüüpi pusledel on hästi määratletud ning neid läheb vaja pusle lahendamiseks. Keeruliseks teeb lahendajale see, et vihjed on erinevalt mõistetavad ning võib võtta aega, kuni jõutakse õige vastuseni (*ibid.*).

								Netja-päev		3 x	Ruumi-meeter	Maa-konna-leht	Hapnik	Ajaleht Eesti ...	Üha, pidevalt
								WASTUS		vokaal					
								Maailma-jagu							
								Võlu, nõidus					Sugulane Kohale-tooja		
								Küllap			Ava selinas Uudiste-saade				
								Esimene täht		Palkneva Audi mudel					
								Maakon-naleht							Eesti luuletaja
								Tonn							
												Eesti luuletaja ... Maria			

vihjetega koostatud ristsõnaga. Vihjete alla kuuluvad näiteks anagrammid, sõnamängud, ümberpööratud jm. Seega saavad blokeeritud ruudustikuga, võrega ja temaatilised ristsõnad olla krüptiliste ristsõnade alamliigiks (vt joonis 1), olenevalt kuidas ristsõna loodud on (*ibid.*).

1.2.2 Blokeeritud ruudustikuga ristsõnad

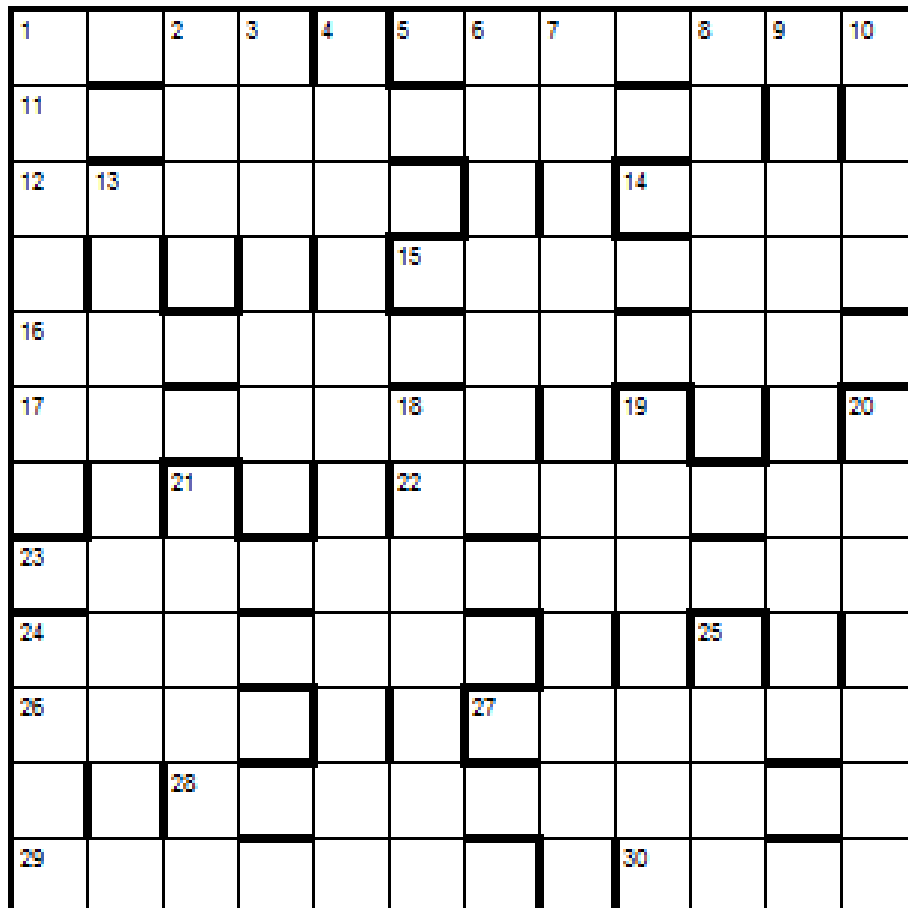
Blokeeritud ruudustiku ristsõnad (ingl *Blocked grid crossword*) on kõige tavalisemad (Fragar, 2018). Sellelaadsed ristsõnad tunneb ära ruudustiku põhjal, kus sõnade lõpud ja sõnade vahelised tühjad alad on märgitud tumeda värviga täidetud ruutudega (vt joonis 2). Tavaliselt on ruudustiku suurus 15 ruutu korda 15 ruutu, kuid mitte alati, ning nendesse ristsõnadesse mahub umbes 30-32 sõna, olenevalt sõnade pikkusest (*ibid.*).



Joonis 2. Blokeeritud ruudustikuga ristsõna näide (Henderson, 2012).

1.2.3 Võrega ristsõnad

Võrega ristsõnad (ingl *Barred grid crosswords*) erinevad eelnevast ristsõnast selle poolest, et täidetud ruutude asemel on sõnade alguste ja lõppude vahel olevad jooned tumendatud, millest tekib võre (vt joonis 3) (Fragar, 2018). Seda tüüpi ristsõna standardsuurus on 12 ruutu korda 12 ruutu, kuid on võimalik ka suuremaid luua. Kuna sõnad on paigutatud väga tihedalt kokku, siis standard suuruses pusle tavaliselt koosneb 32-36 vastusest (*ibid.*).



Joonis 3. Võretatud ristsõna (Henderson, 2012).

Kui võrrelda võrega ristsõnasid blokeeritud ruudustikuga ristsõnadega, siis esimesel on sõnavara segasem ja on tavaline, et selleks on antud viide või fraas. Teise puhul on tihti antud seletus ehk definitsioon ning küsitakse vastet (Fragar, 2018).

1.2.4 Temaatilised ristsõnad

Temaatilised ristsõnad, nagu nimigi viitab, on mingi kindla valdkonna põhised (Amazing Game Room, 2019). Need on tavaliselt võre tüüpi ristsõnad, kuid mitte kindlalt ruudustikus. Kujusid on erinevaid: ristkülik, ring, L-kujuline ja muud. Temaatilisi ristsõnu tehakse tihti

noorematele lastele, kuna mõistatuste kujud on neile atraktiivsed. Kindla kujuga ristsõnu on välja mõeldud ka vanematele lahendajatele. Sedalaadi ristsõnu on võimalik teha erinevate kujundustega, kuid see peab ikka järgima ristsõna reegleid. Kuna iga sõna peab olema ühenduses tervikuga, siis see võib lisada keerukust ristsõnade loomisel (*ibid.*).

1.2.5 Koodsõna ristsõnad

Koodsõna ristsõnade puhul on tegu ruudustikuga, kus on antud iga ruudu sees arv ning sellele arvule vastab kindel täht (Fragar, 2018). Iga täht, mis esineb vähemalt ühe korra, omab kindlat numbrit, olenevalt tähestiku suurusest (*ibid.*).

CODEBREAKER WORD PUZZLE

The same number represents the same letter. Crack the code and fill the grid.
To help you get started some words are already in place.

1		2		3		4	5	6	7	8	9	5		10		11				
12	10	13	14	8	7	6		15		6		7		7	A	P	14	4	6	L
	1		4		16		5	13	1	1	10	14	17		7		8			
17	13	14	17	6	18	5		7		19		16	4	17	3	19	18	8		
	6		4		5	13	14	14	18	7	6		17		4		20			
1	7	4	8	17		8		18		21		1	6	7	17	18	7	13		
	14		20	4	7	8	17		14	18	11	13	18		22		14			
20		5		8		22		5		17		1		14		1		4		
14	18	8	10	2	8		3	4	11	4	6	4	23	7	17	4	10	8		
18		4		7		5		20		3		6		4		1		9		
18	24	1	18	14	4	25		18	8	17	7	6		10	8	6	4	8	18	
8		18		18		4			4		6		21		21		17		24	
	26		5		3	6		4	26	26		17	14	10	10	1	17			
25	4	24	17	13	14	18		4		1		10		2	19	18	7	17		
	20		13		10		10	3	17	7	20	10	8		22		8			
5	13	8	9	10	2	8		7		14		25	18	5	5	7	20	18		
	14		18		8	13	7	8	3	18	5		5		4		6			
5	18	8	8	7		17		3		8		5	17	4	3	16	18	14		
	5		17		7	5	1	18	3	17	5		5		5		9			

1	2	3	4	5	6	7	8	9	10	11	12	13
14	15	16	17	18	19	20	21	22	23	24	25	26

ANSWER: 1=P, 2=W, 3=C, 4=I, 5=S, 6=L, 7=A, 8=N, 9=D, 10=O, 11=V, 12=J, 13=U, 14=R, 15=Q, 16=K, 17=T, 18=E, 19=H, 20=G, 21=B, 22=Y, 23=Z, 24=X, 25=M, 26=F.

Joonis 4. Koodsõna ristsõna näidis (Ratselmeister, s.a.).

Lahendamiseks on mõned tähed alustamiseks antud ja lahendaja peab arvama antud tähtede alusel koodi, mis aitaks ristsõna ära lahendada (vt joonis 4). Kui lahendamisel tekib raskusi, on olemas ka abistavad vihjed (Fragar, 2018).

1.2.6 Sõna otsimise ristsõnad

Sõna otsimise ristsõnad (vt joonis 5) kujutavad endast ruudustikku, mis on tähtedega täidetud, ning lahendaja peab seal leidma sõnad (The360reporters, 2022). Sõnade paigutus võib olla horisontaalselt, vertikaalselt kui ka diagonaalis, nii vasakult paremale kui ülevalt alla ning vastupidi (*ibid.*).

Name: _____ Date: _____

Kindness Crossword

R	K	I	N	D	N	E	S	S	S	J	G	J	Y	L	Z	B
J	Q	Q	O	T	H	O	U	G	H	T	F	U	L	I	Z	B
T	E	P	Q	H	S	E	N	S	I	T	I	V	I	T	Y	S
N	E	O	N	O	I	T	A	R	E	D	I	S	N	O	C	U
J	E	L	O	E	D	Z	P	A	T	I	E	N	C	E	U	O
G	S	I	S	Z	G	E	M	P	A	T	H	Y	G	N	U	R
E	J	T	N	K	Q	N	G	V	C	K	V	H	M	A	P	E
N	M	E	Q	S	R	K	I	O	M	X	K	J	W	G	G	N
E	E	C	A	R	G	O	M	G	T	L	R	T	N	V	V	E
R	C	V	G	J	Q	P	Q	G	A	D	G	I	B	B	B	G
O	P	A	K	O	A	I	V	E	G	R	R	F	J	M	U	C
S	T	I	Y	S	O	H	L	H	I	A	U	N	X	S	A	J
I	K	R	S	C	G	D	L	M	C	H	T	O	W	R	W	I
T	D	I	A	R	R	H	W	V	B	C	Q	E	C	W	A	T
Y	O	S	Q	E	K	E	J	I	K	C	E	H	R	N	R	T
N	H	E	Y	K	H	T	R	G	L	B	O	R	B	N	E	I
T	O	P	O	P	R	U	I	T	N	L	L	Z	Z	N	K	B

Polite
Heart
Sensitivity
Thoughtful
Compassion

Goodwill
Generous
Grace
Consideration
Patience

Encouraging
Empathy
Caring
Generosity
Kindness

Joonis 5. Sõna otsimise ristsõna lahkusega seotud sõnadel (WordMint, 2017).

Sõna otsimise ristsõnad on parimad keele õppimiseks ning neid on kasutatud sõnavara kinnistamiseks koolides väga pikalt (The360reporters, 2022).

1.3 Ristsõnade genereerimise viisid

Projektid, mis käsitlevad ristsõnade genereerimist, on lahendanud probleemi, kuidas luua ristsõna otsitava sõna ning seletuse paari, sedasi, et vaadatakse olemasolevatest sõnaraamatutest sõna definitsioone ning sellest kujunebki programmi võti (otsitav sõna, definitsioon), mille põhjal hakatakse ristsõna looma. Samuti vaadatakse sünonüümsõnastikke või analüüsitakse veebist saadud lauseid, millest pannakse kokku paar, mida kasutada alusena. Ristsõna raamistik on enamasti lahendatud algoritmidega, mis sobitavad sõnad tähthaaval ruudustikku, kuid leidub ka selliseid, mis lisavad sõnad ruudustikku sõnahaaval, kasutades selleks eraldi metoodikat, milleks on sõnade sobitamine.

Rigutini jt. teada esimene täielikult automaatne ristsõna genereeriija loodi aastal 2008 (Rigutini, Diligenti, Maggini, & Gori, 2008). Nende töö eesmärk oli luua programm, mis suudab kasutajate vahelesegamiseta koostada ristsõnu, kasutades infoallikadena eelseatud veebilehtesid nagu itaaliakeelse Wikipedia lehed, sõnastikud jms. Programmi töö jooksul laetakse nendelt lehtedelt alla definitsioonid, kasutades loomuliku keele töötlust ehk NLP (ingl *Natural Language Processing*) meetodeid. Edasi kasutatakse tingimustele kohanduvat programmeerimist (ingl *Constraint Satisfaction Programming*), mille abil paigutatakse internetist saadud otsitavad sõnad ristsõna lauale ehk kastidesse. Lisaks toimub kogu protsess sõnahaaval, mitte tähthaaval. See tähendab, et võetakse veerg või rida ning otsitakse regulaaravaldiste abiga, milline sõna antud vahemikku sobib. Tähthaaval sobitamise puhul alustatakse esimesest lahtrist ning tähtedest käiakse iteratiivselt üle, vaadates, mis sõnad sobivad veel ning kas lahend on juba leitud, et käesolev otsing lõpetada ja alustada uuega. Varasemalt on samuti üritatud teha sõnahaaval protsessi, kuid siis inimeste poolt üle vaadatud sõnastike baasil (*ibid.*).

Proverb (Keim, et al., 1999) on üks selline programm, mis kasutab NLP võtteid vihje probleemi lahendamiseks, töötades eelsätestatud teadmiste baasil, milleks on enamasti sõnade kogumik. Teine näide on WebCrow (Ernandes & Gori, 2005), mis üritab Proverbist parem olla, võttes andmed otse veebist, pannes otsitava sõna otsingumootoris ja seejärel töödeldakse vasted NLP abil, millest tekib võimalike vastuste loend.

Rigutini jt. töö katsetusel saadi tulemuseks, et õigeid definitsioone oli 81%, valesid 3% ja õigeid, kuid kasutamatu 16%. Autorite järgi on selle meetodi suurimaks probleemiks, et suurem protsent sõnu on semantiliselt poole pealt korrektsed, kuid ristsõnas neid kasutada ei saa, sest nad on liiga pikad, sisaldavad väljendeid, mis on liiga lähedased vastusele, või on liiga mitmetähenduslikud (Rigutini, Diligenti, Maggini, & Gori, 2008).

2. Ristsõnade automaatse genereerija rakendus

Käesolevas peatükis tutvustatakse veebirakenduse tööks vajaminevaid teeke, programmi erineva staadiumi algoritme, testimistulemusi ja edasiarendusvõimalusi.

Rakendus kasutab eestikeelseid sõnu, et genereerida ristsõnu. Töötlemisele minevad sõnad saadakse tekstifailist, kuid protsessist räägitakse rohkem peatükis „Andmete eeltöötlus“. Rakendus on programmeerimiskeele Python⁵ baasil loodud ning peamised vajaminevad teegid on samuti mõeldud Pythoni keskkonnas töötama. Kuna projekt vajab eraldi keskkonda allalaetavate moodulite korrektseks käivitamiseks, eeskätt EstNLTK⁶ töötamiseks, laeti alla programmi arendamiseks Anaconda⁷, milles sai luua virtuaalkeskkonna, kuhu moodulid installida.

Kuigi ristsõna rakenduse baas on ülesehitatud Pythoniga, on ülejäänud veebirakenduse loomise pool ehk kasutatakse HTML⁸ (*HyperText Markup Language*) ja CSS⁹ (*Cascading Style Sheets*) keeli programmi ülesseadmiseks ning kujundamiseks. Pythoni ja HTMLi ühendamiseks kasutatakse eraldiseisvat veebiraamistikku nimega Flask¹⁰, mille abil käivitatakse kogu programm.

Olemasoleva rakenduse loomisel oli vaja ka valida, mis tüüpi ristsõnasid see genereerima hakkab. Selle töö puhul sai valitud blokeeritud ruudustikuga ristsõnade genereerimine, kuna nende kuju on kõige lähedasem standardsele ristsõnale, mida kohatakse ka ajalehtedes, lisaks on seda lihtsam luua, sest ristumiskohti on vähem.

2.1 Kasutatud teegid

Rakenduse töötamise tagamiseks on mõned tähtsamad teegid, mida peab eraldi installima, kuna neid ei ole Pythoni baaskoodis. Nendeks on EstNLTK, Estonian Wordnet, Pandas ja Flask. Järgnevalt kirjeldatakse igat teeki lühidalt.

⁵ Pythoni kodulehekül: <https://www.python.org/>

⁶ EstNLTK respositoorium: <https://github.com/estnltk/estnltk>

⁷ Anaconda kodulehekül: <https://www.anaconda.com/>

⁸ HTML õpetused: <https://developer.mozilla.org/en-US/docs/Web/HTML>

⁹ CSS õpetused: <https://developer.mozilla.org/en-US/docs/Web/CSS>

¹⁰ Flaski kodulehekül: <https://flask.palletsprojects.com/en/2.1.x/>

2.1.1 EstNLTK

EstNLTK (ingl *Estonian Natural Language ToolKit*) on peamiselt Pythonis kirjutatud kogumik teek eestikeelsete tekstide tötluseks, mis on loodud Tartu Ülikoolis, projektijuhiks Sven Laur. Bakalaureusetööna loodava rakenduse korrektseks toimimiseks läheb vaja EstNLTK'sse kuuluvat *Estonian Wordnet*¹¹ (EstWN), mida kasutatakse sõnade definitsioonide tagastamiseks (Laur et al., 2020). Kahjuks puudub EstWNs mõningatel sõnadel definitsioon, mis raskendab programmi tööd.

2.1.2 Pandas

Pandase areng sai alguse aastal 2008 AQR Capital Management¹² poolt, mis on ülemaailmne investeringute haldamise ettevõtte. Läks aasta edasi ja pandase lähtekood muutus avatuks, mida on tänapäevani edasi arendatud erinevate kaastööliste panusel. Aastal 2015 muutus pandas *NumFocusi* poolt sponsoreeritud projektiks. See tähendab, et pandas on tänu sellele maailmatasemel arendusse võetud ning liitunud maailmaklassis olevate avatud lähtekoodiga projektidega (The pandas development team, 2020).

Põhjus mooduli kasutamiseks andmetötlusel on kiire ja tõhusa andmeraamistiku (ingl *DataFrame*) objekti olemasolu, mis suudab andmeid töödelda ka integreeritud indekseerimisega. Lisaks omab teek tööriistu, mis laseb lugeda ja kirjutada andmestruktuure erinevate vormingutega: CSV ja tekstifailid, Microsoft Excel, SQL andmebaasid ja kiiret HDF5 formaati. Samuti on võimalik andmekogumikke lihtsalt üheks koguks luua ning programm on kõrgelt optimeeritud jõudlusele.

Käesolev projekt kasutab pandase andmeraamistiku objekti, et luua sõnabaas, milles hoiustada alla laaditavate sõnade andmeid.

2.1.3 Flask

Flask on mikro veebiraamistik, millega on lihtne Pythonis luua veebirakendusi. Nimetus mikroraamistik tuleneb sellest, et ta tuum ei ole mahukas ning raamistik on väga lihtsasti laiendatav ehk teda on lihtne muude moodulitega integreerida. Seda tänu mooduli mitmekülgse rakendamisevõime tõttu, sest Flask ei tee kasutaja eest otsuseid, et mis andmebaasi kasutada, vaid see jääb programmeerija hooleks. Isegi valik, millist mallimootorit (ingl *template engine*)

¹¹ Estonian Wordneti koduleht: <https://www.cl.ut.ee/ressursid/teksaurus/>

¹² Pandase arenduse alustajad: <https://www.aqr.com/>

kasutada, on võimalik lihtsalt muuta. Flaski puhul on võimalik jätta enamus kasutaja otsustada, kuna moodul pakub integreerimisvõimalusi mitmete erinevate laiendustega nagu andmebaasi lahenduse valikud, vormi kinnitamine, üleslaadimise käsitlemine, erinevad avatud autentimistehnoloogiad ja muud (Pallets, s.a.).

Mallimootor on midagi, mida kasutatakse veebirakenduste loomisel. See aitab kiiresti kokku panna veebirakenduse, mis koosneb erinevatest komponentidest, nagu keha, navigeerimine, jalus, armatuurlaud jne. Mallid võimaldavad serveri poolsete andmete kiiret renderdamist, mis edastatakse rakendusele. Flaski puhul on vaikimisi Jinja2 mallimootor (Onyejiaku, s.a.).

2.1.4 Re

Re puhul on tegu Pythoni baaskoodis asetseva mooduliga, mis tegeleb regulaaravaldistega. Enamasti pakub käesolev teek regulaaravaldiste sobitusoperatsioone. Re mooduli tööprotsessil regulaaravaldis täpsustab sõnade komplekti, mida teegi funktsioonide abil on võimalik kontrollida, kas kindel sõne sobitub regulaaravaldisega (Python Software Foundation, 2022). Selle programmi puhul on read ja veerud ristsõnas regulaaravaldisele etteantavad sõned, mille sobitamise käigus leitakse ning tagastatakse kõik ristsõnasse sobivad valikud.

2.2 Andmete eeltöötlus

Käsitletavas töös on ristsõna baasiks, kust programm oma vajalikud sõnad ning nendele vastavad definitsioonid saab, Eesti kirjakeele sagedussõnastiku 10 000 sagedamat lemmat¹³ ja EstWN. Algse andmetöötluse tulemuse saavutamiseks kasutati Pythonis olevat moodulit nimega pandas, mille abil on lihtne luua andmeraamistikku sõnabaasile. Saadud andmed käidi üle ning sealt otsiti, kas ebasobivaid sõnu nende hulgas leidub, kuid autori arvates internetileheküljelt saadud tekstfailis ebasündsad sõnad puudusid. Kuna pole oluline, kas sõna on kasutusel ilukirjanduses või ajalehes, kustutakse sellekohane veerg andmetest ning pärast andmeraamistiku loomist järjestatakse sõnad esinemisjärjekorda, mille abil oleks hiljem programmil lihtsam eraldada eri tasemete vahel käsitletavaid sõnu.

Andmeid töödeldakse iga kord, kui minnakse valminud veebilehe vahelehele „Ristsõna“. Sealses protsessis töödeldakse enne leheküljeni jõudmist andmed ära, mille järgnedes kuvatakse vahelehel ristsõna elementide sisu.

¹³ 10 000 sagedamat lemmat: <https://www.cl.ut.ee/ressursid/sagedused/index.php?lang=et>

2.3 Veebirakenduse algoritmide loogika

Rakenduse algoritmid töötavad enamasti tagaplaanil. Enne ei renderda veebileht sisu, kui sellele vahelehele liikumas pole. Vajutades menüüst uuele vahelehele, genereeritakse jooksvalt nii veebilehekülje sisu kui ka ristsõna puhul tehakse valmis ristsõna laud ja muud andmed, mida vahelehe loomisel kasutatakse info kuvamiseks. Info siinses mõttes on vihjed, ristsõna ise ja ka lahendus. Veebirakenduse HTML faili koodis on enamasti kasutatud Jinja2-e¹⁴ mallina, mille abil väljastati tsükleid kasutades veebilehe sisu nii vihjete kui ka ruudustike puhul.

2.3.1 Ristsõna algoritm

Ristsõna algoritm põhineb re mooduliga rea või veeru kokku sobitamisel olemasoleva sõnabaasiga. Võetakse ette rida regulaaravaldisena, mille puhul kasutatakse `re.search()` funktsiooni sõnastiku peal ning lisatakse kõik sobivad sõned uude listi, mis tagastatakse ja mille seast `random.choice()` funktsiooniga tagastatakse suvaline sobiv sõne. Saadud sõna lisatakse sõnastikku andmetega: alguskoordinaadid matriksis ehk laual, sõne suund, järjekorra number ja küsimus, mille põhjal sõne ära arvata. Seda korratakse seni kuni rohkem sõnu ritta ei mahu või neid ei leidu. Kui kõige lõpuks ikka ei leita vastet, tagastatakse tühi list ning minnakse edasi järgmisele reale või veerule, olenevalt kaugel ollakse.

Algoritmi järgmine etapp on paika panna esimene veerg, kasutades sama metoodikat nagu tehti esimese rea jaoks. Kui esimene rida ja veerg on täidetud, korratakse esimest etappi, kuid jättes ühe rea vahele. Jätkatakse kuni read on täidetud ning alustatakse veergude täitmisega. Veergude puhul korratakse täpselt sama, mis juba tehtud, jättes eelnevale veerule ühe rea vahele ning tsüklis käiakse läbi, kuni rohkem ei ole lauale enam lisada. Sellega lõppeb ristsõna kõige rohkem ressursse võttev protsess.

¹⁴ Jinja tutvustus: <https://palletsprojects.com/p/jinja/>

2.3.2 Veebirakenduse alus

Projekti alusraamistikuks on flask, mille alusprojekt saadi PyCharmi¹⁵, mis on Pythoni IDE (ingl *Integrated development environment*), uue projekti loomisel. HTML koodi jooksutatakse flaskis meetodiga `render_template('veebilehe_nimi.html')`, mille tulemusel tagastatakse veebileht, mida flaski rakendus kuvab.

Selle valmides otsiti viisi, kuidas rakendus jõuaks kasutajani sedasi, et ta ei vajaks programmide alla laadimist ning otsingute tulemusel jõuti järelduseni, et Heroku¹⁶ on keskkond, mis lahendab serveri probleemi. Seda põhjusel, kuna veebirakenduse üleslaadimisel Heroku keskkonda, luuakse veebisait projektist, mida on seejärel võimalik kasutada ilma rakenduse allalaadimiseta.

2.4 Kasutusjuhend

Käesolevas peatükis tutvustatakse kasutusjuhendit veebirakenduse kasutamise jaoks.

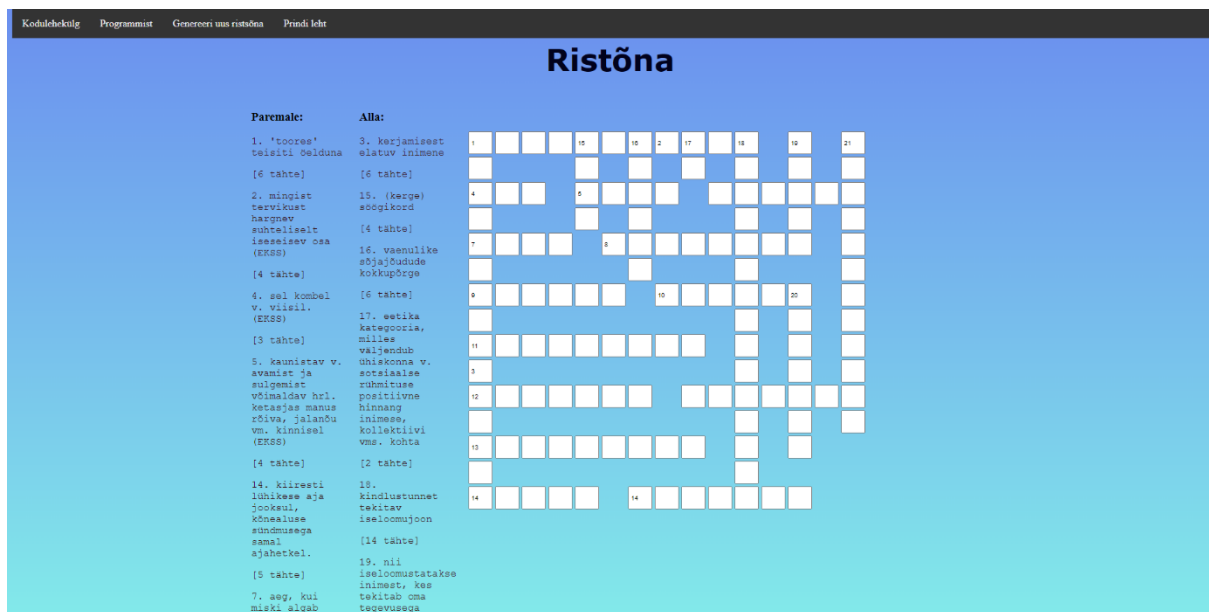
Veebirakenduses alustatakse <https://ristsona-genereerija.herokuapp.com/>, kust on võimalik edasi liikuda programmi tutvustavale leheküljele „Programmist“ ning „Ristsõna“ leheküljele, kus asub põhiprogramm. Koduleheküljel on ka väljatoodud, et kuhu on vaja liikuda ristsõna lahendamiseks.

Ristsõna lehekülje laadimine võtab aega, kuna programm genereerib ristsõna, mida kuvada kasutajale. Uus ristsõna genereeritakse alati, kui vahelehte uuendatakse. Leheküljel on võimalik ka genereerida uus ristsõna, vajutades menüüst „Genereeri uus ristsõna.“. Vajutades menüüs valikule „Prindi leht“, on võimalik ristsõna välja printida või printeri puudumisel salvestada PDF-formaati.

Ristsõna on võimalik veebis täita, selleks on vaja aktiveerida ruut sellele peale toksates telefonis või hiirega vajutades arvutis (vt joonis 6). Kui kast on aktiivne, on võimalik lisada tähti.

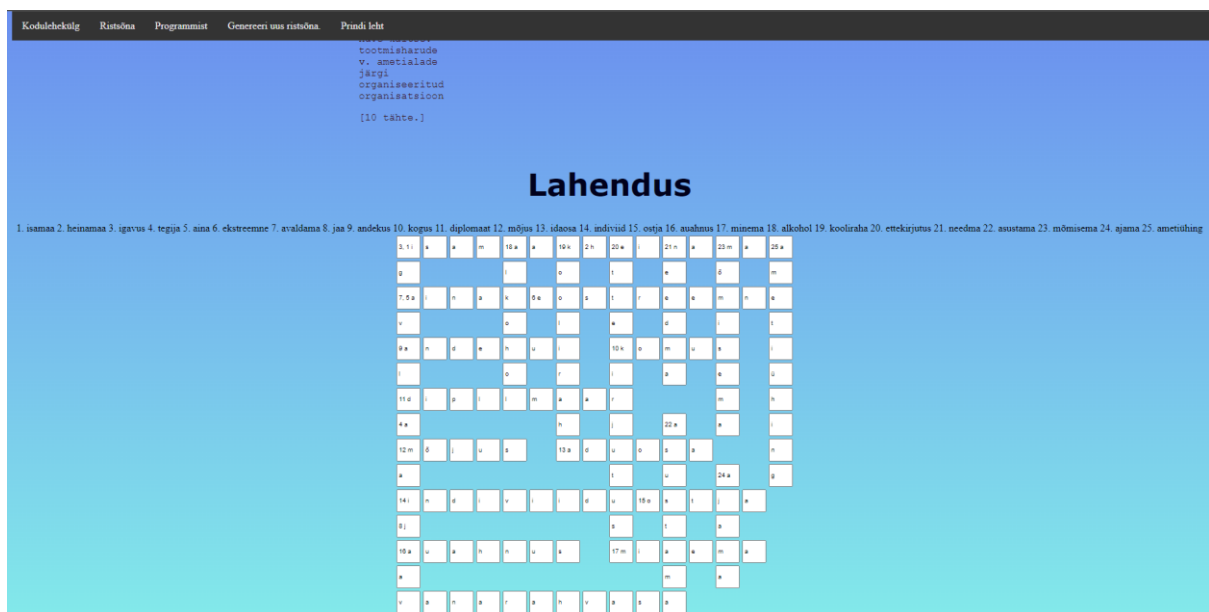
¹⁵ PyCharm'i veebisait: <https://www.jetbrains.com/pycharm/>

¹⁶ Heroku: <https://dashboard.heroku.com/>



Joonis 6. Veebirakenduse „Ristsõna“ vahelehel asuv ristsõna.

Vahelehel „Ristsõna“ on lahendamiseks ristsõna koos vasakul asuvate küsimustega. Sama lehe allosas on võimalik näha vihjeid ning lahendust (vt joonis 7).



Joonis 7. Veebirakenduse „Ristsõna“ vahelehel asuv lahendus.

Telefoniga on võimalik veebilehel tegutseda, kuid programmi hetke versioon ei ole selleks mõeldud ning seepärast tal ka puuduvad nutiseadmetele mõeldud kujundus elementide toeks vajalikud stiilid ning klassid.

2.5 Testimine ja tulemused

Projekti testimine toimus jooksvalt kogu arendamisprotsessi vältel. Programmi valmides ning veebi üles laadides tehti täielik kontroll, kus vajutati üksikshaaval igale lingile, tehes kindlaks, et veebilehtede vahel liikumine ei ole piiratud ning ei tagasta tõrkeid. Saades kodulehekülje ja programmi tutvustava lehekülje kontrollitud, jätkati „Ristsõna“ vahelehega. Sellel lehel olles testiti olemasolevaid funktsionaalsusi, milleks on ristsõna lünkade täitmise võimalus ning kas ristsõna välja printimine toimib, printeri puudumisel selle salvestamine .pdf formaati. Samuti kontrolliti üle uue ristsõna genereerimise link.

Programm laeti üles Heroku repositooriumisse. Tänu sellele sai tööd lasta testida teistel inimestel veebis, ilma et nad peaksid projekti hakkama alla laadima ning installima kolmanda osapoole rakendusi.

Programmi proovisid lisaks autorile ja juhendajale veel 2 inimest. Testimisel tutvustati programmi ning paluti katsetada erinevaid funktsionaalsusi, sealhulgas esmakordse kasutussõbralikkuse kohta ning pärast testimist välja tuua puudused.

Esimese testija poolt saadi tagasisidet, et vihjetel võiks juures olla, kui pikk küsitav sõna on, kuna ruudustikus ei ole see alati arusaadav. Teine testija tõi välja, et ristsõna välja printimisel võiks „Lahendus“-e osa alata uuelt lehelt, see kindlustab, et printimisel ei satu lahendus poolikult lehe lõppu ja järgmise algusesse.

Ristsõna algoritmi testimisel tuli välja, et pärast horisontaalset sõnade sobitamist on vertikaalselt sõnade sobitamine mõnikord vale, kuna sobivate sõnade otsimisel tuleb valepositiivseid, mis kirjutavad ruudustikus juba olemasoleva tähe üle ja teevad ristsõna osaliselt lahendamatuks. Ehk funktsioon tagastab valepositiivse selle suhtes, kas sõna sobib ritta või veergu. See-eest toodi testimise tulemusena välja, et rakendus on selgelt mõistetav, kuna juhendid ja menüüs asuvad lingid on selgelt mõistetavad.

2.6 Edasiarendusvõimalused

Programmil on mitmeid erinevaid võimalusi, kuhu suunas saab seda edasi arendada. Esimene arendusvõimalus on lähtuvalt testimise tulemusest sõnade sobitamise algoritmi muutmine, et regulaaravaldis ei annaks valepositiivseid.

Tänu sõnabaasi suurele mahule on edaspidi võimalik välja töötada erineva raskusastmega ristsõnu, mille jaoks saaks küsida TOP100, TOP200 jne sagedussõnastiku järgi sõnu. Lisaks

saab küsida sõnu teemade kaupa, et rakendust saaks kasutada ka õppetöös sõnade ja nende tähenduste õppimiseks nagu seda on tehtud võõrkeeles ainetes ning terminite õppimiseks.

Täiendada saab ka vihjete süsteemi, kui asendada lahenduse näitamine väikeste vihjete tagastamisega, kui lahendaja seda küsib. Hetkel küsitakse definitsioone, definitsiooni puudumisel teise lähima lemma kaudu ning selle puudumisel antakse pool sõna ette. Vihjete variatiivsust saab täiendada, andes lahendajale sünonüüme, vastandsõnu, lünklausete põhjal lahendi arvamist või pakkudes ühte tähte vastusest.

Edasiarenduseks on ka ristsõna ruudustiku uuendamine. Asendada hetkel olev ruudustiku kujundus kompleksemaga, kus ei ole lahtrite vahel selliseid suuri vahesid või moodustab huvitavama kuju. Võimalus on lisada ka funktsionaalsus, mis laseb ristsõna laua suurust muuta. Hetkel on listina pandud standardsuurused ning kasutatakse kõige väiksemat.

Lisaks on võimalik juurde panna punktisüsteem, mis võib olla seotud lahendamiseks kulunud ajaga või ka ajast sõltumatu. Siin on kaks varianti: üks, kus on teatud ajavahemik, mille jooksul peab ristsõna ära lahendama, ning teine, kus loetakse lahendamiseks kulunud aega.

Kokkuvõte

Töö käigus uuriti ristsõnade ajalugu, leiti nende vajalikkus nii õppetöös kui ka tavaelus ning milliseid erinevaid ristsõnade tüüpe on ja kuidas neid genereeritakse. Selgus, et standardne ristsõna kuju on algoritmiliselt kõige lihtsam genereerida. Eeltööle järgnes veebirakenduse loomine, mis suudab genereerida ristsõnu, mida on võimalik lahendada nii veebis kui ka välja printida paberkujul lahendamiseks. Käesoleva töö eesmärk sai täidetud, loodi interaktiivne veebirakendus, mis suudab automaatselt ristsõnu genereerida eesti keele baasil, kasutades küsimuste moodustamiseks eesti keele keeleressursse Estonian Wordneti põhjal.

Toetudes programmi genereerimise tulemustel, on pigem valminud prototüüp, millel on mitmeid edasiarendusvõimalusi. Seda just funktsionaalsuse variatiivsuse seisukohast, kuna on võimalik luua ajast sõltuv või sõltumatu punktisüsteem. Lisada saab ka vihjepakkuja, mille toimimiseks pakuti töös mitmeid võimalusi. Eelkõige peaks parandama sõnasobitamise algoritmi sedasi, et ta enam valepositiivseid variante ei tagastaks, mille tulemusena muutuks töö prototüübi staatus valmis veebirakenduseks. Ristsõna saab raskemaks teha, kui küsida sõnu valdkondade kaupa, kasutada vähemsagedasi sõnu või muuta ristsõna kuju.

Programmi koostamisel oli kõige raskemateks osadeks, et ristsõna loodaks sedasi, et kordustsüklid maatriksi piiridest välja ei läheks ning tagastaks õige sõna ruudustikus õigesse kohta. Tööd tuleb teha algoritmi ühe osaga, kus valitakse miskipärast ebasobivaid sõnu ristsõnasse, mille tulemusel kirjutatakse olemasolev ristsõna üle. Selle parandamisel muutuks veebirakenduse ristsõna lahendatavaks.

Üks töö etappidest oli leida sobiv veebikeskkond, kuhu saaks projekti üleslaadida ning nii-öelda serverisse panna, mille tulemusel rakenduse kasutajad ei peaks tervet projekti ja muid kolmanda osapoole rakendusi oma arvutisse tõmbama, vaid saaksid seda rakendust kasutada otse veebis. Uuriti erinevaid variante ning otsustati Heroku kasuks.

Autor on tööga enamasti rahul, kuna programm koostab ruudustiku, valib sõnad ja kuvab koos vihjetega salvestataval ning välja prinditaval veebilehel. Kuigi praktilises osas on jätkuvalt vigu, parandati mitmed testimise käigus välja toodud aspektid. Samuti on eesti keelsete ristsõnadega algust tehtud ning edaspidiseks jätkamiseks on olemas alguskoht.

Viidatud kirjandus

- Abuelo, A., Castillo, C., & May, S. A. (1. jaanuar 2016. a.). Usefulness of Crossword Puzzles in Helping First-Year BVSc Students Learn Veterinary Terminology. *Journal of Veterinary Medical Education*, 43(3), 255-262. doi:10.3138/jvme.0915-149R
- Amazing Game Room. (11. aprill 2019. a.). *6 Different Types of Crossword Puzzles - Do You Know Them All?* Kasutamise kuupäev: 13. aprill 2022. a., allikas Amazing Game Room: The Amazing Resource for MTG (Magic the Gathering): <https://amazinggameroom.com/board-games/different-types-of-crossword-puzzles/>
- Bellis, M. (s.a.). *Ristsõnajalgade ajalugu*. Kasutamise kuupäev: 1. jaanuar 2022. a., allikas EFerit: <https://et.eferit.com/ristsonajalgade-ajalugu/>
- Eesti Ajalehtede Liit. (2008). *Eesti Ajalehtede Liit*. Kasutamise kuupäev: 19. aprill 2022. a., allikas Eesti Ajalehtede Liit: <http://vana.meedialiit.ee/ajaleht/ajaleheleht-13.php?c>
- Ernandes, M., & Gori, M. (2005). WebCrow: a WEB-based system for CROssWord solving. *American Conference on Artificial Intelligence (AAAI-05)*, (lk 1412-1417). Allikas: <https://www.aaai.org/Papers/AAAI/2005/AAAI05-224.pdf>
- Est. (2020). EstNLTK 1.6: Remastered Estonian NLP Pipeline. *LREC 2020* (lk 7152-7160). Marseille, France: European Language Resources Association. Kasutamise kuupäev: 10. mai 2022. a., allikas <https://aclanthology.org/2020.lrec-1.884>
- Fragar, M. (31. juuli 2018. a.). *Do you know about these different types of Crossword Puzzles?* Kasutamise kuupäev: 13. aprill 2022. a., allikas Wealth Words: <https://www.wealthwords.com/blog/what-are-the-different-types-of-crossword-puzzles/>
- Gaikwad, N., & Tankhiwale, S. (9. november 2012. a.). Crossword puzzles: self-learning tool in pharmacology. *Perspectives on Medical Education*, 1(5), 237-248. doi:10.1007/s40037-012-0033-0
- Henderson, P. (1. oktoober 2012. a.). *Types of crossword*. Kasutamise kuupäev: 19. aprill 2022. a., allikas Phi Crosswords: <https://phionline.net.nz/types-of-crossword/>
- Keim, G. A., Shazeer, N. M., Littman, M. L., Agarwal, S., Cheves, C. M., Fitzgerald, J., . . . Weinmeister, K. (1999). *PROVERB: The probabilistic cruciverbalist*. Kasutamise kuupäev: 10. jaanuar 2022. a., allikas AAAI-99 Proceedings: <https://www.aaai.org/Papers/AAAI/1999/AAAI99-101.pdf>
- Mazlack, L. J. (21. veebruar 2003. a.). *Computer construction of crossword puzzles using precedence relationships*. doi:[https://doi.org/10.1016/0004-3702\(76\)90019-9](https://doi.org/10.1016/0004-3702(76)90019-9)

- Novaator. (21. detsember 2013. a.). *Eesti Rahvusringhääling*. Allikas: Novaator:
<https://novaator.err.ee/253057/21-detsembril-1913-ilmus-ajakirjanduses-uks-esimesi-ristsonamoistatusi>
- Onyejiaku, T. K. (s.a.). *What are template engines?* Kasutamise kuupäev: 10. mai 2022. a., allikas Educative: Interactive Courses for Software Developers:
<https://www.educative.io/edpresso/what-are-template-engines>
- Pallets. (s.a.). *Foreword*. Kasutamise kuupäev: 10. mai 2022. a., allikas Foreword — Flask Documentation (2.1.x): <https://flask.palletsprojects.com/en/2.1.x/foreword/>
- Python Software Foundation. (26. aprill 2022. a.). *re — Regular expression operations*. Kasutamise kuupäev: 9. mai 2022. a., allikas re — Regular expression operations — Python 3.9.12 documentation: <https://docs.python.org/3.9/library/re.html>
- Ratselmeister. (s.a.). *Puzzle Page With Codebreaker Word Game Or Crossword Puzzle Stock Vector - Illustration of language, level: 113307702*. Kasutamise kuupäev: 19. aprill 2022. a., allikas Dreamstime: <https://www.dreamstime.com/puzzle-page-codebreaker-word-game-crossword-codeword-code-cracker-general-knowledge-some-words-already-place-medium-image113307702>
- Rigutini, L., Diligenti, M., Maggini, M., & Gori, M. (2008). A Fully Automatic Crossword Generator. *2008 Seventh International Conference on Machine Learning and Applications* (lk 362-367). San Diego, CA, USA: IEEE.
doi:10.1109/ICMLA.2008.104
- The pandas development team. (veebbruar 2020. a.). *About pandas*. Kasutamise kuupäev: 10. mai 2022. a., allikas pandas - Python Data Analysis Library:
<https://pandas.pydata.org/about/>
- The360reporters. (12. aprill 2022. a.). *Different Types of Crossword Puzzles*. Kasutamise kuupäev: 13. aprill 2022. a., allikas The360Report: <https://the360report.com/types-of-crossword-puzzles/>
- Weyer, S. (19. aprill 2022. a.). *CROSSWORDS: Terminology & Types*. Kasutamise kuupäev: 19. aprill 2022. a., allikas CommuniCrossings:
<https://communicrossings.com/crosswords-terminology-types>
- Wikipedia. (15. veebruar 2022. a.). *New York World - Wikipedia*, Page Version ID: 1071945835. Kasutamise kuupäev: 10. mai 2022. a., allikas Wikipedia:
https://en.wikipedia.org/w/index.php?title=New_York_World&oldid=1071945835

WordMint. (10. jaanuar 2017. a.). *Kindness Crossword Word Search*. Kasutamise kuupäev:

19. aprill 2022. a., allikas WordMint:

https://wordmint.com/public_puzzles/222949/copy?_puzzle_type=Wordsearch

Lisad

I. Terminid

Loomuliku keele töötlus (NLP)	See on arvutiteaduse, tehisintellekti ja arvutilingvistika ühisosa, mis keskendub inimese ja arvuti vahelisele suhtlusele.
Piirangutega rahulolu programmeerimist (CSP)	Tehisintellekti ja operatsioonide uuringutes on piirangutega rahulolu protsess lahenduse leidmiseks läbi piirangute kogumi, mis seab tingimused, millele muutujad peavad vastama. Lahendus on seega muutujate väärtuste kogum, mis vastab kõigile piirangutele, st punkt teostavas piirkonnas.

II. Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, **Marvin Vihman**,

(autori nimi)

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

„Ristsõnade automaatne genereerija“,

(lõputöö pealkiri)

mille juhendaja on Sven Aller,

(juhendaja nimi)

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.

Tartus, **8.08.2022**