

TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Hans-Kristjan Urm

**Veebilaiendus eesti keele sõnavara õppija
abistamiseks**

Bakalaureusetöö (9 EAP)

Juhendaja: Sven Aller

TARTU 2021

Veebilaiendus eesti keele sõnavara õppija abistamiseks

Lühikokkuvõte:

Käesoleva lõputöö tulemusena valmis veebilaiendus eesti keele sõnavara õppijate abistamiseks. Programm on mõeldud kasutamiseks igas vanuseklassis ning sobib inimestele, kes tunnevad juba kasvõi vähesel määral eesti keelt. Veebilaiendus on hüplikaknas avanev eesti keele keeleressursse kasutatav laiendus, mida saab paralleelselt kasutada teiste veebilehtedega. Info kogumiseks, mida kuvatakse veebilaienduses, kasutati EstNLTK-s leiduvat Eesti Wordneti, eesti keele tesaurust ning sagedussõnastikku. Töös kirjeldatakse info kogumiseks kasutatud algoritmi ning veebilaienduse töökäiku.

Võtmesõnad: Veebilaiendus, e-õpe, keeleressursid, kasutajasõbralikkus, eesti keele sõnavara

CERCS: P175

Browser extension for helping people learn Estonian vocabulary

Abstract:

As a result of this bachelor's thesis, a browser extension was made to help people learn Estonian vocabulary. Program is meant for people of any age class, who are already familiar with a little bit of Estonian vocabulary. Browser extension uses Estonian language resources and is a popup extension that can be used while browsing other websites. Information displayed in the browser extension, was gathered by using Estonian Wordnet in EstNLTK, Estonian thesaurus and frequency dictionary. The thesis also describes the algorithm used for gathering information and the workflow of the extension.

Keywords: Browser extension, digital learning, language resources, user-friendliness, Estonian vocabulary

CERCS: P175

Sisukord

Sissejuhatus.....	4
1. Olemasolevad analoogsed keeleõppe abistamiseks loodud veebilehed ning rakendused .6	
1.1 Eesti keele põhisõnavara sõnastik	6
1.2 Veebirakendus Cloud WordNet Browser.....	7
1.3 Eesti keele tesaurus	7
1.4 Teksaurus 2.0	8
1.5 Olemasolevate analoogsete keeleõppele suunatud veebilehtede ja rakenduste võrdlus ning arvustus	9
2. Kasutatavad keeletehnoloogilised vahendid ning ressursid.....	11
2.1 Eesti Wordnet.....	11
2.2 EstNLTK 1.4.1	11
2.3 Eesti keele tesaurus	11
2.4 Sagedussõnastik	11
3. Veebilaiendused.....	13
4. Veebilaiendus eesti keele sõnavara õppija abistamiseks	15
4.1 Veebilaienduse kirjeldus	15
4.2 Tehnoloogilised vahendid	18
4.2.1 Python	18
4.2.2 PHP	18
4.2.3 JavaScript.....	18
4.2.4 HTML, CSS	19
4.3 Andmete eeltöötlemine	19
4.4 Sõnade töötlusalgoritm.....	19
4.5 Veebilaienduse töövoog	21
4.6 Tagasiside.....	22
4.7 Probleemid	23
4.8 Edasiarendusvõimalused	23
Kokkuvõte.....	24
Kasutatud kirjandus	25
Lisad.....	27
I. Küsimustik	27
II. Litsents	29

Sissejuhatus

Keelt peetakse oluliseks osaks rahvuse ning kultuuri säilitamisel. Inimesi, kes räägivad eesti keelt emakeelena, on maailmas ligikaudu 1,1 miljonit [1], mis on maailmaga võrreldes väike arv ning selle kõnelejate arv väheneb pidevalt [2]. Seetõttu on oluline panustada eesti keele õppesse, et aidata säilitada meie kultuuri ning keelt. Keeleõppe lihtsustamiseks tuleb esmalt mõista, mis aitab inimesel keelt paremini omandada. Varasemates uuringutes on välja toodud, et uue sõna liikumiseks lühiajalisest mälust pikaajalisse mällu, tuleb sellega korduvalt kokku puutuda ning sõnu omandatakse etapi viisi, mille üheks etapiks on sõna sidumine reaalse eluga. Sõna omandamine tähendab seda, et sõna on liikunud lühimälust pikaajalisse mällu ning meenub ka siis, kui seda pole hiljuti nähtud [3]. See kinnitab, et korrates sõnu ning neid seostades teiste sõnade ja tegeliku eluga, mängides või nähes sõnu muu programmi raames, aitab see kaasa sõnavara laienemisele ning selle arendamisele.

Keeleõppe moderniseerimine, mis siinkohal tähendab raamatute asemel arvuti ja interneti kasutamist, toob endaga kaasa erinevaid mõjusid. Nendeks mõjudeks on näiteks õppijate motiveeritus, mida peetakse nii positiivseks kui ka negatiivseks [4]. Positiivse poole pealt on välja toodud, et noored seostavad arvutit tihti mängulisuse ning huvitavusega, mis pakub lisamotivatsiooni õppimise alustamisel. Negatiivse poole pealt tuuakse välja, et arvutiga töötades, tekib õppijatel probleeme pikaajalise tähelepanu säilitamisega [5]. Juhul kui isikul on tõsine soov keelt õppida, siis interneti teel on see palju lihtsam, sest materjal on üldjuhul alati kättesaadav ning tasuta erinevalt paberkandjal olevast õppematerjalist.

Tänapäeval on eriti aktuaalseks teemaks tõusnud sotsiaalne kaugenemine ning e-õpe (suuremalt osalt Covid-19 pandeemia tõttu) ja sellepärast on interneti vahendusel õppimisele kaasa aitamine ning täiustamine vajalikum kui varem. Käesoleva töö puhul keeleõppele kaasa aitamine tähendab uue veebilaienduse loomist, mis aitab e-õppe, kuid ka igapäevase eesti keele kasutamise käigus sõnu paremini mõista ning neid seostada eelnevalt teada olevate sõnadega. Enne pandeemiat on arutletud juba mitmed aasta, kumb on efektiivsem, kas e-õpe või kontaktõpe [6], sest kontaktõpe nõuab individuaalset panust nii õppijalt kui ka õpetajalt ning e-õpe vähendab õpetaja ressursside kasutust, õpilaste ärevust ning ka ebakindlust tänu oma anonüümsusele [7]. Lisaks pakub e-õpe aja paindlikumat kasutusvõimalust ning programmi läbimisel järjekorra vabadust.

Käesoleva rakendusliku uurimistöö käigus luuakse veebilaiendus, mille eesmärk on eesti keele sõnavara õppijatele pakkuda mitte ainult sõnu ja nende tähendusi, vaid ka neile sarnaseid sõnu,

mis aitaks luua seoseid uute, õpitavate ning vanade, teadaolevate sõnade vahel, mille kaudu saab laiendada ning õppida eesti keele sõnavara. Veebilaiendus on suunatud kõigile, kes soovivad õppida eesti keelt ning sellega kaasnevat sõnavara.


Töö on jagatud neljaks peatükiks. Esimeses peatükis kirjeldatakse olemasolevaid analoogseid keeleõppe abistamiseks loodud programme. Teises tutvustatakse programmi loomisel kasutatud keeletehnoloogilisi vahendeid ning ressursse. Kolmandas peatükis kirjeldatakse veebilaiendusi, erinevate laienduste kasutusjuhte ning ohte. Neljas peatükk annab ülevaate valminud programmist, kirjeldab info leidmiseks kasutatud algoritmi, ning laienduse töökäiku. Samuti on kirjeldatud töö käigus tekkinud probleemid ning edasiarendamise võimalusi.


1. Olemasolevad analoogsed keeleõppe abistamiseks loodud veebilehed ning rakendused


Käesolevas peatükis tuuakse ülevaade keeleressursse kasutatavatest veebilehtedest ning rakendustest, mis on loodud keeleõppele kaasa aitamiseks, milliseid võimalusi need pakuvad ning kas ja millised on puudujäägid. Eesmärgiks on luua ülevaade olemasolevatest ressurssidest ning kuidas ja kas nende omadusi veebilaienduse loomisel ära kasutatakse või välditakse.


1.1 Eesti keele põhisõnavara sõnastik


Eesti keele põhisõnavara sõnastik on sõnastik, mis koosneb 5000 eesti keele olulisemast sõnast, mis on lahti seletatud [8]. Kasutamine näeb välja järgnevalt: sisestate sõna, mida uurite, otsinguribale ning seejärel saate kontrollida sõna tähendust, hääldust, leida mitmuse vorme, vaadata illustreerivat pilti ning otsida vastavat sõna ka teistest sõnastikest. Lisaks leidub ka sõnadega vormistatud näitelauseid (Joonis 1).





[PSV] Eesti keele põhisõnavara sõnastik 




Sõnastikust • Kasutusjuhend • Pilt ja heli • Mängime • 

karu  Otsi ja näita Sõna 

 Pildilehed Õppelehed Maad ja rahvad Grammatikatabelid Õpetajale

karu nimisõna (karu , karu , karu ; *mitmus* karud, karude, karusid )

suur metsloom, kellel on paks pruun karv
Karud magavad kogu talve.



karjäär
karm
karp
kartma
kartul
► karu
karumari
karv
karvane
kas
kask

karu teistes sõnastikes

- Seletav sõnaraamat
- Õigekeelsussõnaraamat
- Eesti-vene sõnaraamat
- Inglise-eesti sõnaraamat
- e-keelenõu
- Viipekeele sõnaraamat

Arvamused ja ettepanekud dict.psv@eki.ee

© Eesti Keele Instituut a-ü sõnastike koondleht  veebiliides @veebihaldur

Joonis 1. Eesti keele põhisõnavara sõnastiku päringu vastus sõnale "karu" [8]

Sõnastik töötab probleemideta, kuid veebilaienduse koostamine ning iga sõna juurde veel sarnase sõna ning näidete lisamine aitaks keelt omandada paremini. Selline seostamine sõnade puhul aitab igas eas inimesi, kes keelt õpivad või sõnavara laiendada tahavad, näiteks sõna

„karu“ puhul kuvada alammõisteid ja liike nagu „grisli“, mis võib tuua seoseid inimestel teistest keeltest. Lisaks on vastav programm eraldi veebileht, mis tähendab seda, et juhul kui isik loeb mõnes brauseri aknas teksti ning soovib seal leiduva sõna kohta infot saada, peab täiendava info leidmiseks avama brauseris uue lehe, mis võib osutada tülakaks korduval kasutamisel. Lehel on ka funktsioon selleks, et vahetada keelt eesti, inglise ja vene keele vahel vastavalt soovile, mis keeles lehte kasutada tahetakse.

1.2 Veebirakendus Cloud WordNet Browser

Cloud WordNet Browser [9] on inglise keelne veebirakendus, milles kuvatakse kasutajale ingliskeelsest leksilisest andmebaasist sünonüüme, sarnaseid sõnu, antonüüme ning definitsioone ning selle kaudu aidatakse kaasa sõna tähenduse mõistmisele. Kasutusjuht on järgmine: avatakse veebirakendus uues brauseri aknas, sisestatakse uuritav ingliskeelne sõna lahtrisse ning vajutatakse „Search“ nupule. Seejärel kuvatakse kasutajale eelnevalt kirjeldatud info, mis vastab päritud sõnale. Sünonüümid kuvatakse järjestatult sageduse järgi ning seda ideed kasutatakse ka käesoleva töö käigus loodavas veebilaienduses sõnade sorteerimiseks. Vastav veebirakendus on iseseisev veebileht, mitte hüplikaken, mida saaks kasutada paralleelselt samas aknas teksti lugedes. Käesoleva töö käigus loodava sarnase, kuid hüplikaknas kuvatava veebilaienduse positiivseks küljeks on kasutajasõbralikkus, sest kulub vähem aega teksti ja laienduse vahel navigeerimiseks.

1.3 Eesti keele teaurus

Veebiprogramm eesti keele teaurus [10] on eestikeelne lehekülg, mis on loodud üksnes sünonüümide ning antonüümide leidmiseks. OÜ FiloSoft poolt loodud programmis ühendati automaatselt Asta Õimu sünonüümi- [11] ja antonüümisõnastik [12]. Leheküljele sisenedes on antud väli, kuhu saab sisestada uuritava sõna ning seejärel saab vajutada nupule „Sünonüümid“, mis annab päringu vastuse. Korrektse ja olemasoleva sõna puhul antakse vastuseks loetelu sünonüümidest ning olemasolu korral ka antonüüm, mis on tõstetud esile punaseks värvituna. Vastavat antonüümi eristamis viisi kasutatakse ka töö käigus loodavas laienduses. Välja tuuakse ka sõnad, mis on päritava sõna suhtes osasünonüümid, kuid ühelgi juhul ei täpsustata, kas vastavad sõnad on täis- või osasünonüümid.

jooksma -- Eesti keele teaurus

immitsema lekkima jooksma läbi nõrguma läbi tilkuma läbi tungima välja ajama	jooksma lippama liduma litsuma leekima punuma vudima silkama sibama kihutama tormama kappama tuiskama plagama putkama traavima jooksujalu minema käima	jooksma tilkuma pihkama immitsema läbi laskma pidama	kulgema suunduma jooksma käima	voolama jooksma vulama tulvama lahisema lahvama pahisema valguma
---	---	---	--	--

Copyright © 2013

Joonis 2. Eesti keele teauruse päringu vastus sõnale "jooksma" [10]

Veebileht kuvab sünonüüme ning antonüüme, mis võib kasutajatel aidata luua seoseid erinevate sõnade vahel, kuid selleks peavad nad vähemalt ühe sünonüümi tähendust teadma. Info selgemaks mõistmiseks võiks olla kirjas, millises mõistes päritavat sõna vastavas tabeli veerus kujutatakse või millises kontekstis sõna kasutatakse. Siit ka idee kuvada töö käigus valmivas laienduses päritava sõna juures sünonüüme ning antonüüme, kuid koos lisainfoga.

1.4 Teksaurus 2.0

Teksaurus 2.0 [13] on Eesti Wordneti [14] kasutajaliides, mis on iseseisev veebileht ning võimaldab inimestel kasutada Eesti Wordneti veebilehitsejas. Veebilehe avamisel on ette antud päringuaken, kuhu saab sisestada algvormis sõnu ning seejärel vajutades sisestusklahvi saab päringule vastuse. Veebilehel on ka nupp "Kaja vastu", mis peaks töötama samal põhimõttel kui sisestusklahvi vajutamine, kuid vastav nupp ei tööta. Peale päringu sisestamist antakse esmalt vastuseks kõik Eesti Wordnetist leitud versioonid päritud sõnast. Soovitud sõna peale vajutades kuvatakse vastava sõna definitsioon eesti keeles ning mõnel juhul ka vastav sõna inglise keeles ning selle definitsioon. Lisaks definitsioonidele kuvatakse nupp "Üksikasjad", millele vajutades suunatakse kasutaja uuele veebilehele, kus olemasolu korral tuuakse välja hüperonüümid, hüponüümid(alammõiste) ning "Teised suhted". Veebilehel kuvatav info on sildistatud pealkirjadega, kuid pealkirjad on vahelduvalt eesti keeles ning kohati inglise keeles. Samuti on kasutatud pealkirjana lühendeid nagu näiteks "*eq_synonym*", mille tähendust pole ka lahti seletatud, ehk kasutajale võib jääda segaseks, mida vastava pealkirjaga sildistatud info endast kujutab.

Kõik kuvatav info on sildistatud pealkirjadega ning seda lähenemist kasutab ka käesoleva töö käigus loodav veebilaiendus, kuid pealkirjad on alati eestikeelsed ning täissõnad. Teksauruses kuvatavad sõnad hüpo- ning hüperonüümide seksioonides on interaktiivsed sünohulgad(sünonüümide hulk), mis tähendab, et on kuvatud sõna, sõna liik lühendina ehk tegusõna(v), nimisõna(n), omadussõna(a) või määrsõna(b) ning vastavate sünohulkade peale saab vajutada, et avada vastava sünohulga kohta käiv uus veebileht, kus on sünohulga kirjeldus. Selline interaktiivsus teeb Eesti Wordneti kasutamise mugavamaks, sest saab kiirelt uurida seotud sõnu. Käesolevas töös sellist interaktiivsust sünohulkade puhul ei realiseerita, sest laienduse eesmärk on anda kiirelt põhjalik, kuid võimalikult lühike vastus, mitte võimaldada Eesti Wordneti täielikku kasutamist.

1.5 Olemasolevate analoogsete keeleõppele suunatud veebilehtede ja rakenduste võrdlus ning arvustus

Tehnoloogia arenemise tõttu saab pidevalt luua uuemaid ning tõhusamaid digitaalseid õppevahendeid ning õppimist toetavaid programme. Digitaalsete õppematerjalide plussideks on võimalus kasutada neid erinevates olukordades ja erinevate inimeste poolt, aga ka nende interaktiivsus, kasutajasõbralikkus ning kättesaadavus [15]. Eelmises alapeatükikides käsitletud eestikeelsetel programmidel on nii mõnedki omadused puudulikud. Kõige vähem on pööratud tähelepanu kasutajasõbralikkusele, millele eriti nüüd rohkem kui varem võiks rõhku panna selle pärast, et distantsõpe on igapäevane [16] ning selle käigus kasutatakse aina rohkem digitaalseid õppematerjale.

Cloud WordNet Browser veebirakenduses on kasutatud lähenemist, et tuuakse kõik võimalikud sõna tähendused eraldi välja ning tuuakse ka nende juurde sarnased sõnad ning nende tähendused. See võib aidata täpse vaste leidmisel, kuid samuti võib teha keerulisemaks arusaamist, milline tähendus on vastavas olukorras õige. Sellist segadust proovitakse vältida käesoleva töö loomisel selle kaudu, et kuvatakse ainult kõige sagedamini esinenud vaste, sellisel juhul on tõenäosus suurem, et kuvatud vastus on õige, kuid mitte garanteeritud. Sarnaste sõnade ning sõna tähenduse kuvamist teostatakse ka eesti keelse veebilaienduse loomisel.

Teksauruse veebileht on teiste programmidega võrreldes kõige interaktiivsem, sest saab vajutada kõikide sünohulkade peale, et uurida neid eraldi. Selline interaktiivsus muudab Eesti Wordneti kasutamise internetis kiiremaks ning mugavamaks, sest ei ole vajadust navigeerida tagasi avalehele, et uurida mõnda sünohulka, mis on antud päringu vastuseks. Samuti on Teksauruse veebilehel info kategoriseeritud ning märgistatud pealkirjadega, mis aitab kasutajal

mõista, millist infot kuvatakse. Võrreldes Cloud WordNet Browseriga, on kuvatava info mõistmine selle tõttu palju lihtsam. Tiitritega info tähistamist kasutab ka käesoleva töö käigus loodav rakendus.

Eelnevalt kirjeldatud programmid esinevad internetis ning on eraldiseisvad veebilehed, mis kasutaja vaatepunktist tähendab seda, et juhul kui soovitakse sõna kohta päringut teha, tuleb avada uus veebileht. Töö käigus loodava veebilaienduse peamine eesmärk ongi kasutajasõbralikkust parandada selle kaudu, et veebilaienduse saab avada ühe nupuvajutusega ilma käesolevalt lehelt lahkumata, mis säästab aega, kuid teeb ka näiteks teksti lugemise mugavamaks. Mugavus väljendub selles, et kergem on leida koht, kus eelnevalt lugemine pooleli jäi ning vähendab ka teksti korduvat lugemist, juhul kui eelnevalt loetud osa peaks juba olema meelest läinud.

Teksauruse interaktiivsus on hea eeskuju tulevateks edasiarendusteks loodava veebilaienduse puhul, kuid käesolevas versioonis vastavat interaktiivsust ei realiseerita, sest eesmärk on luua võimalikult minimalistlik, kuid informatiivne veebilaiendus. Ingliskeelses veebirakenduses [9], kuvatakse algselt otsingusse sisestatud sõna vastuses punaseks värvitult, kuid eestikeelse veebilaienduse puhul kuvatakse punases kirjas antonüüm, mis võiks selgemalt näidata, et tegu on vastandsõnaga.

2. Kasutatavad keeletehnoloogilised vahendid ning ressursid

Töö praktilises osas kasutatakse sõnade kuvamiseks eesti keele jaoks loodud tehnilisi vahendeid ning teisi keeleressursse, mille abil on võimalik automaatselt leida seoseid sõnade vahel. Leitud sõnade sorteerimist sooritatakse vastava sõna eesti keeles esinemise sageduse põhjal.

2.1 Eesti Wordnet

Eesti Wordnet [14] on leksiline andmebaas, mis on üles ehitatud Princetoni WordNeti [17] ja EuroWordNeti [18] põhimõtteid järgides. Vastav andmebaas sisaldab üle 86 000 mõiste ja 239 000 suhte ning seal leidub nimisõnu, tegusõnu, omadussõnu, määrsõnu ning mitmesõnalisi ühendeid, mis on omavahel ühendatud keeltevahelise indeksi abil Princetoni WordNetiga. Kõik eelnevalt kirjeldatud sõnaliigid on grupeeritud sünohulkadeks, mille kaudu on saadud tähenduslik sõnade võrgustik. Eesti Wordneti peal tehtud sõnatähenduste ühestamise katsed kinnitavad, et seal on kaetud eesti keele põhisõnavara.

2.2 EstNLTK 1.4.1

Eestikeelsete tekstide jaoks loodud töötlusvahend EstNLTK 1.4.1 [19] on avatud lähtekoodiga teek, milles on järgmised funktsioonid: paragraafide, lausete ja sõnade tokeniseerimine, morfoloogiline analüüsaator, süntaktiline analüüsaator, lemmatiseerimine, wordnet ning muud funktsionaalsused. Käesolevas töös kasutatakse EstNLTK-s leiduvat Eesti Wordneti ning selles leiduvat sünonüümi hulkade võrgustikku, mille abil leitakse ning kuvatakse sõnu, mis on omavahel seotud. Nende seoste vahel saab leida sõnade ülem- ning alammõisteid, kuid leida ka sõnadele vastavaid definitsioone ning näiteid, mida kasutajale kuvatakse.

2.3 Eesti keele teaurus

Veebiprogramm teaurus [10] on eestikeelne lehekülg, mis on loodud üksnes sünonüümide ning antonüümide leidmiseks. Programm on loodud töötamaks spelleri taga, kus oli vaja süno- ja antonüüme. Teaurust kasutatakse lisaks Eesti Wordnetile käesolevas töös sünonüümide ning antonüümide leidmiseks.

2.4 Sagedussõnastik

Sagedussõnastikus [20] esineb 10 000 sagedamat lemmat ning sõnastiku aluseks on üks miljon sõna ajakirjandusest ja ilukirjandusest. Sõnad, mis sagedussõnastiku on valitud peavad esinema nii ilukirjanduses kui ka ajakirjanduses, mis tagab, et valitud sõna on piisavalt tavaline, et seda valida sagedussõnastikku. See ei välista, et sõnad mida ei valitud, ei oleks tavalised. Käesolevas

töös teeb sagedussõnastiku kasutamine võimalikuks valikuvõimaluse korral kuvada ainult sõnu, mida eesti keeles kasutatakse rohkem. Valikuvõimalus tähendab siinkohal seda, et peab esinema rohkem sõnu, kui on laienduses mõistlik ning võimalik kuvada. Vastavat sagedussõnastiku kasutati, sest eesmärk on aidata eesti keele sõnavara õppijaid ning tavaliste sõnade kuvamisel on suurem tõenäosus, et kasutaja oskab neid seostada.

3. Veebilaiendused

Veebilaiendused on programmid, millega on võimalik brauserile lisada uusi omadusi ning funktsionaalsusi lisaks olemasolevatele brauseri omadustele [21]. Laiendusi luuakse kasutades veebipõhiseid tehnoloogiaid nagu HTML, CSS ning JavaScript. Omadused ning funktsionaalsused hõlmavad endas veebilehe muutmist kasutaja kogemuse parandamiseks, reklaamide blokeerimist, küpsiste haldamist vastavalt kasutaja soovile, ajaloo jälgimist jms [22]. Brauserite arendusfirmad haldavad omaenda laienduste poode, kust kasutajad saavad alla laadida soovitud laiendusi ning neid seejärel lisada neid vastavalt brauserile. Laienduste poodi saab igäüks laadida üles enda poolt renditud laienduse ning peale arendusfirma poolt teostavat kontrolli, kas laiendus on pahatahtlik või mitte, juhul kui vastav laiendus pole pahatahtlik, avaldatakse see poes ning on alla tõmmatav kasutajate poolt [23].

Veebilaienduste eesmärgiks üldjuhul on kasutaja jaoks muuta brauseri kasutamist mugavamaks ning kuna inimesed tõlgendavad mugavust erinevalt, on ka väga palju erinevat tüüpi laiendusi loodud. Sõltuvalt laienduse tüübist, on ka nende kasutusjuhud väga erinevad. Laiendused võivad esineda interaktiivse tööriista ribana, mille kaudu funktsionaalsusi kasutada saab, kuid esineb ka laiendusi, millel puudub kasutajaliides ning vastav tarkvara muudab brauseris kuvatavat infot [24]. Üheks enimkasutatud laienduse tüübiks, millel puudub kasutajaliides, on reklaamiblokeerijad, millel näiteks Chrome veebilehitsejas on üle 10 miljoni kasutaja [25]. Üheks eeliseks reklaamiblokeerija kasutamisel on see, et ei esine hüplikakendes reklaame, mis segavad teksti lugemist või hajutavad tähelepanu ning selle tõttu teevad brauseri kasutamise kogemuse ebamugavamaks. Lisaks mugavustele on sellistel laiendustel ka turvalisuse tagamise omadused. Lehekülgedel võivad esineda reklaamid, mis hõlmavad endas pahatahtliku koodi ning reklaami peale vajutades võib kasutaja ohtu seada enda turvalisuse selle kaudu, et suunatakse mõnele mitte turvalisele leheküljele või käivitatakse pahatahtlik kood [26]. Sellised laiendused blokeerivad vastavaid reklaame.

Teiseks enim kasutatud kasutajaliideseta veebilaienduseks on teema-laiendused, mis tähendab, et muudetakse brauseri enda välimust. Enim kasutatakse laiendusi, mis muudavad brauseri värvilahenduse tumedamaks ning neid kutsutakse öö-režiimi laienduseks. Selliste laienduste eesmärgiks on vähendada eredat valgust, mis silmi kahjustada võib pikaajalise kasutamise käigus. Eredad heledad valgused muudetakse ümber tumedaks ning kõrge kontrastiliseks, et muuta brauseri kasutamine mugavamaks nii öösel kui ka päeval [27].

Mugavuse kui ka turvalisuse tagamise eesmärgil on loodud ka laiendusi, mis võimaldavad ühes turvalises kohas hoida kõiki paroole [28]. Sellised laiendused on parem alternatiiv näiteks paroolide paberi peale kirjutamisele, mis pole turvaline. Laiendustel on üldjuhul üks peaparool, millega on kaitstud kõik vastavad laienduses hoitavad paroolid [26].

Lisaks mugavustele ning turvariskidele, on loodud ka laiendusi, mis aitavad kasutajatel paremini jaotada oma aega tööülesannete vahel nagu näiteks Marinara:Pomodoro [29] ning laiendusi, mis aitavad näiteks töö tegemisele otseselt kaasa. Selline laiendus, mis töö tegemisel kaasa aitab on Grammarly [30], mis aitab kasutajatel paremini ingliskeeles kirjutada. Laiendus protsessib kasutaja poolt kirjutatavat teksti ning selle käigus pakub mitmeid erinevaid võimalusi, nagu näiteks õigekirja kontroll, sünonüümide pakkumine, plagiaarismi kontroll ning mitmeid funktsioone veel. Vastav laiendus võib abiks tulla nii teadusartiklite kirjutamisel kui ka igapäevasel suhtlusel.

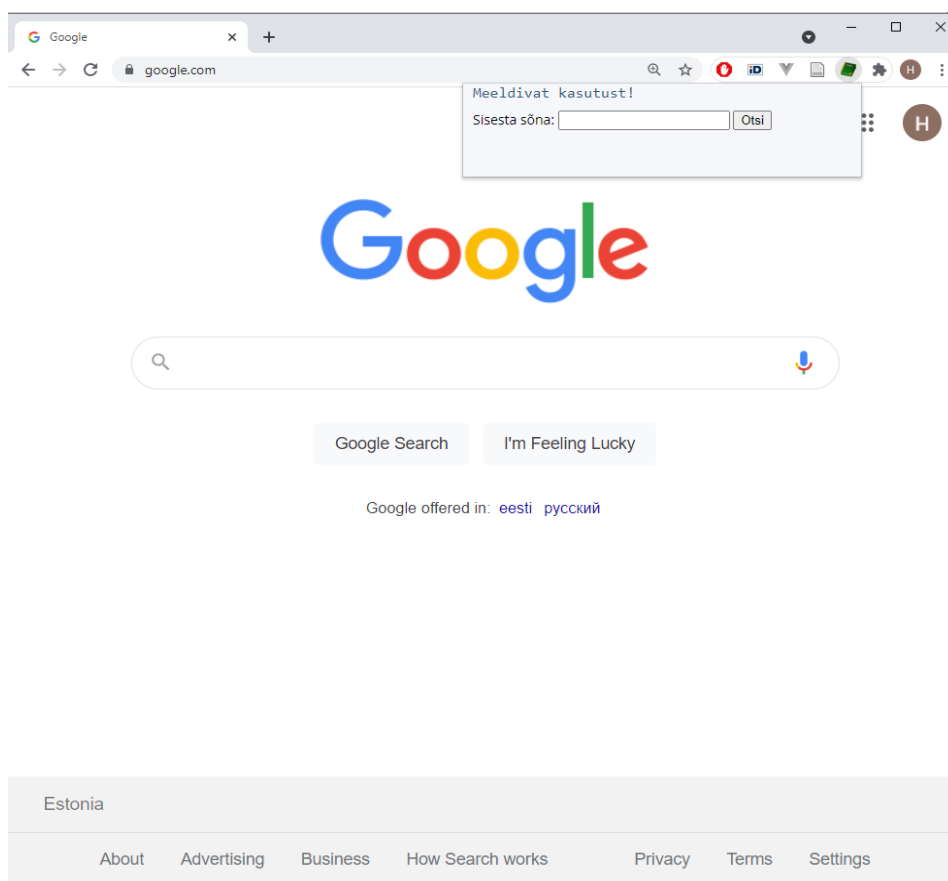
Veebilaiendused on osutunud siiski suureks turvariskiks, sest ligikaudu 10 protsenti vahemikus 2012-2015 Chrome laienduste veebipoodi üleslaetud programmidest kategoriseeriti pahatahtlikeks [21]. Lisaks sellele kümnele protsendile on ka laiendusi, mis pääsesid testist läbi, ilma et oleksid kategoriseerinud pahatahtlikuna. Samuti on pahalased saanud rünnata suurel hulgal inimesi sellisel viisil, et ostetakse juba turul olev laiendus ning muudetakse selle lähtekoodi vastavalt, et see oleks neile kasulik. Lähtekoodi muutmisel uuenduvad laiendused iseenesest ning kasutaja üldjuhul ei saa teadagi, et midagi on muutunud. Pahalaste rünnakuteks on näiteks andmepüük [31], kasutajate poolt sisestava info salvestamine, hajutatud teenuse tõkestamine [32] ning muud ründed.

4. Veebilaiendus eesti keele sõnavara õppija abistamiseks

Bakalaureusetöö käigus loodi veebilaiendus, mis aitab õppida ning laiendada eesti keele sõnavara¹. Selleks kuvatakse kasutajale soovitud sõna kohta sünonüüme, sarnaseid sõnu, hüponüüme, vastavat sõna sisaldav näitelause, sõna definitsioon ning antonüüm, mida leiti tagarakendusega². Veebilaiendus sobib kõigile vanuseklassidele, kes tunnevad juba kasvõi vähesel määral eestikeelseid sõnu ja soovivad õppida või pärida infot eesti keelsete sõnade kohta. Laiendust on võimalik kasutada värskemal Google Chrome veebibrauseriga, käesoleva töö kirjutamise ajal on selleks versioon 92.0.4515.

4.1 Veebilaienduse kirjeldus

Veebilaiendus on hüplikaken, mis tähendab, et laiendus avaneb parasjagu avatud veebilehe ülemises paremas nurgas (Joonis 3) ning sulgub kui kasutaja vajutab hiirega väljaspool laiendust, ehk veebilehel või mujal veebilehitsejas.



Joonis 3. Laienduse avanemine veebibrauseri üleval paremas nurgas

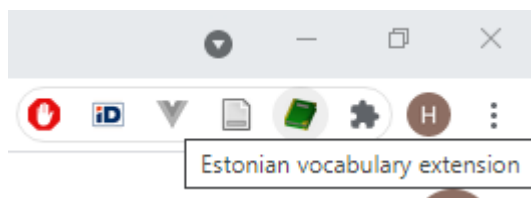
Põhjuseks, miks veebilaiendus avaneb hüplikaknana, mitte eraldi veebikaardil või ei ole eraldi veebileht, on see, et laiendust saaks kasutada paralleelselt veebilehtedega. Vastava lahenduse

¹ Veebilaienduse repositoorium asub aadressil <https://github.com/Hansurm/estonian-voc-extension/>

² Tagarakenduse repositoorium asub aadressil https://github.com/Hansurm/keeleressursside_serveris_kood

eesmärgiks on see, et vähendada aega ning vaeva, mis võib kuluda tekstile uuesti süvenemisele või pooleli jäänud koha leidmisele kui pidevalt navigeerida erinevate veebilehtede vahel.

Laienduse otsetee (ingl *shortcut*) saab kinnitada laienduste ribale veebilehitsejas ning programmi avamiseks tuleb vajutada otsetee ikooni peal. Joonisel 4 on näidatud laienduse otsetee ikoon.

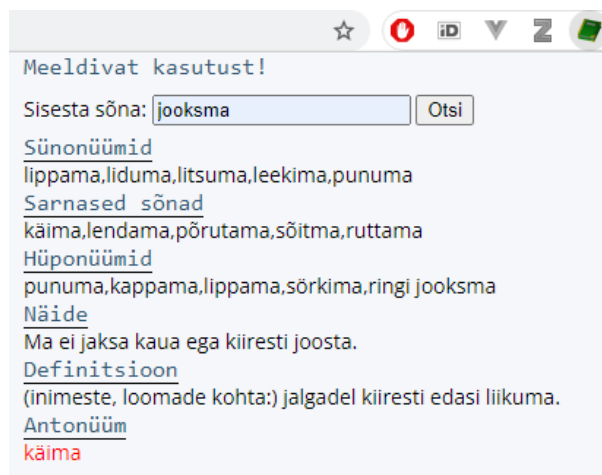


Joonis 4. Laienduse otsetee ikoon hiirekursoriga peal hõljudes

Hiireklõpsuga laienduse otseteel avaneb hüplikaken (Joonis 3). Hüplikaknas on lahter, kuhu saab sisestada soovitava sõna, mille kohta päringut teha soovitakse, ning lahtri kõrval on nupp pealkirjaga "Otsi", millele vajutades annab programm lahtrisse sisestatud sõna kohta päringu vastuse. Otseteel hiirega parema nupuvajutuse puhul on võimalik vastav programm laienduste ribalt eemaldada, uurida programmi konsooli väljundit või eemaldada vastav laiendus veebibrauserist täielikult.

Laiendus annab peale "Otsi" nupul hiirevajutust alati vaste, kui vastavas lahtris ei esine tühisõne, kuid on kahte tüüpi vastuseid:

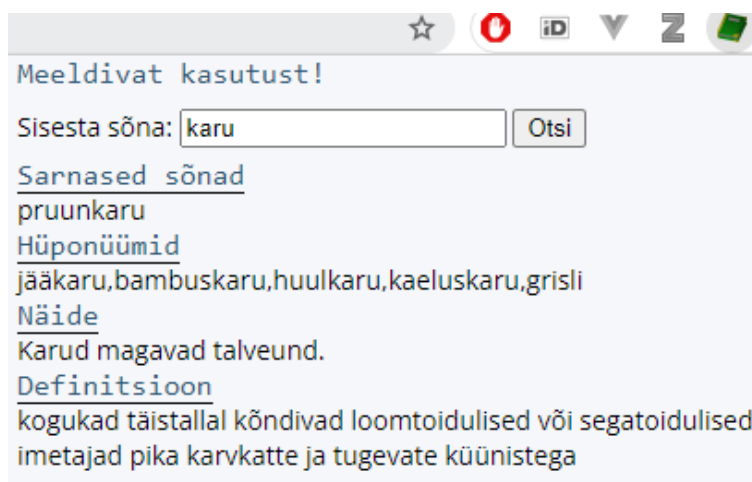
1. Tagarakendus (ingl *back-end*) leidis sõnale vaste ning kuvatakse info, mis kehtib päritud sõna kohta.



Joonis 5. Tagarakendus leidis sõnale "jooksma" vaste ning kuvatakse info

Info, mida kuvatakse hõlmab endas vastavas järjekorras sünonüüme, sarnaseid sõnu, hüponüüme, näidet, definitsiooni ning antonüümi. Info on vastava pealkirja all,

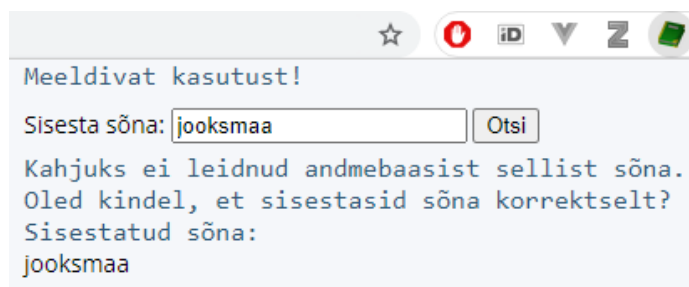
pealkirjad on alla joonitud ning infost erineva šriftiga, et oleks lihtsam leida soovitud välja. Info kuvatakse põhimõtte järgi, et esmalt on sarnasema tähendusega sõnad, seejärel näitelause ning definitsioon ja viimasena antonüüm. Info, mis kasutajale kuvatakse on pealkirjastatud ning antonüümi olemasolu korral on antonüümi kujutatud punase kirjaga, et võimalikult lihtsalt eristada vastandsõna.



Joonis 6. Päringu vastes sõnale "karu" puudub sünonüümid ning antonüüm

Juhtudel kui puudub mõne pealkirja kohta käiv info, siis vastavat pealkirja ei kuvata üldse (Joonis 6). Joonisel 5 on näha, et sõnale "jooksma" antakse antonüümi vastuseks sõna "käima", mis esineb ka sarnaste sõnade rubriigis. Põhjuseks on see, et kuigi on "käima" on sõna "jooksma" vastandsõna, tähendavad mõlemad sõnad jalgadel edasi liikuma, ehk on sarnased sõnad.

2. Tagarakendus ei leidnud sõnale vastet, mis tähendab, et vastav sõna ei esine olemasolevate andmete seas. Sellisel juhul on põhjuseks see, et kas ei leidu sellist sõna kasutatud keeleressursside seas või on teinud kasutaja ise sõna sisestamisel mingit sorti vea.



Joonis 7. Tagarakendus ei leidnud sõnale vastet, kuvatakse teade selle kohta

Kasutajale antakse sellest teada, et on tekkinud probleem ning võimalikud probleemi põhjused. Samuti kuvatakse kasutaja poolt sisestatud sõna, et kasutaja kontrolliks

sisendit, mida näeb ka lahtris, kuid eraldi välja tuues proovitakse sisendi kontrollimisele rohkem tähelepanu pöörata (Joonis 7).

Veebilaiendus ei salvesta varasemat sessiooni, ehk laienduse igal avamisel ei kuvata mingit informatsiooni ning esineb ainult tervitav kiri, sõna sisestamiseks mõeldud lahter ning otsingunupp (Joonis 3).

4.2 Tehnoloogilised vahendid

Töö käigus kasutati erinevaid tehnoloogiaid sõnade eeltöötlemiseks ning veebilaienduse loomiseks. Käesolevas alapeatükis kirjeldatakse vastavaid tehnoloogiaid ning põhjendatakse valikuid.

4.2.1 Python

Sõnade eeltöötlemist teostav programm loodi Python programmeerimiskeeles, sest vastavas keeles, versioon 3.4-s, saab kasutada avatud lähtekoodiga eesti keele loomuliku keele töötlemise liidest, EstNLTK-d [19]. Tänu sellele, saab teha kogu keeletöötlust ühes programmis. Pythoni ja EstNLTK abil teostati kõikide sünohulkade peal eeltöötlus ning seejärel salvestati JSON (*JavaScript Object Notation*) faili sõnastik, mis aitab veebilaienduse programmi tööaega lühendada. Samuti kasutati Pythonit, et teha sõnade kohta päringuid Tesauruse veebilehelt [10].

4.2.2 PHP

Serveris, mille poole veebilaiendus pöördub, et pärida sõna ning kuvada vastus, on PHP (*Hypertext Preprocessor*) programmeerimiskeeles veebileht, mis käivitab Pythoni programmi andes argumentiks päritava sõna ning tagastab seejärel Pythoni programmilt saadud tulemuse. Põhjuseks, miks kasutati PHP-d mitte näiteks HTML-i veebilehe loomisel, on see, et PHP võimaldab käivitada terminali käske, mida ainult HTML-i kasutamisega teha pole võimalik.

4.2.3 JavaScript

JavaScript on veebiarenduses populaarseim ning enim kasutatud programmeerimiskeel, mida kasutavad kõigist internetis leiduvatest veebilehtedest 97.5% [33]. Veebilaiendus kasutab vastavat programmeerimiskeelt, et saata uuritava sõna kohta päring serveripoolsesse PHP programmi ning seejärel saada ning kuvada päringu vastus. Samuti kasutab laiendus JavaScripti, et muuta, lisada ja eemaldada HTML-i elemente.

4.2.4 HTML, CSS

Veebilaienduses esinevate elementide struktuur määratakse HTML-is (*HyperText Markup Language*) [34] ning elementide välimust ja disaini muudetakse CSS-i (*Cascading Style Sheets*) [35] abil. HTML on standardiks veebibrauserites dokumentide kuvamiseks, vastav keel võimaldab kirjeldada veebilehtede struktuuri semantiliselt. Lisaks saab HTML-i sängitada JavaScripti koodi lõike, millega on võimalik veebilehe sisu muuta.

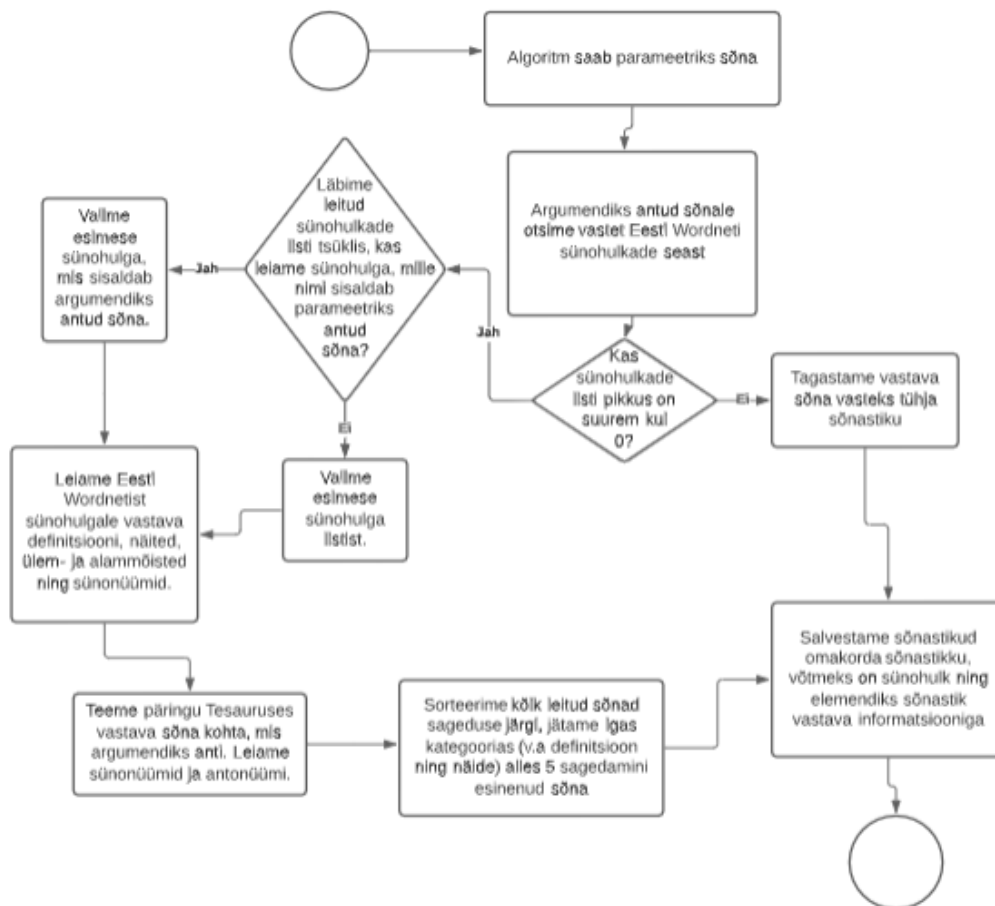
4.3 Andmete eeltöötlamine

Andmed, mida kuvatakse veebilaienduses, on eelnevalt läbi töödeldud, et vähendada ajakulu programmi kasutamise käigus. Eeltöötuse käigus jooksutati tsüklis sõnade töötusalgorithmi (peatükk 4.4) Eesti Wordneti sünohulkade kõikide sünonüümide hulgal ning salvestati sõnastikuna JSON-faili. Programmi reaalse töö käigus saab algoritm argumendiks ainult sõna, mitte sünohulga ning selle tõttu töötab eeltöötlus samal põhimõttel, ehk argumendiks antakse sünohulga nimest eraldatud osa, mis sisaldab ainult sõna.

Keeletöötlus on mahukas protsess ning eelnevalt vastava töö ära tegemine ning programmi töö käigus ainult sõnastiku sisse lugemine vähendas ajakulu ligikaudu 2 sekundi võrra päringu kohta.

4.4 Sõnade töötusalgorithm

Sõnade töötus algoritm (Joonis 8) on loodud selleks, et kuvada soovitava sõna kohta informatsiooni veebilaienduses. Algoritm on kirjutatud programmeerimiskeeles Python ning vastava algoritmiga on teostatud ka eelnevalt kirjeldatud eeltöötlus. Veebilaiendus kasutab päritava sõna eeltöötuse käigus loodud sõnastikus olemasolu korral infot sõnastikust, vastasel juhul jooksutatakse sõna töötuse algoritmi uuesti, andes parameetriks päritav sõna ning kuvatakse sealt saadud vastus. Seda tehakse sellepärast, et juhul kui Eesti Wordneti lisatakse uusi seoseid ning sünohulki, ei pea algoritmi muutma, et vastavat infot kasutajale kuvada.



Joonis 8. Sõnade töötlus algoritm

Algoritm saab parameetris eestikeelse sõna, millele proovitakse leida vaste Eesti Wordneti sünohulkade seast. EstNLTK 1.4 tagastab sünohulkade listi Eesti Wordnetist, mis sõnale vastavad ning tühja listi kui vastet ei leita. Juhul kui sõna kuulub mitmesse sünohulka, valitakse esimene hulk tagastatud sünohulkade listist, mille nimi sisaldab parameetris saadud sõna. Sünohulk, mis sisaldab enda nimes päritavat sõna, võiks olla kõige tüüpilisem vaste sõnale. Juhul kui ei leita sünohulka, mis nimes sisaldaks otsitavat sõna, piisab sellest, et sõna esineb sünonüümide hulgas ning valitakse lihtsalt esimene sünohulk listist. Esimene hulk pole täpsem viis vastava sünohulga valimiseks, kuid töömahu tõttu jäi optimaalseima viisi leidmine edasiarenduse ideeks.

Sarnaste sõnade otsimisel, võetakse eelnevalt leitud sõnale vastav sünohulk Eesti Wordnetis ning leitakse sellele sünohulgale vastavad alammõisted ning ülemmõisted. Kõik leitud sõnad salvestatakse sõnastikku.

Eesti Wordnetis on funktsionaalsus, mis võimaldab leida sünohulgale vastavat sõna sisaldavaid näitelauseid ning selle sõna tähenduse. Näidete puhul tuleb sorteerida välja unikaalsed, sest tihti leidub mitmeid identseid korduvaid näitelauseid. Tulemus salvestatakse sõnastiku.

EstNLTK versioonis 1.4.1, mida kasutatakse käesoleva bakalaureusetöö loomisel, puudub funktsioon leidmaks sünohulga antonüüme. Selle tõttu, tehakse päring Filosoofi tesauruse [10] veebilehele, kus olemasolu korral võetakse sõna antonüüm ning ka selle sõna sünonüümid, mis omakorda salvestatakse taaskord sõnastiku.

Eesti Wordnetist sarnaste sõnade otsimise käigus jäetakse mällu esialgu kõik leitud sõnad. Kuna sõnad kuvatakse veebilaienduses, mille akna mõõtmed on võrdlemisi väikesed, siis ei ole mõtet lõplikult kõiki sõnu salvestada, sest neid ei saa kõiki kuvada. Sagedussõnastikust leitakse iga sõna kohta selle esinemissagedus, mille järgi lõpuks otsustatakse, millised sõnad jäetakse alles ning millised kustutatakse. Alles jäetakse igas kategoorias (ülem-, alammõiste, sünonüümid ja sarnased sõnad) 5 kõige enim esinenud sõna. Kuvatakse 5 kõige sagedamini esinenud sõna sellepärast, et veebilaiendus on mõeldud kasutamiseks paralleelselt veebilehega ning liigse arvu sõnade kuvamisel kataks veebilaiendus liiga suure osa ekraanist.

Lõpuks salvestatakse leitud ja sorteeritud info sõnastikuna, mille elementide võtmeks on sünohulk ning väärtuseks omakorda sõnastik, mis sisaldab eelnevalt kogutud infot, ehk sõna kohta näiteid, definitsiooni, ülemmõisteid, alammõisteid, sarnaseid sõnu, süno- ja antonüüme.

4.5 Veebilaienduse töövoog

Veebilaienduse töö sõltub mitmest programmist. Laiendus ise töötab HTML-i, CSS-i ning JavaScriptiga, kuid nende keeltega pole piisavalt vahendeid, et teostada eesti keele töötlust. Veebilaiendus võtab otsinguväljale sisestatud sõna, moodustab vastava sõnaga URL-i, millega saab pöörduda keeleressursside serveri poole JavaScripti abil. Vastavas serveris töötab PHP programm, millele saab anda argumendiks sõna URL-i kaudu. PHP programm eemaldab üleliigsed tähemärgid(tühikud) ning koostab terminali käsu koos ette antud argumendiga, millega saab käivitada Pythoni programmi. Vastav suhtlus PHP ning Pythoni vahel on vajalik, sest PHP kaudu saab käivitada Pythoni programmi, kus on olemas vajalikud vahendid, et teostada eesti keele töötlust. Terminali käsk kodeeritakse utf-8 kodeeringusse, et kaduma ei läheks eesti keelele kohased täpitähed. Peale käsu käivitamist oodatakse sõnatöötlus programmilt vastust terminalis ning seejärel kuvatakse info taaskord PHP abil keeleressursside serveris jooksva veebilehel.

Veebilehel kuvatakse info sõnastikuna, millest on hiljem lihtne võtmesõnade järgi vastavat infot kätte saada. Kuvatud sõnastik loetakse taaskord sisse laienduses JavaScriptiga. Seejärel kasutatakse JavaScripti, et modifitseerida HTMLi ning kuvada kasutajale veebilaienduses eelnevalt kätte saadud info vastavate pealkirjade all.

4.6 Tagasiside

Veebilaiendust testiti 5 erineva inimese poolt. Peale laienduse testimist tuli täita küsimustik, mida on näha lisas 1. Küsimustikus paluti vabas vormis tagasisidet anda kasutajasõbralikkuse, disaini, esinenud probleemide või vigade ning edasiarendusideede kohta, samuti uuriti, kui tuttavad on kasutajad eesti keele sõnavaraga ning mille raames laiendust kasutati.

Laiendust kasutas brauseris teksti lugemisel 4 inimest 5-st ning üks kasutajatest kasutas haiku kirjutamisel sünonüümide leidmiseks. Samuti oli 4 inimesel 5-st emakeeleks eesti keel ehk olid väga tuttavad eesti keele sõnavaraga ning üks kasutajatest rääkis eesti keelt teise keelena, ehk polnud väga tuttav.

Tehniline viga tekkis ühel korral, kui kasutaja sisestas sõna suure algustähega ning vastet ei leitud, väikese algustähega sama sõna sisestamisel vastus leiti. Vastav viga likvideeriti autori poolt nii, et päring edastatakse väikeste tähtedega sõltumata kasutaja sisendist ning peale seda rohkem vigu ei esinenud. Probleeme tekkis antonüümide kuvamisega, mis kasutaja sõnul kohati olid valed või leidus sagedamini kasutatavaid vastandsõnu.

Kõik kasutajatest pidasid laiendust kasutajasõbralikuks ning 4 inimest tõid põhjuseks, et laiendust sai kasutada samaaegselt teiste lehekülgedega. Disaini kohta tõi üks kasutaja välja, et öö-režiimis brauseri kasutamisel oli laienduse värvilahendus liiga ereda tooniga. Kahele inimesele meeldis värvilahendus, mis oli lihtne ning üks mainis, et meeldis et antonüüm oli kergesti eristatav tänu punasele kirjale.

Kolmel kasutajal ei tekkinud edasiarenduseks ühtegi ideed. Üks kasutajatest arvas, et laiendus võiks kuvada rohkem kui 5 sünonüümi ning teine tõi välja, et päringu vastusele kuvatavaid sõnu võiks saada uurida ilma uue päringu teostamiseta. Viimast soovitus saaks realiseerida tehes kuvatavad sõnad interaktiivseks, millele klõpsates saab automaatselt uue päringu kuvada. Testijad jäid laienduse kasutussõbralikkusega rahule, kuid leiti parandusvõimalusi antonüümide kuvamise ning programmi interaktiivsuse osas.

4.7 Probleemid

Töö eesmärk oli luua veebilaiendus, mis tähendas, et ei olnud võimalik teostada keele töötlust programmi(laienduse) siseselt ehk tuli leida viis, kuidas jooksutada keele töötluks vajalikke programme iseseisvalt. Lahenduseks otsustati kasutada eesti keele keeleressursside serverit, kus jookseb veebileht, mille kaudu saab käivitada Pythoni programmi ning veebilaiendus saab JavaScripti abil vastava veebilehega suhelda. Suurimaks probleemiks osutusid eesti keelsed täpitähed. Nimelt esialgu ei olnud võimalik veebilehel kuvada infot, kui vastuses esines täpitäht, kuigi kõik kodeeringud olid vastavalt sobivad(utf-8). Siinkohal osutus kasulikuks Pythonis leiduv veakäsitlusmeetod “surrogateescape”, mis dekodeerimisel asendab baidid individuaalsete asendustega.

4.8 Edasiarendusvõimalused

Laienduses uuritavale sõnale vastava sünohulga valimiseks leitud lahendus ei olnud kõige optimaalsem, sest hetkel valitakse esimene selline sünohulk, mis sisaldab parameetriks antud sõna ning sellise sünohulga puudumisel valitakse esimene sõnale vastuseks antud sünohulk. Edasiarenduse ideeks sünohulga valimisel oleks see, et kuvada kõik sõnale Eesti Wordneti poolt vastuseks antud sünohulgad, muuta vastavad sünohulgad interaktiivseks, et nende peale saaks vajutada. Seejärel vajutades soovitud sünohulga peale saaks kuvada vastava informatsiooni, mida käesolevas töös kuvatakse.

Eelnevalt kirjeldatud edasiarendus idee inspiratsioon saadi Teksauruse [13] veebilehelt, kus kõik kuvatavad sünohulgad on interaktiivsed. Samasugust lahendust oleks võimalik ka realiseerida käesolevas töös. Selline interaktiivsus võimaldaks inimestel navigeerida erinevate sünohulkade vahel, ilma ise sõna sisestamata. Siiski poleks võimalik kuvada kõiki sünohulki, sest sellisel juhul veebilaienduse akna mõõtmed võiksid mõnel juhul minna niivõrd suureks, et poleks vastavat laiendust enam võimalik teiste veebilehtedega paralleelselt kasutada. Sellisel juhul kaotaks veebilaiendus enda peamise põhimõtte.

Antud hetkel, tagastab laiendus veateate, kui kasutaja poolt sisestatud sõnale vastust ei leita. Vastavat lähenemist saaks muuta selliselt, et leitakse ka sarnaseid sõnu, kui sisestatud sõna täpselt samas vormis ei leitud. Samuti oleks võimalik teostada sisestatud sõnale grammatika kontrolli, nagu tehakse näiteks Google otsingumootoris, kus pakutakse välja variant, mida kasutaja võis silmas pidada. Selline grammatika kontroll nõuaks suurel hulgal eestikeelsete sõnade läbitöötamist ning vastavate sõnade kauguse leidmist sisestatud sõnast.

Kokkuvõte

Käesoleva bakalaureusetöö eesmärgiks oli luua eesti keele õppija abistamiseks veebilaiendus, mida saaks kasutada paralleelselt teiste veebilehtedega. Töö käigus uuriti analoogseid olemasolevaid keeleõppe abistamiseks loodud programme, et õppida nende tugevustest ning nõrkustest ja vastavaid teadmisi programmi arendamisel kasutada. Samuti kirjeldati laienduse loomisel kasutatud vahendeid ning ressursse. Lisaks uuriti veebilaiendusi, nende otstarbeid, liike ning ohte.

Loodud veebilaiendus võimaldab kasutajal kiiresti ja mugavalt pärida infot eestikeelsete sõnade kohta, et luua uusi seoseid teada olevate sõnade vahel ning selle kaudu kaasa aidata eesti keele sõnavara laiendamisele. Laienduse abil on võimalik kasutajal näha päritava sõna kohta sünonüüme, sarnaseid sõnu, hüponüüme, näitelauseid, definitsiooni ning antonüümi. Rakendus loodi minimalistliku ning lihtsa disainiga ning otsustati kuvada hüpikaknas, et säästa aega, mis kulub erinevate lehtede vahel navigeerimiseks.

Laiendust testisid 5 inimest, pärast mida parandati leitud vead ning pandi kirja võimalikud edasiarendusvõimalused. Programmi peeti kõikide kasutajate poolt kasutajasõbralikuks ning 4 inimest tõid põhjuseks selle, et laiendust sai kasutada teiste veebilehtedega samaaegselt.

Rakendust saaks tulevikus täiustada, kuvades kasutajale mitte ainult ühe sünonüümi hulga tähendust, vaid päritava sõnale vastavad kõik sünohulgad, mis teha interaktiivseks, et kasutaja saaks ise valida õige vaste.

Kasutatud kirjandus

- [1] Eesti keel maailma taustal.
http://www.estonica.org/et/%C3%9Chiskond/Eesti_keel/Eesti_keel_maailma_taustal/
(09.06.2021)
- [2] Tomusk I. Eesti keele arengukava 2011-2017.
https://www.hm.ee/sites/default/files/eesti_keeles_arengukava.pdf (09.06.2021)
- [3] Uibu K, Timm M. Sõnasemantika ja teksti mõistmine põhikooli esimeses ja teises kooliastmes. Eesti Rakenduslingvistika Ühingu aastaraamat. 2014. (09.06.2021)
- [4] Teral M. Arvutipõhine eesti keele õpe: vahendid ja hinnangud nende efektiivsusele Tartu ülikooli keelekursuste näitel. 29. oktoober 2015.
<http://ezproxy.utlib.ut.ee/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsair&AN=edsair.od.....1018..7504791a908248da85ad1edf41b6a498&site=eds-live>
(09.06.2021)
- [5] Bangs P, Cantos P. What Can Computer Assisted Language Learning Contribute to Foreign Language Pedagogy? Int J Engl Stud IJES. 2004. (09.06.2021).
- [6] Toots A, Plakk M, Idnurm T. Tiiger luubis 3 : uurimus e-õppes Eesti ülikoolides | DIGAR. <https://www.digar.ee/arhiiv/nlib-digar:56768> (09.06.2021)
- [7] Kruse M. Võõrkeeleeõpe Kui Sotsiaalne, Sünergiline Ja Teadvustatud Protsess. Eesti Ja Soome-Ugri Keeleteaduse Ajak. september 2014. (09.06.2021).
- [8] [PSV] Eesti keele põhisõnavara sõnastik. <https://www.eki.ee/dict/psv/> (09.06.2021)
- [9] Cloud WordNet Browser. cloudapps.herokuapp.com/word/ (09.06.2021)
- [10] Eesti keele teaurus. https://www.filosoft.ee/thes_et/ (09.06.2021)
- [11] [SYS] Sünonüümisõnastik. <http://www.eki.ee/dict/sys/> (09.06.2021)
- [12] Õim A. Antonüümisõnastik. <https://www.eki.ee/dict/antonyymid/antonyymid.html>
(09.06.2021)
- [13] Teksauruse päring. <https://teksaurus.keeleressursid.ee/#> (09.06.2021)
- [14] Eesti Wordnet. <https://www.cl.ut.ee/ressursid/teksaurus/?lang=et> (09.06.2021)
- [15] Digitaalse õppematerjali loomise soovitused – Juhend digitaalse õppematerjali autorile.
<https://oppevara.edu.ee/kvaliteet/> (09.06.2021)
- [16] Koroonaviiruse levik ja haridusvaldkond: soovitused ja korduma kippuvad küsimused | Haridus- ja Teadusministeerium. <https://www.hm.ee/et/koroona> (20.03.2021)
- [17] WordNet | A Lexical Database for English. <https://wordnet.princeton.edu/> (09.06.2021)
- [18] EuroWordNet: Building a multilingual database with wordnets for several European languages. <https://archive.illc.uva.nl/EuroWordNet/> (09.06.2021)

- [19] Estnltk 1.4.1. <https://estnltk.github.io/estnltk/1.4.1/index.html> (09.06.2021)
- [20] Eesti Kirjakeele Sagedussõnastik.
<https://www.cl.ut.ee/ressursid/sagedused/index.php?lang=et> (09.06.2021)
- [21] Varshney G, Bagade S, Sinha S. Malicious browser extensions: A growing threat: A case study on Google Chrome: Ongoing work in progress. 2018 Int Conf Inf Netw ICOIN Inf Netw ICOIN 2018 Int Conf On. 1. jaanuar 2018. (09.06.2021)
- [22] What are extensions? https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions/What_are_WebExtensions (09.06.2021)
- [23] Publish in the Chrome Web Store. <https://developer.chrome.com/docs/webstore/publish/> (09.06.2021)
- [24] Extension and App Types.
https://chromium.googlesource.com/chromium/src/+HEAD/extensions/docs/extension_and_app_types.md (09.06.2021)
- [25] AdBlock. getadblock.com (09.06.2021)
- [26] Br, Tech on V in, September 14 W on, 2016, Pst 10:41 Am. Five types of browser extensions every professional should have. <https://www.techrepublic.com/article/five-types-of-browser-extensions-every-professional-should-have/> (09.06.2021)
- [27] Dark Reader. <https://darkreader.org/> (09.06.2021)
- [28] How LastPass Works. <https://www.lastpass.com/how-lastpass-works> (09.06.2021)
- [29] Marinara: Pomodoro Assistant. <https://chrome.google.com/webstore/detail/marinara-pomodoro%C2%AE-assist/lojgmehidjdhbmpjfamhpkpodfcodef?hl=en> (09.06.2021)
- [30] Grammarly. <https://app.grammarly.com/> (09.06.2021)
- [31] What is phishing. <https://www.imperva.com/learn/application-security/phishing-attack-scam/> (09.06.2021)
- [32] Mis on DDoS rünnak ja kuidas end selle eest kaitsta? <https://www.am.ee/Mis-on-DDoS> (09.06.2021)
- [33] Usage Statistics of JavaScript as Client-side Programming Language on Websites.
<https://w3techs.com/technologies/details/cp-javascript> (09.06.2021)
- [34] HTML For Beginners The Easy Way: Start Learning HTML & CSS Today.
<https://html.com/> (09.06.2021)
- [35] What is CSS? https://developer.mozilla.org/en-US/docs/Learn/CSS/First_steps/What_is_CSS (09.06.2021)

Lisad

I. Küsimustik

Veebilaiendus eesti keele sõnavara õppija abistamiseks

Küsimustik on loodud Hans-Kristjan Urm informaatika eriala bakalaureusetöö "Veebilaiendus eesti keele sõnavara õppija abistamiseks" testimise jaoks.

* Required

Testimise ning tagasiside andmise juhend

Testige veebilaiendust "Estonian vocabulary extension" igapäevase veebibrauseri kasutamise käigus.

Veebilaiendus on alla laetav repositooriumist, mis asub aadressil <https://github.com/Hansurm/estonian-voc-extension>

Vastavas repositooriumis on ka kirjeldus, kuidas veebilaiendust alla laadida ning kasutada.

Kasutage veebilaiendust igapäevaste brauseri tegevuste käigus eesti keele sõnade kohta informatsiooni saamiseks vähemalt 15 minutit ning andke seejärel küsimustiku kaudu tagasisidet.

Kui tuttavad olete eesti keele sõnavaraga? *

- Eesti keel on minu emakeel, ehk väga tuttav
- Valdan eesti keelt teise keelena, päris tuttav
- Valdan eesti keelt väga vähe, olen võõras sõnavara suhtes

Mille käigus kasutasite veebilaiendust? *

Your answer

Andke palun tagasisidet laienduse kasutajasõbralikkuse kohta *

Your answer

Andke palun tagasisidet laienduse disaini kohta, mida võiks muuta? *

Your answer

Kas esines probleeme või vigu laienduse kasutamise käigus? *

Your answer

Kas teil tekkis ideesi, mida võiks laienduse edasiarenduses realiseerida?

Your answer

Submit

II. Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, **Hans-Kristjan Urm**,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „**Veebilaiendus eesti keele sõnavara õppija abistamiseks**“ 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.

Hans-Kristjan Urm, 04.08.2021