



TALLINNA TEHNIKAÜLIKOOL
INSENERITEADUSKOND
Virumaa kolledž

Subtiitrite veebipõhise masintõlkeplatvormi loomine
Creating a web-based machine translation platform for subtitles
ARUKAD SÜSTEEMID JA RAKENDUSINGOTEHNOLOOGIA ÕPPEKAVA
LÕPUTÖÖ

Üliõpilane: Monika Kiik

Üliõpilaskood: 207555EDTR

Juhendaja: Avar Pentel, lektor

AUTORIDEKLARATSIOON

Olen koostanud lõputöö iseseisvalt.

Lõputöö alusel ei ole varem kutse- või teaduskraadi või inseneriplomit taotletud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on viidatud.

LIHTLITSENTS LÕPUTÖÖ ÜLDSUSELE KÄTTESAADAVAKS TEGEMISEKS JA REPRODUTSEERIMISEKS

Mina Monika Kiik (sünnikuupäev: 02.02.1978)

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose Subtiitrite veebipõhise masintõlkeplatvormi loomine, mille juhendaja on Avar Pentel,
 - 1.1. reprodutseerimiseks säilitamise ja elektroonilise avaldamise eesmärgil, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta kolmandate isikute intellektuaalomandi ega isikuandmete kaitse seadusest ja teistest õigusaktidest tulenevaid õigusi.

SISUKORD

LÜHENDITE JA TÄHISTE LOETELU	6
SISSEJUHATUS	7
1. SUBTIITRID	8
1.1 Allalaetavad tõlkeplatvormid	9
1.1.1 FAB Subtitler	9
1.1.2 SubtitleEdit	10
1.1.3 AegiSub	11
1.2 Veebipõhised tõlkeplatvormid	11
1.2.1 Subtitles Translator	11
1.2.2 Kapwing	11
1.3 Tõlkeplatvormide kokkuvõte	12
2. SUBTIITRITE VEEBIPÕHISE MASINTÕLKEPLATVORMI LOOMINE	13
2.1 Veebipõhise masintõlkeplatvormi struktuur	13
2.1.1 Avaleht ja sisselogimine	14
2.1.2 Tööde staatus	14
2.1.3 Tõlkimine	14
2.1.4 Kasutajate lisamine	14
2.1.5 Filmide lisamine	15
2.2 Figma kasutamine kujunduse loomisel	15
2.3 <i>JavaScript</i> 'i kasutamine koodi kirjutamisel	16
2.4 AWS serverteenuse uurimine võimaliku teenusepakkujana	17
2.5 Microsoft SQL Serveri kasutamine kasutajate loomiseks	18
2.6 DeepL ja sellega ühendamine	19
2.7 Veebipõhise masintõlkeplatvormi kokkuvõte	21
3. SUBTIITRITE VEEBIPÕHISE MASINTÕLKEPLATVORMI ÕPETAMISE VÕIMALIKKUS	
23	
3.1 Masintõlke arengud	23
3.1.1 Tehisnärvivõrgud	23
3.1.2 Loomulik keele töötlemine	24
3.2 Masina õpetamine	24
3.3 Ühendamine ja kasutamine platvormil	26
KOKKUVÕTE	27

SUMMARY	28
KASUTATUD KIRJANDUSE LOETELU	29

LÜHENDITE JA TÄHISTE LOETELU

AI	tehisintellekt (Artificial Intelligence)
API	rakenduse programmiliides (Application Programming Interface)
AWS	Amazoni veebiteenused (Amazon Web Services)
AWS S3	Amazoni pilveserver (Amazon Simple Storage Service)
BRNN	kahepoolne rekurrentne närvivõrk (Bidirectional Recurrent Neutral Network)
GRU	värvatega rekurrentsed ühikud (<i>Gated Recurrent Unit</i>)
HTML	hüpertexti märgistuskeel (<i>HyperText Markup Language</i>)
LSTM	pikk lühiajaline mälu (Long Short-Term Memory)
NLP	loomulik keele töötlemine (Natural Language Processing)
NMT	neurotõlge (Neural Machine Translation)
RNN	rekurrentne närvivõrk (Recurrent Neutral Network)
SQL	struktuurpäringukeel (Structured Query Language)

SISSEJUHATUS

Lõputöö eesmärgiks on luua subtiitrite veebipõhine masintõlkeplatvorm, mida hiljem magistriastmes õpetada saaks.

Veebipõhised ettevõttesisesed tõlkeplatvormid aitavad piirata piraatlust ning tõsta tõlgete tootmist ja kvaliteeti. Enamus tänaseid platvorme eeldavad filmi- ja subtiitrifaili saatmist tõlkijale ja toimetajale, mis võimaldavad filmi, teleseriaali või valmis tõlgete keelatud levitamist.

Idee ei ole uus, suurematel tõlkebüroodel on oma platvormid olemas. Autor ise on töötanud juhendite tõlkimisel masintõlkega, kus masinat on õpetatud spetsiifilisel alal kasutama õigeid väljendeid ja seeläbi aitab tõlkijal kiiremini edasi edeneda. Subtiitrite puhul on tegemist olnud lihtsalt platvormiga, kus inimene peab ise kogu teksti tõlkima.

Tänapäevane masin ei suuda veel inimest täielikult asendada. Keelte iseärasused, kohalikud väljendid ja lausete mitmeti mõistmisi saab õigesti tõlgendada ja tõlkida ainult inimene. Kui aga õpetada masinat konkreetse teleseriaali keelekasutusega, võiksid tulemused olla juba märkimisväärselt paremad. Siiski säilib ka inimese roll tõlgetes, vähemalt seni, kuni tehisintellekt õpib videot vaatama, teksti kuulama ja emotsioone eristama.

Käesolevas töös võrreldakse mõningaid olemasolevaid tööluarakendusi ja veebipõhiseid rakendusi. Lõputöö suurem eesmärk on luua veebipõhine tõlkeplatvorm, rakendada masintõlget ja uurida masinõpet. Platvorm peab oluliselt vähendama inimese tööaega ühe tõlkega, õpetatud masina eesmärk on jätta inimesele ainult keelekorrektuuri osa, see kõik vähendab tootmisaega ja teenuse hind muutub tellijatele soodsamaks. See on oluline ühtlustamiseks Baltikumis küsitud hinda, sest filmide ja seriaalide õigused omandatakse tavaliselt kõigi kolme Baltiriigi turule, et muuta nende film meie tarbijatele taskukohasemaks.

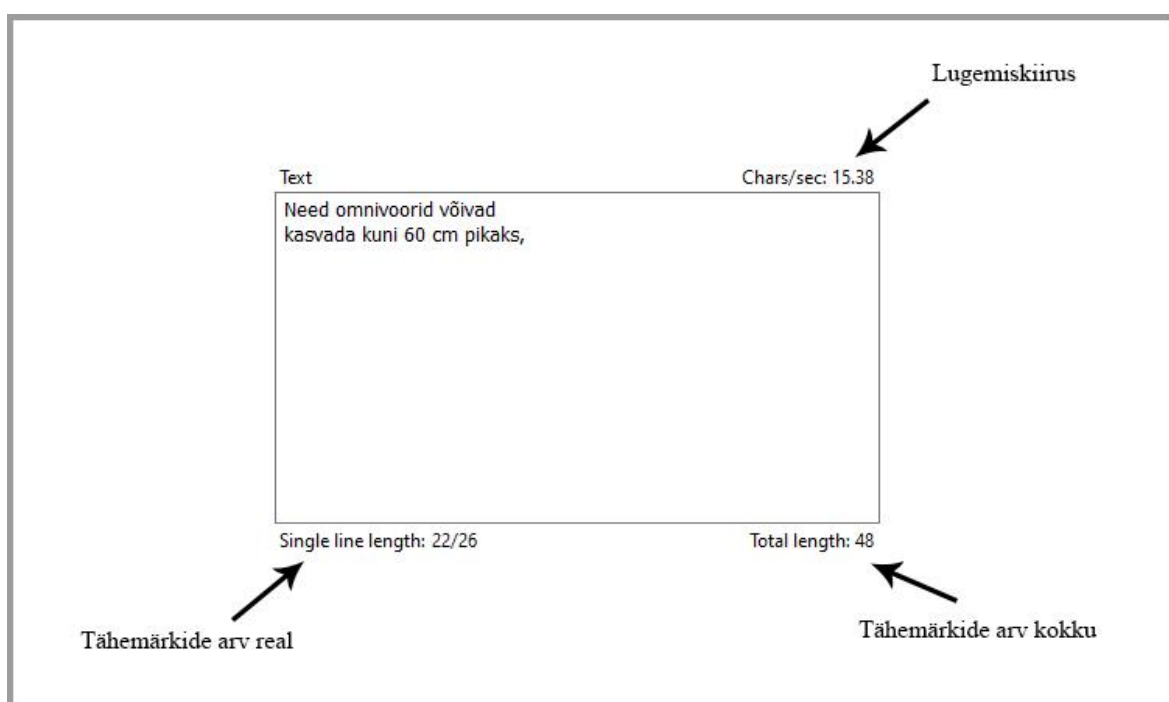
Võtmesõnad: subtiiter, masintõlge, veebiplatvorm, bakalaureusetöö

1. SUBTIITRID

Subtiitrite tõlkimine ja masintõlge on keerukad valdkonnad, mis hõlmavad erinevaid tehnikaid ja lähenemisi. Subtiitrite tõlkimisel on mitmeid aspekte, mida tuleb arvesse võtta, et tagada tõlke kvaliteet ja ajakohasus.

Oluline on kultuuriline tõlkimine ehk lokaliseerimine. Subtiitrite tõlkimine ei ole tavaline sõna-sõnalt tõlkimine, vaid nõuab kõnekäändude ja väljendite kohandamist sihtkeele kultuuriruumiga.

Arvestama peab ruumi- ja ajapiirangutega. Tähemärkide arv ja ajakood sunnivad tihti tõlkijat teksti sisu oluliselt lühendama (vt Joonis 1.1). Seetõttu peab tõlk oskama tekstist olulisema välja valida ja sellest normaalse(d) lause(d) looma.



Joonis 1.1 Subtiitri ülesehitus <https://interlex.ee/subtiitrite-tolkimise-telgitagustest/>

Kahjuks on neid, kes kasumi teenimise nimel jagavad valeinformatsiooni. /... Korraliku algmaterjali puudumisel on väga töömahukas subtiitrite ajastamine ehk ajakoodide paikapanek, mis võib võtta kordades rohkem aega kui tõlkimine ise.../ [1]. Õigesti ülesehitatud platvormiga on see töö kiire ja lihtne.

Tähtis on tõlkimisel originaalkeeles subtiitrifaili ja video olemasolu. Ideaalis on tõlkijal ligipääs ka filmi käsikirjale, mis aitab mitmeti mõistetavaid olukordi õigesti tõlkida.

Interlex OÜ kodulehe artikkel "Subtiitrite tõlkimise telgitagustest" käsitleb subtiitrite tõlkimise keerukust ja eripära. Artiklis tuuakse välja, et subtiitrite tõlkimine on sarnane ilukirjanduse tõlkimisega, kuna see hõlmab tihti kõnekäändude ja väljendite

lokaliseerimist ja kohandamist sihtkeele kultuuriruumiga. Lisaks rõhutatakse artiklis, et subtiitritõlkijal peab olema lai silmaring, et mõista ja tõlkida erinevaid teemasid ja kontekste, milles sõnu ja väljendeid kasutatakse. [2]

Masintõlke valdkonnas on AI kasutamine laialdaselt levinud. AI-põhised masintõlketööriistad on muutumas üha populaarsemaks. Paljud professionaalsed tõlkijad on endiselt skeptilised, kuigi tunnistatakse, et masintõlge on teinud olulisi edusamme. Lõpuks ei ole enam tõlki vaja, piisab ainult keeleteoimetajast.

Algasemel masintõlge asendab üldiselt ühes keeles sõna teise keele sõnaga. Rahuldava tõlke saamiseks peab masin mõistma terveid fraase. Nt *Google* tõlge kasutab protsesse nagu „statistika, et valida antud fraasi jaoks parim tõlge“ või „struktureeritud reeglid kõige tõenäolisema tähenduse valimiseks“. Suurimad edusamme teeb AI neuronmasintõlke kasutamisega. Siin kasutatakse süvaõpet, kus masin töötleb terveid lauseid ja kasutades „murdosa statistiliste meetodite jaoks vajalikust mälust, nii saab töö palju kiiremini tehtud. [3]

1.1 Allalaetavad tõlkeplatvormid

1.1.1 FAB Subtitler

FAB Subtitler (F.A. Bernhardt GmbH) pakub terviklikku subtiitrite süsteemi, mis katab laia valikut subtiitrite rakendusi.

- *FAB Subtitler STD* sobib subtiitrite ettevalmistamiseks ja toetab üle 90 erineva failiformaadi importi ja eksporti.
- *FAB Subtitler PRO* toetab subtiitrite käsitsi edastamist, SDI-kaarte ja automaatseid failipõhiseid operatsioone.
- *FAB Subtitler LIVE* on mõeldud otse-eetris subtiitrite loomiseks.
- *FAB Subtitler BCAST/XCD* on automatiseeritud süsteemide poolt juhitud subtiitrite edastamiseks.
- *FAB Subtitler MPEG* ja *FAB DVB Interface* on samuti subtiitrite edastamiseks, vastavalt MPEG ja DVB standarditele.
- *FAB SDI PCIe* kaarti kasutatakse avatud subtiitrite, teleteksti subtiitrite ja suletud pealkirjade sisestamiseks SDI-sse.
- *FAB Subtitle Database* on mõeldud subtiitrifailide salvestamiseks ja töövoogude korraldamiseks. [4]

Failide transkribeerimise osas toetab *FAB Subtitler PRO/LIVE/MPEG* pilvepõhiseid kõnetuvastusteenuseid. Seda teenust kasutatakse videofailidest audio

transkribeerimiseks subtiitrifailidesse teksti ja ajakoodidega. Protsess hõlmab heli edastamist videofailidest pilve. *FAB Subtitler* ühildub mitme pilve transkribeerimisteenusega, sealhulgas *Microsoft Video Indexer*, *Scriptix Speech to Text* ja *Google Speech To Text*. Tuvastatud teksti kvaliteet nõuab tavaliselt mõningaid käsitsi parandusi, kuid nende teenuste kasutamine võib oluliselt vähendada subtiitrite ettevalmistamiseks vajalikku aega. Näiteks *Microsoft Video Indexeri* puhul võib *FAB Subtitler* kuvada heligraafikus stseenimuutusi ja erinevate kõnelejate teksti erinevates värvides, kui see on vastavalt seadistatud. [5]

FAB Subtitler on integreerinud AI oma subtiitrite loomise protsessi läbi pilvepõhiste kõnetuvastusteenuste, hõlbustades nii failide kui ka reaalajas transkribeerimist ning toetades laia keelte valikut. See tehisintellekti integratsioon parandab subtiitrite loomise ja edastamise efektiivsust ja paindlikkust.

1.1.2 SubtitleEdit

SubtitleEdit on mitmekülgne tarkvara, mis on loodud nii algajatele kui ka ekspertidele subtiitrite redigeerimise valdkonnas. Tegemist on vabavaralise tööriistaga, mis hõlbustab subtiitrite loomist, redigeerimist ja konverteerimist. *SubtitleEdit* toetab kõiki populaarseid subtiitriformaate, võimaldades kasutajatel hõlpsalt lisada subtiitreid mis tahes videofailile. Tarkvara pakub täiustatud funktsioone, nagu automaatne sünkroniseerimine ja õigekirjakontroll, mis aitavad kokku hoida aega ja tagavad täpsuse.

SubtitleEdit pakub automaatset tõlget erinevate API-dega nagu *Google Translate V1* ja *V2 API*, *Bing Microsoft Translator*, *DeepL translate*, *LibreTranslate* ja *ChatGPT*. *Google Translate V1 API* on koheselt kasutatav, teiste jaoks on vaja API võtit, mille saab ise sisestada ja on üldiselt tasuline.

SubtitleEdit on väärtuslik tööriist erinevatele professionaalidele, kes sõltuvad oma töös või äris subtiitritest. Tarkvara pakub mitmeid funktsioone, mis on eriti kasulikud subtiitrite automaatseks sünkroniseerimiseks videofailidega, säästes oluliselt aega ja tagades, et subtiitrid vastaksid täpselt vastavale kõnele. *SubtitleEdit* võimaldab kasutajatel hõlpsalt luua ja redigeerida subtiitreid mitmes formaadis.

Peale nende funktsioonide võimaldab *SubtitleEdit* kohandada seadeid, pakkudes kasutajatele vabadust isikupärastada oma subtiitrite redigeerimise kogemust. *SubtitleEdit* on võimas ja kasutajasõbralik tööriist, mis on hädavajalik kõigile, kes vajavad subtiitrite loomist või redigeerimist. Selle intuitiivse liidese ja täiustatud funktsioonidega on *SubtitleEdit* asendamatu tööriist subtiitrite kiireks ja täpseks lisamiseks videotele.[6][7]

1.1.3 AegiSub

Aegisub on tasuta ja avatud lähtekoodiga subtiitrite loomise ja redigeerimise programm, mida kasutatakse laialdaselt kogu maailmas. Selle peamises aknas on kaks olulist ala: subtiitrite ruudustik ja redigeerimisaken. Ruudustikus kuvatakse kõik subtiitri failis olevad read erinevate värvidega, mis näitavad nende omadusi. Redigeerimisaknas saab subtiitrite teksti ja muid omadusi otse muuta. Aegu saab sisestada ajakastidesse, kuid tavaliselt on kiirem ajastada read kasutades helikasti, mis võimaldab täpset ja kiiret sünkroniseerimist heliga. Videoaken võimaldab näha, kuidas subtiitrid videol välja näevad ning võimaldab valida konkreetseid videokaadraid, mille peale subtiitrid algavad/lõppevad, kasutada hiirt subtiitrite rea positsioneerimiseks, pööramiseks ja kärpimiseks. [8]

1.2 Veebipõhised tõlkeplatvormid

1.2.1 Subtitles Translator

Subtitles Translator on tasuta online-tööriist, mis võimaldab tõlkida kuni 20 subtiitrifaili korraga 10. erinevasse keelde. Keeli, millesse tõlkida saab, on üle 80. Toetab kõige tähtsamaid failiformaate. *Subtitles Translator* pakub ka subtiitrite redigeerimise ja konverteerimise võimalusi, võimaldades kasutajatel teha muudatusi tõlgitud subtiitrites või muuta subtiitrifailide formaate. [9] Lehel ei pea end kasutajaks registreerima, see tekitab turvalisuse küsimuse.

1.2.2 Kapwing

Kapwing on subtiitrite lisamise ja redigeerimise platvorm, mis kasutab AI-tehnoloogiat, et muuta subtiitrite lisamine igale videole lihtsaks. Subtiitrideid ei pea käsitsi sisestama ega üles laadima faile, kuigi see võimalus on olemas. Subtiitrite loomiseks peab lihtsalt video üles laadima ja valima „Auto subtitles“. Seejärel loob *Kapwing* automaatselt sõnasõnalt subtiitrid video dialoogi või jutustuse põhjal.

Kapwing lubab redigeerida teksti, ajastust ja subtiitrite välimust nagu dokumenti, võimaldades muuta fonte, värve ja lisada animeeritud efekte. *Kapwing* toetab ka mitmekeelseid subtiitrideid, võimaldades neid tõlkida üle 70. keelde. Subtiitrideid saab alla laadida transkripti või subtiitrifailina. [10]

Platvorm tundub olema mänguline. Kui ei pea seda tööks kasutama, on kindlasti lõbus proovida ja katsetada erinevaid võimalusi. Platvorm on tasuline. Tasuta versioonis on väga väikesed mahud.

1.3 Tõlkeplatvormide kokkuvõte

Kuigi on olemas väga häid ja kvaliteetseid tööriistu subtiitrite tootmiseks, eeldab iga programmiga töötamine videote edastamist tõlkijatele ja kui on olemas, siis ka keelekorrektoritele. Subtiitri tellijad vähendavad videofailide suurust miinimumini ja kasutavad ka vesimärki videotel piraatluse piiramiseks. Siiski on failide saatmine mahukas ja aeganõudev ning iga uus inimene, kellele failid saadetakse, on risk piraatlusele.

Veebipõhised platvormid on rohkem tavakasutajale, nendega saab luua lustakaid subtiitreid ja tekste videotele.

Töölauaprogrammidest autori enda kogemuse põhjal on FAB Subtitler ja SubtitleEdit subtiitritööstuse parimad tööriistad. Nendega töötamine on lihtne, loogiline ja kiire. On võimalik kasutada masintõlget erinevatelt pakkujatelt või lisada oma tõlkemudel.

Just need 2 programmi on aluseks subtiitrite veebipõhise masintõlkeplatvormi loomisel.

2. SUBTIITRITE VEEBIPÕHISE MASINTÕLKEPLATVORMI LOOMINE

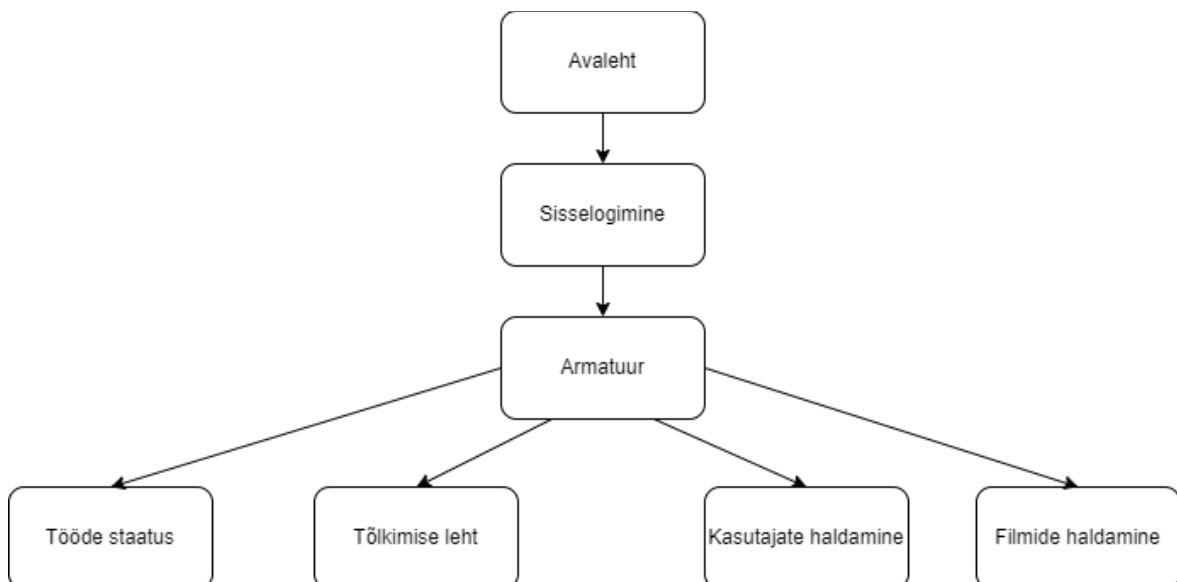
Autor lähtus platvormi loomisel Eesti turu vajadustest. Puudus on kvaliteetsetest tõlkijatest, tähtajad on liiga lühikesed ja tõlgete kvaliteet kannatab.

Veebipõhise platvormi suurteks plussideks on masineeltõlgitud tekst ja puudub vajadus failide liigutada. Platvormi haldajaks on ettevõtte, kes tegeleb nii videofailide kui ka algkeele ja sihtkeele subtiitrifailide üles ja allalaadimisega platvormile. Tõlkija/keelekorrektor ei pea failidega enam vaeva nägema ning väheneb oht failide ebaseaduslikule levitamisele. Autori kogemusest aitab selline platvorm ka hilinenud tõlgetega õigeaegselt tegeleda.

2.1 Veebipõhise masintõlkeplatvormi struktuur

Veebipõhise masintõlkeplatvormi struktuur (vt Joonis 2.1):

- avaleht;
- sisselogimine;
- tööde staatus;
- kasutajate lisamine / kustutamine;
- tööde lisamine / allalaadimine / kustutamine
- tõlkimise lehekülg.



Joonis 2.1 Subtiitrite veebipõhise masintõlkeplatvormi ülesehitus

2.1.1 Avaleht ja sisselogimine

Kuna veebipõhine masintõlkeplatvorm ei ole avalikuks kasutamiseks, puudub igasugune vajadus avalehel tutvustada lehe sisu või eesmärki, piisab ettevõtte kontaktidest ja sisselogimise nupust. Sisselogimise nupp suunab kasutajad autentimise lehele.

2.1.2 Tööde staatus

Peale edukat sisselogimist avaneb vaade, kus on näha tegemata ja tehtud tööd. Vastavalt kasutaja rollile näeb ta kas kõiki töid, pooleli olevaid töid ja uusi töid, millele otsitakse tegijat, kinnitusnupp lisab selle tegija vaatesse. Lõpetatud tööd seevastu kinnitatakse ja liiguvad lõpetatud tööde hulka, mida see kasutaja enam muuta ei saa, hiljem kustutab administraator ka tehtud tööd.

Välja on vaja töötada tööde arhiiv. Koostöös kasutajatega peab otsustama arhiivi pikkuse: kuni tehtud tööde eest tasu saamiseni või saab vaadata ajaloost kõiki kasutaja tehtud töid. Samuti ka tööde maksumuse kuvamine.

2.1.3 Tõlkimine

Filmide redigeerimisele on kõigil kasutajatel ligipääs vastavalt määratud filmidele, administraatoril on juurdepääs kõikidele filmidele.

Redigeerimise lehel on videovaade ja heliriba, nupud video/heli mängimiseks, pausile panekuks, stopp nupp viib tagasi filmi algusesse, kerimisnupud kaadri haaval edasi ja tagasi ning sekundi kaupa edasi ja tagasi kerimiseks. Kõik nupud on juhitud klaviatuuril numbriklahvistiku abil, nii on töö kiirem ja sujuvam.

Videol jooksevad subtiitrid nii, nagu need välja peavad nägema: tavatekst, kursiivis, subtiitri poolitus. Heliribal on näha subtiitri asukohta ja pikkust. Vastavalt kasutaja õigustele saab muuta subtiitri asukohta, pikkust, liita või poolitada subtiitreid.

Eraldi väljad on subtiitri algusajale, lõppajale ja kestvusele. Kestvusaeg muudab vastavalt värvi, kas subtiiter on liiga pikk, vastab ajale või ajakood ise on liiga pikk.

Kaks suurt tulpa on originaalsubtiitri vaateks ja tõlkimiseks. Redigeerida saab kõiki ridu tõlkesubtiitri tulbas ühe subtiitri kaupa lubades näha nii eelmisi kui järgmisi ridu teksti paremaks mõistmiseks.

Lehevaate täpne asetus ja funktsionaalsus on katsetusjärgus.

2.1.4 Kasutajate lisamine

Kasutajate lisamine ja kustutamine on määratud rolliga. Administraator annab uuele kasutajale ligipääsu veebipõhise masintõlkeplatvormile ning määrab tema õigused. Kasutajanimi on meiliaadress ja salasõna genereeritakse automaatselt. Esimesel

sisselogimisel kasutaja muudab salasõna. Salasõna peab olema vähemalt kuni 8 tähemärki, sisaldama suurt ja väikest tähte, numbrit ning erisümboleid.

Kasutaja kustutamiseks küsib veebipõhine masintõlkeplatvorm administraatorilt kinnitust salasõnaga, et vältida juhuslikku kasutaja kustutamist.

2.1.5 Filmide lisamine

Administraator saab lisada veebipõhisele masintõlkeplatvormile filmi ja selle originaalsubtiitri. Uued tehnoloogiad on võimelised ka ise helifaili (video) järgi looma originaalkeeles suhteliselt täpsed subtiitrid. Originaalsubtiitrite puudumisel peab olema platvormil võimalus need luua.

Peale video ja originaalsubtiitri lisamist tõlgib masin teksti sihtkeelde ja see salvestatakse tõlgile redigeerimiseks.

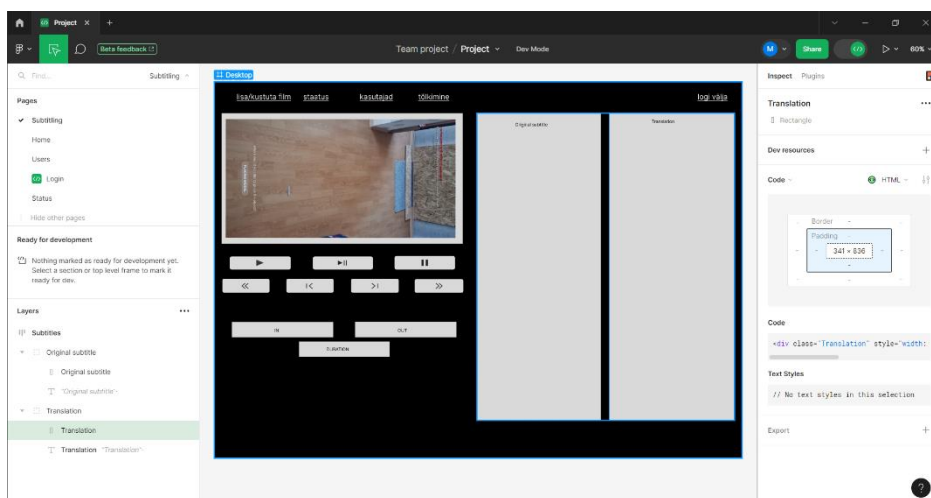
Kui tööd on valmis, saab administraator tõlked alla laadida ja saata tööde tellijale.

2.2 Figma kasutamine kujunduse loomisel

Figma on visuaalse disaini tööriist, mis võimaldab luua kasutajaliidese kujundusi, prototüüpe ja graafikat.

Lõputöös on kasutatud *Figma* rakendust, et kujundada veebirakenduse visuaalset poolt. Lõputöö autoril oli visioon veebirakenduse väljanägemisest (vt Joonis 2.2). *Figma* aitab konteinerite paigutuste ja koodiga.

Figma kõiki võimalusi veebirakenduse loomisel kasutatud ei ole, aga edasiarendamisel saab kasutada nt meeskonnatööd disaini täiendamisel ja funktsioonide väljatöötamisel võimalike partnerite ja/või kasutajatega.



Joonis 2.2 *Figma* abil tehtud veebipõhise masintõlkeplatvormi kujundus

2.3 JavaScript'i kasutamine koodi kirjutamisel

JavaScript on dünaamiline programmeerimiskeel, mida kasutatakse peamiselt veebirakenduste arendamiseks.

Selles lõputöös kirjutatakse JavaScripti abil nii *front-end* (vt Joonis 2.3) kui ka *back-end* loogika (vt Joonis 2.4). Tööd hõlmavad funktsioonide kirjutamist, andmete haldamist, API-dega suhtlemist ja kasutajaliidese interaktiivsuse lisamist.

```
function displaySubtitles(subtitles, containerId) {
  const container = document.getElementById(containerId);
  if (!container) {
    console.error("Konteinerit ei leitud:", containerId);
    return;
  }
  container.innerHTML = "";

  subtitles.forEach((subtitle, index) => {

    const textarea = document.createElement("textarea");
    textarea.className = "textarea-fixed-height";
    textarea.readOnly = true;
    textarea.dataset.index = index;
    textarea.dataset.startTime = subtitle.startTime;
    textarea.dataset.endTime = subtitle.endTime;
    textarea.dataset.duration = calculateDuration(
      subtitle.startTime,
      subtitle.endTime
    );
    textarea.value = `${subtitle.id}   ${subtitle.startTime} - ${subtitle.endTime}\n${subtitle.text}`;

    textarea.id = `subtitle-textarea-${index}`; // Unikaalne ID
    textarea.addEventListener("dblclick", function () {
      if (containerId === "trans_sub") {
        temp = `${subtitle.id}   ${subtitle.startTime} - ${subtitle.endTime}`;
        this.value = `${subtitle.text}`;
        this.readOnly = false;
      }
    });
    textarea.addEventListener('blur', function() {
      if (containerId === "trans_sub" && temp.length!=0) {
        // Prevent default behavior of moving to the next element
        event.preventDefault();
        this.value = temp + `\n${this.value}`;
        this.readOnly = true;
        temp=''
      }
    })
    textarea.addEventListener('keyup', function(event) {
      if (event.keyCode === 'Tab' && containerId === "trans_sub" && temp.length!=0) { // Check if the key is the Tab key

        // Prevent default behavior of moving to the next element
        event.preventDefault();
        this.value = temp + `\n${this.value}`;
        this.readOnly = true;
        temp=''
      }
    })
  });
}
```

Joonis 2.3 JavaScript *front-end* koodinäidis lõputöös


```

const express = require('express');
const cors = require('cors');
const path = require('path');
const sequelize = require('./db.config.js'); // Import Sequelize configured instance

const app = express();

//Use CORS
app.use(cors());

// Use express.static to serve static files from the 'Production' directory
// Test the connection to the database
sequelize.authenticate()
  .then(() => {
    console.log('Connection to the SQL server has been established successfully.');
```

Joonis 2.4 JavaScript back-end koodinäidis lõputöös

2.4 AWS serverteenuse uurimine võimaliku teenusepakujana

AWS pakub laia valikut pilveteenuseid, nagu serverite hostimine, andmebaasid, ladustamine ja palju muud [11].

Lõputöös kasutati AWS S3-e, mis pakub objektipõhist salvestusteenust (vt Joonis 2.5). Kuna S3 on mõeldud erinevat tüüpi ja suurusega andmete salvestamiseks, arhiveerimiseks ja varundamiseks, on see igati sobiv sellele lõputööle.

Kuna teenus on tasuta, saab tegeliku maksumuse S3-e kasutamisel alles mõne aja möödudes tegelike andmete ja filmidega.

```

import { PutObjectCommand, CreateBucketCommand } from "@aws-sdk/client-s3";
import { s3Client } from "../libs/s3Client.js";

// Set the parameters
const params = {
  Bucket: "sample-bucket-101", // The name of the bucket. For example, 'sample-bucket-101'.
  Key: "esimeneproovaws", // The name of the object. For example, 'sample_upload.txt'.
  Body: "SVID_20230503_170836_1.mp4", // The content of the object. For example, 'Hello world!'.
};

const run = async () => {
  // Create an Amazon S3 bucket.
  try {
    const data = await s3Client.send(
      new CreateBucketCommand({ Bucket: params.Bucket })
    );
    console.log(data);
    console.log("Successfully created a bucket called ", data.Location);
    return data; // For unit tests.
  } catch (err) {
    console.log("Error", err);
  }
  // Create an object and upload it to the Amazon S3 bucket.
  try {
    const results = await s3Client.send(new PutObjectCommand(params));
    console.log(
      "Successfully created " +
      params.Key +
      " and uploaded it to " +
      params.Bucket +
      "/" +
      params.Key
    );
    return results; // For unit tests.
  } catch (err) {
    console.log("Error", err);
  }
}

```

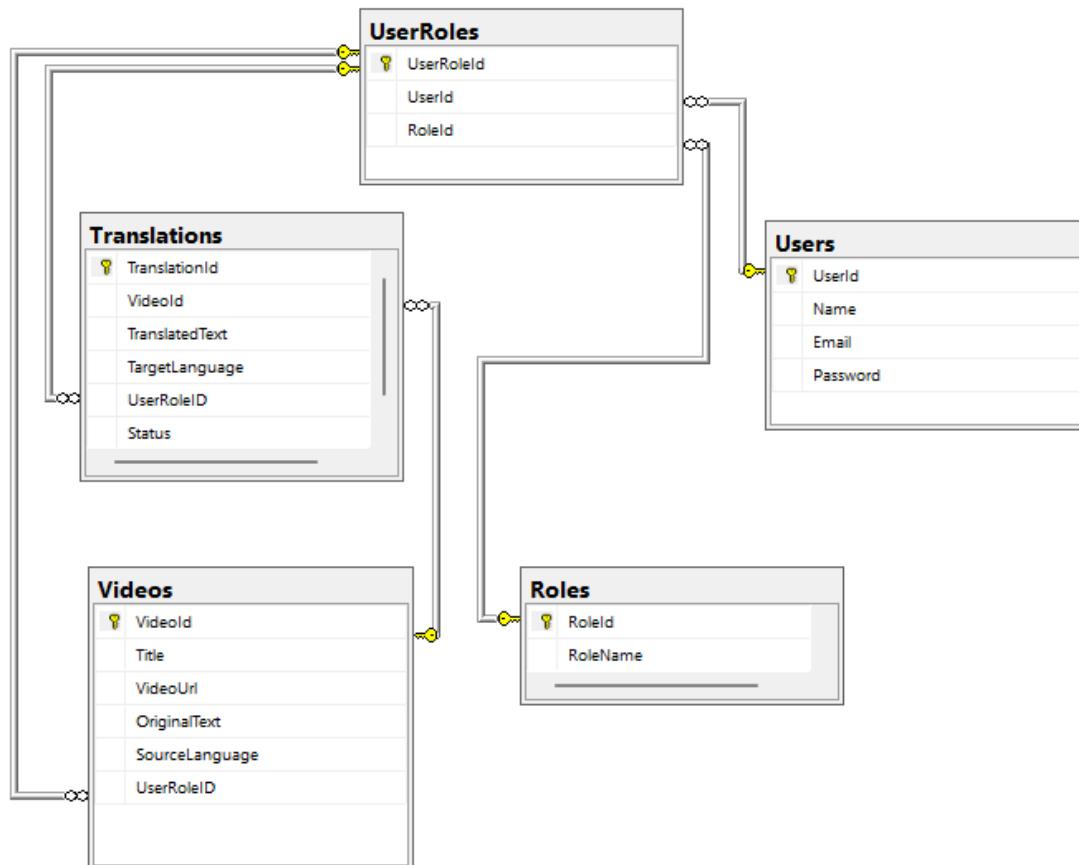
Joonis 2.5 AWS S3 koodinäidis lõputöös

2.5 Microsoft SQL Serveri kasutamine kasutajate loomiseks

SQL server on relatsiooniline andmebaasisüsteem, mida kasutatakse andmete salvestamiseks, pärimiseks ja haldamiseks.

Lõputöös kasutati SQL Serverit kasutajate registreerimiseks, õiguste andmiseks ja haldamiseks. Tööd hõlmavad andmebaasi skeemi loomist, tabelite ja seoste kujundamist ning päringute kirjutamist andmete manipuleerimiseks (vt Joonis 2.6).

Kindlasti tuleb uurida veel teisi võimalusi kasutajate hostimiseks. Lõputöös kasutati selleks TalTech Virumaa kolledži SQL serverit.



Joonis 2.6 SQL tabelid kasutajate salvestamiseks ja rollide jagamiseks

2.6 DeepL ja sellega ühendamine

DeepL kuulub masintõlke tehnoloogia valdkonda. See on masintõlke teenus, mille on arendanud Saksamaal asuv ettevõtte *DeepL GmbH*. *DeepL* kasutab süvaõppe (*deep learning*) algoritme ja tehisintellekti, et pakkuda kõrgetasemelist tõlget erinevate keelte vahel. *DeepL*-i masintõlke süsteem on tuntud oma täpsuse, loomulikkuse ja kontekstipõhise tõlkevõime poolest, mis sageli ületab teiste masintõlkeplatvormide, nagu *Google Translate* ja *Microsoft Translator*, tulemusi.

DeepL kasutab NMT tehnoloogiat (vt Joonis 2.7), mis on keerukam kui varasemad reeglipõhised või statistilised masintõlke süsteemid. NMT tehnoloogias arvestab programm tervet sisendlauset iga sammu juures, kui loob väljundlauset, mitte ainult mõnda sõna tõlgitava termini ümber. Tänu süvaõppe algoritmidele toodab *DeepL* inimlikumaid tõlkeid kui statistilised masintõlke mootorid. Selle võrgu arhitektuur võimaldab õppida suurtest andmehulkadest ja kohaneda uute kontekstidega. [12]

Just inimlikuma tõlke lubadus oli põhjuseks valida lõputöösse just DeepL tõlketeenus. Ühendamine oli lihtne. Koodis on näha (vt Joonis 2.8) ajakoodi kustutamist ja uuesti lisamist üleliigsete tähemärkide kulutamise tõkestamiseks.

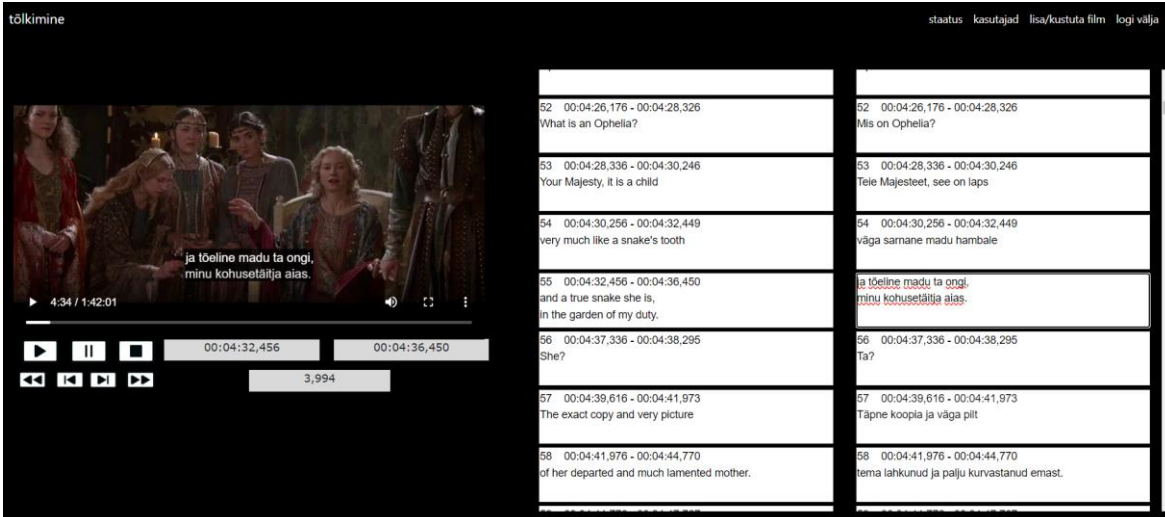
Lõputöö alguses on kasutatud DeepL tasuta versiooni, mis on piiratud tähemärkide arvuga, mida on kuus 500 tuhat, mis omakorda on olenevalt filmi pikkusest ja sõnalisest sisust umbes 1 filmi jagu. DeepL API Pro pakub kasutamispõhist hinda, millele lisandub kuutasu. Ning muidugi on olemas eraldi pakkumine ettevõtetele. [14]

Autor lõputöö tegemise käigus pidi kahjuks nentima, et tõlke kvaliteediga eriti rahule ei saanud jääda.

2.7 Veebipõhise masintõlkeplatvormi kokkuvõte

Veebipõhise masintõlkeplatvormi struktuur ei ole kindlasti veel valmis, kasutajate kogemused ja tagasiside puuduvad. Arendada on vaja materjalide edastamist tellijatega. Kas on olemas võimalus, et tellija laadib materjalid otse serverisse ja kas tellijad saavad ise valmis tõlked alla laadida või saadab administraator need otse veebiplatvormi vormilt tellijatele?

Platvorm peab peale valmimist suutma töötada sujuvalt ka suuremate failidega ega tohi olla mõjutatud töö tegija internetikiirustest või arvuti suutlikkusest. Veebipõhiste platvormide miinuseks võivad olla suured mahud, mis segavad sujuva töö tegemist. Välja tuleb arvutada miinimumnõuded, et vältida raisatud tööaega.



Number	Start Time	End Time	English Text	Estonian Text
52	00:04:26,176	00:04:28,326	What is an Ophelia?	Mis on Ophelia?
53	00:04:28,336	00:04:30,246	Your Majesty, it is a child	Teie Majesteet, see on laps
54	00:04:30,256	00:04:32,449	very much like a snake's tooth	väga sarnane madu hambale
55	00:04:32,456	00:04:36,450	and a true snake she is, in the garden of my duty.	ja tõeline madu ta ongi, minu kohusetäia aias.
56	00:04:37,336	00:04:38,295	She?	Ta?
57	00:04:39,616	00:04:41,973	The exact copy and very picture	Täpne koopila ja väga pilt
58	00:04:41,976	00:04:44,770	of her departed and much lamented mother.	tema lahkunud ja palju kurvastanud emast.

Joonis 2.9 Lõputöö tõlkelehe kasutajaliides

Lõputöö tegemise käigus on selgunud, et subtiitrite masintõlge on veel unistus. Subtiitrite ja tavateksti tõlge erineb teineteisest märgatavalt.

Kuna lõputöös kasutatakse masintõlkeks tähemärgipõhist tasulist tõlkeprogrammi, siis enne tõlkimist eemaldatakse subtiitri järjekorranumber ja ajakood. Tekst jääb ridade vahel poolitatuks nagu subtiitris, mitte ladusas tekstis ja seepärast ei mõista tõlkeprogrammid lausete täielikku ulatust ja tegelikku mõtet (vt Joonis 2.9). Lauseid lihtsalt kokku liita mõistliku tõlke saavutamiseks ei saa, sest siis pole masinal võimalik sellest uuesti subtiitrit kokku panna.

Selle töö suurim eesmärk saabki olema subtiitrikeeles õpetamine. Heaks algmaterjaliks on teleseriaalid, mis kestavad aastaid. Nii saab õpetada masinat tõlkima samaväärseid lauseid, masin teab vanemate osade järgi konkreetse sõna või lause tõlget.

Edasine töö oleks luua oma tõlkeprogramm, mis ei lõpeta lause lugemist enne kirjavahemärki. Nii suudab masin tõlkida terviklikku lauset, aga arvestades subtiitri eripära.

3. SUBTIITRITE VEEBIPÕHISE MASINTÕLKEPLATVORMI ÕPETAMISE VÕIMALIKKUS

Kogu lõputöö idee oli panna isiklik masin õppima parandatud lauseid.

Peale subtiitri kinnitamist salvestatakse algtekst ja sihttekst masinõppe programmi. Programmi eesmärk on õppida terveid lauseid. Hiljem saab masinalt ka lause tõlkesoovitusi. Nt tõlgitakse seriaali. Seriaalides vaadatakse tihti tagasi möödunud sündmustele, aga tekstist võetakse vaid lühike lõik. Tõlgid vahetuvad tihti ning osalisest tekstist on raske teha õigeid tõlkeid. Siin tuleb appi masin, kes annab soovitusi teksti tõlkimiseks eelneva sisu põhjal.

3.1 Masintõlke arengud

Masintõlke turg on märkimisväärses kasvus ja arengus. Selle turu laienemine peegeldab mitmekülgse sisu suurenevat nõudlust erinevates tööstusharudes nagu e-kaubandus, meedia, tervishoid ja reisimine, samuti edusammud loomuliku keele töötlemises ja masinõppes.

Turuväärtuse ja finantsinvesteeringute osas oli globaalne masintõlke turg 2021. aastal hinnanguliselt umbes 812,6 miljonit USA dollarit ja *GlobeNewswire*'i andmetel prognoositakse kasvavat umbes 4 069,5 miljoni USA dollarini aastaks 2030, kasvades 19,9% aastase kasvumääraga. Selle kasvu üheks ajendiks on pilvepõhiste masintõlketeenuste kasutuselevõtt ja mitmekeelse sisu suurenev nõudlus paljudes tööstusharudes. [15] Teised kanalid jäävad oma ennustuses veidi tagasihoidlikumaks, kuid suurt kasvu prognoosivad nemadki.

Võrreldes teiste valdkondadega on masintõlke turg muutumas üha konkurentsivõimelisemaks ja uuenduslikumaks, kusjuures ettevõtted ja valitsused investeerivad selle arendamisse märkimisväärselt. Need investeeringud ei ole ainult rahas, vaid ka uurimis- ja arendustegevuses, et parandada masintõlke täpsust, tõhusust ja rakendatavust erinevates keeltes ja valdkondades.

3.1.1 Tehisnärvivõrgud

RNN on tehisnärvivõrkude tüüp, mis sobib eriti hästi järjestikuste andmete, nagu keel või ajaseeriad, töötlemiseks. Erinevalt tavalistest *feedforward*-võrkudest, kus informatsioon liigub ainult ühes suunas, võimaldavad RNN-id informatsiooni tagasiulatuvat liikumist. See tähendab, et RNN võtab arvesse mitte ainult praegust sisendit, vaid ka sellele eelnenud sisendeid.

RNN-i põhielement on selle mälu, mis võimaldab võrgul säilitada teavet varasemate andmepunktide kohta. See mälu saavutatakse läbi tagasisidestatud ühenduste, mis

toovad väljundi osaliselt tagasi sisendkihti. See omadus muudab RNN-id ideaalseks valikuks selliste ülesannete jaoks nagu keeletõlge, kõnetuvastus ja ajast sõltuvate andmete analüüs.

Siiski on RNN-idel ka oma puudused. Näiteks on nad alati nn "lühiajalise mälu" probleemile, kus võrk kaotab võime meenutada varasemaid sisendeid. Selle probleemi lahendamiseks on välja töötatud täiustatud RNN tüübid nagu BRNN, LSTM ja GRU, mis on paremini võimelised säilitama informatsiooni pikaajaliselt. [16]

3.1.2 Loomulik keele töötlemine

NLP on arvutiteaduse, eriti AI valdkonna haru, mis võimaldab arvutitel mõista ja töödelda inimkeelt, nii teksti kui ka kõnet, viisil, mis sarnaneb inimese võimega keelt mõista ja tõlgendada.

NLP kombineerib arvutilingvistika, mis on inimkeele reeglipõhine modelleerimine, statistiliste, masinõppe ja süvaõppe mudelitega. Need tehnoloogiad võimaldavad arvutitel töödelda inimkeelt teksti või häälandmete kujul ja "mõista" selle täielikku tähendust, sealhulgas kõneleja või kirjutaja kavatsust ja emotsioone.

NLP on oluline tehnoloogia paljudes rakendustes, nagu masintõlge, kõnetuvastus, teksti kokkuvõtmine ja klienditeeninduse vestlusrobotid. NLP võimaldab arvutitel teha järgmist:

- Teksti ja kõne tuvastamine ja töötlemine: arvutid suudavad tuvastada ja mõista inimkõnet ning muuta selle tekstiks või vastupidi.
- Tõlkimine: automaatne tõlkimine ühest keelest teise.
- Emotsioonide analüüs: mõistmine, mis emotsiooni tekst väljendab.
- Olulise teabe ekstraheerimine: tuvastamine ja klassifitseerimine nimega üksustest, nagu isikunimed, kohanimed ja organisatsioonid.
- Küsimustele vastamine: süsteemid, mis suudavad tekstist leida vastuseid konkreetsetele küsimustele.
- Teksti kokkuvõtmine: loob teksti lühikese ja kokkuvõtliku esituse.

NLP mängib üha suuremat rolli nii tarbijate igapäevaelus kui ka ettevõtete ärioperatsioonide lihtsustamises, töötajate tootlikkuse suurendamises ja missioonikriitiliste äriprotsesside haldamises. [17]

3.2 Masina õpetamine

Lõputöö käigus loodud kood hõlmab andmete eeltötlust ja ettevalmistust masinõppe mudelite jaoks, eriti tehishälvõrkude ja süvaõppe kontekstis.

Funktsioon `clean_and_normalize_subtitles` (vt Joonis 3.1) on loodud subtiitrifailide puhastamiseks ja ettevalmistamiseks. See eemaldab HTML-laadsed sildid ja üleliigsed märgid, ignoreerib ainult numbreid või suurtähti sisaldavaid ridu ning ühendab tekstid ajatemplite põhjal blokkidesse. Eeltöötlus on iga masinõppe ülesande jaoks hädavajalik, et tagada mudelisse söödetaivate andmete ajakohasus ja puhtus.

```
def clean_and_normalize_subtitles(file_path):
    with open(file_path, 'r', encoding='utf-8') as file:
        lines = file.readlines()

    cleaned_text = []
    current_block = []

    for line in lines:
        # Eemalda HTML-sarnased sildid
        line = re.sub(r'<.*?>', '', line).strip()

        if re.match(r'\d{2}:\d{2}:\d{2},\d{3} --> \d{2}:\d{2}:\d{2},\d{3}', line):
            # Kui jõuad uue ajakoodini ja current_block ei ole tühi, ühenda ja lisa cleaned_text listi
            if current_block:
                cleaned_text.append(' '.join(current_block))
                current_block = [] # Tühjenda current_block uueks kasutamiseks
            elif line.isdigit():
                continue # Jäta reanumbrid vahele
            else:
                # Ignoreeri ridu, mis on kirjutatud ainult suurtähtedega
                if line.isupper():
                    continue
                if line: # Kui rida ei ole tühi, lisa see current_block listi
                    current_block.append(line)

    # Lisa viimane tekstiblokk, kui see eksisteerib
    if current_block:
        cleaned_text.append(' '.join(current_block))

    # Tagasta ühendatud tekstid, iga subtiitriblokk eraldi real
    return '\n'.join(cleaned_text)
```

Joonis 3.1 Andmete puhastamise koodinäidis lõputöös

Kood jaotab inglise ja eestikeelsed failid, et luua andmekogumeid masintõlkeks. Tekstid laaditakse ja `assert` avaldise abil kontrollitakse, et inglise ja eesti tekstiridade arv ühtiks. Seejärel jaotatakse andmed koolitus- ja testimisandmeteks (vt Joonis 3.2), kuigi hetkel on andmekogum väike ja tegelikke koolitus ja testandmeid kasutada ei saa.

```

from flask import Flask, request, jsonify
from flask_cors import CORS
from transformers import MarianMTModel, MarianTokenizer

app = Flask(__name__)
CORS(app)

model_name = 'Helsinki-NLP/opus-mt-en-et'
tokenizer = MarianTokenizer.from_pretrained(model_name)
model = MarianMTModel.from_pretrained(model_name)

@app.route('/translate', methods=['POST'])
def translate_text():
    data = request.get_json()
    input_text = data['text']
    batch = tokenizer([input_text], return_tensors="pt", padding=True, truncation=True, max_length=512)
    translated = model.generate(**batch)
    translated_text = tokenizer.batch_decode(translated, skip_special_tokens=True)[0]
    return jsonify({'translated': translated_text})

```

Joonis 3.2 *Python*'i koodinäide masintõlkeks

3.3 Ühendamine ja kasutamine platvormil

Andmete edastamiseks platvormil on kasutatud veebiteenust (API), see on üks levinumaid viise *Pythonis* loodud mudeli kasutamiseks ja andmete vahetamiseks *Javascriptis* või mõnes muus keskkonnas.

```

async function showTranslationSuggestion() {
    const textArea = document.getElementById('subtitle-edit-box');
    const suggestionBox = document.getElementById('suggestion-box');
    const translationSuggestion = document.getElementById('translation-suggestion');
    const text = textArea.value;

    if (text.length < 1) {
        suggestionBox.style.display = 'none'; // Peida soovituskast, kui tekstikast on tühi
        return;
    }

    // Kuvab soovituskasti õiges kohas
    suggestionBox.style.display = 'block';
    suggestionBox.style.left = `${textArea.offsetWidth + textArea.offsetLeft}px`;
    suggestionBox.style.top = `${textArea.offsetTop}px`;

    try {
        const response = await fetch('http://localhost:5501/translate', {
            method: 'POST',
            headers: {
                'Content-Type': 'application/json',
            },
            body: JSON.stringify({ text: text })
        });
        const data = await response.json();
        translationSuggestion.textContent = data.translated; // Kuvab tõlke soovituskastis
    }
}

```

Joonis 3.3 *JavaScripti* koodinäidis soovitusakna kuvamisest ja andmevahetusest

KOKKUVÕTE

Lõputöö "Subtiitrite veebipõhise masintõlkeplatvormi loomine" käsitleb innovaatilist tööd, mille eesmärk on luua veebipõhine masintõlkeplatvorm spetsiaalselt subtiitrite jaoks.

See lõputöö on oluline samm audiovisuaalse sisu tõlkimise valdkonnas, kuna keskendub subtiitrite ainulaadsetele nõudmistele ja väljakutsetele.

Lõputöö analüüsib olemasolevaid tõlkeplatvorme, nende võimalusi ja puudujääke, pakkudes ülevaadet masintõlke tehnoloogia arengust ja rakendamisest subtiitrite tõlkimisel.

Autor rõhutab vajadust kohandada tõlkeprotsessi, et see oleks kiirem ja efektiivsem, säilitades samal ajal tõlkimise kvaliteeti.

Lõputöö tehniline aspekt hõlmab masinõppe ja masintõlke tehnoloogiate integreerimist, et automatiseerida ja optimeerida tõlkimisprotsessi.

Lõputöö pakub tehnilist kirjeldust platvormi arhitektuurist, kasutatavatest tehnoloogiatest ja arendusprotsessi etappidest. See lõputöö esindab olulist sammu keeletehnoloogia ja audiovisuaalse sisu tõlkimise valdkonna arengus, pakkudes väärtuslikke lahendusi tõlkidele ja laiemale publikule, kes tarbib mitmekeelset sisu.

Andmete vähesuse tõttu jääb platvormi eesmärgiks teiste masintõlkeprogrammide tõlgitud tekstide kohandamine ja masina õpetamine. Selleks, et loobuda teiste tasuliste tööde kasutamisest, läheb vaja veel suurt hulka andmeid. Kiiremaks arenguks tuleb autori arvates alustada teleseriaalide tõlkimisega. See suurendab konkreetsete sõnade/lausete kasutamist ning tõhustab õpetamist.

SUMMARY

Bachelor's thesis: "Creating a web-based machine translation platform for subtitles"

Author: Monika Kiik

The bachelor's thesis deals with an innovative project aimed at developing a specialised web-based machine translation platform for subtitles. This thesis represents a significant advancement in the field of audiovisual content translation, focusing on the unique requirements and challenges associated with subtitling.

The thesis examines existing translation platforms, identifies their shortcomings and potential improvements, and provides insights into the development and application of machine translation technology specifically for subtitling. The author highlights the need to adapt the translation process to be faster and more efficient, while maintaining translation quality.

The technical aspect of the thesis involves the integration of machine learning and machine translation technologies to automate and optimise the translation process. The thesis provides a detailed description of the platform architecture, the technologies used and the stages of the development process. This thesis represents a significant step forward in the field of language technology and the translation of audiovisual content, providing valuable solutions for translators and a wider audience of multilingual content consumers.

The author highlights the limitations of modern machine translation, which cannot fully replace human translators due to the nuances and local idioms of languages. However, training the machine with the specific language used in the TV series can yield significantly better results. Human oversight remains crucial, particularly to ensure that translations capture the right context and emotion.

The thesis compares existing desktop and web-based subtitle translation applications. The main objective is to create a platform that integrates machine translation and machine learning to automate and optimise the translation process, reducing human work to primarily linguistic editing. This approach can reduce production time and costs, making translated content more affordable, especially in the Baltic market.

Due to a lack of data, the platform will initially adapt texts translated by other machine translation programs and train the machine. A significant amount of data will be needed to move away from the use of other paid work. The author suggests starting with the translation of TV series to increase the volume of words/phrases used and improve the training process. This approach is crucial for unifying service prices across the Baltic States.

KASUTATUD KIRJANDUSE LOETELU

1. „Filmide tõlkimine ja subtiitrid“. [Online] <https://www.unicom.ee/teenused/filmide-tolkimine-ja-subtiitrid/> (18.12.2023)
2. „Subtiitrite tõlkimise telgitagustest“. [Online] <https://interlex.ee/subtiitrite-tolkimise-telgitagustest/> (18.12.2023).
3. „Masinõpe aitab kaasa keerulisele tõlkekunstile“. [Online] <https://www.unite.ai/et/masin%C3%B5pe-tungib-keerukasse-t%C3%B5lkimiskunsti/> (18.12.2023).
4. „FAB Subtitling Products“. [Online] <https://www.fab-online.com/subtitling/> (18.12.2023).
5. „Using file transcription in FAB Subtiter“. [Online] <https://www.fab-online.com/subtitling/> (18.12.2023).
6. „Subtitle Edit“. [Online] <https://www.nikse.dk/subtitleedit> (18.12.2023).
7. „Subtitle Edit: A Comprehensive Review and Feature Analysis“. [Online] <https://ai-archive.org/ai-tools/subtitle-edit-review-and-features-of-this-powerful-subtitle-tool/> (18.12.2023).
8. „What is Aegisub?“ [Online] <https://aegisub.org/docs/latest/about/> (18.12.2023).
9. „Free online subtitle translation tool“. [Online] <https://subtitlestranslator.com/en/> (18.12.2023).
10. „Creat more content in less time“. [Online] <https://www.kapwing.com/> (18.12.2023).
11. „What is Amazon S3“. [Online] <https://docs.aws.amazon.com/AmazonS3/latest/userguide/Welcome.html> (18.12.2023).
12. „Exploring DeepL for Machine Translation: How It Works, and How Accurate it is“. [Online] <https://phrase.com/blog/posts/deepl/> (18.12.2023).
13. „What is Neural Machine Translation (NMT)?“ [Online] <https://omniscien.com/faq/what-is-neural-machine-translation/> (18.12.2023).
14. „Find your perfect plan“. [Online] <https://www.deepl.com/pro?cta=header-prices> (19.12.2023).
15. „Machine Translation Market Size to Touch USD 4,069.5 Million By 2030“. [Online] <https://www.globenewswire.com/news->

[release/2023/01/24/2594778/0/en/Machine-Translation-Market-Size-to-Touch-USD-4-069-5-Million-By-2030.html](https://www.ibm.com/topics/machine-translation-market-size-to-touch-USD-4-069-5-Million-By-2030.html) (18.12.2023).

16. „What are recurrent neural networks?“ [*Online*]
<https://www.ibm.com/topics/recurrent-neural-networks> (18.12.2023).
17. „What is natural language processing?“ [*Online*]
<https://www.ibm.com/topics/natural-language-processing> (18.12.2023).
18. OpenAI, „ChatGPT-4“ [*Online*] <https://openai.com> (2023, 2024).