REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO

EVROPSKA UNIJA
EVROPSKI SKLAD ZA REGIONALNI RAZVOJ
NALOŽBA V VAŠO PRIHODNOST

Razvoj slovenščine v digitalnem okolju – RSDO

Rezultat R2.6.1 in R2.6.2
Načrt za izdelavo robustnega splošnega razpoznavalnika (R2.6.1) in razpoznavalnika za področje izobraževanja s podporo delovanju v realnem času (R2.6.2)

verzija 1

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

Projekt / aktivnost / objava	RSDO, aktivnost A2.6, rezultat R2.6.1 in R2.6.2 (M28)
Vrsta dokumenta	Poročilo
Rezultat dostopen na	
Pripravili	Marko Bajec, UL-FRI Iztok Lebar Bajec, UL-FRI Simon Dobrišek, UL-FE Andrej Žgank, UM-FERI Darinka Verdonik, UM-FERI Mitja Rizvič, VITASIS
Datum priprave	Januar 2023

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

Uvod

V okviru projekta RSDO smo razvili štiri razpoznavalnike govora, dva splošna in dva specializirana za izbrane domene. Splošna razpoznavalnika dosegata bistveno nižjo napako pri razpoznavi (WER < 10%) kot prejšnji javno dostopni razpoznavalnik (WER ≈ 27%). Napaka pri domenskih razpoznavalnikih je še nižja, kar je pričakovano, saj imamo opravka z omejenim besediščem.

V tem dokumentu podajamo predloge za nadaljnji razvoj razpoznavalnikov slovenskega govora in podpornih orodij. Ker so nadaljnje aktivnosti v zvezi z razvojem robustnega splošnega razpoznavalnika precej prepletene z aktivnostmi razvoja razpoznavalnika za področje izobraževanja, podajamo predloge v skupnem dokumentu.

Odprti izzivi na področju razpoznave slovenskega govora

Kljub dobrim rezultatom projekta RSDO ostaja na področju razpoznave slovenskega govora še veliko odprtih izzivov. Ti so predvsem v povezavi s splošno razpoznavo. Domenske razpoznavalnike lahko namreč gradimo s specializacijo besedišča splošnega razpoznavalnika, s čimer se približamo jezikovnim specifikam domene. Bolj kot je domena jezikovno omejena (v smislu besedišča), boljša bo razpoznava. Pri gradnji splošnih razpoznavalnikov imamo težjo nalogo. Ne moremo se namreč zamejiti na specifične okoliščine, ampak moramo razpoznavalnik izdelati tako, da bo kos čim širšemu naboru možnih okoliščin in ne bo odvisen od govorca, značilnosti govora, kvalitete zvočnega zapisa in drugih, predvsem akustičnih okoliščin. Splošni razpoznavalnik mora biti robusten in sposoben natančne razpoznave tudi v situacijah, ki so za strojno razpoznavo zelo zahtevne (hkratni govor, govor v narečju ipd.).

Eden od izzivov pri razvoju tovrstnih razpoznavalnikov je vzdrževanje *slovarja* besed in izgovorjav, ki mora zajemati celotno besedišče. Pri splošnih razpoznavalnikih je besedišče tako-rekoč neomejeno in težko zagotovimo, da bodo vse besede v slovarju. Zato se pojavi vprašanje, ali pri splošni razpoznavi uporabljati tehnike, ki temeljijo na slovarjih, ali raje pristope, ki izvajajo razpoznavo na osnovi črk ali besednih podenot in kot taki ne potrebujejo slovarjev. Čeprav se zdi, da je odgovor na dlani – če ne moremo zagotoviti slovarja, je bolje uporabljati pristope, ki ga ne potrebujejo – vendar ni tako preprosto, saj se slovarski modeli v nekaterih primerih izkažejo kot bolj natančni.

Poleg zahteve po splošnosti in robustnosti velja kot izziv pri razvoju splošnih razpoznavalnikov omeniti tudi zahtevo po razpoznavi v realnem času. V praksi se pogosto srečamo s primeri, ko je potrebna takojšnja razpoznava, pri kateri nimamo časa ali možnosti, da bi govor najprej v celoti zajeli (posneli) in ga potem obdelali ter tako pridobili transkripte. Primeri, kjer pričakujemo takojšnjo razpoznavo, so denimo a) klepetalniki z govornim vmesnikom (angl. *Chatbot*), s katerimi se lahko pogovarjamo, b) govorno upravljanje v avtomobilih, c) narekovalniki (npr. v zdravstvu za narekovanje izvidov), d) podnaslavljanje govora itd.

Podnaslavljanje govora je z vidika pravočasne razpoznave še posebej zahtevno. Podnapis moramo namreč prikazati čim prej oziroma sinhronizirano z govorom ali video vsebino. Posledično imamo na voljo zelo malo časa in informacije (kratek časovni okvir govora) za izvajanje razpoznave. Pri večini drugih primerov uporabe razpoznave v realnem času nas ne moti, če se transkripti prikazujejo z manjšim zamikom (npr. ko končamo poved ali naredimo večji premor), pri podnaslavljanju v živo pa to pomeni preveliko oviro. Pri govoru včasih tvorimo zelo dolge povedi in v takih primerih bi s podnapisom zelo zamujali, če bi s formiranjem podnapisa čakali na zaključek povedi, sam podnapis pa bi bil posledično zelo dolg in bi ga težko prikazali v predpisanem formatu (podnaslavljanje ima predpisan format, dolžina podnapisov je omejena). Razpoznava za tvorjenje podnapisov je zato še precej zahtevnejša kot razpoznava v realnem času, saj se oddaljujemo od okoliščin, v katerih razpoznavalniki delujejo najbolje, to je takrat, ko izvajajo razpoznavo nad celotnimi povedmi. Zahtevnost razpoznave za potrebe podnaslavljanja se še poveča, če

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

upoštevamo, da pri ročni pripravi podnapisov običajno govora ne transkribiramo dobesedno, ampak ga pogosto prilagodimo vsebini in optimiziramo za bralca z željo, da bi pripomogli k razumevanju podnaslovljene vsebine. To je še posebej pomembno, ko podnaslavljamo za gluhe in naglušne osebe. Razpoznavalnik bi zato veljalo graditi tako, da bi znal obenem izvajati tudi prilagajanje in optimiziranje vsebine. Ali poskusiti z razvojem enojnega modela, ki deluje kot prevajalnik in govorni signal pretvarja neposredno v podnapise (z ločili in velikimi začetnicami). Da bi kaj takega lahko poskusili, potrebujemo velike učne množice.

Razpoznavalnike govora navadno uporabljamo skupaj z drugimi podpornimi orodji, kot sta na primer orodje za napovedovanje ločil in velikih začetnic ter orodje za denormalizacijo besedila (razpoznavalniki so večinoma učeni na korpusih, ki so predhodno normalizirani, tako da v njih ni s številkami zapisanih pojavnic ali okrajšav, akronimov in merskih enot.). V RSDO smo takšna orodja razvili in dobro delujejo v kombinaciji s t.i. *offline* razpoznavo. Za potrebe *sprotne razpoznave* ali razpoznave v realnem času pa jih ne moremo uporabiti, saj delujejo dobro le nad večjimi enotami besedila in kadar pri uporabi nismo časovno omejeni. Uporaba tovrstnih orodij v kombinaciji z razpoznavo v realnem času je precej zahtevnejša in zahteva povsem drugačne pristope pri razvoju.

Predlagane aktivnosti za nadaljnji razvoj

Pri nadaljnjem razvoju na področju razpoznave govora je smiselno nasloviti izzive, ki so opisani v prejšnjem poglavju in s tem zagotoviti pogoje za pospešen razvoj čim širšega spektra aplikacij z uporabo govornih tehnologij.

Pomemben korak na tem področju predstavlja že projekt RSDO, ki daje na voljo ključne gradnike za gradnjo tovrstnih aplikacij. Z njegovimi rezultati vsaj nekoliko zmanjšujemo zaostanek, ki ga beležimo na področju digitalizacije jezika in odpiramo možnosti za razvoj novih in inovativnih informacijskih rešitev, za katere poprej, zaradi zahtev po velikih začetnih investicijah in/ali ekspertnem znanju slovenska IT podjetja preprosto niso bila zainteresirana. Če želimo slediti trendom, ki napovedujejo, da bo v prihodnosti govor temeljni komunikacijski vmesnik med človekom in napravo, moramo poskrbeti, da se razvoj na tej točki ne ustavi in tudi sami, kot država, pospešeno in osredotočeno vlagamo v nadaljnji razvoj na tem področju. V nadaljevanju podajamo nekaj predlogov za nadaljevanje razvoja na področju razpoznave govora. Ti zajemajo:

1. Nadgrajevanje govorne baze
2. Dopolnjevanje slovarja in korpusov
3. Pridobivanje in dopolnjevanje virov na področju izobraževanja
4. Nadgradnja podpornih orodij
5. Izdelava novih modelov razpoznave
6. Razvoj podpore za razpoznavo v realnem času
7. Prilagoditev podpornih orodij za razpoznavo v realnem času
8. Razvoj dodatnih podpornih orodij
9. Razvoj algoritmov za podnaslavljanje

1. Nadgrajevanje govorne baze

Čeprav je pri razvoju razpoznavalnikov možno uporabljati tehnike *prenosa znanja* in s tem nekoliko omiliti problem majhnega obsega govornih virov v slovenščini, je za visoko natančnost razpoznave še vedno ključno, da imamo na voljo veliko govorno bazo. V okviru RSDO je bila pripravljena govorna baza v obsegu 1000 ur, kar je ogromno v primerjavi s tem, kar je bilo na voljo poprej. V primerjavi z jeziki z veliko govorci pa je to še vedno zanemarljivo malo - na primer zadnji znani model razpoznave govora za angleški jezik,

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

Whisper (https://openai.com/blog/whisper), je bil naučen na 438.000 urah označenega govora. Zato je pomembno, da se govorni viri za slovenščino še naprej ves čas nadgrajujejo. V ta namen je potrebno:

1. izdelati podroben izvedbeni načrt zagotavljanja posnetkov in transkripcij na ekonomično vzdržen, a učinkovit način (npr. prek množičenja in spodbujanjem uporabnikov za sodelovanje z nagradami),
2. zagotoviti zanj potrebno tehnično infrastrukturo (npr. portal za oddajanje posnetkov, strojna in programska oprema za hranjenje in urejanje posnetkov in transkripcij),
3. definirati in vzpostaviti delotok, ki bo zagotavljal pregledne in varne procese dela (preglednost podatkov in varnost pred njihovo izgubo ali vnosom napak sta bistvena elementa pri razvoju govornega vira),
4. zagotoviti vzdržen, a stabilen dolgoročni vir financiranja, ki se ne bo zaključil po nekaj letih.

V procesu pridobivanja novih posnetkov je potrebno tudi sprotno preverjanje kategorijo posnetkov, tako z vidika govorcev (vrsta govora, značilnosti govorcev,...), kot tudi akustičnih značilnosti (zvočna ozadja, karakteristike zajema,…). Na takšen način lahko zagotavljamo ustrezno heterogeno zastopanost posnetkov v korpusu, ki je pomembna za zagotavljanje robustnost delovanja razpoznavalnikov govora v različnih realnih situacijah. V nasprotnem primeru obstaja nevarnost, da bi sicer povečevali razpoložljive ure zajetega govora, ki pa bi bil po svojih značilnostih zelo podoben že zajetemu gradivu (npr. velik obseg branega govora). Poleg zagotavljanja novih posnetkov se je potrebno osredotočiti tudi na pridobivanje posnetkov, ki že obstajajo ali imajo celo na voljo tudi zapise govora (npr. zapisniki in posnetki obravnav s sodišč, podnapisi TV oddaj, parlamentarne razprave in razprave delovnih teles DZ ipd.), vendar zaradi avtorskih pravic ali GDPR do sedaj niso bili dostopni. V tej skupini velja še posebej izpostaviti zvočne knjige, ki bi jih bilo tudi možno uporabiti kot stalen viri novega transkribiranega govornega materiala. Že če bi zaradi zaščite avtorskih pravic v bazo vključili samo polovico vsake zvočne knjie (npr. vsako drugo stran), bi lahko na takšen način relativno preprosto zbrali več sto ur dodatnega govora. Primer podobne govorne baz zvočnih knjig v angleškem jeziku je LibriSpeech, ki danes velja za referenčno angleško bazo za razpoznavalnike govora.

V RSDO smo se ukvarjali s *standardizirano slovenščino*, kar pomeni, da so narečja zajeta zgolj naključno, večinoma le pri spontanem govoru. To ne zadošča za razvoj robustnega splošnega razpoznavalnika, za katerega pričakujemo, da bo kos tudi narečnemu govoru, vključno z govorom zamejskih Slovencev. Z vidika zagotavljanja robustnega delovanja razpoznavalnika govora z različnimi narečji je potrebno zagotoviti tako njihovo boljšo pokritost, kot tudi povečanje količine narečnega govornega materiala. Prav tako je treba usmerjeno zbirati gradiva za slovenščino kot drugi/tuji jezik. Pri tem je treba skrbeti za ustrezno pokritost materinih jezikov oz. jezikovnih skupin govorcev slovenščine kot drugega/tujega jezika, saj to ključno vpliva na delovanje razpoznavalnika govora. Tako narečja kot slovenščino kot drugi/tuji jezik je potrebno v govornih virih sistematično pokriti znotraj zgoraj navedenih postopkov nadgradnje virov.

2. Dopolnjevanje slovarja in korpusov

Pri številnih aplikacijah razpoznave govora se uporaba slovarja in jezikovnega modela še vedno izkaže kot dobrodošla. Še posebno to velja pri specializaciji na specifične domene, ki je brez uporabe slovarja in jezikovnega model bistveno zahtevnejša. V okviru RSDO smo za slovar uporabili *Sloleks* za korpus jezikovnega modela pa *Gigafido*. Oboje, terminološki slovar in korpus, je potrebno še naprej dopolnjevati, in pri korpusu močno povečati delež zapisov spontanega govora.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

3. Pridobivanje in dopolnjevanje virov na področju izobraževanja

Posebej velja omeniti področje izobraževanja, kjer je želja, da bi z uporabo strojne razpoznave in prevajanja podprli scenarije, kot so sprotno prevajanje slovenskih predavanj v tuje jezike, pomoč učencem v nižjih razredih osnovne šole, katerih materni jezik ni slovenščina ipd.

Sprotno strojno prevajanje je zelo zahtevna naloga, saj se vse napake pri razpoznavi in avtomatskem postavljanju ločil prenašajo naprej in vplivajo na kakovost strojnih prevodov. Zato so napake v razpoznavi zelo problematične in je potrebno zagotoviti, da jih je čim manj. Smiselno se je osredotočiti na dopolnjevanje virov, ki so potrebni za specializacijo jezikovnih modelov za posamezna izobraževalna področja. Na primer z rednim zbiranjem posnetkov predavanj (na vseh ravneh izobraževanja), njihovim transkribiranjem (ročnim in strojnim) ter sistematičnim dopolnjevanjem posameznih področnih jezikovnih modelov.

4. Nadgradnja podpornih orodij

Podporna orodja, ki se uporabljajo v fazi učenja razpoznavalnikov in/ali v fazi uporabe razpoznavalnikov je potrebno nadgraditi z novimi funkcionalnostmi.

1. **Grafemsko fonemski pretvornik**: grafemsko fonemski pretvornik je potrebno nadgraditi, da bo znal tvoriti izgovorjave tudi za posebne vrste pojavnic, kot so tujke, kratice in okrajšave. Pri pretvorbi tujk se moramo zanašati na strojno razpzonavo jezika, kar je trd oreh. Za daljša besedila lahko z veliko natančnostjo napovemo, v katerem jeziku so napisana, ko gre za posamezno besedo pa natančnost strmo pade. Razširitev je potrebna tudi za tvorjenje izgovorjav v različnih narečjih (glejte tudi 1. Nadgrajevanje govorne baze).
2. **Normalizacija in inverzna normalizacija ali denormalizacija**: tako za normalizacijo kot denormalizacijo velja, da ju lahko implementiramo s pomočjo pravil (s pravili opišemo pretvorbe). Alternativa je strojno učenje nevronskih mrež, ki se naučijo pretvarjanja. Čeprav še vedno velja, da so pretvorniki na osnovi pravil natančnejši, je njihovo vzdrževanje zaradi prepletanja pravil bistveno težje, kot vzdrževanje modelov nevronskih mrež. Zato velja za normalizacijo in denormalizacijo izdelati pretvornike, ki temeljijo na uporabi globokega učenja. Obenem je potrebno preizkusiti novejše načine za izdelavo normalizatorjev in inverznih normalizatorjev besedila na osnovi pravil, ki temeljijo na uporabi WFST (angl. weighted finite-state transducer) gramatik.
3. **Orodje za avtomatsko postavljanje ločil**: postavljanje ločil je bistveno težje za transkripte spontanega govora kot za transkripte branega govora. Učni korpus je potrebno povečati s transkripti spontanega govora in izdelati orodje za postavljanje ločil, ki bo specializirano za spontani govor.

5. Izdelava novih modelov razpoznave

Skupaj z nadgrajevanjem govorne baze, širjenjem slovarja in korpusov ter nadgradnjo podpornih orodij je potrebno občasno izdelati nove splošne razpoznavalnike in jih objaviti na javno dostopnem portalu ter v repozitoriju CLARIN.

6. Razvoj podpore za razpoznavo v realnem času

Razpoznava v realnem času je precej zahtevnejša kot *off-line* razpoznava in tudi postopki gradnje razpoznavalnikov za ti dve področji se pogosto razlikujejo. Zagotovo pa za izvajanje razpoznave v realnem času potrebujemo drugačne arhitekture programskih rešitev ali ogrodja, v katerih modeli razpoznave tečejo. V okviru te aktivnosti je potrebno razviti celovito rešitev za razpoznavo v realnem času (model razpoznave ter ogrodje informacijske rešitve za poganjanje modela).

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

7. Prilagoditev podpornih orodij za razpoznavo v realnem času

Tudi pri razpoznavi v realnem času bo potrebno transkripte opremiti z ločili in velikimi začetnicami ter besedilo denormalizirati. Orodja, ki se uporabljajo v kombinaciji z *off-line* razpoznavo, ne bodo več zadoščala, saj bodo okoliščine pri razpoznavi v živo drugačne. Postavljanje ločil je lahko zelo natančno, če imamo na voljo celotno besedilo, ki ga lahko podrobno analiziramo. Pri razpoznavi v realnem času pa žal te možnosti nimamo, saj želimo ločila postavljati sproti. Prav tako si pri razpoznavi v realnem času ne moremo privoščiti časovno potratnih analiz, ki se uporabljajo pri *off-line* razpoznavi (npr. stavčna analiza za potrebe denormalizacije), ampak se moramo zadovoljiti s pristopi, ki so sposobni dati rezultate že v delčku sekunde.

8. Razvoj dodatnih podpornih orodij

Razpoznava govorca

V kombinaciji z razpoznavo govora se pogosto uporablja tudi *razpoznava govorca*, katere naloga je napovedati, kdo v nekem trenutku (v posnetku govora) govori. Pri aplikacijah razpoznave govora nas najbolj zanimata dve tehniki: *razpoznava govorcev* in *detekcija spremembe govorca*.

Pri razpoznavi govorcev poskušamo ugotoviti, kdo so govorci, ki nastopajo v posnetku. Kadar gre za znane govorce (njihove *govorne odtise* poznamo), od algoritma pričakujemo, da jih bo identificiral, transkript, ki ga vrne razpoznavalnik, pa bo opremljen s podatki o govorcih, in sicer za vsako besedo transkripta posebej. Pri posnetkih z nepoznanimi govorci pa od algoritma pričakujemo, da bo ugotovil, koliko govorcev v posnetku nastopa in kdaj nastopa posamezen govorec.

Detekcija spremembe govorca

Detekcija spremembe govorca je drugačna naloga kot razpoznava govorcev. Pri detekciji spremembe govorca nas zanima le, kdaj se govorec zamenja, to je, kdaj začne govoriti drugi govorec. Izziv se zdi lažji kot razpoznava govorcev, vendar temu ni tako, saj nas detekcija spremembe govorca navadno zanima v povezavi z razpoznavo v realnem času (na primer pri podnaslavljanju, ko bi želeli v podnapisu označiti trenutek, ko začne govoriti drugi govorec). Detekcija spremembe govorca je tako zelo težka naloga, vendar izjemnega pomena pri podnaslavljanju in podobnih aplikacijah.

Akustična klasifikacija zajetega zvoka

V mnogih uporabniških scenarijih je pomembno ločiti med zvočnim posnetkom, v katerem je prisoten govor in posnetki, kjer ne gre za govor, ampak za drugačne, morda pomembne akustične dogodke. Če je v zvočnem posnetku na primer glasba s petjem, bo razpoznavalnik poskušal razpoznavati govor, četudi bo verjetnost pravilnosti napovedi majhna. Zato je pomembno, da znamo ločiti, kdaj govor nastopa in kdaj ne. Nakatera ogrodja, ki so na voljo za učenje razpoznavalnikov, imajo že sama vgrajen algoritem, ki detektira govor (angl. VAD - Voice activity detection), vendar se na to ne moremo zanašati. Smiselno bi bilo razviti ločen klasifikator, ki bi znal ločevati med govorm in “ne-govorom”. Za namene podnaslavljanja v živo pa je smiselno takšen klasifikator ne-govora razširiti še na različne podkategorije zvokov iz okolja (npr. zvok vozila, živali, stroja…). Njihovo označevanje v podnapisih lahko namreč gluhim in naglušnim doda pomembne informacije o dogajanju na posnetku.

Identifikacija sentimenta in čustev

Identifikacija sentimenta v besedilu je že precej raziskan izziv na področju procesiranja naravnega jezika. Za nekatere jezike pa so na voljo tudi modeli, ki sentiment napovedujejo iz govornih posnetkov s skupno analizo govornih in tekstovnih značilnosti. Modeli razpoznave, ki so nadgrajeni z modeli za identifikacijo sentimenta za vsako vrnjeno hipotezo (enoto razpoznanega teksta) podajo tudi podatek o identificiranem

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

sentimenu (npr. *pozitiven*, *negativen*, *nevtralen*). Za slovenski jezik takih modelov še nimamo in bi jih veljalo razviti.

9. Razvoj algoritmov za podnaslavljanje

Za učinkovito podnaslavljanje govora je potrebno razviti posebne modele razpoznave, ki so sposobni delovanja v realnem času in so s stališča razpoznave govora čim bolj *stabilni*. Modeli razpoznave v realnem času namreč delujejo tako, da analizirajo posamezne segmente govora in napovedujejo, katere besede so bile izgovorjene. Napovedi niso *fiksne*, ampak se z daljšanjem zajetega govora spreminjajo in popravljajo, dokler govorec na naredi dovolj dolge pavze oz. algoritem razpoznave na osnovi drugih kriterijev določi, da je govorna enota zaključena. Tedaj poda zadnjo oziroma *končno hipotezo* razpoznave. Končna hipoteza se od vmesnih hipotez pogosto razlikuje, kar je pri podnaslavljanju, kjer skušamo podnapise formirati s čim manjšim časovnim zaostankom, zelo moteče. Zelo velik izziv pri podnaslavljanju predstavlja tudi post-obdelava transkriptov (npr. postavljanje ločil ter velikih začetnic in denormalizacija besedila), saj je natančnost teh postopkov odvisna od količine besedila, ki ga lahko naenkrat obravnavamo. Pri podnaslavljanju je izjemno pomembno zasledovanje čim višje kvalitete razpoznave, in post-obdelave v čim krajšem možnem času.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.