



Ljubljana, 15. december 2022

Letno poročilo (2022)

Projekt Online Notes (ON)

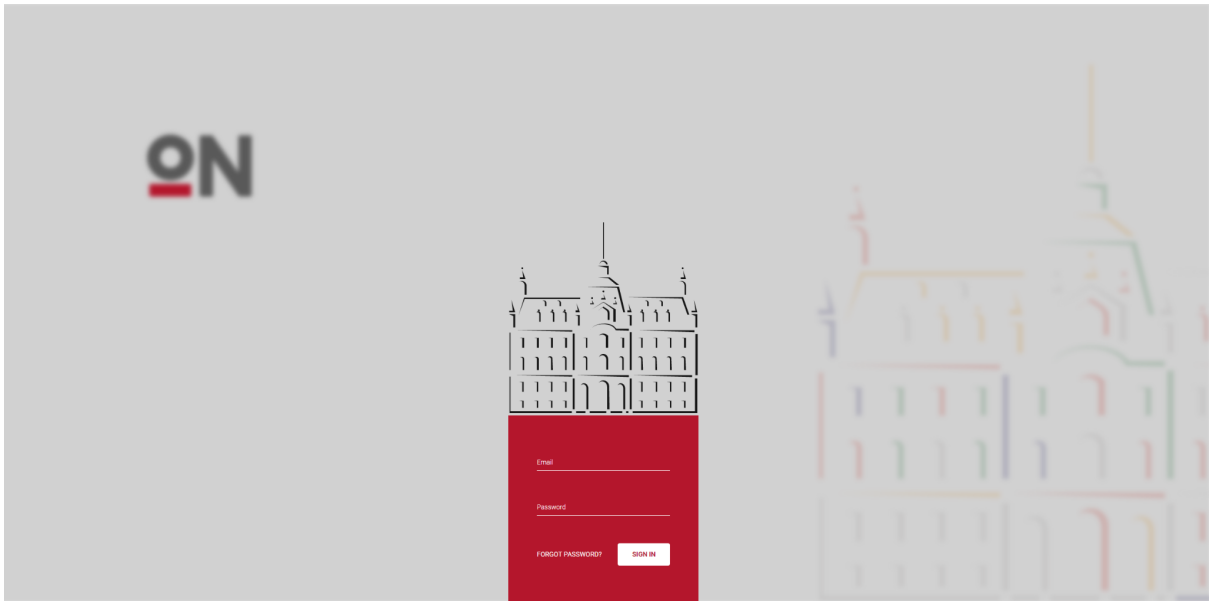
1 Uvod

V letu 2022 so na projektu Online Notes (ON) potekale aktivnosti tako na področju razvoja in izboljšave portala Online Notes (tehnični del; poglavje 2) kot tudi na področju izvajanja pilotnih predavanj na različnih članicah Univerze v Ljubljani (vsebinski del, poglavje 3). Izvedli smo tudi prve evalvacije med uporabniki sistema ON (profesorji, študenti in skupine oseb z omejitvami), kar predstavljamo v poglavju 4. V letnem poročilu povzemamo poglobitve dejavnosti in rezultate, v zaključku (poglavje 5) pa nakažemo točke za nadaljevanje dela v prihodnjem letu.

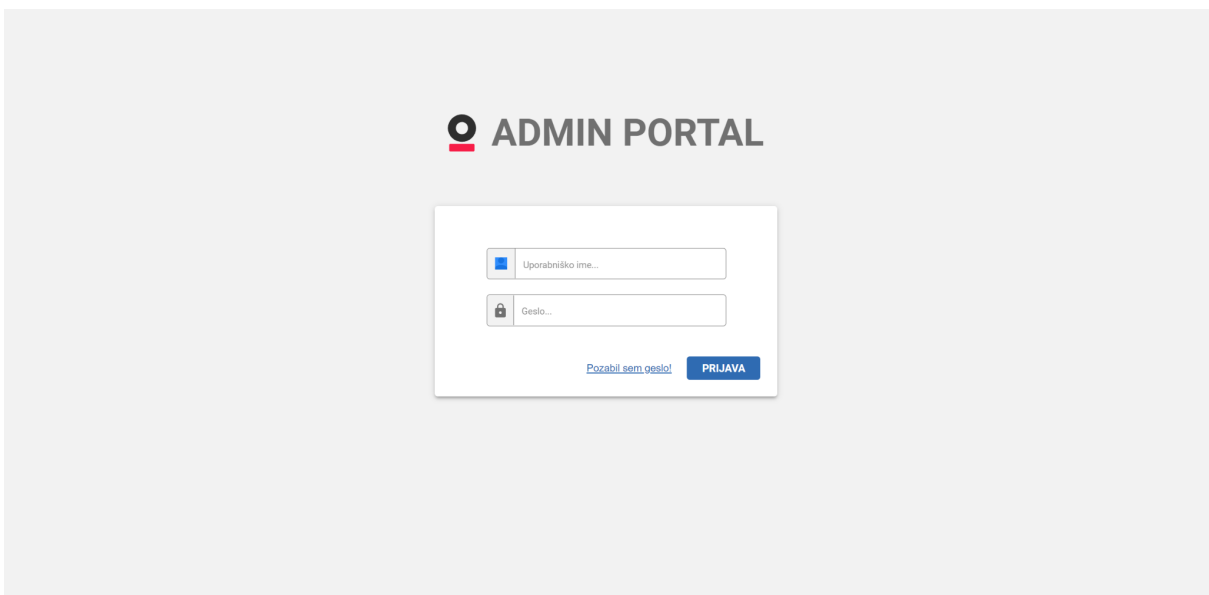
2 Tehnične izboljšave sistema ON

V letu 2022 smo pripravili študentski portal, profesorski portal in administratorski portal. Gre za tri komponente sistema ON: študentski portal med drugim omogoča študentom, da spremljajo transkripcijo razpoznavne govora predavanj; profesorski portal je namenjen predavateljem, da npr. snemajo predavanja, urejajo transkripcijo in dodajajo študijsko gradivo. Administratorski portal je namenjen tehničnim nastavitvam portala.

Študentski in profesorski portal sta dostopna prek spletne strani <http://on.uni-lj.si> (Slika 1), stran nas na enega ali drugega usmeri glede na vpisne podatke; študentske račune na študentski portal, profesorske na profesorski. Administratorski portal se zaenkrat nahaja na začasni strani <https://dev-on-admin.true-bar.si/login> (Slika 2).



Slika 1: Vstopna stran portala ON; študentski in profesorski portal.



Slika 2: Vstopna stran administratorskega portala.

Za boljšo uporabniško izkušnjo smo pripravili tudi uporabniško dokumentacijo, ki je za študente na voljo v [slovenskem](#) in [angleškem](#) jeziku, za profesorje pa le v [slovenskem jeziku](#).

2.1 Razvoj študentskega portala

V okviru tehničnih izboljšav sistema ON smo v letu 2022 implementirali več novosti in nadgradenj delovanja in vmesnika. Za študentski portal smo uvedli paginacijo, tj. delitev predavanja na strani, kar omogoča hitrejšo premikanje zlasti v primeru daljših

predavanj. Dodali smo tudi možnost, da študenti s portala prenesejo dodatno gradivo (zapiske, prosojnice, izročke ipd.), ki ga k predavanju naloži profesor. Popravljen je bil tudi prikaz časovnih oznak segmentov transkripcije predavanja, izboljšana pa je bila tudi robustnost delovanja ter odpravljene manjše napake pri prikazovanju transkripcije. Na pilotnih testiranjih so bili navadno prisotni le študentje, ker pa se je pojavila potreba po tem, smo v letošnjem letu pripravili tudi t. i. račune za goste. Gre za uporabniške račune, ki imajo funkcionalnosti študentskih, torej omogočajo sledenje predavanju prek študentskega portala. Novost bo omogočila sledenje za vse goste, ki nimajo študentskega statusa.

2.2 Razvoj profsorskega portala

Pri profsorskem portalu (tj. delu sistema ON, v katerem profesor ureja transkripcije in prevode svojih predavanj) smo implementirali možnost ponovnega prevajanja posameznih segmentov (če torej profesor popravi le en del besedila, ima zdaj možnost, da ga ponovno strojno prevede in nato po potrebi dodatno uredi). Profesorji lahko zdaj k svojim predavanjem nalagajo tudi dodatna gradiva.

2.3 Razvoj administratorskega portala

Na administratorskem portalu je bil dodan urnik, na katerem je vizualni prikaz predavanj, omogočeno je filtriranje po predavalnicah, predavateljih in predmetih. Uporabnik lahko izbira časovne okvire, dodaja, ureja in briše predavanja. Ureja lahko imena predavalnic in predavateljev, na ta način je mogoče dodati tudi gostujoče predavatelje, ki sicer niso na uradnem urniku. Administrator lahko dodaja in briše predavanja, spreminja imena predavanj in ureja predmete.

2.4 Nadzorna plošča

Nadzorna plošča je namenjena sistemskim administratorjem za lažji nadzor in vzdrževanje sistema ON. Zasnovana in implementirana je na podlagi standardov s področja opazovanja sistemov. Na pregleden način omogoča celovito opazovanje - od stanja systemske infrastrukture do interakcij uporabnikov aplikacij. To med drugim vključuje: zbiranje in pregledovanje dnevniških zapisov mikrostoritev, analiziranje uporabniške izkušnje, spremljanje sistemskih metrik in dosegljivosti sistema ter poglobljeno razumevanje delovanja posameznih komponent in sistema kot celote. S pomočjo strojnega učenja se avtomatsko zaznavajo in kategorizirajo anomalije, kar olajša odkrivanje hroščev in pohitri odpravljanje težav. V pripravi so specifični pogledi po meri in opozarjanje skrbnikov ob kritičnih napakah.

2.5 Jezikovni model

Online Notes vsebuje dva modela, družboslovni in naravoslovni model. Oba sta se med letom sproti posodabljala in nadgrajevala, dodanega je bilo veliko novega besedja. Novi modeli so bili testirani na pilotnih predavanjih. Nadgrajen je bil tudi model za strojno prevajanje.

2.6 Stresni test

Ker se je pri izvedbi pilotnih predavanj v letu 2021 izkazalo, da pri večjem številu prijavljenih udeležencev na platformi Online Notes lahko prihaja do tehničnih težav, smo v letošnjem letu ponovno izvedli stresni test. Testiranje je potekalo pri predavanju dr. Rozmana, in sicer 17. novembra 2022.

Na platformo se je v začetku ure naenkrat prijavilo 84 študentov. Večjih težav ni bilo, platforma je normalno delovala naprej, opaziti je bilo le nekaj drobnih napak. Te so se kazale kot rahla zakasnitev izpisa vmesnih izpisov na enem izmed računov, na drugem pa kot neenakomerno izpisovanje segmentov. Napake niso bile preveč očitne, saj študentje zakasnitev niso omenili.

3 Pilotna predavanja

3.1 Organizacija pilotnih predavanj

V okviru projekta Online Notes (ON) je bilo v koledarskem letu 2022 na Univerzi v Ljubljani izvedenih 12 pilotnih predavanj, na katerih se je v praksi testiral sistem za strojno prevajanje slovenskih predavanj v angleški jezik. Prvi sklop pilotnih testiranj se je odvijal v letnem semestru študijskega leta 2021/2022, drugi sklop pa v zimskem semestru študijskega leta 2022/2023. Predavanja so zajemala tako družboslovna kot tehnična področja na osmih fakultetah:

1. Fakulteta za socialno delo Univerze v Ljubljani (FSD UL)
2. Pravna fakulteta Univerze v Ljubljani (PF UL)
3. Biotehnična fakulteta Univerze v Ljubljani (BF UL)
4. Fakulteta za družbene vede Univerze v Ljubljani (FDV UL)
5. Veterinarska fakulteta Univerze v Ljubljani (VF UL)
6. Fakulteta za gradbeništvo in geodezijo Univerze v Ljubljani (FGG UL)
7. Fakulteta za računalništvo in informatiko Univerze v Ljubljani (FRI UL)
8. Fakulteta za elektrotehniko Univerze v Ljubljani (FE UL)

Za predavanja v prvem sklopu smo se dogovorili že v začetku leta 2022, vabilo za sodelovanje pri pilotnih predavanjih v drugem sklopu pa je bilo konec junija 2022 z Rektorata Univerze v Ljubljani poslano dekanjam in dekanom na vseh fakultetah

članicah. V vabilu je bila prošnja, naj na svojih fakultetah izberejo nekoga, ki bi želel sodelovati pri pilotnih predavanjih in ki obenem izpolnjuje pogoje za sodelovanje pri projektu (tj. ima na voljo vsaj eno uro posnetkov svojih predavanj ter v semestru). V ta namen smo pripravili poseben elektronski naslov, ki smo ga uporabljali za komunikacijo s profesorji (predavajalnik@cjvt.si). Za sodelovanje se je javilo 15 profesorjev¹, 3 pa so pozneje odstopili. Od teh 2 profesorja oz. kandidata nista vedela, da za sodelovanje potrebujeta posnetke oz. nista imela predavanj v slovenščini, zato se za sodelovanje nista odločila. 1 profesor je bil neodziven.

Izvedena pilotna predavanja so bila naslednja (v oklepajih so navedeni fakulteta, učilnica in datum predavanja):

Prvi sklop:

- dr. Samo Kropivnik: **Raziskovanje medijev in komuniciranja** (FDV UL, P7, 13. marec 2022)
- dr. Lea Šugman Bohinc: **Dialoški pristop k soustvarjanju podpore in pomoči** (FSD UL, predavalnica 11, 29. marec 2022)
- dr. Barbara Rajgelj: **Temelji prava** (FDV UL, Velika dvorana, 13. april 2022)

Drugi sklop:

- dr. Goran Turk: **Prostorska statistika** (FGG UL, III/3, 4. 10. 2022)
- dr. Darja Zaviršek: **Ljudje z ovirami: teorije in metode** (FSD, predavalnica 1, 5. 10. 2022)
- dr. Robert Rozman: **Računalniška arhitektura** (FRI, PA, 13. 10. 2022)
- dr. Modest Vengušt: **Osnove klinične diagnostike** (VF, Zoohigijska predavalnica, 26. 10. 2022)
- dr. Aleksandra Kanjuo Mrčela: **Ekonomska sociologija** (FSD, predavalnica 2, 2. 11. 2022)
- dr. Iztok Prislan: **Kemija** (BF, predavalnica Janeza Hribarja, 9. november 2022)
- dr. Katarina Zajc: **Ekonomija** (PF, Siva učilnica, 15. november 2022)
- dr. Matevž Pogačnik in Klemen Pečnik: **Uvod v multimedijske sisteme** (FE, predavalnica 6, 25. 11. 2022)
- dr. Patricio Bulič: **Organizacija računalniških sistemov** (FRI, P22, 30. november 2022)

¹ Izrazi, ki se v poročilu nanašajo na osebe in so zapisani v moški slovnični obliki, so uporabljeni kot nevtralni za ženski in moški spol.

Okvirna tematika predavanj je bila s predavatelji, ki so s projektno ekipo delili posnetke svojih predavanj, dogovorjena. Če so imeli predavatelji na voljo posnetke svojih predavanj z enako ali podobno tematiko, so nam jih posredovali. Posnetke smo transkribirali in transkripcije uporabili, da smo v slovar sistema Online Notes dodali manjkajoče besedišče in poskrbeli, da ga bo razpoznavalnik govora lahko izpisoval. Transkripcije posnetkov so bile uporabljene tudi za gradnjo jezikovnega modela za razpoznavalnik govora.

Količina prejetega gradiva je bila od predavatelja do predavatelja različna - pri posameznem predavatelju smo pridobili največ 4.100 minut posnetkov (tj. 68 ur), najmanj pa 30 minut (z izjemo enega predavatelja, ki ni imel posnetkov). Pri treh predavateljih, ki bodisi niso imeli posnetkov bodisi so posredovali manjšo količino posnetkov, ki niso bili neposredno povezani z njihovim predavanjem, smo besedišče dodali iz njihovih akademskih člankov.

3.2 Tehnična opremljenost učilnic

V načrtu za leto 2022 je bil tudi popis zvočne opremljenosti vseh predavalnic UL, v katerih so bila izvedena pilotna predavanja z uporabo sistema ON. Proces dela je bil zasnovan tako, da smo pred izvedbo testiranj najprej kontaktirali tehnične ekipe posameznih fakultet in pridobili informacije o opremljenosti posameznih učilnic. Ker je sistem ON namenjen dolgoročni uporabi, je opremljenost učilnic pomembna informacija, saj lahko le na ta način predvidimo, koliko dodatne opreme bi potrebovali za nemoteno delovanje aplikacije ter za uspešno izvajanje in podnaslavljanje predavanj. Preverjali smo, ali v posameznih učilnicah že obstaja sistem ozvočenja, ali je učilnica opremljena z računalnikom in (žično) internetno povezavo, ki je za kvalitetno delovanje sistema zelo pomembna – nestabilna brezžična povezava namreč vodi v prekinitve in slabše delovanje aplikacije (na tovrstne težave smo npr. naleteli pri pilotnem testiranju št. 8).

Tabela 1 prikazuje fakultete in predavalnice, v katerih so potekala pilotna predavanja, in opremo, ki je bila uporabljena za izvedbo predavanj.

Fakulteta	Predavalnica	Oprema v učilnici	Opis
FSD	Pred. 11	računalnik, ETH	Snemanje smo izvajali z zunanjim prenosnim računalnikom in mikrofonom RWGII.
FSD	Pred. 1	ozvočenje, računalnik, ETH	Zajem se je izvajal s pomočjo ročnega in namiznega mikrofona, aplikacija delovala na računalniku, na katerega je že priključeno tudi ozvočenje.
VF	Zoohigijska pred.	računalnik, ETH	Snemanje smo izvajali z zunanjim prenosnim računalnikom in mikrofonom RWGII.

Fakulteta	Predavalnica	Oprema v učilnici	Opis
FDV	Pred. 7	ETH	Snemanje smo izvajali z zunanjim prenosnim računalnikom in mikrofonom RWGII.
FDV	Velika pred.	ozvočenje, računalnik, ETH	Zajem zvoka iz ozvočenja (USB Audio vmesnik) in zunanji prenosnik.
FDV	Pred. 2	računalnik, WLAN	Snemanje smo izvajali z zunanjim prenosnim računalnikom in mikrofonom RWGII. WLAN ni dobro deloval, zato priporočilo za žični priključek.
BF	Velika pred.	ozvočenje, računalnik, ETH	Ozvočenje je v učilnici, vendar ni izhoda. Snemanje smo izvajali z zunanjim prenosnim računalnikom in mikrofonom RWGII.
PF	Siva učilnica.	ozvočenje, računalnik, ETH	Starejše ozvočenje, a ni kakovostno. Snemanje smo izvajali z zunanjim prenosnim računalnikom in mikrofonom RWGII.
FE	Pred. 6	računalnik, ETH	Snemanje smo izvajali z zunanjim prenosnim računalnikom in mikrofonom RWGII.
FRI	PA	ozvočenje, računalnik, ETH	Ozvočenje je v učilnici. Snemanje smo izvajali z zunanjim prenosnim računalnikom in mikrofonom RWGII.
FRI	P22	ozvočenje	Ozvočenje je v učilnici. Snemanje smo izvajali z zunanjim prenosnim računalnikom in mikrofonom RWGII.
FGG	III/3	računalnik, ETH	Semestrsko testiranje se izvajajo s pomočjo mikrofona RWGII, ki ga je ekipa projekta ON posodila profesorju na FGG.

Tabela 1: Uporaba tehnične opreme v predavalnicah med pilotnimi predavanji.

Ugotovili smo, da je prednost učilnic, ki že imajo integriran sistem ozvočenja, predvsem v ponovljivi kvalitetni izvedbi zajema in prenosa zvoka. Predavatelju v tem primeru ni treba skrbeti za opremo in z izvedbo nima skoraj nobenega dela, za začetek snemanja le vpiše ustrezen spletni naslov in štirimestno kodo. Slabosti se kažejo v stacionarnosti uporabe: oprema je v tem primeru vezana na specifično učilnico in ni prenosljiva v primeru spremembe lokacije predavanja.

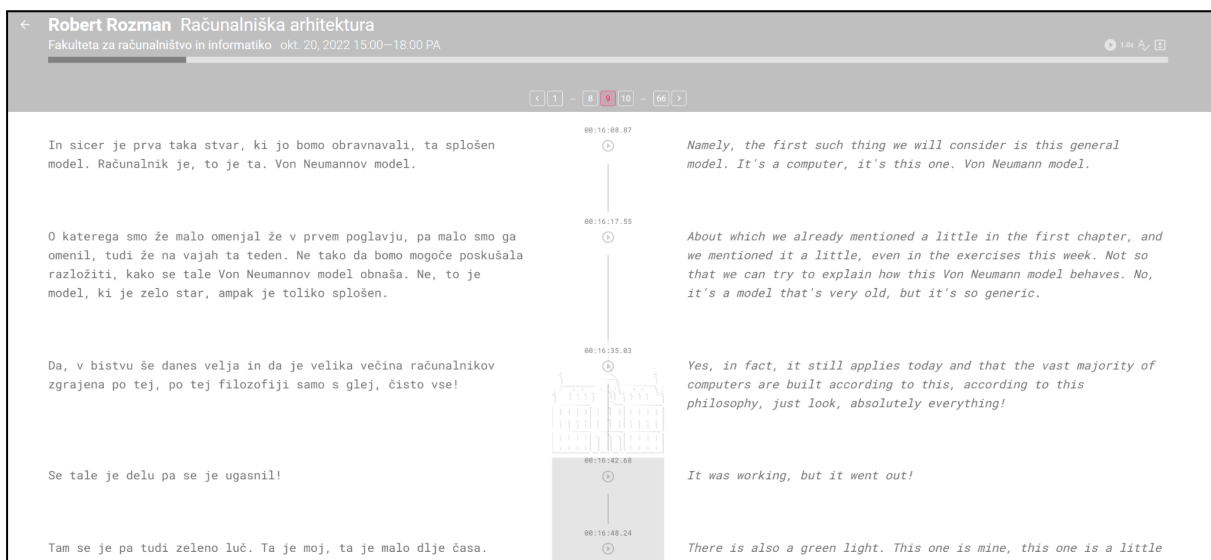
Po drugi strani je bila povezava na obstoječe ozvočenje v primeru pilotnih testiranj tehnično zahtevnejša, saj je zahtevala sodelovanje tehnične podpore posamezne fakultete. Kvaliteta zajema zvoka je bila v tem primeru običajno slabša.

Za optimalno delovanje sistema ON je ključen kvaliteten mikrofoni, ki je pravilno nameščen. Hitrost internetne povezave na pilotnih testiranjih ni bila pomembna (izvedba tudi s pomočjo mobilne dostopne točke (tj. mobilnih podatkov)), ključnega pomena pa je stabilnost omrežja.

3.3 Potek pilotnih testiranj

Predavanja so potekala običajno - predavatelji so bili opremljeni z mikrofonom, ki je zaznaval njihov govor in ga pošiljal sistemu za razpoznavo govora in prevajalniku.

Študenti so imeli možnost razpoznavo govora in strojni prevod spremljati na pilotni platformi Online Notes. Prosili smo jih tudi za mnenja in jim ponudili povezavo do anketnih vprašalnikov (več o tem v poglavju 4). Izpis slovenske transkripcije predavanja in njenega angleškega prevoda, kot je bil na voljo študentom med predavanji, je prikazan na Sliki 3.



Slika 3: Vmesnik za spremljanje slovenske razpoznavo govora in angleškega prevoda.

3.4 Analiza napak

Da bi ocenili, kako uspešna je bila razpoznavo govora med pilotnimi predavanji, smo s pomočjo evalvacijskega programa Evaluator 1.0, ki ga znotraj projekta *Razvoj slovenščine v digitalnem okolju* (<https://www.slovenscina.eu/>) razvijata Fakulteta za računalništvo in informatiko ter podjetje Vitasis d.o.o., analizirali vse slovenske izpise razpoznavo govora, Program primerja izvorni izpis s popravljenimi izpisi in poda

številsko oceno, v kolikšni meri se besedili razlikujeta (npr. po številu zamenjanih besed oz. pojavnic, po številu izbrisanih ali dodanih pojavnic).

Najprej smo analizirali besedila prvega sklopa. Primerjali smo avtomatsko izpisano besedilo, ki je nastalo na pilotnem testiranju, s popravljenim besedilom, ki so ga za namen evalvacije popravili študentje.

V prvem sklopu (Tabela 2) prilagamo rezultate primerjave prvih treh besedil, in sicer:

- dr. Samo Kropivnik: **Raziskovanje medijev in komuniciranja** (FDV, P7, 13. marec 2022)
- dr. Lea Šugman Bohinc: **Dialoški pristop k soustvarjanju podpore in pomoči** (FSD, predavalnica 11, 29. marec 2022)
- dr. Barbara Rajgelj: **Temelji prava** (FDV, Velika dvorana, 13. april 2022)

Predavanje	Št. 1	Št. 2	Št. 3
Delež napak	14,31 %	19,93 %	14,11 %
Število pojavnic v izvirnem besedilu	11.076	14.765	12.132
Število pojavnic v popravljenem besedilu	11.111	14.948	12.305
Število prekrivnih pojavnic	9.703	12.265	10.724
Število izbrisanih pojavnic	177	260	131
Število nadomeščenih pojavnic	1.082	1.669	996
Število dodanih pojavnic	326	1.014	585

Tabela 2: Evalvacija slovenskih izpisov razpoznavne govora za pilotna predavanja iz prvega sklopa.

Iz drugega sklopa (Tabela 3) smo analizirali naslednja predavanja:

- dr. Goran Turk: **Prostorska statistika** (FGG UL, III/3, 4. oktober 2022)
- dr. Darja Zaviršek: **Ljudje z ovirami: teorije in metode** (FSD, predavalnica 1, 5. oktober 2022)

- dr. Robert Rozman: **Računalniška arhitektura** (FRI, PA, 13. oktober 2022)
- dr. Modest Vengušt: **Osnove klinične diagnostike** (VF, Zoohigienska predavalnica, 26. oktober 2022)
- dr. Aleksandra Kanjuo Mrčela: **Ekonomska sociologija** (FSD, predavalnica 2, 2. november 2022)
- dr. Iztok Prislan: **Kemija** (BF, predavalnica Janeza Hribarja, 9. november 2022)

Predavanje	Št. 4	Št. 5	Št. 6	Št. 7	Št. 8	Št. 9
Delež napak	20,24 %	16,46 %	23,12 %	22,65 %	32,16 %	15,66 %
Število pojavnic v izvornem besedilu	10.277	5.688	14.962	12.004	16.641	11.545
Število pojavnic v popravljenem besedilu	10.215	5.412	13.579	11.814	16.021	11.401
Število prekrivnih pojavnic	8.562	4.869	12.149	9.672	11.959	10.130
Število izbranih pojavnic	353	72	328	345	471	371
Število nadomeščenih pojavnic	1.175	383	1.074	1.770	3.555	858
Število dodanih pojavnic	540	436	1.738	562	1.127	557

Tabela 3: Evalvacija slovenskih izpisov razpoznavne govora za pilotna predavanja iz drugega sklopa.

V času pisanja poročila predavanja dr. Zajc ter dr. Pogačnika in Pečnika še niso bila popravljena, predavanje dr. Buliča pa se še ni izvedlo, zato v poročilo niso vključena.

Pilotna testiranja, ki smo jih izvajali pri posameznih predavanjih, so bila dolga od 57,09 do 150,57 minute. Izpisana besedila so se razlikovala po povprečni dolžini segmentov.

Najdaljši segment najdemo v predavanju št. 8, traja kar 83,4 sekunde in obsega 167 besed. Tovrstne deviacije se pojavijo le občasno, kar potrjuje tudi povprečna dolžina posameznega segmenta, ki pri predavanju št. 8 traja 12,4 sekunde, splošna analiza vseh predavanj pa pokaže, da je povprečen čas trajanja posameznega segmenta pri pilotnih testiranjih 8,84 sekunde.

Na podlagi primerjave lahko ugotovimo, da se je najvišje število napak (32,16 %) pojavilo pri predavanju št. 8. Rezultat je pričakovan, saj za predavanje nismo prejeli nobenih posnetkov, zaradi česar je bilo težje pokriti besedišče, ki se je uporabljalo v predavanju. Najboljše rezultate dosega predavanje 9, pri katerem je predavatelj_ica zagotovil_a zadostno število posnetkov (1720 minut).

Avtomatski evalvator je kot napako zaznal vsako spremenjeno pojavnico, in sicer ne glede na to, ali je šlo za popolnoma drugo besedo, ki je vplivala na prevod, ali je bil nadomeščen le del besede (npr. pripona). Najvišjo vrednost dosega število nadomeščenih besed, kar pomeni, da je razpoznavalnik večinoma pravilno ločeval med posameznimi pojavniciami, število manjkajočih in dodanih besed je relativno nizko.

Zanimalo nas je, kolikšen del napak, ki jih je zaznal avtomatski evalvator, je takšnih, ki ne vplivajo na prevod. Za namen natančnejše analize smo izvedli še krajšo ročno analizo. Iz drugega dela pilotnih predavanj smo izbrali dve, ki sta potekali na družboslovnem modelu (predavanji 5 in 8) in dve, ki sta potekali na naravoslovnem modelu (6 in 9). Segmente smo nato pregledali in jih ocenili glede na uspešnost razpoznave govora in strojnega prevoda.

Pri razpoznavi govora smo označevali več kategorij napak, ki so predstavljene v Tabeli 4.

Vrsta napake	Razlaga
napaka pri izpisu imena	Razpoznavalnik napačno izpiše lastno ime.
napačna segmentacija	Sistem ON napačno segmentira besedilo, kar lahko vpliva na pomen in na prevod besedila.
napačna delitev	Napačna delitev povedi, postavitve končnega ločila in velike začetnice na mestu, kjer se poved ne zaključí.
napačna oblika	Razpoznavalnik izpiše napačno obliko besede, lahko gre za napačen sklon, število, napačno obliko glagola ipd., ob

	tem pa še vedno izpiše pravo besedo (manjši vpliv na prevod).
napačna beseda	Razpoznavalnik izpiše popolnoma napačno besedo.
šum	Razpoznavalnik zazna delno izgovorjeno besedo ali mašilo, kar povzroči izpis napačne besede.
diskurzni označevalec	Osredotočili smo se na izpis diskurznega označevalca "(a/ali) ne", ki pri prevajanju predstavlja težave, saj napačno zanika pomen.
manjkajoča beseda	Razpoznavalnik ne izpiše besede (izpust besede, izpust pomožnega glagola ipd.).
pisanje skupaj/narazen	Beseda je napačno zapisana skupaj oz. dve besedi sta napačno zapisani narazen.

Tabela 4: Kategorije napak.

V Tabeli 5 so navedene napake razpoznavalnika govora. Navedene količine predstavljajo število napak določene kategorije, ki jih razpoznavalnik zagreši v obsegu tridesetih segmentov različnih dolžin.

Tip napake	Zajc	Rozman	Kanjuo Mrčela	Zaviršek
napaka pri izpisu imena	2	0	3	2
napačna segmentacija	6	7	17	11
napačna delitev	2	4	6	4
napačna oblika	25	31	85	7
napačna beseda	24	11	67	10
šum	7	6	21	12
diskurzni označevalec	24	37	2	9
manjkajoča beseda	1	0	0	1

pisanje skupaj/narazen	1	2	2	0
------------------------	---	---	---	---

Tabela 5: Število napak razpoznavalnika govora v primerjanih segmentih.

Najpogostejša napaka, ki se pojavi na pilotnih testiranjih, je izpis napačne oblike besede, kar pomeni, da je program sicer izpisal pravo besedo, le v napačni obliki (napačen sklon, število, spol ipd.).

Kot druga najpogostejša napaka se pojavlja izpis napačne besede, pogost razlog je bližnja homonimija (več o tem v poglavju 3.4.3), včasih pa je razlog za napačen zapis ta, da besede še ni v slovarju, program zapiše tisto, ki je po zvočni podobi najbližje (poglavje 3.4.2). Program redko izpusti določeno besedo.

Pogosta je napačna segmentacija, ki besedilo prelomi na nepričakovanem mestu in ustvari dve ločeni povedi. Napaka včasih vpliva tudi na pomen (*“Torej prekarnost, jaz sem // relativno brez ocene govorila o tem, kako vse bolj // selimo, se selimo in uporabljamo različne, prožne oblike dela”*). Napačna segmentacija vpliva tudi na prevod (*“So that the precarity, I, am. // Relatively no reviews talked about how more and more?// We move, relocate and use different, flexible forms of work”*).

Zaradi diskurznihih označevalcev prihaja do negacij (*“v podrobnosti, ne, potem pa tisto, kar je še mogoče”*, ki v prevodu postane *“not in detail, but then what is possible.”*). Pojavlja se tudi napačna delitev znotraj posameznega segmenta, (*“okej, potem še bi rekel, naslednja! Taka ideja je višja, zmožljiva z pomočjo paralelizma”*), ki otežuje razumevanje in vpliva na prevod. Postavljanje ločil v poročilu ni analizirano.

Iz tabele je razvidno, da je pri določenih predavateljih prisotnih manj napak. Do odstopanj prihaja zaradi razlik v načinu govora posameznikov (diskurzni označevalci, mašila, vpliv narečja ipd.). Med ročno analiziranimi besedili se najmanj napak pojavi pri št. 5, in sicer le 56. Število prejetih posnetkov je pri omenjenem_i predavatelji_ici sicer nizko (71,9 minute), vendar je poleg zvoka priložil_a tudi članke in na ta način zagotovil_a pokritost besedja. Največ napak je v izsekih iz predavanja št. 8, in sicer 203, rezultat je pričakovan, saj za namen pilotnih testiranj nismo prejeli nobenih posnetkov.

3.4.1 Napačno ali nekonsistentno razpoznane besede

Čeprav so bile določene besede prisotne v slovarju izgovorjav, ki so ga pripravili študentje s transkribiranjem posnetkov, nekatere niso bile razpoznane pravilno oz. so bile razpoznane nekonsistentno (različno od segmenta do segmenta). Pri izpisovanju besedila je tako včasih prišlo do pomanjkljivih oziroma napačnih prevodov terminološkega besedja (Tabela 6).

Slovensko	Prevod	Pravilno
Porazdelitvena funkcija	Distribution function	Cumulative distribution function
Verjetnostna funkcija	Probability function	Probability mass function
Medsebojno nezdružljivi	Mutually incompatible	Mutually exclusive

Tabela 6: Primerjava avtomatskih prevodov terminološkega besedja.

Pregled pogosto rabljene terminologije (Slika 4) pokaže, da se navadno program dobro odziva in prevaja iste termine na enak način.

The screenshot shows a video player with three lines of Slovenian subtitles on the left and three lines of English subtitles on the right. The Slovenian text discusses mapping 'porazdelitvene funkcije' (distribution functions) and plotting them. The English text translates this as 'distribution functions' and 'graphs one after the other'. The term 'porazdelitvene funkcije' is highlighted in yellow in the Slovenian text, and 'distribution functions' is highlighted in yellow in the English text. A central video player interface is visible, showing a play button and a progress bar.

Slika 4: Konsistentnost prevajanja terminološkega izrazja.

Sistem ON običajno pravilno zazna besedo znotraj konteksta, a se včasih pojavi nekonsistentnost rabe, največkrat zaradi večpomenskosti. Primer je denimo beseda "zvezni" (npr. "zvezna porazdelitev"), kjer v angleškem prevodu najdemo bodisi "continuous distribution" (kar je ustrezno) bodisi "federal distribution" (kar je v kontekstu neustrezno). Pričakujemo, da se bodo tovrstne napake odpravile, ko zberemo dovolj podatkov za prilagoditev jezikovnih modelov.

3.4.2 Napačna razpoznavna besede kot posledica manjkajoče besede v slovarju

Ker so nam predavatelji poslali zelo različne količine posnetkov, iz katerih smo pridobivali besedišče, ki ga lahko pokrije razpoznavnik govora, zlasti pri tistih predavateljih brez posnetkov nismo mogli v slovar dodati vsega (npr. lastnih imen in specifične terminologije). To je vplivalo tudi na kakovost razpoznavne govora, saj program besede, ki je nima v slovarju, ne more pravilno razpoznati. V takih primerih izpiše drugo besedo (oz. kombinacijo besed), ki je po zvočni podobi najbližje (Tabela 7).

Primer izpisa	Pravilno
Pa trebušni organi, srček, takšen konj sem , Poglejte, ko grejo, vrta ven!	Pa trebušni organi, srček, takšen konusen , pogledajte, kako gre aorta ven!
Ketoprofen je nesteroidni Antifologistik , ki ga uporabljajo tudi v humani medicini; uporabljamo ga veliko pri teh produkcijski živalih.	Ketoprofen je nesteroidni antiflogistik , ki ga uporabljajo tudi v humani medicini; uporabljamo ga veliko pri teh produkcijski živalih.
Ker sem vodil del klinike, ne sploh ne Nataliji del sem 14 dni imeli žrebet ...	Ker sem vodil del klinike, ne, sploh neonatalni del, sem 14 dni imel žrebet ...
Se vem, da sem že povedal, ja, vas je opozorila kolegica Cehtejov! Ure pa antibiotik?	Se vem, da sem že povedal, ja, vas je opozorila kolegica, ceftiofur je pa antibiotik.
... to je Foks, to me zdaj. Še . Ali lažna umetna inteligenca. // ² To je nekaj ta fauxtomation ali lažna umetna inteligenca, to je nekaj ...

Tabela 7: Primeri napačne razpoznavne, ki so posledica manjkajoče besede v slovarju ON.

3.4.3 Homonimija in kratke enote

Pogosta je težava pri razpoznavanju kratkih imen, kratic in enot. Zlasti pri predavanjih iz naravoslovnih tematik so pogosto v rabi različne enačbe, ki se zapisujejo s pomočjo črk in simbolov. Kratke enote povzročajo težave pri razpoznavi govora, saj je krajši zvočni izpis navadno prekriven z več različnimi enotami, ki pripadajo drugim besedam oz. daljšim zvočnim enotam.

V prikazanem primeru (Tabela 8) se ob črki A pojavi število nič, kombinacija pa zvočno sovpaše s fonetično podobo priimka Anić. Program se odloča med obema in izpiše priimek Anić. Sčasoma bo jezikovni model izboljššan - z večjo količino gradiva se bo lažje odločal o izbiri ustrezne besede, trenutno pa so izbire včasih naključne.

² Tu pride tudi do napačne segmentacije.

Izpis	Pravilno
Ani Anić , plus A, ena krat XI od B. Kaj torej, tale stvar naredi. Vrednosti XI primeži a ena prišteje a nič delili z BB, in ostanek vzame kot naslednjo vrednost. To to pomeni, to pomeni ta ostal.	A ni(), A nič plus A ena krat X I od B. Kaj torej tale stvar naredi? Vrednosti XI primnoži a ena, prišteje a nič, deli z B in ostanek vzame kot naslednjo vrednost. To() ³ to pomeni, to pomeni ta ostal().

Tabela 8: Primer težave s homonimijo.

3.4.4 Raba angleškega jezika

V govornem slovenščini (tudi akademski) je pogosta prisotnost angleških besed, kar se je pokazalo tudi na pilotnih testiranjih. Včasih gre zgolj za dodatno navajanje angleških različic slovenskih izrazov, včasih pa je angleški izraz uporabljen namesto slovenskega. Ker je sistem ON prilagojen za razpoznavo slovenskega govora, besed v tujem jeziku ne prepozna ustrezno in poišče slovensko ustreznico, ki je po izgovoru najbližje. Zaradi teh opažanj smo v letu 2022 v model dodali še nekaj pogostih angleških besed, ki so jih predavatelji uporabljali, kar je razpoznavo nekoliko izboljšalo.

Številka predavanja	Izpis	Pravilno
8	Ni hard kavarno , ne vem, ko ste recimo slovenska ne, in potem ...	Ni hard cover , ne vem, kako se to reče po slovensko, ne, in potem ...
4	V ILO hočem , če je tukaj x x nič in potem je tudi tukaj naprej vrednosti slej ko prej bo prišlo nazaj do x, nič.	Wheel of fortune , če je tukaj X nič in potem je tudi tukaj naprej vrednosti, slej ko prej bo prišlo nazaj do X nič.
12	Ki bo neko vezje ki bo aktiviralo izbrani, izbrano tisto, linija word line , torej, da bo odprla vsa stikala da steče naboj na bitne linije?	What will be some circuit that will activate the selected word line , so that it will open all the switches so that the charge flows to the bit lines?

³ Prazen oklepaj označuje na pol izgovorjeno besedo, v teh primerih razpoznavnik pogosto zapiše napačno besedo, saj prave profesor ne izgovori do konca.

8	In zato, ja, potem se je uveljavilo nekaj, čemur rečemo people Frst. Mogoče ste v prvem letniku že za to slišal. To je načelo: najprej ljudje, potem neke diagnoze, a ne zato rečemo: recimo, ne vem.	And that's why, yes, then something we call people Frst took hold. Maybe you heard about it in your first year. This is the principle: first people, then some diagnoses, but that's not why we say: let's say, I don't know.
---	--	--

Tabela 9: Primeri razpoznavne angleških besed na pilotnih testiranjih.

3.4.5 Stalne besedne zveze, frazeologija in metaforika

Za prevajalnik predstavljajo težavo stalne besedne zveze in frazemi, ki so pogosto prevedene dobesedno (Tabela 10). Napaka je pričakovana in ne predstavlja zgolj pomanjkljivosti sistema ON, temveč splošni problem pri strojnih prevajalnikih. Za izboljšanje je potrebna večja količina podatkov za učenje strojnega prevajalnika.

5	"Tako, mogoče na prvo žogo, je ja, Ana"	"o, maybe on the first ball, yes, Ana,"
---	---	---

Tabela 10: Primer problema z dobesednim prevajanjem stalnih besednih zvez.

3.4.6 Diskurzni označevalci

Ker gre pri predavanjih za spontani govor, se v njem pogosto pojavljajo diskurzni označevalci (Tabela 11), npr. "ne", "ane", "ali ne", "ne vem", "se pravi", "skratka". Precejšen problem predstavlja diskurzni označevalec '(a/ali) **ne**', ki se v prevodu izrazi kot nikalnica in negira trditve. Ker smo se s podobnimi težavami srečevali že lani, smo letos poseben poudarek posvetili redukciji izpisovanja teh diskurzni označevalcev, specifično (a/ali) ne, rezultati so boljši od lanskih, ponekod sistem težavo reši tako, da "ne" postavi v samostojno poved in tako ne vpliva na prevod.

Seveda pa še vedno ostajajo primeri, kjer diskurzni označevalec vpliva na prevod in ga napačno zanika.

	Izpis	Diskurzni označevalec v popravljenem besedilu	Prevod	Vpliv na pomen
8	To je med pandemijo v glavnem dotikala prostorska prožnost, ne Ana , postale smo namesto tega, da smo sedel ...	Ta se je med pandemijo v glavnem dotikala prostorska prožnost, ne, a na , postali smo, namesto tega, da smo sedeli ...	This was mainly touched by spatial flexibility during the pandemic, not Ana , we became instead of sitting in ...	NE
8	... ne samo o kakovosti delovnega življenja, ampak o dobrem življenju. Ne ne samo o kakovosti delovnega življenja, ampak o dobrem življenju. Ne not only about the quality of working life, but about the good life. No .	NE
6	Ne pa seveda tudi lahko ne .	Ne , pa seveda tudi lahko, ne .	Of course, you can't .	DA
9	Iz osebne izkušnje ne izhajam, koncentracija mi ni nikoli bil problem.	Iz osebne izkušnje, ne , izhajam, koncentracija mi ni nikoli bil problem.	I am not speaking from personal experience, concentration has never been a problem for me.	DA
9	... govorimo o številu Molov. Ne , kot je reaktantov, potem Ja potem, če zvišamo tlak govorimo o številu molov, ne , kot je reaktantov, potem, ja, potem, če zvišamo tlak we are talking about the number of Moles. No , as there are reactants, then Yes, then if we increase the pressure	NE

	ki jih določena dobrina potem na trgu lahko dobi, a ne.	ki jih določena dobrina potem na trgu lahko dobi, a ne.	that a certain good can then get on the market, but not.	?
--	--	--	--	---

Tabela 11: Vpliv diskurznihih označevalcev na prevod.

3.5 Semestrski predavanja

V študijskem letu 2022/2023 smo v sklopu testiranj sistema ON pripravili tudi semestrski testiranja, in sicer na Fakulteti za gradbeništvo in geodezijo UL ter na Fakulteti za računalništvo in informatiko UL.

Za razliko od ostalih pilotnih testiranj, kjer smo izvajali posamezne teste in le enega pri vsakem izmed profesorjev, so semestrski testiranja potekala ves semester. Ker v času pisanja tega poročila testiranja še niso zaključena, bo vsebina in analiza obeh semestrskih predavanj zajeta v naslednjem poročilu.

4 Odzivi uporabnikov

Da bi bil sistem Online Notes koristen za čim več različnih ciljnih skupin, smo od uporabnikov zbirali tudi povratne informacije o delovanju sistema. V ločenih razdelkih predstavljamo odzive študentov in profesorjev, v času pisanja tega poročila pa poteka tudi pridobivanje povratnih informacij oseb z omejitvami.

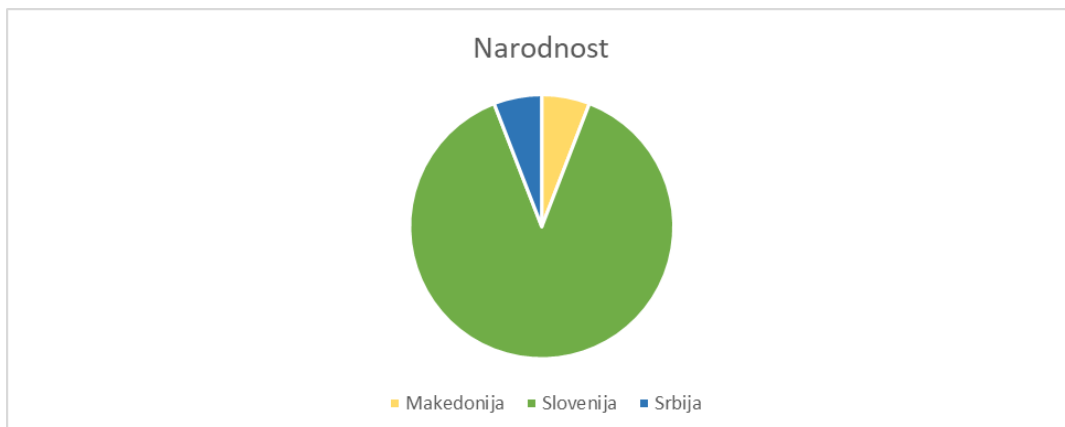
4.1 Odziv študentov

Študentski portal je bil uporabljen na vseh pilotnih testiranjih, študente smo pred začetkom usmerili na spletno stran in prosili za sodelovanje. Na pilotnih testiranjih smo prosili študente, da nam s pomočjo anonimne spletne ankete podajo svoje mnenje in predloge za izboljšavo.

Na vsakem izmed pilotnih testiranj smo prosili študente, da rešijo anketo in s tem prispevajo svoje mnenje. Anketo je rešilo v prvem delu 17 študentov, od tega 11 žensk in 6 moških. V drugem delu je sodelovalo 69 študentov, od tega 28 žensk, 27 moških in 2 posameznika, ki se identificirata drugače. 2 študenta se nista želela opredeliti. Največjo pripravljenost za reševanje anket so izkazali študentje prvega letnika dodiplomskih študijev.

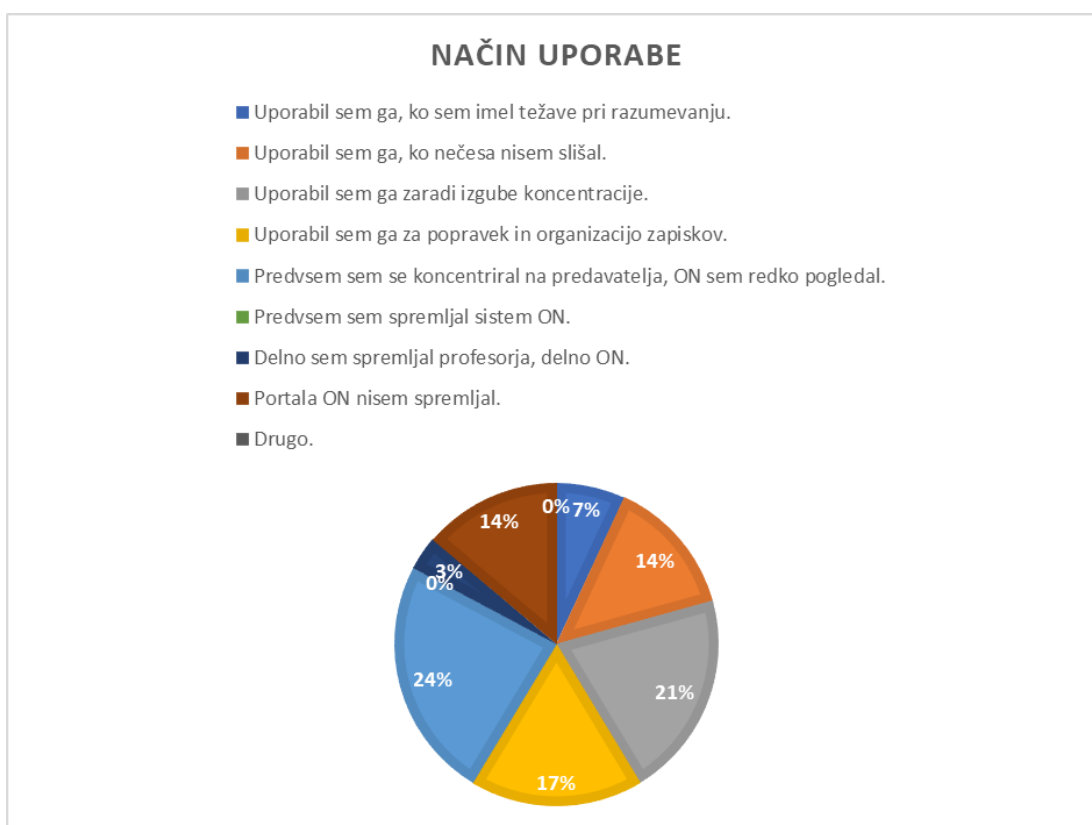
4.1.1 Prvi sklop pilotnih predavanj

V prvem sklopu pilotnih predavanj so na vprašanja odgovarjali študentje in študentke prvega (5), drugega (2) in tretjega letnika (9) dodiplomskih študijev (Slika 5).



Slika 5: Razporeditev sodelujočih študentov glede na narodnost v prvem sklopu pilotnih predavanj.

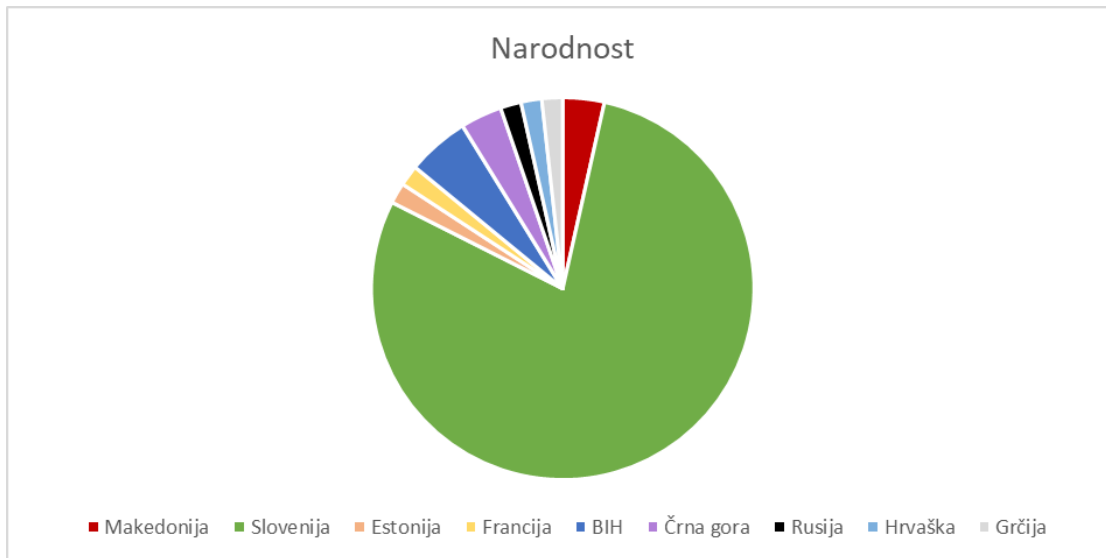
Med študenti, ki so na ankete odgovarjali v prvem delu testiranja, sta se dva opredelila kot slabša govorca slovenskega jezika. V prvem delu testiranja je 9 posameznikov ocenilo, da imajo pri poslušanju predavanj težave s koncentracijo, vidom (oziroma oddaljenostjo table) ali sluhom (oziroma glasnostjo predavanj). Izbrali so eno ali več možnosti. Študentje so po testiranju aplikacije odgovarjali na vprašanje, kdaj so (in bi tudi v prihodnje) aplikacijo uporabljali, v katerih primerih bi jim prišla prav (Slika 6).



Slika 6: Načini uporabe sistema Online Notes v prvem sklopu pilotnih predavanj.

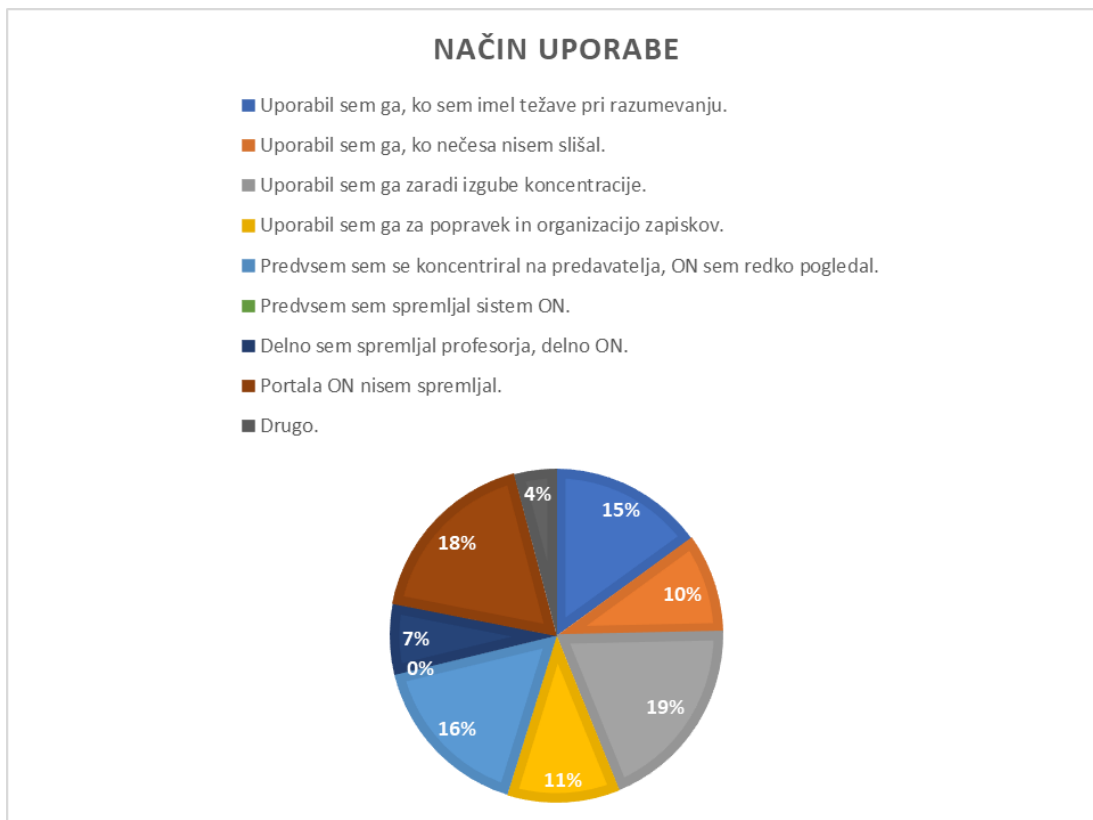
4.1.2 Drugi sklop pilotnih predavanj

V drugem sklopu pilotnih predavanj (Slika 7) so na vprašanja odgovarjali študentje in študentke prvega (52), drugega (1) in tretjega letnika (3) dodiplomskih študijev ter prvega (2) študija magistrskega programa. En študent_ka (1) se pri vprašanju ni opredelil.



Slika 7: Razporeditev sodelujočih študentov glede na narodnost v drugem sklopu pilotnih predavanj.

V drugem delu testiranja je 25 posameznikov ocenilo, da imajo pri poslušanju predavanj težave s koncentracijo, vidom (oziroma oddaljenostjo table) ali sluhom (oziroma glasnostjo predavanj). Izbrali so eno ali več možnosti (Slika 8).



Slika 8: Načini uporabe sistema Online Notes v drugem sklopu pilotnih predavanj.

Študente smo vprašali tudi, ali se jim zdi sistem ON uporaben oz. ali je izboljšal kvaliteto spremljanja predavanj. Na vprašanje je pozitivno odgovorilo 80 % sodelujočih. Prav tako so tudi ocenili, da bi vsebine uporabili za učenje oziroma pomoč pri izdelavi zapiskov (prav tako 80 % vprašanih).

Študentje so imeli tudi priložnost, da v odprtem vprašanju izpostavijo vidike, ki so jih med spremljanjem predavanja zmotili. Nekateri so izpostavili pomanjkljivosti v izpisovanju segmentov transkripcije, npr. zamudo med predavanjem in izpisovanjem prevoda:

thing that disturbed me the most was delay between lecture and translation in english.

—

i am not sure it goes here but : the fact that long sentences take a long time to be written is annoying. because of that, sometimes i don't have the time to read it, it is already gone replaced by the next one

—

the time when he said things was not correct- 40 min delay

—

the delay.

Motilo jih je tudi, da ni mogoče drsati naprej in nazaj po transkripciji predavanja:

i'd prefer if it would be easier to scroll back in class, since the tempo is quite fast.

—

na telefonu je mogoče videti le angleški prevod, če se umes ugasne zaslon, se besedilo neha izpisovati na ekranu, besedilo se ne pomika samo od sebe ampak moraš sam premikati navzdol do najnovejšega. všeč bi mi bilo, če bi lahko izbral ali boš spremljal slovenščino ali angleščino na telefonu. če greš med predavanjem ven, se takrat ko prideš nazaj noter izpišejo samo neki naključni deli besedila.

—

i would suggest that when new text appears the screen automaticly scroles to the bottom where the new text is

Predlagali so tudi uvedbo funkcije kopiraj-prilepi, kar bi jim omogočilo, da bi transkripcijo oz. dele transkripcije lahko kopirali v lastne zapiske:

copy paste function

—

it would be nice if you could copy trascription to paste it to your notes

Nekateri so omenili, da v nekaterih primerih transkripcija in prevod nista ustrezna:

sometimes the translation can say something entirely different and now even close to the thing that the professor says.

—

the most disturbing were the incorrect translations and the ai not being able to follow slovene correctly and transcribing something really incorrect

Ostale predlagane izboljšave zadevajo izpisovanje transkripcije (manj mašil in pravilno izpisovanje kratic) ter možnost, da na telefonu odstranimo tipkovnico z zaslona:

odstraniti mašila, uvesti zapis črk pri črkovanju in ne sestavljanje novih besed

—

da ni tipkovnica ves cas na enkranu na telefonu

Študentje so lahko izrazili tudi splošno mnenje o sistemu Online Notes - odzivi so bili večinoma pozitivni in so v sistemu prepoznali potencial:

it was intensive, but very useful

usefull to follow a course in another language. but i think it is easier to understand the video with correct subtitles than the course in live. it is really hard to read the on translation and to follow what the teacher is wriiting on the board (in slovene). and it is sometimes really hard to understand the meaning because of mistakes in the translation.
it is helpfull for people who do not know slovene , but after few months students usually get used to slovenian so i think it is easier to follow lectures live.
it is a great idea, but is a big project and it probably needs 100s of people working on it to make on a really good real time translator. but the transcripts are a big bonus.
bad but funny
ON mi je bil všeč, vendar je trenutno še neuporaben.
i was impressed.
uporabno za tuje študente
it is impressive how good the translations are and how fast the translation works and its quality
it is useless because of incorrect transcription and therefore translation
it's amazing
pretty amazing.
ima velik potencial, se posebej ce bi bil delay manjsi
goodi
it has potential

4.2 Odziv profesorjev

Po predavanjih smo se s profesorji dogovorili za kratke intervjuje, na katerih bodo sporočili svoje mnenje glede uporabnosti platforme, zanimalo nas je, ali bi bili sami pripravljeni uporabljati platformo, katere funkcionalnosti bi sami uporabljali (podnaslavljanje, izpisana besedila v slovenskem in angleškem jeziku, posnetki predavanj, dodajanje priponk za študente ipd.).

Po izvedbi prvega sklopa pilotnih testiranj smo vsem trem profesorjem iz prvega sklopa posredovali podatke, s katerimi so se lahko vpisali na platformo. Vsi trije profesorji na fakultetah poučujejo več kot 20 let. Vprašali smo jih npr., kaj jih je v transkripciji najbolj motilo, ali so kaj opazili oz. kaj bi spremenili. Opazili so, da je transkripcija slovenskega dela boljša kot angleški prevod, kar je pričakovano, ker je slovenski del lažje razumeti tudi z napakami, ki v angleškem delu postanejo večji problem.

V transkripciji so težave povzročali diskurzni označevalci ("ne" v pomenu "kajne"), ki so se v angleški jezik prevajali kot nikalnice. Eden izmed sodelujočih je komentiral, da je opazil, da v govoru tvori veliko podrednih stavkov.

Sodelujočim profesorjem_icam smo po prvem delu testiranja ponudili vstopne podatke za profesorski portal. Prosili smo jih, da nam posredujejo svoja mnenja in predloge.

Na vprašanje, ali se jim zdi ON koristen, so odgovarjali:

- "Da. Sam vidim prednost predvsem za Erasmus študente in druge študentke in študente iz tujine, ki jim nudimo le nekaj predmetov v angleščini, s pomočjo te naprave pa bi lahko poslušali glavnino predmetov, ki jih izvajamo. Stvar se mi zdi uporabna tudi za domače študente, ki lahko dobijo transkript predavanja ... slednje pa je odlično in potencialno zelo uporabno tudi za učitelje."
- "Da, vendar ima sedanja različica po moji oceni preveč pomanjkljivosti; najbolj primeren se mi zdi za študente s posebnimi potrebami (senzorno ovirani) in tuje študente; za ostale pa zna prispevati k temu, da se bodo še bolj kot doslej učili samo iz zapiskov (kar učitelji_ce sicer lahko reguliramo z vprašanji iz literature) ali se celo izogibali predavanj."

Profesorji_ice menijo, da bi jim transkript predavanj prišel prav pri pripravi učbenikov in drugih pedagoških gradiv za študente, za organizacijo lastnih zapiskov, koristni se jim zdijo tudi prevodi. Uporaba portala se vprašanim zdi enostavna. Največjo težavo predstavlja popravljanje transkriptov, to bi vzelo preveč časa, zato se za deljenje transkriptov na portalu ON zaenkrat še ne bi odločili, z veseljem pa bi uporabljali možnost podnaslavljanja v živo.

4.3 Posamezniki z omejitvami na različnih področjih

Želeli smo si, da bi platforma koristila tudi posameznikom s senzornimi oviranostmi (gluhi, naglušni, slepi, slabovidni), bralno-napisovalnimi težavami (disleksija) in ostalimi specifičnimi učnimi težavami (skotopični sindrom, diskalkulija).

V letu 2022 smo v razvoj sistema Online Notes začeli aktivneje vključevati tudi predstavnike ranljivih skupin, predvsem slepih in slabovidnih ter gluhih in naglušnih oseb. Vzpostavili smo stik z Društvom študentov invalidov Slovenije (DŠIS) in se dogovorili za sodelovanje na dveh nivojih: (a) študenti DŠIS bodo pomagali evalvirati sistem Online Notes in njegov vmesnik ter podali povratne informacije o morebitnih izboljšavah in spremembah, ki bi aplikacijo napravile bolj prijazno uporabnikom z omejitvami; in (b) v naslednjem koledarskem letu bomo - če bo mogoče - organizirali pilotno predavanje z uporabo sistema Online Notes tudi pri predmetih, ki jih

obiskujejo študentje DŠIS, in tako v realnem času preizkusili delovanje aplikacije ter pridobili povratne informacije o tem, kako delovanje sistema dojemajo osebe z omejitvami.

Za dodatne povratne informacije, kako sistem ON približati posameznikom z omejitvami, smo se dogovorili tudi z dr. Darjo Zaviršek, profesorico na Fakulteti za socialno delo, ki se ukvarja s področjem posameznikov z oviranostmi. Skupaj bomo preučili zasnovo in delovanje sistema ter se dogovorili, kako ga čim boljše opremiti, da bo uporaben za vse. Glede na informacije, ki jih bomo prejeli, bomo v prihodnjem letu prilagodili platformo in pripravili poročilo, ki bo vključevalo opis sprememb.

5 Povzetek ključnih ugotovitev

Razvoj sistema Online Notes je v letu 2022 potekal po zastavljenem načrtu. Analiza 9 pilotnih testiranj je pokazala, da je povprečen delež napak, ki so se pojavljale pri izpisovanju slovenskega besedila s sistemom ON, 19,85 %. Glede na pilotna testiranja je jasno, da se z dodajanjem besedja v slovar ON izboljšuje kvaliteta izpisovanja slovenskih transkripcij, posledično pa tudi prevodi: pri predavateljih, ki so zagotovili zadostno količino posnetkov, so bili rezultati očitno boljši, najslabša razpoznava pa je bila pri tistem predavanju, ki je bilo izvedeno brez predhodno dodanih transkripcij in posnetkov. V prihodnjem koledarskem letu nameravamo s pomočjo razširjanja modelov in dodajanja besedja delež napak še dodatno zmanjšati. Prevodi so pri tistih delih, ki ne vsebujejo napak v slovenskem izpisu, večinoma ustrezni, v naslednjem koledarskem letu pa bomo nadgradili tudi model za strojno prevajanje.

Tudi odzivi uporabnikov na uporabo sistema ON so pozitivni – kljub nekaterim pomanjkljivostim (tako tehničnim kot vsebinskim) je večina anketiranih študentov in profesorjev poudarila, da ima aplikacija potencial in da bi jim koristila pri delu. V naslednjih nadgradnjah bomo upoštevali pridobljene pripombe, implementirali pa bomo tudi spremembe, ki jih bodo predlagali predstavniki skupin oseb z omejitvami.

V tekočem letu smo izpeljali več pilotskih predavanj, na katerih smo testirali delovanje trenutnega sistema ON. Cilj je bil pridobiti čim več podatkov o tem, kako zadovoljivo sistem deluje v realnih okoljih. Poleg običajnih triurnih predavanj, ki smo jih izvajali že lansko leto, smo tokrat s sistemom ON spremljali dva celosemestrski predmeta, in sicer enega na FGG in drugega na FRI. Dodaten vir informacij je predstavljala poglobljena diskusija s predstavniki gluhih, naglušnih, slepih in slabovidnih ter z eksperti za dostopnost medijev. Načrt za 2023 smo oblikovali na osnovi pridobljenih informacij. Načrt vključuje: 1) Izboljšave študentskega/profesorkega portala. Ključne funkcionalnosti, ki jih uporabniki še pogrešajo oz. ne delujejo optimalno, so: a) hitro iskanje po vsebini celotnega predavanja, b) strukturiranje vsebine za lažje premikanje po predavanju, c)

dosnemavanje, d) bližnjice na tipkovnici za hitro editiranje, e) uporaba sintetizatorja za slepe in slabovidne kot pomoč za detekcijo pozicije na portalu, f) podpora za večjezičnost. 2) Nadgradnja zalednega dela. Na zalednem delu bomo v 2023 poskrbeli za integracijo nekaterih izročkov projekta RSDO, predvsem želimo izkoristiti SLOLEKS leksikon ter storitve za avtomatično tvorjenje besednih oblik in izgovorjav pri dodajanju besedišča v slovar. V okviru nadgradnje zalednega dela želimo preučiti varnostne vidike ter poskrbeti za ustrezno raven varnosti. Produktijsko delovanje bo zahtevalo tudi večjo robustnost in razširljivost, zato želimo v 2023 izvesti stresne teste (simulacija velikega števila sočasnih uporabnikov). 3) Dodatne funkcionalnosti za pomoč pri postavitvi sistema v učilnicah. Glede na pridobljene izkušnje smo si za leto 2023 zadali nalogo, da razvijemo nekaj dodatnih funkcionalnosti, ki bodo v pomoč pri uporabi sistema. Prva je t. i. "preview" način delovanja, ki bo namenjen hitremu preskusu sistema pred začetkom predavanja. Druga je detekcija prekrivanja predavanj, s čimer želimo pokriti nekatere robne primere. 4) Zamejitev strojne opreme. Že v letošnjem letu smo izvedli pregled opremljenosti učilnic na različnih fakultetah oziroma tam, kjer smo izvajali pilotne projekte. Na osnovi izkušenj smo izbrali avdio/video opremo in nekaj te opreme kupili ter preskusili v okviru pilotskih predavanj. V 2023 bi želeli napisati priporočila, kakšna avdio/video oprema je potrebna za optimalno uporabo sistema. Priporočila bomo posredovali tudi Digitalni UL. 5) Izboljšanje uporabniške izkušnje. Dobra uporabniška izkušnja je zelo pomembna za pravo produktijsko delovanje, zato se želimo posvetiti tudi tej temi. S te plati bomo pregledali in po potrebi nadgradili vse komponente sistema ON. 6) Implementacija nadzorne plošče. Nadzorno ploščo za nadzor nad delovanjem sistema smo letos že postavili. V letu 2023 jo želimo nadgraditi z dodatnimi vpogledi, ki omogočajo "top-down" pregled nad delovanjem sistema. 7) Dokumentacija sistema. Uporabniška dokumentacija nastaja sproti in se dopolnjuje vzporedno z razvojem. Manjka pa sistemsko-tehnična dokumentacija ter skrbniška dokumentacija.

prof. dr. Marko Bajec

vodja projekta Online Notes