

# Duboko učenje i Prepoznavanje govora

Minijaturan voice assistant

Milan Bojić

Decembar 2022.

# Contents

<b>1</b>	<b>Uvodna ideja</b>	<b>3</b>
<b>2</b>	<b>Funkcionalni zahtevi</b>	<b>4</b>
2.1	Prepoznavanje govora . . . . .	4
2.2	Transformator govora . . . . .	4
2.3	Spoljašnji sistem . . . . .	5
2.4	Sekundarni model . . . . .	5
2.5	Text-2-Speech model . . . . .	6
2.6	Detektovanje nedefinisanih komandi . . . . .	6
<b>3</b>	<b>Tehnički zahtevi</b>	<b>7</b>
3.1	Ograničenja alata . . . . .	7
3.2	Dokumentacija . . . . .	7
3.3	Timski rad . . . . .	7
<b>4</b>	<b>Bodovanje i predaja projekta</b>	<b>9</b>

# 1 Uvodna ideja

U ovom projektu radićete nad mini sistemom *voice assistant-a*. Sistem će se sastojati od više komponenti, nad kojima ćete koristiti više ideja koje smo prošlo kroz kurs Dubokog učenja.

Projekat može da bude timski, odnosno, ideja je da se radi u timu do troje. Implementacija sistema treba da podržava sledeće funkcionalnosti:

- Razaznavanje zvuka od glasa
- Voice-2-Text
- Razaznavanje različitih komandi
- Pozivanje spoljašnjeg Sistema
- Obrada zadatih komandi
- Otkrivanje nesmislene komande
- Text-2-Voice

Projekat će biti vezan za finalni projekat iz Prepoznavanja govora, deo gradiva će se proći na tim vežbama.

Funkcionalni zahtevi za sistem su opisani u sekciji 2.

Tehnički zahtevi za sistem su opisani u sekciji 3.

Bodovanje zadatka, kao i instrukcije za predaju zadatka su dati u sekciji 4.

## 2 Funkcionalni zahtevi

Na slici se nalazi idejni dijagram sistema koji ćete da implementirate.

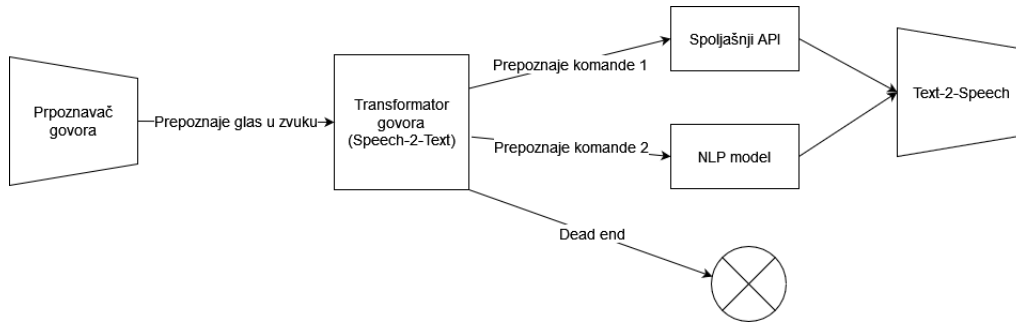


Figure 1: Idejni dijagram

Proći ćemo kroz svaku komponentu u sledećim pod sekcijama.

### 2.1 Prepoznavanje govora

Prepoznavanje govora (eng. Voice activity detection) je bitna komponenta svakog *voice assistant-a*, služi tome da oslobodi glavni "duboki" model od obrađivanja svakog zvuka. U najvećim sistemima uglavnom se koristi ključni izraz kao aktivator asistenta (npr. "Hello Alexa"). Ova komponenta je uglavnom jedini deo sistema na uređaju (npr. telefon, IOT uređaj).

Vi će te da implementirate komponentu sa ovom funkcionalnošću. Ona može da se implementira na jedan od dva načina:

- Jednostavan algoritam bez mašinskog učenja
- Jednostavna neuronska mreža

Kod neuralne mreže možete da koristite prepoznavanje glasa u zvuku ili da koristite ključni izraz da bi aktivirali vaš kompleksan model (npr. "Zdravo RAF").

### 2.2 Transformator govora

Jezgro sistema, kao što na dijagramu piše, služi za pretvaranje zvuka u tekst. Vi trebate da implementirate sekvencijalni model nad limitiranim rečnikom,

dovoljno je da vam se rečnik sastoji samo od komandu, koje prepoznajete, uz par permutacija na isto. Takođe možete da koristite pre trenirane modele kao osnovu vašeg modela, pa da vi samo doučite model sa vašim podacima.

Takođe, deo ove komponente je klasifikacija komandi i prosleđivanje istih ka sledećem sloju sistema. Ovaj deo komponente treba da prepozna komandu i ekstraktuje potrebne podatke za sledeći sloj.

## Koje je vreme u Beogradu

*komanda*      *informacija*

Figure 2: Primer komande

Komponenta treba da prepozna tri različita scenarija:

- Komada koje se prosleđuje spoljašnjem sistemu (Spoljašnji sistem)
- Komada koje se prosleđuje sekundarnom dubokom modelu (Sekundarni model)
- Tekst koji nije deo predefinisanih komandi

### 2.3 Spoljašnji sistem

Ako dolazna komanda odgovara ovom sistemu, izvučeni podaci se prosleđuju sistemu koji može da bude neki API ili neki mini program koji vi možete da implementirate koji će raditi odvojeno od vašeg sistema. Primer spoljnog sistema je API za vreme za date geografske lokacije (Primer komande) Izlaz iz sistema treba da se transformiše za sledeći sloj, neka to bude u oblika rečenice (npr. U Beogradu je 17 stepeni celzijusa, U Beogradu trenutno je oblačno...)

### 2.4 Sekundarni model

Ova komponenta treba da bude duboki model, gde za neku komandu i model treba da odradi podatke iz komande (npr. Sentimental analysis za rečenicu, ekstrakcija subjekta iz rečenice, spell checker...) Za vaš projekat najlakše

je da imate NLP model (Primer komande za NLP model), ali možete da koristite i druge arhitekture koje smo radili na vežbama (npr. Konvolucione mreže, Reinforcement learning, Diffusion modeli...)

Šta znači dan as je bio lep dan  
*komanda* *informacija*

Figure 3: Primer komande za NLP model

## 2.5 Text-2-Speech model

Funkcionalnost svakog *voice assistant* jeste zvučni izlaz koji daje tražene informacije korisnicima. Vi trebate da implementirate Sekvencijalni model koji radi ovaj zadatak. Možete da koristite pretrenirane modele kao osnovu, ali da ih referencirate u svom radu.

## 2.6 Detektovanje nedefinisanih komandi

*Voice assistant* treba da zna kada nešto ne razume ili ne zna, i da vrati informaciju korisniku o tom stanju (npr. Ne razumem vas, molim vas da probate...). Vaš sistem treba da pri klasifikaciji ulazne komande, prepozna da komanda nije definisana i da korisniku vrati zvučnu informaciju. Izlaz može da bude snimljen van Text-2-Speech model.

## 3 Tehnički zahtevi

Pošto sistem ima više komponenti, možete da koristite različite alate i jezike. Uz projekat očekuje se i kratka dokumentacija.

### 3.1 Ograničenja alata

Za vaše duboke modele dozvoljeno je da koristite Tensorflow (+ Keras) i PyTorch. Vaše modele treba da čuvate u **h5** formatu, kao što smo radili na vežbama. Pojedinačne model treba da čuvate u Google Collab-u i da se mogu pokrenuti na platformi. Kod komponenti koje nisu duboki modeli, dozvoljeni jezici su sledeći: C, C++, Python, Java, JavaScript, Golang. Ako radite u Python-u, bilo bi poželjno i da taj kod imate u Google Collab-u.

### 3.2 Dokumentacija

U vašoj dokumentaciji, treba da za svaku komponentu imate sledeće:

- Koji algoritam ste koristili
- Arhitekturu vaših dubokih modela
- Bazni model, ako ste ih koristili
- Vreme treniranja
- Parametri koji ste vi odabrali u sistemu

Takođe u dokumentaciji treba da napišete osnovne informacije studenata u timu (ime, prezime, indeks).

Format dokumenta treba da bude ili pdf ili docx.

### 3.3 Timski rad

Rad se može raditi u timovima. Tim se sastoji do tri člana. Imaćete period za prijavu projekta. Pri prijavi dobićete ime tima, i za svaki mejl oko projekta treba da prosledite ime tima u subject-u mejla. Pri slanju projekta, potrebno je da pošaljete samo jedan mejl.

Na odbrani očekuje se da svi članovi tima imaju znanja o svim komponentama u sistemu. Ako ne pokažu adekvatno znanje o sistemu biće negativno bodovani.

Ako želite da raditi individualno, za Duboko učenje, od vas se očekuje da implementirate sledeće komponente: Transformator govora, Sekundarni model, Text-2-Speech model, Detektovanje nedefinisanih komandi.

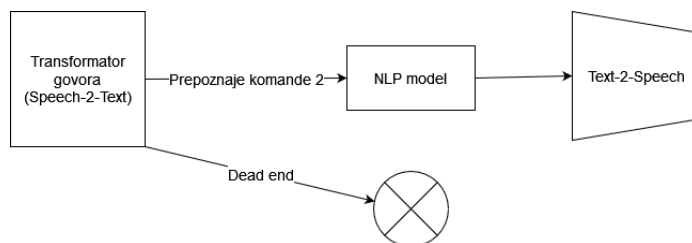


Figure 4: Dijagram sistema za individualan rad



## 4 Bodovanje i predaja projekta

zahtevi	bodovanje
Prepoznavanje govora	5
Transformator govora	7
Spoljašnji sistem	3
Sekundarni model	7
Text-2-Speech model	7
Detektovanje nedefinisanih komandi	2
Spajanje sistema	4
Dokumentacija	5

Tabela bodovanja za individualan rad

zahtevi	bodovanje
Transformator govora	9
Sekundarni model	9
Text-2-Speech model	9
Detektovanje nedefinisanih komandi	3
Spajanje sistema	5
Dokumentacija	5

Rok za prijavu timova je **11.12.2022 23:59**. Rok za slanje projekta je **08.01.2023 23:59**(moguće izmene datume). Prijava timova i projekat šaljite na mejl mbojic@raf.rs.