

PRIMENA VEŠTAČKE INTELIGENCIJE U ANALIZI I OBRADI JAVNIH GOVORA: TRANSKRIPCIJA, PROVERA ISTINITOSTI I KONZISTENTNOSTI

Apstrakt - U svetu gde se informacije šire izuzetnom brzinom, često je izazovno razlikovati tačne od netačnih podataka koje dobijamo iz medija. Ručna analiza javnih govora zahteva mnogo vremena i resursa, što je čini nepraktičnom za obradu velikog broja informacija. Kako bismo odgovorili na ovaj izazov, razvili smo sistem koji kombinuje moć velikih jezičkih modela i pretragu na internetu, omogućavajući automatsku proveru činjenica i pružanje pouzdanih rezultata. Ovakva tehnologija ima široku primenu – od medijskih kuća i istraživačkih timova do svih onih koji žele da dobiju pouzdane informacije i bolje razumeju sadržaj koji im se plasira.

Ključne reči: tačnost informacija, javni govor, veliki jezički modeli, veštačka inteligencija, automatizovana provera

I UVOD

U savremenom informacionom društvu, sve veći značaj pridaje se razvoju inteligentnih sistema za obradu i analizu javno dostupnih informacija. Sa pojavom naprednih tehnika obrade prirodnog jezika (NLP) i velikih jezičkih modela (LLM), stvoreni su uslovi za automatizovano prepoznavanje, analizu i evaluaciju izgovorenih informacija, što otvara prostor za primenu u različitim domenima – od novinarstva i politike do akademskih i regulatornih okvira.

Polazeći od pretpostavke da pouzdana analiza govora mora obuhvatiti sve njegove ključne komponente, u ovom radu se istražuje integrisani pristup koji objedinjuje tri funkcionalna podsistema: transkripciju govora, verifikaciju tačnosti tvrdnji i analizu konzistentnosti izjava. Poseban akcenat stavljen je na javni govor na srpskom jeziku, koji, zbog svoje jezičke specifičnosti i manje zastupljenosti u globalnim modelima, predstavlja izazovan, ali značajan istraživački kontekst.

Funkcionisanje celokupnog sistema zasniva se na hijerarhijskoj obradi podataka – proces započinje automatskom transkripcijom audio/video zapisa pomoću modela za pretvaranje govora u tekst, čime se formira ulaz za naredne faze analize. U sledećim koracima, sistem koristi velike jezičke modele za identifikaciju potencijalno netačnih tvrdnji kroz

upoređivanje sa relevantnim izvorima informacija, kao i za detekciju nekonzistentnosti u iskazima analizom prethodno datih javnih izjava iz odgovarajuće baze podataka.

U okviru ovog rada biće predstavljena arhitektura razvijenog sistema, primenjeni modeli i alati, kao i evaluacija efikasnosti svakog od podsistema. Kroz analizu dobijenih rezultata, biće razmotren potencijal ovakvog pristupa za unapređenje transparentnosti i odgovornosti u javnom govoru, kao i mogućnosti za dalji razvoj sistema u kontekstu šireg jezičkog okruženja.

II TRANSKRIPCIJA I DIJARIZACIJA GOVORA

Automatska transkripcija govora predstavlja ključni prvi korak u procesu analize audio i video zapisa, jer omogućava pretvaranje govora u tekstualni format, što je preduslov za sve dalje obrade. U okviru ovog projekta, poseban akcenat stavljen je na tačnost transkripcije i jasno razgraničenje među govornicima.

Za automatsku transkripciju korišćen je Whisper model. Prvobitno je testiran model Whisper Medium, zbog njegove brzine i nižih zahteva za resursima. Međutim, ispitivanja su pokazala da ovaj model ne pruža zadovoljavajuću tačnost na srpskom jeziku, naročito u segmentima koji uključuju politički govor, zbog čega je odlučeno da se pređe na Whisper Large model, koji je omogućio značajno preciznije rezultate.

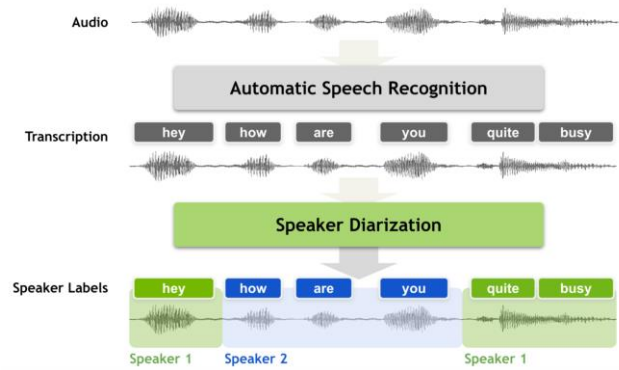
Nakon transkripcije, sprovedena je dijarizacija, odnosno proces identifikacije i odvajanja govornika u audio zapisu. Za ovu fazu korišćen je Whisper-Diarization alat, koji kombinuje automatsku transkripciju sa dodatnim modelima za prepoznavanje govornika, čime je omogućeno precizno označavanje govora svakog učesnika u transkriptu.

S obzirom na to da sirova transkripcija nije uvek optimalna za dalje analize, bilo je neophodno dodatno obraditi transkripte. Za ovaj zadatak testirani su modeli DeepSeek i GPT-4o Mini. Model DeepSeek je pokazao zadovoljavajuću tačnost, ali nedovoljnu

brzinu za potrebe projekta. Sa druge strane, GPT-4o Mini je omogućio i visoku tačnost i znatno veću brzinu obrade, što je dovelo do njegovog izbora za završno uređivanje transkripta.

Model	Accuracy	Speed
DeepSeek	Dobro	Slabo
GPT-4o Mini	Dobro	Dobro

Radi dodatnog unapređenja strukture transkripta, sproveden je proces segmentacije teksta na logičke celine (chunking). Testirane su različite metode segmentacije, uključujući semantičko grupisanje na osnovu tematskih promena, kao i jednostavnije pristupe poput segmentacije na osnovu dužine rečenica i pauza u govoru.



Konačni rezultat transkripcije i diarizacije predstavlja strukturirani JSON izlaz, u kojem su za svaki segment definisani govornik i pripadajuća tema razgovora. Ovaj format omogućava jednostavno prosleđivanje podataka sistemima zaduženim za proveru činjenica i konzistentnosti izjava, čime se značajno ubrzava analiza i smanjuje potreba za manuelnim radom.

```
"date": "2/3/2025",
"topics": [
  {
    "topic_name": "Rušenje Srbije",
    "description": "Diskusija o političkoj i društvenoj situaciji u Srbiji, uključujući kritiku vlasti, prelaznu vladu i rušenje same države.",
    "speaker_segments": [
      {
        "speaker_name": "Interviewers",
        "start_time": 0.00,
        "end_time": 2.596,
        "text": "Da li razmišljate o ostavci ili ekspertskoj prelaznoj vladi?"
      },
      {
        "speaker_name": "Speakers",
        "start_time": 3.403,
        "end_time": 35.827,
        "text": "Nismo se ovde skupili da čuvamo vlast bilo čiju, da čuvamo nečiju fotelju, nismo došli da govorimo o tome koliko žele izbore..."
      },
      {
        "speaker_name": "Speakers",
        "start_time": 36.231,
        "end_time": 62.084,
        "text": "A onda je došlo vreme za njihov pokušaj da Srbiju sruše. Naš je posao da svim tim ljudima pružimo ruku, da pokušamo..."
      }
    ]
  }
]
```

Proces transkripcije i diarizacije može se predstaviti sledećim koracima:

1. **Ulazni podaci:** audio/video zapis
2. **Transkripcija** korišćenjem Whisper modela
3. **Dijarizacija** i razdvajanje govornika
4. **Klasifikacija govornika**
5. **Poliranje i uređivanje transkripta**
6. **Chunking** (podela na logičke celine)
7. **Generisanje JSON izlaza**

Automatizacijom ovog procesa uspostavljen je čvrst i pouzdan temelj za dalje korake u analizi govora.

III PROCES PROVERE ČINJENICA (FACT CHECKING)

U ovom projektu, automatski sistem za proveru činjenica koristi velike jezičke modele za obradu prirodnog jezika i pretragu na internetu kako bi se verifikovale tvrdnje izrečene u audio i video zapisima.

Početak procesa provere činjenica je prijem podataka u JSON formatu od sistema za transkripciju. Ovaj format sadrži tekstualne transkripte govora, uključujući tvrdnje koje je potrebno proveriti, kao i relevantne

informacije o kontekstu, poput datuma govora i imena govornika. Na osnovu ovih podataka, veliki jezički model (LLM) analizira sadržaj i izdvaja ključne činjenice iz svake tvrdnje. Na ovaj način, sistem generiše listu činjenica koje je potrebno proveriti putem pretrage na internetu.

Nakon što se činjenice izdvoje, sistem generiše odgovarajuće upite za pretragu na internetu. Upiti se šalju pretraživaču korišćenjem SERPAPI servisa, koji omogućava prikupljanje rezultata sa pouzdanih i kredibilnih izvora, koji uključuju članke iz vesti, naučne publikacije i vladine izvore, dok se izbegavaju tabloidi i druge nesigurne platforme. Prikupljeni podaci se potom preuzimaju sa relevantnih veb stranica koristeći Jina API servis, koji omogućava ekstrakciju sadržaja.

Kada su podaci prikupljeni, sistem koristi LLM za analizu i upoređivanje sa originalnim tvrdnjama, kako bi procenio tačnost informacija. Sistem generiše izveštaj, koji pruža detaljan pregled svih činjenica koje su bile predmet provere, zajedno sa oznakama koje ukazuju na njihov status tačnosti (tačne, netačne ili neodređene). Pored toga, za svaku netačnu tvrdnju, izveštaj sadrži tačne informacije, kao i izvore sa kojih su tačne tvrdnje preuzete. Ovaj izveštaj se generiše u CSV formatu, čime se omogućava brzo pregledanje i dalja obrada rezultata.

Ovaj proces omogućava efikasnu i brzu proveru činjenica, smanjujući potrebu za manuelnim radom i povećavajući preciznost u analizi. Automatizacija ovog procesa čini ga pogodnim za velike količine podataka, kao što su transkripti govora političara i drugih javnih ličnosti, čime se omogućava brza detekcija netačnih ili manipulativnih tvrdnji. Konačan rezultat ovog procesa je tačna, objektivna i verifikovana lista činjenica koja je dostupna za dalje analize ili javnu upotrebu.

Proces provere činjenica može se predstaviti sledećim koracima:

1. **Ulazni podaci:** JSON format sa transkriptima govora.
2. **Ekstrakcija ključnih činjenica** korišćenjem LLM modela.
3. **Generisanje upita za pretragu** koji su optimizovani za pronalaženje relevantnih informacija.

4. **Pretraga na internetu** korišćenjem SERPAPI servisa.
5. **Preuzimanje sadržaja** sa relevantnih veb stranica koristeći Jina API.
6. **Analiza i procena tačnosti** prikupljenih podataka sa LLM-om.
7. **Generisanje izveštaja u CSV formatu**, koji uključuje verifikovane činjenice i njihove izvore.

Automatizacija ovog procesa omogućava bržu, precizniju i efikasniju analizu govora, posebno u kontekstu političkih i društvenih debata, gde tačnost informacija ima ključnu ulogu u oblikovanju javnog mišljenja.

IV PROCES PROVERE KONZISTENTNOSTI (CONSISTENCY CHECKING)

Consistency checking segment, zadužen za automatsku proveru konzistentnosti javnih govora, temelji se na RAG (Retrieval-Augmented Generation) arhitekturi.

U izradi celog sistema pa tako i consistency checking segmenta korišćen je Python kao osnovni programski jezik, prvenstveno zbog svoje čitljivosti, jednostavne sintakse i širokog spektra biblioteka koje omogućavaju efikasan rad sa podacima, implementaciju modela veštačke inteligencije, kao i jednostavnu integraciju sa različitim API servisima.

Izlaz sistema na kojem je rađeno u početnom delu sistema transkripcije je upravo ulaz sistema zaduženog za proveru konzistentnosti govora, to je JSON fajl koji sadrži transkribovani tekst izdvojen na segmente, kao i sve neophodne informacije o sagovornicima, dužini trajanja razgovora itd.

I faza - klasifikacija

U početnoj fazi rada, korišćenjem velikog jezičkog modela izvršena je klasifikacija segmenata JSON fajla kako bismo utvrdili da li transkript govora sadrži informacije za koje je potrebno ispitati konzistentnost ili ne. Na primer, izjave poput "dobar dan" ili "kako ste" ne podležu daljoj analizi jer ne sadrže suštinske informacije koje bi mogle biti predmet provere.

Pre samog procesa klasifikacije neophodno je podeliti svaki segment JSON fajla na manje celine, kako bismo obezbedili što efikasniji dalji rad. Za ovo je korišćena tehnika fixed length chunking sa overlapom, pri čemu smo u početku fiksnu dužinu celine podesili na 200 karaktera i 50 karaktera za preklapanje. Na samom kraju impementacije sistema smo podešavali ove dužine kako bismo dobili što tačnije rezultate. Za potrebe klasifikacije segmenata iz transkribovanih govora korišćen je OpenRouterAPI, koji omogućava fleksibilno povezivanje sistema sa više velikih jezičkih modela.

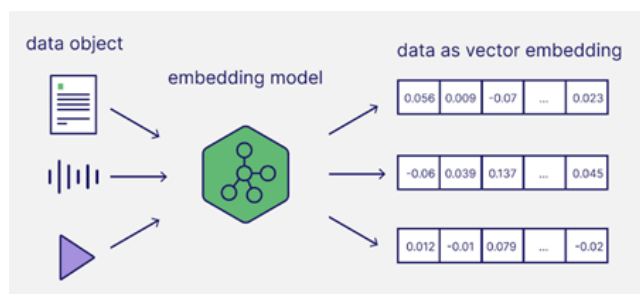
U okviru ovog sistema, primenjen je model Claude 3 Haiku kompanije Anthropic, poznat po efikasnosti u obradi i razumevanju prirodnog jezika.

Klasifikacija je izvršena prompt-based metodom, u zero-shot režimu, pri čemu model na osnovu jasno definisanih uputstava (engl. *prompt*) odlučuje da li određeni segment sadrži proverivu tvrdnju relevantnu za političku funkciju.

Ova procedura omogućava visoku fleksibilnost i skalabilnost bez potrebe za dodatnim treniranjem modela na domenskim podacima.

II faza - embedovanje

Ukoliko je neophodno ispitati konzistentnost segmenta vršimo njegovo embedovanje.



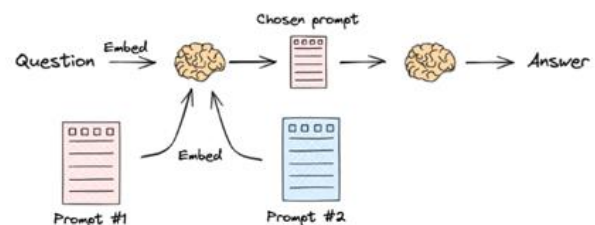
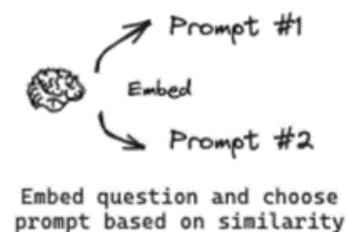
Embedovanje je ključni proces pretvaranja originalnih podataka, u našem slučaju pretvaranje transkribovanog govora, u numeričke vektore u višedimenzionalnom prostoru. (prikaz na slici iznad) Ovi vektori su od suštinskog značaja jer sažimaju semantičko značenje originalnih podataka na način koji omogućava računarskim modelima da lakše uporede i analiziraju sadržaj. Za proces ugradnje (embedding) tekstova u vektorski prostor korišćen je COHERE API, sa modelom embed-

multilingual-v3.0. Izabrali smo ovaj model jer je posebno specijalizovan za rad sa srpskim jezikom i time omogućava očuvanje semantičkog značaja tekstova prilikom vektorisanja, što je korisno za dalji rad.

III faza - stvaranje baze, pretraga(similarity search), response

Sledeći korak podrazumeva pretraživanje baze podataka koja u sebi sadrži embedinge svih prethodnih govora neophodnih za proveru konzistentnosti.

U ovom koraku je cilj pronaći prvih k embedinga koji su najrelevantniji za ispitivanje posmatranog govora (KNN- k nearest neighbor search). Odabrani najrelevantniji embedinzi će se potom koristiti za generisanje odgovora kojim popunjavamo kolonu csv fajla .



Ovaj odgovor može biti:

- "Konzistentno" / "Kontradiktorno" uz prethodno datu izjavu i datum kada je to rečeno
- "Ne znam" uz dodatno objašnjenje u slučaju kada nema relevantnih podataka za proveru

	A	B	C
1	Sentence;	"Response";	"Relevant";
2	Rekli ste da Srbija g	"Kontradiktorno";	"Srbija ima najveće
3	Vise od pola miliona	"Ne znam. Nema r	"Nema relevantnih
4	Tokom 2018. godine	"Konzistentno";	"Da li znate da sm

U određenim delovima sistema, gde je bilo potrebno dodatno zaključivanje i generisanje odgovora, korišćen je i model GPT-4o-mini, poznat po brzini i efikasnosti u obradi prirodnog jezika.

Kombinacijom ovih servisa i modela razvijen je pouzdan sistem sposoban da automatski detektuje kontradiktorne tvrdnje u tekstovima, uz mogućnost proširenja i primene u različitim domenima.

Jedna od glavnih poteškoća tokom razvoja bila je vezana za korišćenje OpenRouter API-ja, budući da su API ključevi brzo trošeni zbog ograničenog broja poziva, što je zahtevalo pažljivo planiranje i optimizaciju korišćenja modela tokom testiranja i razvoja.

V ZAKLJUČAK

Istraživanje predstavljeno u ovom radu doprinosi dubljem razumevanju potencijala velikih jezičkih modela u obradi i analizi govora. Korišćenjem savremenih tehnologija, uključujući modele poput Whisper i GPT-4o Mini, kao i integracijom eksternih API servisa, razvijen je funkcionalan i modularan sistem sposoban za obradu kompleksnih audio i video sadržaja.

Poseban akcenat stavljen je na izazove dijarizacije, generisanja semantički relevantnih upita, evaluacije informacija iz dostupnih izvora i izvođenje pouzdanih zaključaka o verodostojnosti tvrdnji. Postignuti rezultati pokazuju visok nivo efikasnosti sistema u realnim uslovima i pružaju vredne uvide za buduća istraživanja u oblasti političke analitike i medijske pismenosti.

Perspektive daljeg razvoja obuhvataju širenje sistema na veće i jezički raznovrsnije korpuse, unapređenje algoritama za rad sa akcentovanim i dijalekatskim govorima, kao i integraciju sa bazama podataka zvaničnih izvora informacija. Takav razvoj bi dodatno doprineo povećanju tačnosti, robusnosti i upotrebljivosti sistema u analizi javnog diskursa u savremenom digitalnom okruženju.