

Preizkus Googlovega govornega programskega vmesnika pri samodejnem razpoznavanju govornjene slovenščine

Simon Dobrišek, David Čefarin, Vitomir Štruc, France Mihelič

University of Ljubljana
Faculty of *Electrical Engineering*



Pregled predstavitve



- Razvoj razpoznavalnikov govora
- Googlov govorni programski vmesnik
- Preizkus govornega vmesnika
- Rezultati na treh govornih zbirkah
- Zaključni komentar

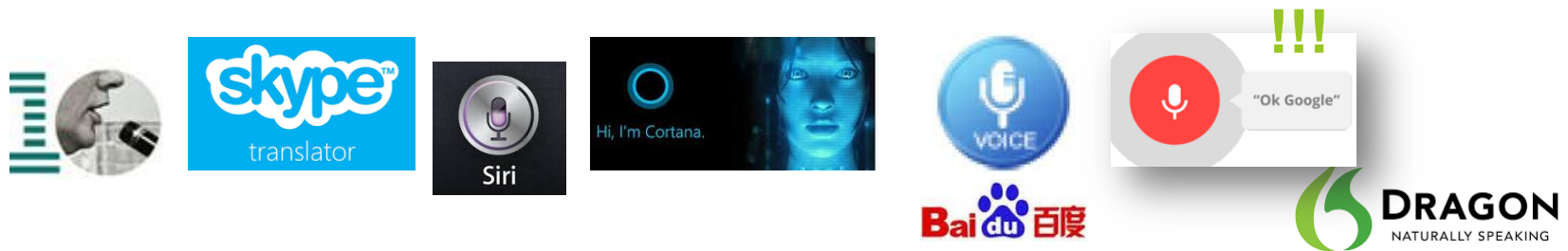
Razvoj razpoznavalnikov govora

- Razvoj samodejnih razpoznavalnikov govora je zahteven izziv, ki ga spremljajo **velika pričakovanja**, žal **pa tudi razočaranja**.
- Zaradi okoljskih, fizioloških, družbenih in drugih vplivov govor izkazuje **precejšnjo spremenljivost**, s katero se razpoznavalniki govora **s težavo spopadajo**.
- Še posebej to velja za govorjeno slovenščino, ki ima **veliko število pregibnih oblik** besed in bolj **prost besedni red**.

Razvoj razpoznavalnikov govora



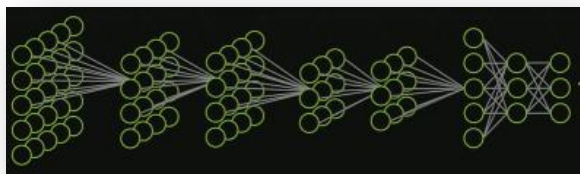
- Velika svetovna podjetja so razvila vrsto komercialnih računalniških programskih rešitev, ki vključujejo že solidno zanesljive razpoznavalnike govora.



- Po črnogledih pričakovanjih pa pri teh rešitvah podpora govornjeni slovenščini še **izostaja ali vsaj zaostaja**.

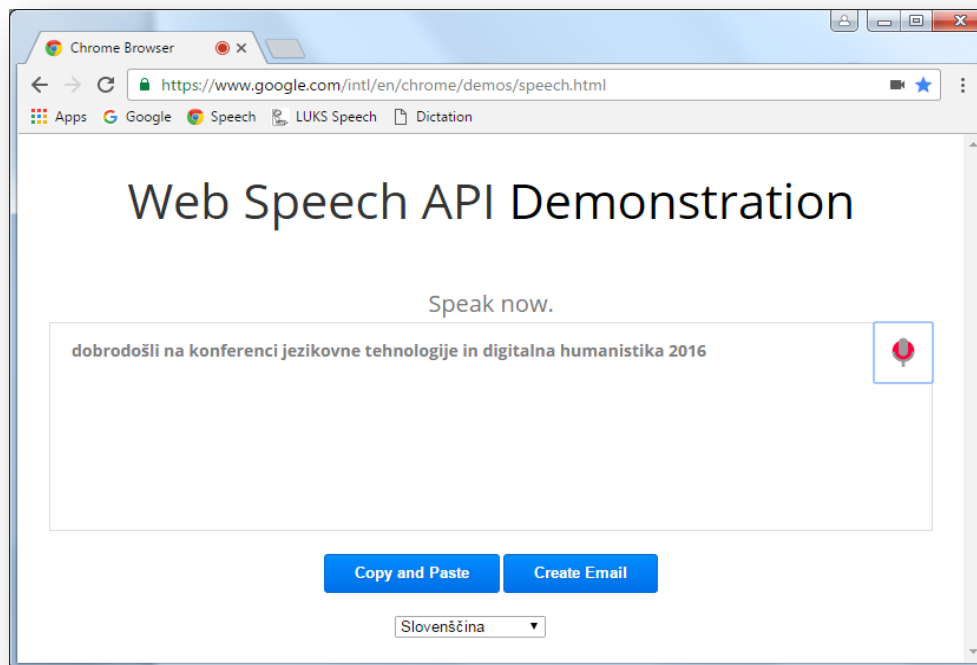
Razvoj razpoznavalnikov govora

- Prikriti Markovovi modeli (HMM) in modeli mešanic Gaussovih porazdelitev (GMM) so leta prevladovali kot najbolj uspešen zgled akustičnega in jezikovnega modeliranja govora.
- V zadnjih nekaj letih pa je bil dosežen velik napredek z novim zgledom modeliranja govora, ki temelji na uporabi t.i. globokih nevronskega omrežij (DNN).
- Nadaljnji napredek je bil dosežen z uporabo konvolucijskih nevronskega omrežij (CNN) in povratnih nevronskega omrežij (RNN), ki modelirajo dolgi kratkoročni spomin (LSTM).



Podpora govornjeni slovenščini

- Med najbolj razvitimi komercialnimi govornimi vmesniki je zaenkrat uspelo le **Googlu** povsem podpreti tudi govornjeno **slovenščino**.



Googlov govorni programski vmesnik

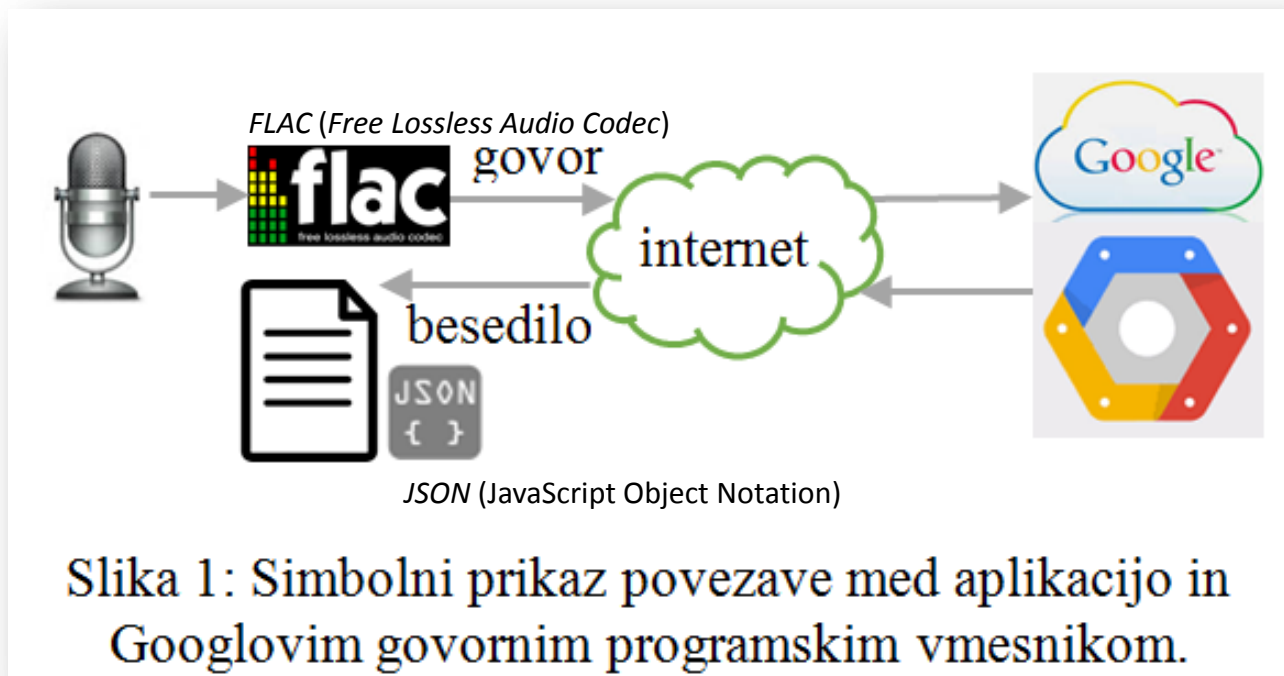
- Googlov govorni programski vmesnik je **slabo dokumentiran** in omejeno dosegljiv za širšo skupnost razvijalcev novih aplikacij.
- Vmesnik se najbolj enostavno uporablja z uporabo programskega jezika **Javascript** in programskega vmesnika Google Javascript Speech API.
- V tem primeru se aplikacije, ki uporabljajo Googlov govorni vmesnik, razvijejo kot **običajne spletne aplikacije**, ki se naložijo v spletni brskalnik.

Googlov govorni programski vmesnik

- Za neposredno uporabo Googlovega govornega programskega vmesnika potrebujemo **ključ**, ki je registriran v Googlovem oblračnem sistemu.
- Po pridobitvi ključa, ki istoveti našo aplikacijo, lahko Googlov govorni programski vmesnik brezplačno uporabimo do **50-krat na dan** in z omejitvijo na zvočne posnetke, ki so lahko dolgi do **15 sekund**.
- Za intenzivnejšo uporabo govornega programskega vmesnika je potrebno **plačilo po posebnem ceniku**.

Googlov govorni programski vmesnik

- Google govorni programski vmesnik lahko uporabljamo na izmenjujoči **enosmerni** ali hkratni **dvosmerni** način.



Googlov govorni programski vmesnik

- Parametri internetne povezave z govornim program-skim vmesnikom omogočajo različne nastavitve načina delovanja razpoznavalnika govora, kot so na primer:
 - `maxAlternatives` - število manj verjetnih rezultatov razpoznavanja, ki jih želimo pridobiti;
 - `continuous` - stikalo za razpoznavanje neprekinjenega tekočega govora;
 - `interim` - stikalo, ki omogoči pošiljanje vmesnih rezultatov razpoznavanja še pred koncem izrečenih povedi
 - `pFilter` - stopnjo zakrivanja/filtriranja neprimernih besed (psovke ipd) v rezultatu razpoznavanja

Googlov govorni programski vmesnik

- Rezultati razpoznavanja so v JSON formatu:

```
{"result":[]}  
{"result":[{"alternative":[  
{"transcript":"koliko časa traja polet iz amsterdama do 4 a","confidence":0.88918835},  
{"transcript":"koliko časa traja polet iz amsterdama do vrha"},  
{"transcript":"koliko časa traja polet iz amsterdama"},  
{"transcript":"koliko časa traja polet iz amsterdama do 4"},  
{"transcript":"koliko časa traja polet iz amsterdama določili ha"}],"final":true}],  
"result_index":0}
```

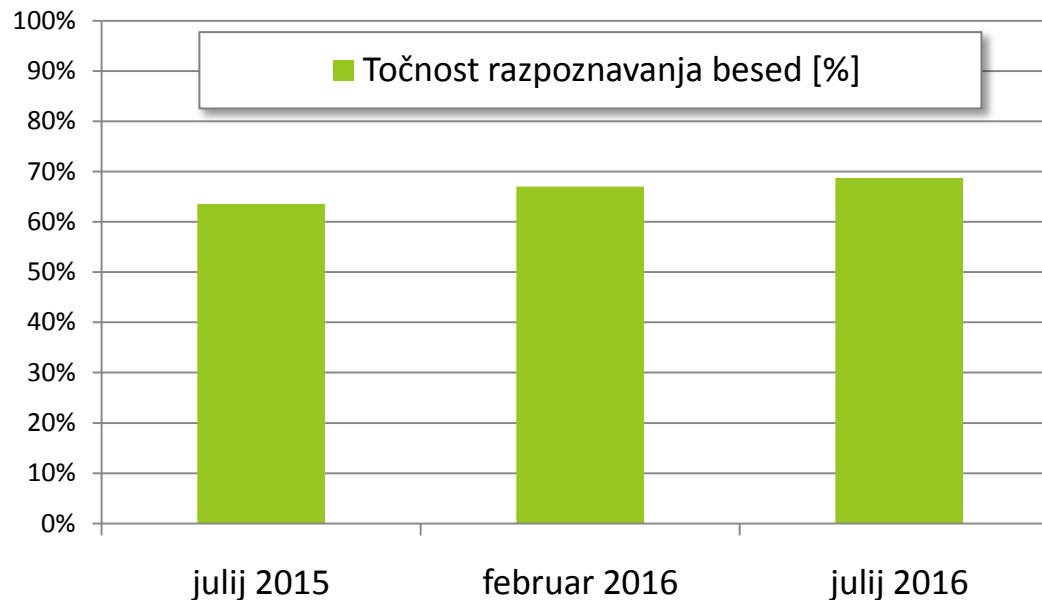
Preizkus govornega vmesnika

- Preizkus Googlovega govornega vmesnika smo izvedli predvsem z uporabo lastnih zbirk govornih posnetkov GOPOLIS in VNTV.
- Uporabili smo tudi nekaj dodatnih zvočnih posnetkov prebiranja elektronskih pisem v slovenščini (dolžine okoli 100 besed).
- Elektronska pisma so se prebiral najprej počasi in razločno, nato normalno hitro in manj razločno ter nato še spontano, manj razločno, z medmeti in s prekinitvami.

Rezultati na govorni zbirki GOPOLIS

- Skupaj 1925 testnih posnetkov povedi v skupnem trajanju dobre 1 ure in 37 minut se je poslalo Googlovemu govornemu programskemu vmesniku in pridobilo rezultate razpoznavanja.
- Preizkus smo v zadnjem letu dni izvedli trikrat, ker nas je zanimalo, če se bo rezultat zaradi morebitnega prilagajanja Googlovega razpoznavalnika govora že obdelanim govornim posnetkom v vmesnem času kaj izboljševal.

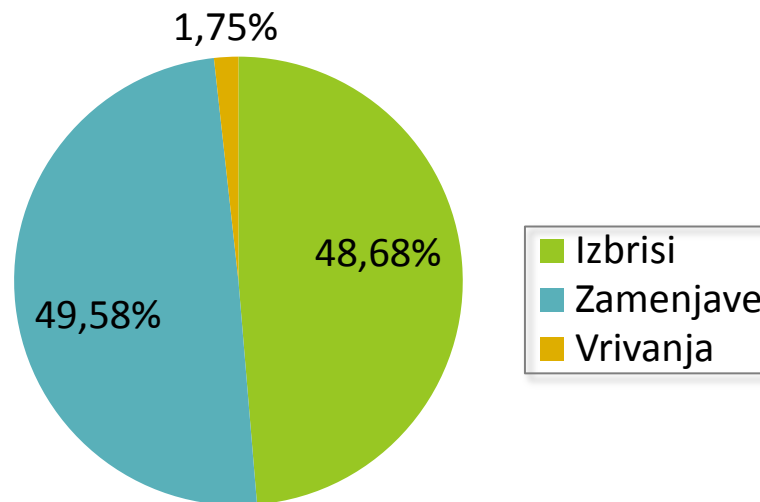
Rezultati na govorni zbirki GOPOLIS



- Zaradi večjega števila lastnih imen krajev, letališč in letalskih prevoznikov je del napak tudi posledica napak pri njihovem ortografskem zapisu, na primer *Sheremetyevo – Šeremetjevo*, *Zuerich – Cirihi*, ipd.

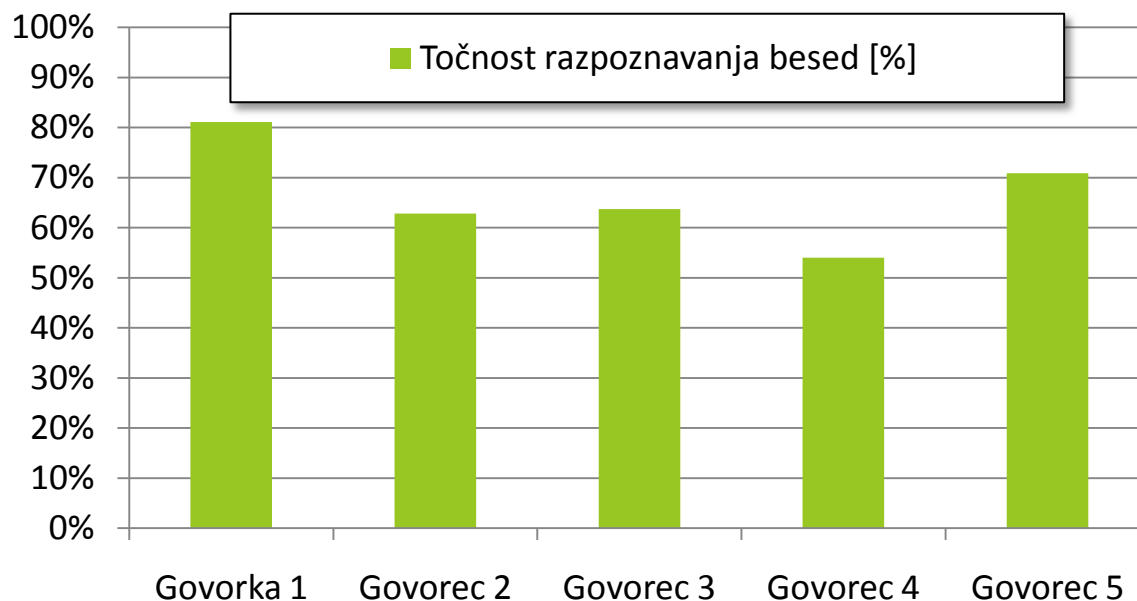
Rezultati na govorni zbirki GOPOLIS

- Analizirali smo tudi, kakšna je pogostost različnih vrst napak, torej napak izbrisa, zamenjave in izbrisa besed.
- Googlov razpoznavalnik je večkrat vrnil prazno besedilo (napake izbrisa besed) in to se je navadno dogajalo, kadar posnetek povedi ni bila zaključen z dovolj premora oz. tišine.



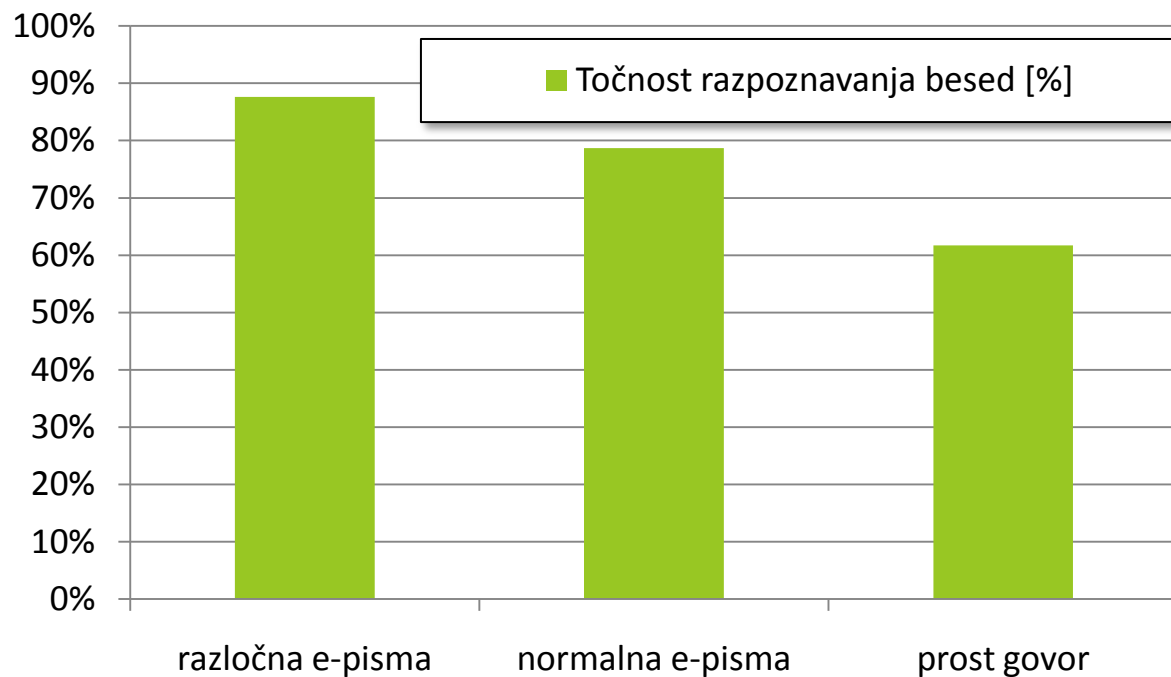
Rezultati na govorni zbirki VNTV

- Uporabili smo še 1548 televizijskih posnetkov povedi iz vremenskih napovedi petih govorcev iz govorne zbirke VNTV v skupnem trajanju 1 ure in 48 minut.



Rezultati na elektronskih pismih

- Rezultati razpoznavanja prebiranja elektronskih pisem, narekovanih na tri načine.

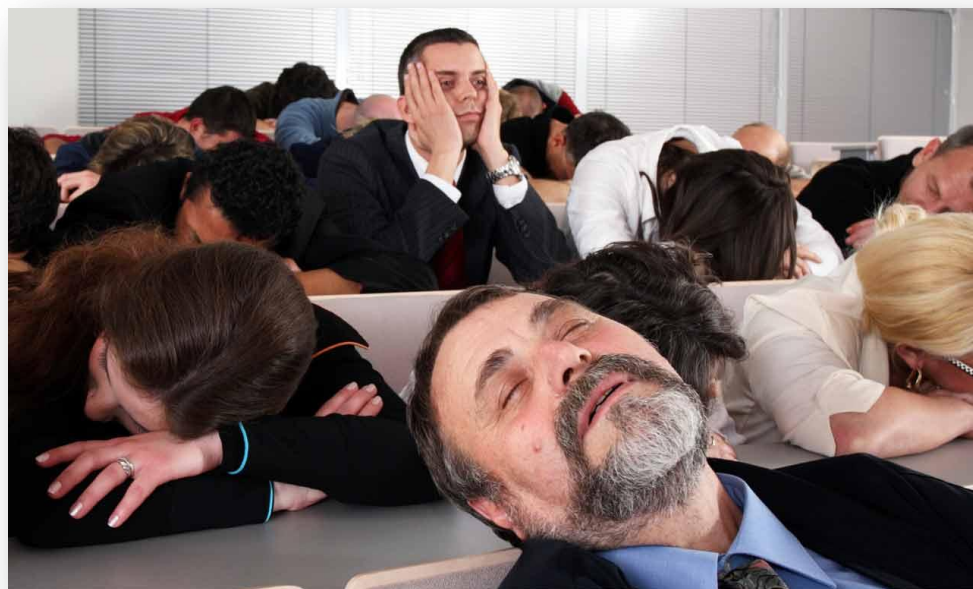


Zaključni komentar



- Napredek pri razvoju govornih tehnologij
- Podpora govorjeni slovenščini se izboljšuje
- Govorne tehnologije so vse bolj uporabne
- Nadaljnje izboljšave so vse bolj zahtevne
- Potreba po iskanje ožjih raziskovalno-razvojnih niš

Hvala za pozornost!



Vprašanja?