UDK: 025.43:808.61:681.32(047)=861
COSATI: 05-02, 09-02

Prepoznavanje ključnih reči u kontinualnom govoru srpskog jezika

Mr Ljiljana Stanimirović, dipl.inž.[1)]

Prikazani su najnoviji rezultati dobijeni u realizaciji algoritma za prepoznavanje određenog broja ključnih reči u kontinualnom govoru srpskog jezika. Suština primenjenog postupka je da se ključne reči u izgovorenoj rečenici modeluju skrivenim Markovljevim modelima (SMM) a da se u fazi dekodovanja računa mera pouzdanosti na celoj rečenici. Poređenjem mere pouzdanosti sa pragom, koji se utvrđuje u fazi obučavanja, utvrđuje se da li je u izgovorenoj rečenici izgovorena ili nije jedna od ključnih reči. Ključne reči se modeluju kontinualnim skrivenim Markovljevim modelima (KSMM), koji se dobijaju povezivanjem KSMM modela sa po tri stanja za svaki slog iz kojih se sastoji ključna reč. U radu je uvedena *mera kvaliteta realizovanog sistema - MSQ (measure_of_system's quality)* u cilju određivanja optimalnog koraka i optimalnog praga za meru pouzdanosti u fazi dekodovanja.

Ključne reči: **Prepoznavanje ključnih reči, algoritam za prepoznavanje, skriveni Markovljev model, srpski jezik, dijalog čovek-računar.**

Uvod

NAGLI razvoj govornih tehnologija je uzrok sve većeg broja raznih servisnih usluga zasnovanih na interaktivnoj komunikaciji čovek – računar, pre svega za engleski jezik. Kroz dijalog, govorom, korisnik dobija razne informacije od računara. Kod nas se, međutim, tek počelo sa razvojem sličnih servisa. Ovaj rad je rezultat rada na projektu čiji je cilj realizacija dijaloga čovek – računar za srpski jezik. U osnovi dijaloga je sistem za prepoznavanje određenog broja ključnih reči, koje se utvrđuju unapred zavisno od toga o kojoj se aplikaciji, odnosno servisnoj usluzi radi.

Za realizaciju uspešne aplikacije u ovoj oblasti potrebno je pažljivo razviti govorni dijalog koji će na najbolji način pružiti korisniku osećaj komfornosti i efikasnosti iskorišćenja napredne govorne tehnologije. Dijalog u sebi, pre svega, podrazumeva da sistem prepozna jednu od prethodno definisanih ključnih reči u kontinualnom govoru i da preduzme odgovarajuću akciju, odnosno pruži korisniku određenu informaciju.

Fokus istraživanja iznetog u ovom radu* se odnosi na realizaciju algoritma za prepoznavanje ključnih reči u kontinualnom govoru srpskog jezika koji se bazira na statističkoj metodi - skrivenim Markovljevim modelima (SMM), (Hidden Markov Models - HMMs). Ovo istraživanje je prirodan nastavak rada sa skrivenim Markovljevim modelima u sistemima za prepoznavanje izolovano izgovorenih reči srpskog jezika [1-6].

Dijalog

U ostvarivanju dijaloga čovek - računar, veoma je važno prevazići problem koji nastaje kada korisnik priča spontano i pritom koristi razne oblike iskazivanja emotivnog stanja (uzdasi, psovke, poštapalice). Cilj dijaloga jeste ostvarenje krajnjeg cilja korisnika, što znači da korisnik treba da dobije željenu informaciju, odnosno da iskoristi servisnu uslugu koja se kroz dijalog pruža. Nije neophodno da sistem razume svaku izgovorenu reč, već da razume suštinu korisnikove namere, odnosno da pruži korisniku željenu informaciju. Zato u realizaciji dijaloga treba ostvariti prepoznavanje ključnih reči u izgovorenoj rečenici, a ne svih izgovorenih reči.

Kao osnovne ciljeve, koje treba ostvariti u realizaciji dijaloga čovek - računar, treba navesti sledeće:

- *lako korišćenje servisa koji se kroz dijalog nudi*

dijalog treba da obezbedi što konkretniju realizaciju korisnikove želje da dobije neku od servisnih usluga koje pruža sistem.

- *prijatnost dijaloga*

dijalog treba da bude prijatan za korisnika, što podrazumeva prijatan glas kao odziv sistema i usredsređenost i vođenje dijaloga na takav način da se što brže ostvari korisnikova želja za nekom od servisnih usluga sistema. Dijalog treba da izgleda što spontanije uz samo povremeno traženje od korisnika da potvrdi da li je sistem dobro prepoznao ključnu reč, kao i uz simpatične, nekada kompromisne odgovore sistema, a ne samo kao šturo postavljanje pitanja i davanja odgovora.

- *mogućnost učenja*

dijalog treba da obezbedi sistemu za prepoznavanje ključnih reči, koji je sastavni deo dijaloga, da lakše dođe do ispravno prepoznate ključne reči, što znači da treba da sadrži u sebi i elemente ekspertnog sistema, odnosno elemente ugrađenog znanja sa mogućnošću učenja.

- *tečnost dijaloga*

dijalog treba da što više liči na dijalog između ljudi, koji pitaju i odgovaraju naizmenično, u želji da što pre dođu do zahtevanog cilja.

*Ovaj rad je rezultat rada na projektu br. S.1.04.10.0137 Ministarstva za nauku i tehnologiju Republike Srbije, za period 1998-2001. godina.

[1)] Institut "Mihajlo Pupin", 11000 Beograd, Volgina 15

- *nenapornost dijaloga*

važno je ostvariti takav dijalog da korisnik ne čeka dugo na odgovor sistema, da sistem odgovori na postavljeno pitanje što direktnije, odnosno da korisnik za što kraće vreme dobije traženu informaciju. Smatra se da je potrebno obezbediti dijalog u ne više od 20 rečenica, a da ceo dijalog ne traje duže od 3 minuta.

- *korektan odgovor sistema*

sistem treba veoma dobro da prepoznaje ključne reči, odnosno treba da ima što je moguće veću verovatnoću prepoznavanja da bi dijalog tekao u željenom pravcu. Smatra se da je neophodno da verovatnoća prepoznavanja ključnih reči bude bar 80 %.

- *odziv sistema - brzina odgovora*

veoma je važno da korisnik ne čeka dugo na odgovor sistema. Ovo vreme se skraćuje promišljeno organizovanim i vođenim dijalogom, kada se sistemu za prepoznavanje ključnih reči postavlja zadatak da uvek prepoznaje manji broj ključnih reči, npr. pet, a ne celokupan fond ključnih reči. Takva organizacija dijaloga se može shvatiti kao stablo, kada se na određenim nivoima dijaloga, odnosno granama stabla, prepoznaje samo po nekoliko ključnih reči. Sistem tako brže dolazi do prepoznate reči, odnosno pruža odgovor korisniku za kraće vreme nego u slučaju da stalno treba da prepoznaje ceo fond ključnih reči koji sistem podržava.

Kvalitet ostvarenog dijaloga, koji se procenjuje na osnovu toga koliko je korisnik njime zadovoljan, može se utvrditi na osnovu odgovora na sledeća pitanja:

- da li je korisnik dobio traženu informaciju od sistema?
- da li je sistem komforan za korisnika?
- da li je korisnik dobio informaciju za neko prihvatljivo vreme?

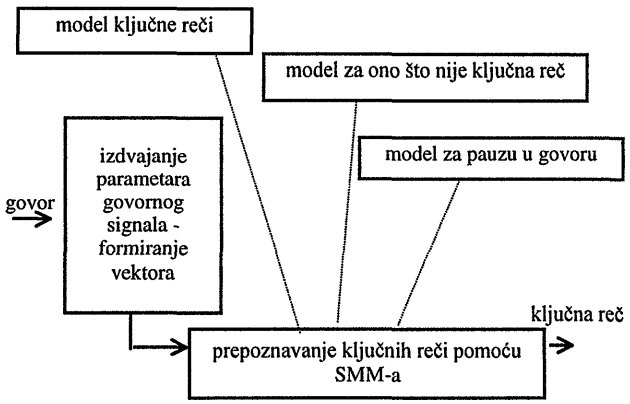
Suština dobro ostvarenog dijaloga je svakako dobar sistem za prepoznavanje ključnih reči u kontinualnom govoru, pa će njegovoj rezlizaciji biti posvećena pažnja u nastavku rada.

Prepoznavanje ključnih reči

Tehnika, koja je u engleskoj terminologiji poznata pod nazivom *keyword spotting* ili *word spotting,* predstavlja prepoznavanje nekih, unapred definisanih ključnih reči u spontanom govoru. U poređenju sa prepoznavanjem izolovano izgovorenih reči [6], kada se od korisnika očekuje da sarađuje, odnosno da u određenom trenutku kaže jednu od reči iz rečnika, prepoznavanje ključnih reči je fleksibilnije jer dozvoljava korisniku da može normalno, spontano da govori, a da pri izgovoru jedne od ključnih reči sistem za prepoznavanje odreaguje. U tom smislu, sistem za prepoznavanje ključnih reči se može posmatrati kao fleksibilniji sistem za prepoznavanje izolovano izgovorenih reči. Prepoznavanje ključnih reči u spontanom govoru se može posmatrati kroz dva različita pristupa [9]:

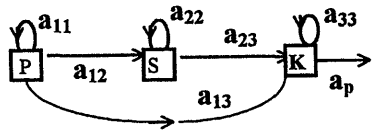
Prvi pristup

Blok-šema sistema za prepoznavanje ključnih reči može da izgleda kao na sl.1. Polazi se od toga da se posebno modeluje ključna reč, posebno ono što nije ključna reč i posebno pauza u govoru. Za modelovanje pojedinih delova izgovorene rečenice koriste se skriveni Markovljevi modeli. Osnovna govorna celina koja se modeluje je slog [10]. Poznato je, da je najveći koartikulacioni efekat unutar sloga, dok je on mnogo manji između dva sloga. Slogovi predstavljaju prirodne celine za obradu govora, kako zbog artikulacije, koja brine o ritmici govora otvaranjem i zatvaranjem vokalnog trakta, tako i zbog percepcije govora jer je dokazano da uvo ne čuje odvojeno suglasnik i vokal, već da ih obrađuje kao celinu. Kratke pauze se mogu pojaviti samo na krajevima sloga i one doprinose jasnijem i sporijem izgovoru.



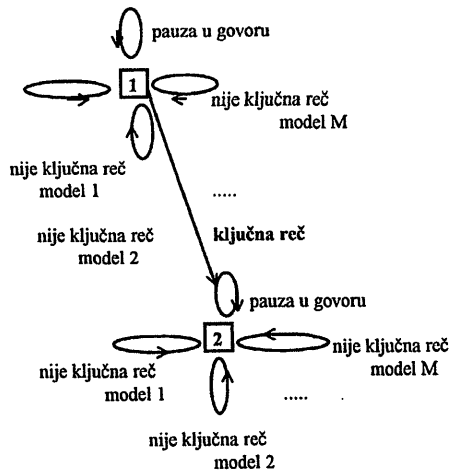
Slika 1. Sistem za prepoznavanje ključnih reči – prvi pristup

Model ključne reči se formira povezivanjem modela za slogove iz kojih se ključna reč sastoji. Slogovi se modeluju skrivenim Markovljevim modelima (SMM) sa tri stanja - početno (P), srednje (S) i krajnje (K), kao što je prikazano na sl.2 [10]. Verovatnoće prelaza iz jednog stanja u drugo stanje ili verovatnoće ostajanja u istom stanju su: $a_{11}, a_{12}, a_{13}, a_{22}, a_{23}, a_{33}$, dok je verovatnoća prelaza na sledeći skriveni Markovljev model a_p, model za sledeći slog. Koriste se kontinualni skriveni Markovljevi modeli [1-3].



Slika 2. Skriveni Markovljev model za jedan slog

Uloga modela za ono što nije ključna reč je da odvoji ključnu reč iz ulaznog govornog signala od ostatka rečenice. Zato je veoma važno da se napravi dobar model za ono što nije ključna reč u rečenici. Polazi se od toga da se u okviru jedne rečenice može pojaviti samo jedna od ključnih reči, što znači da se ceo sistem može nalaziti u sledeća dva stanja: 1. ili nije ključna reč (u pitanju je jedan od M modela) ili je pauza u govoru, 2. izgovorena je ključna reč. Dijagram stanja sistema je prikazan na sl.3



Slika 3. Dijagram stanja sistema za prepoznavanje ključnih reči – prvi pristup

Pauze u govoru se modeluju SMM modelima sa pet stanja. Modeli za ključne reči se formiraju u fazi obučavanja [2] od govorne baze sa kojom se raspolaže i od govornika koji su te reči izgovorili izolovano. Za modelovanje onoga što nije ključna reč, potrebno je napraviti M različitih modela, bar M=5, od govorne baze koju čine govornici koji su izgovorili rečenice u kojima nema ključnih reči.

Drugi pristup

Nedostatak prvog pristupa je što je potrebno veoma dobro modelovati ono što nije ključna reč u rečenici, što predstavlja složen i teško ostvariv cilj. S obzirom na to da se u prepoznavanju govora koriste skriveni Markovljevi modeli, nameće se ideja da bi bilo dobro kada bi se paralelno sa izračunavanjima koja su inače potrebna u fazi dekodovanja računao i neki parametar, koji bi ukazivao na to da li je ključna reč izgovorena ili ne, bez potrebe za detaljnim modelovanjem svakog dela izgovorene rečenice. Znači, u svakom trenutku trebalo bi da se ima odgovarajuća mera koja ukazuje na prisustvo ključne reči.

Drugi pristup se zasniva na izračunavanju mere, tzv. mere pouzdanosti (u engleskoj terminologiji: *confidence measure C(W/O)*)[7]. Ova mera pokazuje da je detektovano prisustvo reči W u sekvenci vektora O, ako je ispunjen sledeći uslov:

$$C(W|O) \geq prag \tag{1}$$

pri čemu se *prag* određuje eksperimentalno u fazi obučavanja modela. Treba ponovo naglasiti, da nije potrebno posebno modelovati ono što nije ključna reč unutar izgovorene rečenice, što je svakako prednost ovog pristupa u odnosu na prvi.

Prema [7] mera pouzdanosti se računa kao u (2) tj. kao negativan logaritam verovatnoće da je ključna reč W izgovorena u okviru sekvence vektora O.

$$C = -\log P(W \setminus O) \tag{2}$$

pri čemu je $P(W \setminus O)$ verovatnoća da je u sekvenci vektora O detektovana ključna reč W. Ako se primeni Bayesovo pravilo i pređe na nivo frame-a, izraz (2) se može napisati kao u (3), tako da se za lokalnu meru pouzdanosti $c(O_t \setminus s_j)$, za vektor O_t u stanju s_j, dobija:

$$c(O_t \setminus s_j) = -\log \frac{P(O_t \setminus s_j) P(s_j)}{P(O_t)} \tag{3}$$

U izrazu (3), verovatnoća da je generisana sekvenca vektora O_t, $P(O_t)$ računa se uzimajući u obzir sva stanja modela ključne reči u obzir, kao u izrazu (4):

$$P(O_t) = \sum_k P(O_t \setminus s_k) P(s_k) \tag{4}$$

Svako stanje s_k modela ključne reči sada emituje lokalnu meru pouzdanosti u konvencionalnom Viterbijevom algoritmu [2]. U fazi dekodovanja, potrebno je izračunati ukupnu meru pouzdanosti *ISc*, koja se dobija kao suma lokalnih mera pouzdanosti u vremenskom intervalu koji odgovara trajanju ključne reči, kao u (5), gde t_1 *i* t_2 treba da budu granice u kojima se nalazi ključna reč, po [7].

$$ISc(O) = \sum_{t=t_1}^{t_2} c(O_t \setminus s_j) \tag{5}$$

Posebno će se razmatrati određivanje granica t_1 *i* t_2 tzv. optimalnog koraka za svaku ključnu reč, kao i praga za meru pouzdanosti prisustva ključne reči u izgovorenoj rečenici.

Određivanje optimalnog koraka

Opis baza govornih reči

Za potrebe istraživanja snimljene su tri baze govornih uzoraka [8]. Snimanje je obavljeno preko standardnog mikrofona na sound blasteru na standardnom PC-u u radnom ambijentu. Učestanost odmeravanja je 8 kHz. Prva baza govornih uzoraka SDB (the sentence database) sastoji se od 60 rečenica koje sadrže ili ne sadrže 4 ključne reči koje su izgovorili 20 govornika. Izabrane su sledeće ključne reči: *Beograd, Beopetrol, krstaši* i *pobednik.* Ovakve ključne reči su izabrane zato što pojedine reči slično zvuče (*Beograd, Beopetrol, pobednik)*, pa se algoritam može testirati pod težim okolnostima. Druga baza govornih uzoraka KWDB (the keyword database) sadrži izolovano izgovorene ključne reči koje su izgovorili 20 govornika. Treća baza govornih uzoraka TSDB (test sentence database) se sastoji od 100 rečenica, različitih od onih u SDB bazi, koje sadrže ili ne sadrže ključne reči, a izgovorili su ih 20 govornika. Ova baza govornih uzoraka je korišćena za testiranje.

Modelovanje ključnih reči

Kao što pokazuje izraz (3), mera pouzdanosti se računa za svaku rečenicu iz SDB baze za svaki vremenski trenutak, tako što se, za svaku ključnu reč njen skriveni Markovljev model pomera kroz izgovorenu rečenicu i računa se ukupna mera pouzdanosti. Model za ključnu reč je dobijen tako što je svaki slog ključne reči modelovan jednim skrivenim Markovljevim modelom sa tri stanja, kao na sl.2, pa su modeli za slogove nadovezani jedan na drugi.

U predobradi govornog signala računa se 12 kepstralnih koeficijenata duž MEL skale dobijenih pomoću FFT-a [11] na intervalu od 32 ms, sa Hammingovim prozorom sa preklapanjem na pola u frekventnom opsegu telefonskog kanala. Koristi se preemfazirani signal sa koeficijentom 0.95.

Prema (5), računa se ukupna mera pouzdanosti *ISc* za svaki vremenski trenutak na sledeći način. U bazi SDB je utvrđen vremenski interval u kome je izgovorena ključna reč za svaku ključnu reč. Različiti govornici su ključnu reč izgovorili za različito vreme. Za svako moguće trajanje ključne reči, tzv. korak, računa se ukupna mera pouzdanosti prema (5), podrazumevajući da korak odgovara intervalu $[t_1, t_2]$. Na primer, za ključnu reč *Beograd*, moguće trajanje, odnosno korak se nalazi u intervalu od 30 do 50. Zatim se određuje minimalna vrednost ukupne mere pouzdanosti MIN1 za svaku rečenicu iz SDB baze, za svaki korak u cilju da se odredi optimalni korak i optimalni prag za meru pouzdanosti za svaku ključnu reč. Pošto se unapred zna u kojim se rečenicama nalazi ključna reč, može se postaviti cilj da se maksimizira mera kvaliteta sistema - **MSQ** (*measure-of-system's quality)*, razmatrajući različite korake i pragove. Mera kvaliteta sistema je uvedena kao u (6), kao kriterijum koliko je sistem dobar:

$$MSQ = \frac{n\_g\_d\_kw}{n\_kw} * \frac{n\_nkw - n\_g\_d\_nkw}{n\_nkw} \tag{6}$$

gde su:

- **n_g_d_kw**- broj ispravno detektovanih ključnih reči (u onim rečenicama gde ih stvarno i ima),

- **n_kw**- ukupan broj ključnih reči,
- **n_g_d_nkw**- broj pogrešno detektovanih ključnih reči (u onim rečenicama u kojima ih nema) i
- **n_nkw** je ukupan broj rečenica koje ne sadrže ključne reči.

Cilj je postići što bolji sistem, tj. imati što bolju meru kvaliteta sistema **MSQ** [12]. To znači da sistem treba da prepozna što je moguće više ključnih reči u rečenicama u kojima ih ima, a istovremeno sistem ne treba da prepoznaje ključne reči u rečenicama koje ih ne sadrže. Postupak je sledeći: posmatra se minimalna vrednost mere pouzdanosti MIN1 za svaku rečenicu u bazi SDB koja sadrži ključnu reč. Za svaki mogući korak, prag se računa kao minimalna vrednost od svih MIN1 vrednosti za ove rečenice.

Eksperimentalni rezultati

Sistem je testiran sa bazom TSDB. Dobijeni rezultati su dati u tabeli 1. Može se zaključiti da sistem veoma dobro prepoznaje ključne reči. U svakoj od po deset rečenica u kojima se nalazi ključna reč, ključna reč se ispravno prepoznaje. Sistem je grešio kada je prepoznao ključnu reč u rečenicama koje je nisu imale (npr., za ključnu reč: *Beograd*, sistem je pogrešno prepoznao 3 od 90 rečenica).

Vredno je napomenuti da je sistem pokazao dobre rezultate uprkos činjenici da se tri ključne reči: *Beograd*, *Beopetrol* i *pobednik*, mogu smatrati konfuznim (zvuče slično).

Tabela 1. Rezultati prepoznavanja po ključnim rečima

ključna reč	n_g_d_kw / n_kw	$\left(1 \approx \frac{n-g-d-nkw}{n-nkw}\right)$	MSQ
Beograd	10 / 10	87 / 90	96 %
Beopetrol	10 / 10	90 / 90	100 %
pobednik	10 / 10	81 / 90	90 %
krstaši	10 / 10	84 / 90	93.33 %

Zaključak

Cilj istraživanja je bio da pokaže da je moguće realizovati dobar sistem za prepoznavanje ključnih reči u kontinualnom govoru srpskog jezika i pored toga što su ključne reči izabrane kao konfuzne i činjenice da se nije raspolagalo većom bazom govornih uzoraka za treniranje modela za ključne reči. Bolje modelovani modeli ključnih reči sigurno bi doprineli i većoj verovatnoći prepoznavanja celog sistema.

Korišćeni su skriveni Markovljevi modeli na taj način da je svaki slog jedan skriveni Markovljev model sa tri stanja. Sledeći korak u istraživanju bi bio da se modeluju manji delovi reči, odnosno da se koriste skriveni Markovljevi modeli sa tri stanja za fonemu u kontekstu tzv. trifon. Takođe, interesantno bi bilo testirati sistem za slučaj većeg broja ključnih reči.

Literatura

[1] STANIMIROVIĆ,LJ. *Prepoznavanje izolovano izgovorenih reči iz ograničenog rečnika srpskog jezika.* magistarski rad, Elektrotehnički fakultet, Beograd, 1997.

[2] ĆIROVIĆ,Z. *Obučavanje skrivenih Markovljevih modela za prepoznavanje izolovano izgovorenih reči iz ograničenog rečnika.* magistarski rad, Elektrotehnički fakultet, Beograd, 1996.

[3] STANIMIROVIĆ,LJ., ĆIROVIĆ,Z. Sistemi za prepoznavanje pojedinačno izgovorenih reči bazirani na skrivenim Markovljevim modelima. *Naučnotehnički pregled*, 1996, vol.XLVI, no.1-2, p.67-77.

[4] RABINER,L., JUANG,B-H. *Fundamentals of speech recognition.* Prentice Hall, 1993.

[5] STANIMIROVIĆ,LJ., ĆIROVIĆ,Z., SAVIĆ,M. *Isolated Serbian word recognition system.* In: Proc.of the Int. Conference of Signal Processing and Communication - ICSPC'98, Las Palmas, Spain, 1998.

[6] ĆIROVIĆ,Z., STANIMIROVIĆ,LJ. *Man-Machine Communication: An Isolated Word Recognition System Based On Hidden Markov Models.* In: Proc. of the DMMS'97, Budapest, Hungary, 1997, p.111-117.

[7] JUNKAWITSCH,J., RUSKE,G., HOEGE,H. *Efficient methods for detecting keywords in continuous speech.* In: Proc. of the IEEE ICASSP'96, Vol. II, Munich, Germany, 1996.

[8] STANIMIROVIĆ,LJ., STANKOVIĆ,N. *Prepoznavanje ključnih reči u kontinualnom govoru srpskog jezika.* ETRAN'98, zbornik radova, vol.II, p.399-401, Vrnjačka Banja, Jugoslavija, 1998.

[9] LEE,K.H., KIM,H.S. *Computational Reduction in the Continuous HMM Based Keyword Spotting System.* In: Proc. of the ICSP'97, Seoul, Korea, 1997, p.475-478.

[10] STANIMIROVIĆ,LJ. *Serbian connected digit recognition*, In: Proc.of the IWSSIP'97, Poznan, Poland, 1997, p.117-119.

[11] STANIMIROVIĆ,LJ. *Optimalni vektor parametara govornog signala u sistemu za prepoznavanje govora zasnovanom na skrivenim Markovljevim modelima.* TEHNIKA, 1998, no.5, E7-E11.

[12] STANIMIROVIĆ,LJ., ĆIROVIIĆ,Z. *Keyword spotting system for Serbian language.* In: Proc.of the ICT' 99, Korea, 1999, p.234-236.

Rad primljen: 31.8.1999.god.