

AUDIO BIBLIOTEKA ZA SLEPE I SLABOVIDE OSOBE KAO VID PRIMENE TTS TEHNOLOGIJE

Dragiša Mišković, Nataša Vujnović, Milan Sečujski, Vlado Delić
Fakultet tehničkih nauka, Novi Sad

Sadržaj –U radu je prikazana audio biblioteka kao novi resurs za slepe i slabovide osobe na južno-slovenskom govornom području. Audio biblioteka je aplikacija izgrađena na bazi tehnologije sintetizovanja govora na osnovu teksta na srpskom jeziku (TTS) koja je prethodnih godina razvijana na FTN u Novom Sadu. U radu su opisani moduli i arhitektura aplikacije koja on-line pretvara tekst u govor. Na serveru su pohranjeni brojni tekstovi iz knjiga, časopisa, udžbenika i drugih tekstualnih sadržaja, kojima se pristupa pomoću aplikacija organizovanih u klijent-server arhitekturi. Pored tehničkih aspekata ukazano je i na prva iskustva u primeni i društveni značaj izrade jednog ovakvog resursa.

1. UVOD

Razvoj govornih tehnologija na jednom jeziku omogućuje pojavu velikog broja korisnih primena i novih aplikacija na tom govornom području. Govorne tehnologije su za razliku od mnogih drugih novih tehnologija zavisne od jezika i ne mogu se jednostavno uvesti iz inostranstva, odnosno moraju se razvijati gotovo iz početka za svaki novi jezik. Potrebno je objediniti znanja iz lingvistike, fonetike, akustike, programiranja, digitalne obrade signala i drugih disciplina, povezati ih sa specifičnostima tog jezika i implementirati u okruženje reapoložive računarske konfiguracije da bi taj računar mogao naglas da čita bilo koji tekst, razumljivim i koliko-toliko prirodnim sintetizovanim govorom (TTS), odnosno da bi mogao da razume čovekov govor i prevodi ga u tekst (ASR).

Jedna od najhumanijih i trenutno najznačajnijih primena TTS tehnologije je kao pomagalo za slepe i slabovide osobe kojima u izvesnoj meri pomaže da nadomeste svoj hendikep i ravnopravnije se uključe u društvene tokove života i rada. Naime, pomoću računara koji govore, slepe i slabovide osobe mogu samostalno da čitaju i pišu, pristupaju pisanim informacijama, pismeno komuniciraju, rade i tako se ravnopravnije uključe u proces obrazovanja i socijalne integracije. Zahvaljujući dostignućima u razvoju govornih tehnologija za srpski jezik, sada se otvara mogućnost definisanja novih pristupa u realizaciji socijalne podrške i na srpskom govornom području, tj. razvoja pomagala i resursa za osobe oštećenog vida.

Najnoviji i do sada ubedljivo najkvalitetniji sintetizator govora na osnovu teksta na srpskom jeziku, *anReader*, slepim i slabovidim korisnicima računara omogućuje lakše razumevanje, duže i prijatnije korišćenje govorne sinteze nego bilo koji drugi sintetizator koji se koristio na prostorima bivše Jugoslavije. Izuzetan kvalitet sintetizovanog govora motivisao je izradu novih resursa i pomagala za osobe oštećenog vida u našoj zemlji, kao što su računarski rečnik za slepe i slabovide, govorni portal za slepe i slabovide i, najzad, audio biblioteka za slepe i slabovide osobe (ABSS).

U Srbiji i Crnoj Gori ima blizu 15.000 slepih i slabovidih osoba, a najveći centar za školovanje učenika oštećenog vida je škola „Veljko Ramadanović” u Zemunu. Pisane informacije neophodne za obrazovanje ove dece bile su do sada dostupne pomoću sintetizatora govora skromnog kvaliteta, ili na Brajevom pismu, ili u vidu zvučnih zapisa. Slušanje neprirodnog sintetizovanog govora sa pogrešnom intonacijom i akcentima je zamorno. Izdavanje knjiga na Brajevom pismu je skupo, sporo i uglavnom u malom broju primeraka, a za njihovo arhiviranje potreban je poseban prostor i održavanje. Snimanje i arhiviranje zvučnih zapisa je takođe vredan, ali skup resurs ograničenih mogućnosti u pogledu obrazovanja i informisanja.

Cilj izrade audio biblioteka za slepe i slabovide je da prevaziđe sve pobrojane probleme. Arhitektura klijent-server aplikacije omogućuje stvaranje ogromne baze udžbenika, knjiga, časopisa i drugih pisanih informacija u tekstualnom formatu, koja se jeftino arhivira na disku računara, lako proširuje novim naslovima, a pomoću *anReader*-a po potrebi pretvara u glasan, mnogo prijatniji i razumljiviji sintetizovani govor. Istom sadržaju može istovremeno da pristupa veći broj učenika, a pristup nekim servisima audio biblioteke biće omogućen i preko Interneta, kao što će i tekstualni sadržaji na Internetu biti dostupni iz računarske mreže na kojoj je instalirana ova audio biblioteka. Izgradnja audio biblioteke za slepe i slabovide obezbeđuje uslove da čitava populacija slepih i slabovidih osoba može da pristupa pisanim informacijama pomoću računara (neposredno kroz audio biblioteku) ali i bez računara – preko snimljenih audio CD-ova sa sintetizovanim govorom. Pri tome treba reći da pravljenje audio zapisa određenog tekstualnog sadržaja pomoću govornog softvera može da se uradi neuporedivo brže i jeftinije nego da ga čita živ spiker.

Krajnji cilj je da slepim i slabovidim osobama postanu dostupni svi tekstovi koji se za videće osobe štampaju, a za slepe i slabovide osobe automatski pretvaraju u govor.

U nastavku ćemo opisati tehničke aspekte izrade, organizaciju aplikacije, module i arhitekturu audio biblioteke.

2. ORGANIZACIJA APLIKACIJE

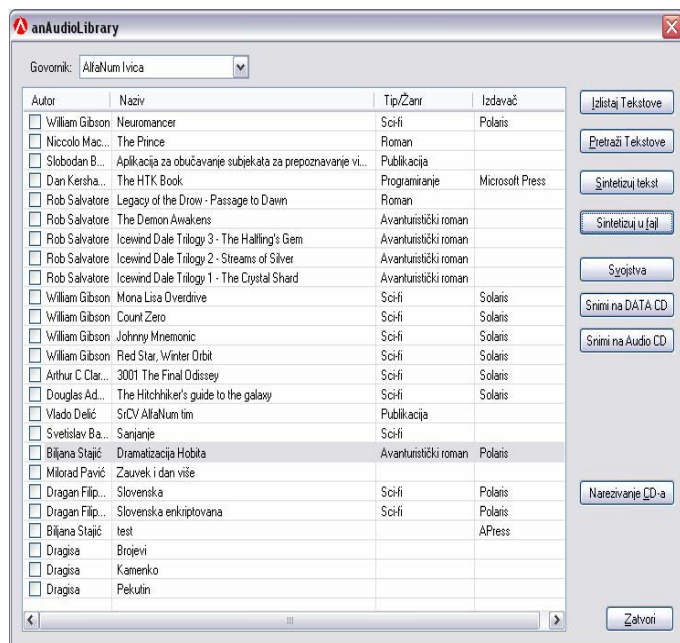
Generisanje govora u okviru audio biblioteke zasnovano je na *anReader*-u, softveru koji vrši sintezu govora iz teksta nezavisno od načina na koji se tekst unosi u računar. Funkcionisanje ovog modula se u potpunosti oslanja na *anTTS C++* biblioteku koja vrši sintezu pronalazenjem i spajanjem optimalnih govornih odsečaka iz posebno pripremljene govorne baze. Za razliku od dosadašnjih sintetizatora na ovim prostorima govorni segmenti ne mogu biti samo dvoglasni, nego i čitavi slogovi pa i cele reči, što značajno doprinosi kvalitetu računarski generisanog govora. Kvalitetu sintetizovanog govora pored daleko bogatije govorne baze posebno doprinosi akcenatsko-morfološki

rečnik srpskog jezika, koji sadrži veliki broj reči u skoro svim svojim pojavnim oblicima i u kome su označeni akcenti i još neke morfološko-sintaksičke kategorije, koje pomažu u pronalazenju optimalne intonacije reči i rečenica. Ovako precizan i obiman akcenatsko-morfološki rečnik predstavlja osnov za pravilno čitanje čak i latiničnih tekstova bez dijakritičkih znakova, koji se najčešće koriste pri pisanju *e-mail* i SMS poruka.

U organizacionom smislu, audio biblioteka predstavlja model klijent-server aplikacije. Na serverskoj strani nalazi se administratorska aplikacija, baza knjiga i servis zadužen za prihvatanje zahteva sa klijentske strane i pristup bazi. Da bi se korisnicima omogućilo lakši pristup pojedinim delovima knjige one su snimljene u HTML formatu. Za razliku od tekstualnog, ovaj format omogućava obradu sadržaja pre same sinteze u smislu lociranja pojedinačnih paragrafa i pasusa.

Klijentska strana sadrži aplikaciju namenjenu direktnoj interakciji sa slepim korisnicima kao i zasebno implementirane funkcije koje omogućavaju mrežnu komunikaciju sa serverom. Korisnički interfejs čine dva modula – **anAdministrator** (smešten na serveru) i **anKlijent** (na klijentskoj strani).

AnAdministrator je interfejs namenjen za administriranje audio biblioteke. Omogućava dodavanje novih knjiga i pretraživanje postojećih, pri čemu kriterijum može biti autor, naziv dela, žanr, izdavač. Funkcionalnost aplikacije odnosno konektovanje na bazu i izvršavanje potrebnih upita se ostvaruje preko ODBC (*Open Database Connectivity*) drajvera za MySQL.



Slika 1. *anKlijent*

AnKlijent je deo audio biblioteke sa kojim slepi korisnici jedino i dolaze u dodir. Projektovan je sa posebnim osvrtnom na njihov hendikep, a rezultat toga je aplikacija sa kojom se lako, uz pomoć tastature, ostvaruje interakcija. Potpuna ozvučenost ovog dela biblioteke znači da je postojeća TTS tehnologija iskorišćena za automatsko čitanje sadržaja prozora, menija i drugih tekstova bitnih za rukovanje

aplikacijom. Na ovaj način je eliminisana potreba za dodatnim čitačem ekrana (npr. *JAWS for Windows*) koji praktično i ne funkcionišu dobro ukoliko su meniji i nazivi komandi na srpskom jeziku, kao što je ovde slučaj.

Korisniku su pored informacija o postojećim knjigama u bazi dostupne brojne mogućnosti koje se mogu podeliti u dve grupe:

- 1) Opcije TTS-a vezane za način čitanja:
 - Izbor TTS govornika pri čemu to ne moraju biti samo govornici vezani za AlfaNum TTS *engine*.
 - Podešavanje parametara svakog govornika kao što su brzina govora, visina tona, čitanje interpukcije, skraćena, stranih izraza i reči bez naših slova (č,ć,đ...)
- 2) Opcije same audio biblioteke:
 - Preslušavanje određene knjige uz mogućnost navigacije kroz pasuse i paragrafe. Nakon izbora knjige i startovanja *on-line* sinteze korisniku su na raspolaganju komande za pauziranje, prelazak na prethodni ili sledeći pasus (paragraf)...
 - Sinteza u *wav* fajl sa mogućnošću izbora kvaliteta zvučnog zapisa odnosno stepena kompresije.
 - Sinteza i snimanje na CD sa mogućnošću izbora formata (audio ili "data" CD)
 - Dodatni prikaz bitnih informacija o knjizi kao što su ukupna veličina teksta, broj paragrafa i pasusa, procena trajanja sintetizovanog govora i sl. Ove informacije su veoma bitne prilikom snimanja sintetizovanog govora na neki od prenosnih medijuma.

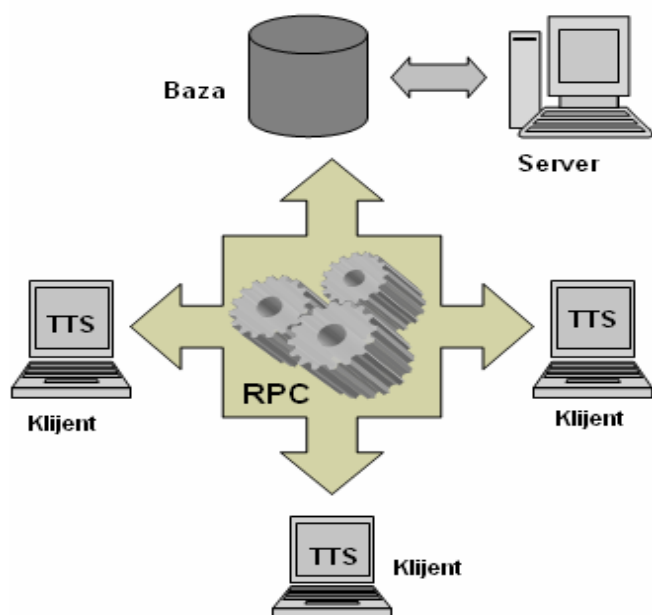
3. ARHITEKTURA SISTEMA

AlfaNum audio biblioteka organizovana je kao modularan sistem čime se postiže zadovoljavajući nivo apstrakcije u smislu olakšanog proširivanja postojeće funkcionalnosti. Korelacije između pojedinih modula prikazane su na slici br. 2.

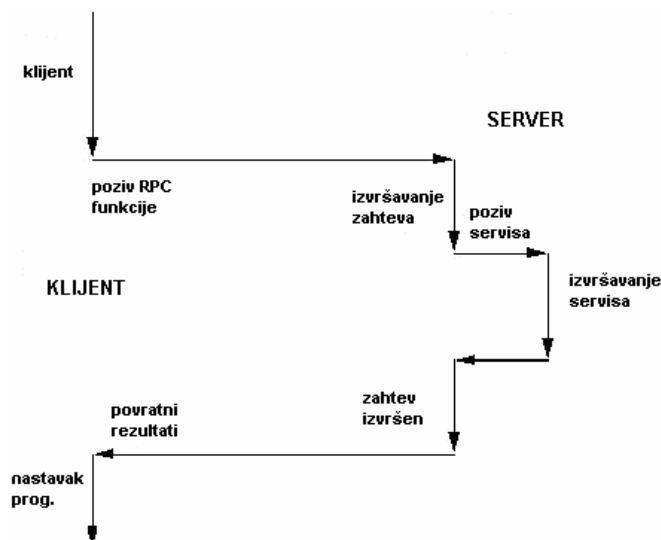
Izvršni modul celokupne aplikacije je anKlijent. On se sa jedne strane oslanja na komunikacioni modul a sa druge na SAPI 5.0 (*Speech Application Programming Interface*) preko koga obezbeđuje pristup nekom od Windowsovih virtuelnih TTS govornika. SAPI interfejs nudi niz olakšica vezanih za automatsko rukovanje audio uređajem, *multi-threading*, izbor govornika i sl. Ove prednosti omogućavaju anKlijentu da pored AlfaNum TTS *engine*-a, implementiranog u vidu dva virtuelna govornika, koristi i inicijalnu TTS sintezu ugrađenu u Windows koja je prikladnija za obradu tekstova pisanih na engleskom jeziku.

Najviši nivo apstrakcije obezbeđuje RPC (*Remote Procedure Call*) mehanizam. RPC obezbeđuje da aplikacija, koja u posmatranoj arhitekturi komunikacije ima ulogu klijenta, koristi API (*Application Programming Interface*) definisan za pristup aplikacionom interfejsu servera. RPC mehanizam omogućen je preko *dll*-a koji se registruje na klijentskim računarima i u kome su implementirane rutine koje u sprezi sa serverom omogućavaju pribavljanje podataka iz baze kao što su: lista knjiga, rezultati pretraživanja kao i sam sadržaj određene knjige (u HTML formatu). Servis koji se izvršava na serveru osluškuje RPC pozive koji dolaze sa

klijentskih računara i izvršava pozvane funkcije za pristup bazi podataka. Uloga ovih funkcija je da generisanjem SQL upita obezbede povratne rezultate klijentu. Slika 3. prikazuje osnovne korake u funkcionisanju ovog mehanizma.



Slika 2. Arhitektura sistema



Slika 3. RPC mehanizam

Ovakva organizacija pretvara audio biblioteku u jedan sistem praktično nezavisan od lokacije servera i baze knjiga, te dozvoljava upotrebu Interneta za klijent-server komunikaciju.

4. ZAKLJUČAK

Uspesahan razvoj govornih tehnologija za srpski jezik omogućio je izradu inovativnih aplikacija i resursa kao što je audio biblioteka za slepe i slabovide, opisana u ovom radu.

Govorne tehnologije predstavljaju logičan nastavak razvoja interakcije između čoveka i računara. AlfaNum audio biblioteka namenjena je slepim i slabovidim osobama, kojima je govor praktično i jedini način da praktično ostvare pravo na obrazovanje i samostalno informisanje. U skladu sa tom činjenicom, nastavak rada na ovom projektu biće usmeren na ugrađivanje komponente za automatsko prepoznavanje govora čime će se omogućiti dvosmerna govorna komunikacija između čoveka i računara.

Pored toga, mehanizam za komunikaciju između klijenata i servera koncipiran u ovom radu otvara vrata daljem razvoju aplikacije u smeru obezbeđivanja Internet podrške.

U ovoj ili sličnoj formi audio biblioteka može da posluži i za stara i invalidna lica kojima je teško da drže i listaju knjige i časopise.

ZAHVALNICA

Izgradnju Audio biblioteke u školi za učenike oštećenog vida "Veljko Ramadanović" u Zemunu, finansiralo je odeljenje za međunarodni razvoj Velike Britanije u okviru opštinskog akcionog programa socijalnih inovacija opštine Zemun.

LITERATURA

- [1] MSDN Library Visual Studio 6.0
- [2] M.Sečujski, R.Obradović, D.Pekar, Lj.Jovanov, V. Delić: *AlfaNum System for Speech Synthesis in serbian Language*, TSD 2002, Brno, Czech Republic, 2002
- [3] T.Dutoit: *An Introduction to Text-To-Speech Synthesis*, Kluwer Academic Publishers, London, 1997.
- [4] Microsoft Corporation Speech SDK. [CD_ROM]

Abstract – This paper presents an application created for blind and visually impaired people. The application reads custom books on the client machine from a collection on the server. Software is based on client-server model and existing TTS technology developed for Serbian language.

AUDIO LIBRARY FOR BLIND PEOPLE AS AN APPLICATION OF TTS TECHNOLOGY

Dragiša Mišković, Nataša Vujnović, Milan Sečujski, Vlado Delić