

Digitalizacija jezika i razvoj jezičkih tehnologija u funkciji digitalizacije kulturnog naslijeđa

doc.dr.sc. Alma Granov

E-mail: alma.granov@dujit.ba

Abstrakt: *U digitalnom okruženju suvremenog čovjeka od izuzetne je važnosti razviti tehnologije za pristup informacijama, proizvodima i uslugama informacijskog društva na prirodnom jeziku. U procesu digitalizacije kulturne baštine neke zajednice osobito je važno da taj digitalni materijal bude pretraživ i dostupan na matrenjem jeziku zajednice kako bi se osigurao ravnopravan i slobodan pristup za sve korisnike zajednice, ali je isto toliko važno i obezbijediti funkcioniranje i povezivanje digitalizirane građe za funkcioniranje u multilingvalnom i multigrafijskom okruženju digitalnog prostora. To se isključivo može postići primjenom jezičkih tehnologija, te u tom smislu razvoj jezičkih tehnologija i digitalizacija jezika predstavljaju strateški faktor u postizanju punih ciljeva digitalizacije kulturne baštine neke zajednice i njenog integriranja u svjetsku kulturnu baštinu.*

Ključne riječi: *digitalna paradigma, jezičke tehnologije, obrada prirodnog jezika, automatsko prevođenje*

1 Uvod

Digitalna paradigma postojanja realnost je u kojoj živi suvremeni svijet. Digitalizacija je stvorila, odnosno, profilirala jedan poseban način života-digitalni život. U današnjem razvijenom svijetu u procesima digitalizacije posebno mjesto se poklanja digitalizaciji jezika. S pravom, jer jezik je jedan od najdominantnijih sistema za komunikaciju ab hominibus relegare.

U današnjem digitalnom okruženju mi svakodnevno stupamo u interakciju sa kompjuterima i prirodna je naša težnja da tu komunikaciju obavljamo sistemom znakova koji nam je najbliži, naprirodniji, za koji posjedujemo anatomsku i kognitivnu specijalizaciju. Taj sistem znakova je prirodni jezik.

Čovjek jezikom verbalizira svoj um i svoju svijest, ono što je on, putem jezika saopćava, odnosno prenosi ljudima oko sebe.

U jeziku je pohranjeno, između ostaloga i kulturno naslijeđe zajednice koja se koristi tim jezikom tako da je jezik nešto što je nezaobilazan faktor u razvojnoj paradigmi jedne zajednice.

Virtualni prostor digitalnog postojanja postaje danas ključnim elementom afirmacije i glavni garant opstojnosti. Tadić ističe da se do nedavno u novinarskim krugovima znala pojavljivati uzrečica da ono što nije bilo prikazano na televiziji, kao da se nije ni dogodilo, a da je danas tu ulogu preuzeo web prostor i da je prisutnost nacionalnog jezika na web prostoru dovedeno gotovo do razine nacionalnog simbola (Tadić, 2003:47).

Jasno je već odavno da širenje digitalnog prostora neće biti moguće u tako velikoj mjeri ukoliko ono bude nastojalo egzistirati primarno na jeziku na kojem su kompjuteri 'progovorili'. Iako se možemo složiti da je engleski jezik lingua franca naše civilizacije, ipak da bi se postigla puna upotrebnost vrijednost digitalnih sistema oni moraju biti

realizirani na jezicima zajednice kojoj ti digitalni sistemi trebaju služiti. Tek masovnom i efikasnom upotrebom digitalni sistemi svojim daljim udruživanjem postižu povezanost i svoju punu efikasnost u povezivanju na globalnom nivou. LIP softveri bili su prvi koji su ukazali na potrebu afirmacije nacionalnih jezika a u svrhu omasovljenja digitalnih sistema i njihovih potencijala.

2 Jezičke tehnologije

Jezičke tehnologije bi se mogle definirati kao zbirni naziv za skup metoda i postupaka za informatičku obradu prirodnog jezika. Jezičke tehnologije se često nazivaju tehnologije prirodnog (ljudskog) jezika (HLT) ili se za njih koriste odrednice kao što su procesiranje prirodnog jezika ili kompjuterska obrada prirodnog jezika (NLP). Jezičke tehnologije predstavljaju zaokruženu granu IT tehnologija koja počiva na spoznajama lingvistike i predstavljaju izrazito interdisciplinarnu naučnu granu čije su dvije osnovne sastavnice lingvistika i informatika. Jezičke tehnologije imaju za cilj digitalizaciju čovjekove govorne i jezičke sposobnosti. Postavlja se, tako i logično i nužno pitanje pripadaju li jezičke tehnologije lingvistima ili informatičarima. Odgovor bi mogao glasniti ni jednima ni drugima, odnosno i jednima i drugima i to iz razloga što sve ključne informacije za razvoj jezičkih tehnologija dolaze upravo iz spoznaja i dostignuća lingvistike, a informatičke tehnologije su te koje služe kao sredstvo za obradu tih nalaza lingvistike.

Prema Tadiću, jezičke tehnologije se načelno dijele na: jezične resurse, jezične alate i komercijalne proizvode (Tadić, 2003:27).

Jezični resursi jesu zbirke jezičkih podataka koje su kompjuterski pribavljene, pohranjene i podržane. Jezičke resurse prvenstveno čine digitalni korpusi,

tj. korpusi koji su kodirani s ciljem da budu kompjuterski (mašinski) pretraživi i digitalni rječnici, tj. oni koji su digitalno obrađeni i pohranjeni te kompjuterski (mašinski) pretraživi. Jezičke alate čine aplikacije koje obrađuju ili se služe već postojećim jezičkim resursima (korpusima i rječnicima) u digitalnom obliku, te su jezički alati upravo ti koji omogućavaju bržu, jednostavniju, efikasniju i jeftiniju upotrebu prirodnog jezika u informatičkom okruženju, dok su komercijalni proizvodi oni koji su nastali na temelju istraživanja jezičkih resursa jezičkim alatima (Tadić, 2003:27).

Omogućiti pristup servisima od javnog značaja na jeziku zajednice koja ga koristi trebao bi biti jedan od strateških ciljeva zajednice. Treba imati na umu da digitalizacija jezika može omogućiti i harmoniziranje i rješavanje čak i nekih bitnih političkih pitanja.

Tako npr. u višejezičnim zajednicama, kakva je i BiH, sa tri službena jezika i dva pisma gotovo je uvijek, u većoj ili manjoj mjeri opterećena pitanjem službenog jezika, jednakim tretmanom tih jezika i pisama od državnog vrha, parlamenta pa sve do nižih razina vlasti i dalje u sve pore društva.

Jezičke tehnologije i digitalizacija jezika mogu relaksirati ovu situaciju razvojem sistema za automatsko (strojno) prevodenje za službene jezike BiH. Time bi se ostvarivala zagarantirana sloboda pojedinca da se opredjeljuje za upotrebu jezika i pisma koji želi, znatno bi se pojednostavilo pripremanje mnogih administrativnih radnji koje su opterećene sadašnjim postupcima lektoriranja i pripremanja materijala za našu administraciju u sve tri jezičke norme, a konačno svi bi se dokumenti mogli imati i na tako potrebna sva tri jezika i oba pisma.

To je nešto što bi direktno bilo u funkciji očuvanja kulturne baštine BiH, omogućavajući ljudima da bez ikakvih formalnih i političkih barijera mogu baštiniti ono što je kulturološka tekovina i raznoliko bogato kulturno naslijeđe BiH neodričući se niti jednog dijela svoje kulturne historije zato što to neko drugi umjesto njih želi. Isto tako, sama mogućnost jednakomjerne i same mogućnosti neometane upotrebe svih jezika BiH, spriječila bi mnoge neugode i zloupotrebe koje su trenutno prisutne samo zbog mogućnosti pojedinca da diskriminira upotrebu nekog od standarda.

Obrada prirodnog jezika u digitalnim sistemima nužna je jer se bez digitalnih korpusa, rječnika, ne može kvalitetno prezentirati ni mnoga arhivska građa. Jezičke zapise iz pojedinih perioda treba obraditi na takav način da oni budu u punom informacijskom kapacitetu dostupni i običnom istraživaču, a ne samo specijalisti iz oblasti, a to digitalni medij može omogućiti. Time, nacionalna građa i informacija koju ona nosi postaje nešto što lahko može ući u naše domove i više nego do sada

profilirati naš odnos prema našem kulturnom naslijeđu.

Kako su među osnovnim ciljevima i zadacima jezičkih tehnologija, između ostalih, izrada programa za optičko prepoznavanje pisma, rukopisa, izrada programa za obradu informacija pohranjenih u prirodnojezičkom obliku, obrada arhiviranih dokumenata, kompjuterski potpomognuto stvaranje i uređivanje teksta, ekstrakcija informacija i generiranje sažetaka, jasno je od kolike su važnosti postupci digitalizacije jezika u procesu digitalizacije kulturnog naslijeđa.

Konačno i prema nalazima UNESCO-a smatra se da će razvoj jezičkih alata biti ključ u potpori većoj prisutnosti multilingvizma i multigrafije na internetu što treba biti faktorom budućeg rapidnog razvoja procesa u digitalizaciji kulturnog naslijeđa upravo u onim sredinama koje su manje ili više pogođene današnjom predominacijom engleskog jezika na internetu.

Zaključak

Digitaliziranu kulturnu baštinu čine kompjuterski fundirani materijali koji bi trebali biti sačuvani za generacije koje dolaze. Jeziku kao tzv. nematerijalnom aspektu kulturne baštine svakako u procesu digitalizacije kulturne baštine pripada posebno mjesto.

O značaju i ulozi koju jezik ima za čovjeka možda na najjednostavniji i najilustrativniji način govori priča o Vavilonskoj kuli.

Digitalizaciji kulturne baštine cilj je i omogućiti neometan, slobodan, ravnopravan pristup ovoj građi svim korisnicima. To nikako nije moguće ostvariti bez razvijenih jezičkih tehnologija kako za jezik zajednice iz kojeg potiče građa tako ni bez razvoja alata koji će omogućiti prevazilaženje jezičke barijere i omogućiti komforno multijezičko okruženje kako bi digitalizirana građa postala globalno dobro.

Jezik je do sada bio bitan ograničavajući element u afirmaciji kulturnog naslijeđa neke zajednice. Čak i kada govorimo o 'velikim' zajednicama i 'velikim' jezicima, uvijek je trebalo mnogo, mnogo truda i znanja da se elementi kulturne baštine prezentiraju i spoznaju.

Mogućnosti koje pružaju jezičke tehnologije danas bitno relativiziraju poteškoće jezičke barijere, ako možda ne bi bilo preslobodno konstatirati, one ih poništavaju. Ovo će svakako imati utjecaja na problikovanje dosadašnjih paradigmi i gledišta kada je ova problematika u pitanju.

Sve ovo danas otvara jedan drugi problem, a to je imperativni razvoj jezičkih tehnologija za svaki jezik pojedinačno, jer u današnjem digitalnom svijetu to znači postojanje ili nepostojanje, funkcionalna pismenost ili funkcionalna nepismenost neke zajednice, njeno uključanje u

razvijeni umreženi svijet ili izostanak iz njega i vegetiranje i propadanje na dokovima civilizacije. Stoga je razvoj jezičkih tehnologija i digitalizacija nezaobilzan i bitan faktor digitalizacije kulturne baštine i priključivanja i povezivanja sa svjetskom kulturnom baštinom.

Literatura:

- [1] Tadić, Marko, "Jezične tehnologije i hrvatski jezik", Ex Libris, Zagreb, 2003.
- [2] Granov, Alma, "Kompjuterska lingvistika", Sarajevo, Lica, 5-6, 54-57, 1998.
- [3] Granov, Alma, "Kompjuterska semiotika" (Gradovrh, Matica Hrvatska – Tuzla, Tuzla, godište IV, br. 4, studeni 2007, 44-52)

- [4] Lash Scott, "Tehnološki oblici života", u Theory, Culture & Society 18(1): 105-120, prev. Goran Vujasinović, 2001.
- [5] Jurafski, Daniel & Martin, James, "Speech and Language Processing", Prentice Hall, Inc. 2000.
- [6] <http://www.dujit.ba/>, 19.03.10.