Označavanje teksta na glagoljici pomoću TEI-a

Maganjić, Dino

Master's thesis / Diplomski rad

2015

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Faculty of Humanities and Social Sciences / Sveučilište Josipa Jurja Strossmayera u Osijeku, Filozofski fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:142:491953

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

FILOZOFSKI FAKULTET
SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU

Repository / Repozitorij:

FFOS-repository - Repository of the Faculty of Humanities and Social Sciences Osijek

zir.nsk.hr

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

Sveučilište J.J. Strossmayera u Osijeku

Filozofski fakultet Osijek

Diplomski studij informatologije

Dino Maganjić

Označavanje glagoljičkih tekstova pomoću TEI standarda

Diplomski rad

Mentor: doc. dr. sc. Boris Bosančić

Osijek, 2015.

Sadržaj

SAŽETAK ... 2
1. UVOD ... 3
2. OSNOVNO O TEI-u ... 4
2.1. Osnovno o digitalnoj humanistici ... 4
2.2. TEI standard ... 5
2.3. TEI označiteljski jezik ... 7
2.4. O XML-u ... 9
3. OZNAČAVANJE NESTANDARDNIH ZNAKOVA POMOĆU TEI STANDARDA ... 11
3.1. O Unicode-u ... 11
3.2. Prikazivanje nestandardnih znakova i glifova ... 12
3.3. Svojstva znakova i glifova ... 14
3.4. Anotiranje (označavanje) znakova i glifova ... 14
3.5. Dodavanje novih znakova ... 18
3.6. Korištenje kodnih vrijednosti iz PUA-e ... 19
3.7. Načini (stilovi) pisanja ... 19
3.8. Primjeri stilova pisanja ... 20
3.8.1.Vertikalni stil pisanja ... 20
3.8.2. Vertikalni tekst s umetnutim horizontalnim tekstom ... 21
3.8.3. Vertikalna orijentacija u horizontalnim pismima ... 22
3.8.4. Pisanje od dolje prema gore ... 24
3.8.5. Mješoviti horizontalni smjer ... 24
3.8.6. Primjeri rotacije teksta ... 24
4. PRIMJER OZNAČAVANJA GLAGOLJIČKOG TEKSTA POMOĆU TEI STANDARDA ... 26
4.1. Općenito o glagoljici, vrstama glagoljice i posebno o uglatoj glagoljici ... 26
4.2. Svrha, ciljevi i metodologija rada ... 26
4.3. Opis postupka označavanja ... 28
4.4. Provedba postupka označavanja glagoljičkog teksta putem TEI standarda ... 30
4.4.1. Označavanje transliteriranog glagoljičkog teksta Bašćanske ploče putem TEI standarda ... 30
4.4.2. Označavanje glagoljičkog teksta Bašćanske ploče čiji znakovi su mapirani s kodnim vrijednostima PUA-e ... 34
4.4.3. Prikaz znakova glagoljičkog teksta u pregledniku ... 35
5. ZAKLJUČAK ... 38

SAŽETAK

U ovom radu bit će govora o TEI standardu te njegovoj primjeni u kontekstu označavanja nestandardnih znakova, konkretno znakova uglate glagoljice. U skladu s tim, u teorijskom dijelu rada prikazan je okvirni razvoj digitalne humanistike, od početaka rada na digitalnoj inačici *Indexa Thomisticus*-a, pa sve do danas. Nadalje, prikazan je osnutak i razvoj Inicijative za označavanje teksta (*Text Encoding Initiative - TEI*) te njezinih principa, smjernica, tehničke specifikacije, projekata i sl. Uz TEI, rad se dotiče i razvoja XML-a, označiteljskog jezika koji predstavlja polaznu točka za postupak označavanja teksta, i *Unicode*-a, standarda koji se koristi za prikazivanje znakova i grafičkih reprezentacija znakova, tzv. glifova, većine svjetskih pisama u digitalnom okruženju. Što se tiče TEI smjernica, u radu je dan osvrt na peto poglavlje TEI smjernica koje opisuje načine označavanja nestandardnih znakova. U TEI smjernicama su navedene metode označavanja nestandardnih znakova i glifova te njihova svojstva, zatim, mogućnost dodavanja komentara ili anotiranja znakova, mogućnost dodavanja novih znakova, korištenje kodnih vrijednosti iz Privatnog korisničkog područja (*Private Use Area - PUA*) *Unicode*-a, te stilovi pisanja koji se mogu primijeniti na tekstove na raznim jezicima. U praktičnom dijelu rada označena je transliterirana verzija teksta Bašćanske ploče na latinici te isti tekst napisan prijelaznim oblikom glagoljice (iz oble u uglatu) prema TEI smjernicama za označavanje nestandardnih znakova. Ujedno, pokazani su i načini prikaza tako označenog teksta u pregledniku. U završnom dijelu rada ukratko su opisani svi postupci rada, ciljevi te zapažanja do kojih se došlo prilikom njihove provedbe.

Ključne riječi: TEI, XML, označavanje teksta, glagoljica, uglata glagoljica, Bašćanska ploča

1. UVOD

Inicijativa za označavanje teksta (*Text Encoding Initiative* – TEI) nastala je u sklopu digitalne humanistike kao rješenje za označavanje i prikazivanje tekstova u digitalnom okruženju. Jednostavnost i fleksibilnost su osnovne karakteristike koje ističu ovaj standard, te on ubrzo postaje norma i poveznica istraživača iz humanističkih znanosti diljem svijeta. Funkcionirajući kao tipični opisni označiteljski jezik, TEI standard predstavlja unaprijed definiran skup elemenata i atributa koji se koriste za označavanje teksta.

Osnovna ideja i svrha ovog rada je prikaz označavanja glagoljičkog teksta pomoću TEI standarda. Naglasak je stavljen na hrvatsku uglatu glagoljicu, a to znači, pismo koje se, za razliku od oble glagoljice, još uvijek ne nalazi u *Unicode*-u, te se za potrebe ovoga rada smatra nestandardnim pismom. Ovim radom se u najvećoj mjeri nastoji prikazati mogućnost korištenja nestandardnih znakova u digitalnom okruženju pomoću TEI standarda. Prvi cilj praktičnog dijela rada je označavanje transliteriranog teksta uglate glagoljice; drugi cilj je označavanje istog teksta s privremenim kodnim vrijednostima prema TEI smjernicama za označavanje nestandardnih znakova, dok se treći cilj rada tiče propitivanje mogućnosti prikazivanja označenog teksta u pregledniku. Prethodno, u radu će se teorijski ukratko obraditi razvoj digitalne humanistike, XML-a, TEI-ja te dati kratki uvod u glagoljičko pismo u vrlo općenitom smislu. U drugom poglavlju, bit će govora o TEI standardu, odnosno što on, u osnovi, predstavlja. U potpoglavljima će se prikazati razvoj digitalne humanistike i općenito TEI-ja, TEI tehničke specifikacije i XML-a. U trećem poglavlju opisat će se označavanje nestandardnih znakova putem TEI-ja. U potpoglavljima dan je prikaz *Unicode* standarda te mogućnosti prikazivanja nestandardnih znakova i njihovih svojstava. Nadalje, govori se o postupku anotiranja i načinu označavanja novih znakova, korištenju privremenih kodnih vrijednosti koji se nalaze u rasponu kodnih brojeva nazvanom Privatno korisničko područje (*Private User Area - PUA*) u okviru *Unicode*-a te stilovima pisanja uz navođenje primjera stilova. Četvrto poglavlje sadržava pregled glagoljice i njezinih vrsta, te praktični dio rada. Na kraju, u zaključku će se prikazati primjenjivost i praktičnost TEI standarda po pitanju označavanja glagoljičkih tekstova.

2. OSNOVNO O TEI-u

2.1. Osnovno o digitalnoj humanistici

Digitalna humanistika, poznata i pod nazivom računalna humanistika, znanstvena je grana u okviru humanističkih znanosti, a podrazumijeva istraživanje potpomognuto računalnom tehnologijom.[1] Termin 'digitalna' u nazivu odnosi se na ključnu ulogu digitalnih alata i izvora u području humanističkih znanosti.[2]

Veliki broj istraživača koristi digitalne izvore pri istraživačkom radu. Digitalna tehnologija ima moć transformiranja humanističkih znanosti, omogućavajući jednostavnost i efikasnost, nove načine djelovanja, kao i mogućnost pokretanja novih istraživanja ili odgovaranja na postojeća pitanja.

Digitalna humanistika svoj početak bilježi 1949. godine posjetom jezuitskog svećenika Roberta Buse tvrtki *IBM* s ciljem stvaranja indeksa djela Sv. Tome Akvinskog, kao i drugih srodnih srednjovjekovnih autora. Prvi svezak je objavljen 25 godina kasnije, a poznat je pod nazivom *Index Thomisticus*.[3]

Rad na indeksu su obilježila dva postupka:

- Izrada konkordancija – popisa riječi korištenih u tekstu u njihovom kontekstu.
- Lematiziranje naziva – grupiranje svih mogućih oblika jednog naziva koji se javljaju u tekstu.[4]

Daljnjem razvoju digitalne humanistike doprinio je još jedan svećenik, A. Morton, koji je u suradnji s G. H. C. MacGregorom odlučio napraviti stilističku analizu Novog zavjeta, s naglaskom na provjeri autorstva poslanica Svetog Pavla. Prema provedenoj računalnoj analizi proizlazi da je Sveti Pavle autor svega četiri poslanice od ukupno trinaest.[5] Godine 1966. počinje izlaziti časopis *Computers and the Humanities* J. Rabena, a sama "konsolidacija

---

[1]Usp. Bosančić, Boris. Osnovna terminologija i pojmovi označavanja teksta i povijest digitalne humanistike. Digitalna humanistika i označavanje teksta. Sveučilište J. J. Strossmayer, Filozofski fakultet, Odsjek za informacijske znanosti. Osijek, 2. 11.2012.

[2] Usp. Hockey, Susan. The History of Humanities Computing. // A companion to digital humanities / Susan Schreibman, Raymond George Siemens and John M. Unsworth. Wiley-Blackwell, 2004. URL: http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1&toc.depth=1&toc.id=ss1-2-1&brand=9781405103213_brand (2014-04-20)

[3] Usp. Index Thomisticus. URL: http://www.corpusthomisticum.org/it/index.age (2015-09-03)

[4] Usp. Bosančić, Boris. Nav. dj.

[5] Usp. Hockey, Susan. Nav. dj.

područja" događa se između 1970-ih i 1980-ih godina prošlog stoljeća. Važna godina u razvoju digitalne humanistike je 1973. kada nastaje Udruženje za literarno i jezično računarstvo (*The Association for Literary and Linguistic Computing)* koje organizira nekoliko ALLC simpozija, a 1986. prerasta u časopis pod nazivom *Literary and Linguistic Computing.*

Za razvoj ovog područja važno je spomenuti L. Bournarda koji 1976. uspostavlja Arhiv tekstova Sveučilišta Oxford (*Oxford Text Archives).* OTA je repozitorij namijenjen pohranjivanju tekstova koje su proučavali istraživači iz humanističkih znanosti, a sve u svrhu njihova očuvanja za buduće generacije.[6] Tijekom 1980-ih dolazi do krize u području digitalne humanistike, a uzroci krize su prepoznati u usmjeravanju istraživača na razvoj dijelova računalnih sustava koji su od manje važnosti za obradu humanističkih tekstova (npr. *font,* tipkovnice, printeri i sl.). Pored toga, nije se obraćala pozornost na razlikovanje dva aspekta elektroničkog dokumenta, odnosno na njegov prikaz na ekranu i njegov izvornik koji se sastoji od običnog teksta. Veliki broj znanstvenika sve se manje bavio tekstom, trošeći većinu vremena na formatiranje dokumenata za završnu prezentaciju. U to vrijeme, glasno se govorilo i o nedostatku kvalitetnog označiteljskog standarda.[7]

2.2. TEI standard

Nastanak TEI standarda se smatra jednim od najvažnijih događaja u sklopu razvoja digitalne humanistike. Početak stvaranja standarda vezan je uz konferenciju koja se organizirala na sveučilištu *Vassar* u *Poughkeepsiju* (*New York*) pod sponzorstvom Udruženja za računala i humanistiku (ACH) i financijsku podršku Nacionalne zaklade za humanističke znanosti (NEH). Na konferenciji su sudjelovali arhivisti, akademici i istraživači, te se raspravljalo o mogućnosti stvaranja standardne označiteljske sheme za označavanje humanističkih tekstova u digitalnom okruženju. Uz ACH, u projekt su se također uključili ALLC (*the Association for Literary and Linguistic Computing*) i ACL (*the Association for Computational Linguistics*), što je rezultiralo izradom posebnih TEI smjernica pod nazivom *Poughkeepsie Principles*. Projekt je s vremenom dobio potporu Uprave Komisije Europske zajednice (*Directorate XIII of the Commission of the European Communities - CEC/DG-XIII*), Kanadskog vijeća

[6] Usp. Isto
[7] Coombs, J. H.; Renear, A. H.; DeRose, S. J. Markup systems and the future of scholarly text processing // Communications of the ACM. 30, 11(1987), str. 933-947. URL: http://xml.coverpages.org/coombs.html

socijalne znanosti i humanističkog istraživanja te Zaklade Andrew W. Melloon.[8] Uspostavljeno je nekoliko komisija koje su se okrenule prema rješavanju slijedećih problema:

- Problem dokumentiranja teksta
- Problem reprezentacije teksta
- Problem analize i interpretacije teksta
- Problem metajezika i sintakse.[9]

Prva verzija TEI smjernica (P1) objavljena je 1990. godine pod nazivom *Guidelines for the Encoding and Interchange of Machine-Readable Texts.* TEI smjernice su koristile označiteljski jezik SGML, a do danas je objavljeno četiri verzije smjernica. Od četvrte verzije (P4) počinje se koristiti i XML koji predstavlja, u neku ruku, pojednostavljenu verziju SGML-a. Prije uspostavljanja TEI standarda, stručnjaci na većini projekata koji su uključivali digitalnu reprezentaciju teksta, kao i druge lingvističke aspekte, morali su izraditi i specifičnu shemu označiteljskog jezika. To je značilo da se označiteljska shema može koristiti isključivo za označavanje teksta za koji je dizajnirana. U većini slučajeva nije postojala prethodna analiza obaveznih kategorija i sadržaja, kao ni povezanost između njih za dani tip, odnosno vrstu teksta.[10] Pojavom TEI standarda, postignuto je da zaživi kompleksni proces spomenute analize za veliki broj vrsta tekstova. Postignuća TEI standarda su slijedeća:

- Odluka da *Standard Generalized Markup Language* (SGML) bude okvir za izradu TEI smjernica.
- Specificiranje ograničenja, kao i preporuka za korištenje SGML-a na način koji omogućuje što bolju razmjenu humanističkih tekstova, a time i da uključuje maksimalnu općenitost i fleksibilnost pri izradi aplikacija.
- Analiza i identifikacija kategorija i sadržaja za označavanje različitih vrsta tekstova.
- Efektivne, fleksibilne, opsežne i proširive tekstualne strukture.
- Specificiranje metoda za dokumentiranje elektroničkih tekstova kompatibilnih s knjižničnim kataložnim standardima u samom dokumentu.

[8] Usp. TEI P5: Guidelines for Electronic Text Encoding and Interchange / TEI Consortium eds. URL: http://www.tei-c.org/Guidelines/P5/ (2014-03-01)
[9] Usp. Isto
[10] Usp. Ide, N. Encoding standards for large text resources: the Text encoding Initiative. str. 575 URL: http://clair.eecs.umich.edu/aan/paper.php?paper_id=C94-1094#pdf (2015-03-01)

- Specificiranje označiteljskih TEI smjernica za posebne vrste i aspekte teksta poput:
    - skupine znakova
    - jezičnih cjelina
    - jezikoslovlja
    - rječnika
    - terminologije
    - transkribiranih govora
    - *hipermedije*
    - proze
    - poetskih tekstova
    - dramskih tekstova
    - starih knjiga
    - kritičkog aparata.[11]

Nadalje, 1999. predloženo je osnivanje TEI konzorcija, točnije međunarodne organizacije čija bi glavna zadaća bila razvoj, održavanje i promoviranje TEI standarda. Osim toga, konzorcij je ustanovljen radi osiguravanja stalne skrbi za TEI koji je po prirodi neprofitna udruga, te djeluje na principu načela jednakosti, a također je ekonomski i akademski nezavisan. Poglavlje koje slijedi bavit će se opisom TEI elemenata, atributa i modula, odnosno tehničkom specifikacijom TEI standarda.[12]

2.3. TEI označiteljski jezik

Shema TEI označiteljskog jezika sastoji se od velikog broja modula, a uloga modula je opisivanje TEI elemenata i njihovih atributa. Osnovni moduli su TEI infrastruktura (modul *tei*), TEI zaglavlje (modul *header*), TEI osnovni modul (*core* modul) s elementima koji se mogu koristiti u svakom TEI dokumentu, te TEI strukturni modul (modul *textstructure*) s elementima osnovne strukture teksta.

TEI infrastruktura, odnosno modul *tei* opisuje i identificira osnovne elemente i atribute koji se koriste pri postupku označavanja teksta. U okviru spomenutog modula definirani su neobavezni globalni atributi za svaki TEI element, a to su:

[11] Usp. Isto

[12] Usp. TEI P5. Nav. dj.

- @xml:id – definira jedinstveni identifikator elementa.
- @n – pridruženi broj određenom elementu.
- @xml:lang – jezik vrijednosti elementa.
- @rend – način prikaza elementa u izvornom tekstu.
- @style – stilski prikaz elementa.
- @rendition – opis prikaza ili prezentacije korištene za ovaj element u izvornom tekstu.
- @xml:base – URI adresa.
- @xml:space – definiran način tretiranja praznina u tekstu.[13]

Nadalje, modul TEI zaglavlje opisuje elemente za bibliografski opis građe. Modul se sastoji od četiri osnovna dijela: bibliografskog opisa elektroničke građe, kao i digitalnih faksimila izvorno tiskanih ili rukom pisanih tekstova (<*fileDesc*>), podataka i informacija koji opisuju postupke označavanja teksta (<*encodingDesc*>), ne-bibliografskih podataka i informacija vezanih uz tekst (<*profileDesc*>) i informacija o izmjenama koje su nastale tijekom postupka označavanja teksta (<*revisionDesc*>). *Core* modul čine elementi koji se mogu pojaviti u svakom TEI dokumentu, poput imena, adresa, datuma, brojeva i sl. Na kraju, modul *textstructure* se sastoji od elemenata i atributa čiji je zadatak opisivanje teksta, kao i definiranje strukture samog TEI dokumenta *(*<*text*>,<*front*>, <*body*>, <*back*> i dr.). Uz spomenute module još se koriste i moduli *dictionaries, msdescription, namesdates, textcrit, gaiji, verse, drama, spoken, transcr, figures, corpus, linking, analysis, iso-fs, nets, certainty, tagdocs.* Uz module i atribute, važno je spomenuti i mrežne alate TEI standarda. Prije svega tu je mrežna stranica TEI inicijative, wiki stranica za TEI standard te *TEI by example* online priručnik. Upravo ti alati omogućuju zainteresiranim korisnicima TEI-ja da se pobliže upoznaju s povijesnim razvojem standarda, pregledavaju upute za označavanje teksta i projekte, a ujedno mogu i sami doprinositi razvoju standarda vlastitim sadržajem. Također, TEI ima i časopis pod nazivom *Journal of the Text Encoding Initiative.* Časopis se objavljuje jednom godišnje, a sadrži članke o sastancima, konferencijama i djelovanju članova. Radovi

[13] Usp. Isto

su recenzirani od strane uglednih stručnjaka.[14] Neki od projekata koji rabe TEI standard su *Brown University Women Writers Project, Oxford Text Archive*, *Perseus Project*, *Croatian Language Corpus*, te *Croatiae Auctores Latini*. *Brown University Women Writers Project* okuplja u svoju kolekciju više od 200 tekstova spisateljica i autorica na engleskom jeziku iz perioda od 1450. do 1850. godine.[15] Projekt *Oxford Text Archive* je pokrenut 1976. godine s ciljem prikupljanja, katalogiziranja, očuvanja i distribucije elektroničkih dokumenata raznih istraživanja iz svih područja znanstvenog djelovanja.[16] *Perseus Project* označava digitalnu knjižnicu koja prikuplja tekstove vezane uz svijet antičke Grčke. Većinom su to transkripcije i prijevodi brojnih grčkih i latinskih izvora, kao i renesansnih djela na engleskom jeziku. Također, projekt *Perseus* okuplja umjetničke i arheološke izvore, a sveukupno sadrži oko 25000 slika iz raznih područja djelovanja, te poveću zbirku atlasa, planova i karti.[17] Među projektima se nalaze i dva hrvatska: *Croatian Language Corpus* i *Croatiae Auctores Latini*. *Croatian Language Corpus* je projekt koji okuplja novinske članke i literaturu na hrvatskom jeziku[18], a *Croatiae Auctores Latini* je digitalna zbirka koja nastoji stvoriti pristupačni okvir za prikupljanje i izdavanje hrvatskih neo-latinskih tekstova koji su napisani od 1100. do danas.

2.4. O XML-u

Vraćajući se na povijesni razvoj digitalne humanistike, 1986. godina predstavlja prekretnicu u njenu razvoju, prije svega zbog objave novog standarda za tekstualno označavanje SGML-a. SGML osigurava mehanizam odnosno sintaktička pravila za postupak označavanja teksta te rješava probleme poput označavanja metapodataka, heterogenih vrsta tekstova, strukture dokumenta i sl.[19] Iako fleksibilan i izražajno moćan, SGML je bio izuzetno kompliciran, što je rezultiralo nastojanjima da se razvije nova, jednostavnija inačica SGML-a.

Cilj je ostvaren u vidu XML-a, čija specifikacija obuhvaća daleko manji broj stranica (XML – 25 str., SGML – 155 str.). Osnovna razlika između SGML-a i XML-a mogla se primijetiti u mogućnosti obrade XML dokumenata bez pripadnog DTD (*Document Type Definition*)

---

[14] Usp. Journal of the Text Encoding Initiative. URL: http://journal.tei-c.org/journal/index (2015-03-05)
[15] Usp. Brown University Women Writers Project. URL: http://www.tei-c.org/Activities/Projects/br04.xml (2015-03-07)
[16] Usp. Oxford Text Archive. URL: http://www.tei-c.org/Activities/Projects/ox01.xml (2015-03-04)
[17] Usp. Perseus Project. URL: http://www.tei-c.org/Activities/Projects/pe01.xml (2015-03-03)
[18] Usp. Croatian Language Corpus (Hrvatski jezični korpus). URL: http://www.tei-c.org/Activities/Projects/cr01.xml (2015-03-03)
[19] Usp. Bosančić, B. Uloga opisnih označiteljskih jezika u razvoju digitalne humanistike. // Libellarium: časopis za povijest pisane riječi, knjige i baštinskih ustanova. 4 (2001). Str. 73. URL: http://hrcak.srce.hr/92394?lang=en (2015-03-07)

dokumenta, odnosno dokumenta koji propisuje strukturu, te elemente i atribute od kojih će se dani dokument sastojati.

Prva verzija XML-a, XML 1.0, objavljena je 1998. godine, koja svoje posljednje ažuriranje bilježi 2008. godine. Druga verzija je XML 1.1., a osnovna razlika među njima neovisnost je druge verzije o *Unicode*-u, standardu koji se koristi za predstavljanje znakova i simbola većine svjetskih pisama.

XML omogućava kreatorima sadržaja i kreatorima programa da zasebno rade na svojim zadatcima, bez potrebe za komunikacijom, a dokumente obrađuje aplikacija koja je napravljena za tu namjenu, a to je *XML Parser*. Spomenuti *XML Parser* se oslanja na stroga sintaktička pravila XML dokumenta. Također, u XML dokumentima nisu dozvoljena preklapanja elemenata. Pored toga, vrijednosti svakog atributa moraju biti navedene pod navodnim znacima, a velika slova se razlikuju od malih slova.[20]

[20] Usp. Isto

3. OZNAČAVANJE NESTANDARDNIH ZNAKOVA POMOĆU TEI STANDARDA

3.1. O Unicode-u

Kao što je već spomenuto, *Unicode* je standard koji se koristi za prikaz slova i znakova većine svjetskih pisama u digitalnom okruženju. Naime, za potrebe prikaza znakova u digitalnom obliku, računala im pripisuju određene brojeve koji tvore kodnu vrijednost (engl. *code points*) a prije nego što je uspostavljen *Unicode* postojalo je na stotine takvih sustava i standarda odnosno kodnih stranica. Niti jedan standard za jezično označavanje nije mogao sadržavati sve znakove, tj. ni jedan označiteljski standard nije bio dovoljan da prikaže sva slova i simbole. Također, dva različita standarda su mogla koristiti istu kodnu vrijednost za označavanje različitih znakova, kao i dvije različite kodne vrijednosti za isti znak. To se sve promijenilo pojavom *Unicode*-a koji je osigurao jedinstvenu kodnu vrijednost za svaki znak, neovisno o platformi, programu ili jeziku.[21] Standard je osnovan zbog potrebe za univerzalnošću, odnosno jednoznačnog prikaza znakova, po mogućnosti, svih svjetskih pisama. Nadalje, *Unicode* nastoji osigurati ujednačenost znakova, tj. nepromjenjivu širinu (*fixed-width*) kôdova radi učinkovitijeg pristupa te jedinstvenu interpretaciju znakova.[22]

Usprkos njegovoj dostupnosti, stručnjaci koji se bave označavanjem nerijetko se nađu u situaciji da im nedostaje *Unicode* kodna reprezentacija određenih znakova. Takav slučaj je čest sa znakovima starih pisama koja se više ne koriste, za koje označiteljski standardi ili ne postoje ili se nastoje u digitalnom okruženju prikazati jedino putem fonta odnosno 'sličica' znaka ili glifova (grafičkog oblika znaka).[23] XML dokument u principu može sadržavati bilo koji znak koji je zastupljen u *Unicode*-u. Standard omogućava prikazivanje znakova putem pripadajuće verzije označitelja (najčešće UTF-8), što predstavlja direktni način prikazivanja, ili indirektno putem brojčanog znakovnog upućivanja (engl. *numeric character reference* - NCR). Na primjer, indirektni oblik prikazivanja jednog znaka može izgledati ovako: &#196. Korisnici koji se bave jezičnim označavanjem imaju mogućnost ograničenja obima znakova koji su prikazani u tekstu putem deklaracije. Na primjer, ako dokument započinje sljedećom XML deklaracijom <?xml encoding="iso-8859-1"?>, u dokumentu će svi znakovi koji nisu dio ISO-8859-1 kodne stranice biti prikazani putem NCR-a.[24]

---

[21] Unicode: What is Unicode? URL: http://www.unicode.org/standard/WhatIsUnicode.html (2015-03-10)
[22] Unicode: Summary Narrative. URL: http://www.unicode.org/history/summary.html (2015-03-10)
[23] Usp. TEI P5. Nav. Dj.
[24] Usp. Isto

3.2. Prikazivanje nestandardnih znakova i glifova

Pri izradi XML dokumenta, korisnici i stručnjaci nerijetko se nađu u situaciji kada trebaju određeni znak prikazati u digitalnom okruženju a koji nije zastupljen u *Unicode*-u. U takvoj situaciji, uvijek se treba zapitati radi li se doista o novom nestandardnom znaku, koji zahtjeva svoj vlastiti *Unicode* kodnu vrijednost? Prvo što treba napraviti u toj situaciji je provjeriti mrežnu stranicu *Unicode* standarda.[25] Nakon toga, potrebno je provjeriti novopredložene znakove koji čekaju na prihvaćanje na *Pipeline* stranici.[26] Također, moguće je obratiti se mailing listi *Unicode*-a.[27]

Za nestandardne znakove namijenjen je modul *gaiji* TEI smjernica koji nudi određene mehanizme za njihov prikaz u digitalnom okruženju. Što je važno za ovaj rad, *gaiji* modul osigurava mogućnost prikaza i označavanja i nestandardnih znakova poput uglate glagoljice, koji još uvijek nisu zastupljeni u *Unicode* standardu.[28] Pomoću *gaiji* modula korisnici mogu razlikovati znakove i glifove (grafičke reprezentacije znakova) koje *Unicode* tretira kao identične kako bi se mogao dodati novi nestandardni glif ili znak. Mehanizam označavanja nestandardnog znaka putem TEI standarda zasniva se na sljedećim elementima:

- <*charDecl*> u zaglavlju dokumenta osigurava detalje o nestandardnim znakovima i glifovima;
- <*g*> koristi se za označavanje nestandardnih znakova i glifova u tekstu;
- <*desc*> sadrži sažeti opis znaka;
- <*char*> opisne informacije o znaku;
- <*glyph*> opisne informacije o glifu;
- <*charName*> naziv znaka;
- <*glyphName*> naziv glifa;
- <*gloss*> identificira frazu ili riječ korištenu da se osigura citat ili definicija za tu istu određenu frazu ili riječ;
- <*charProp*> osigurava naziv ili vrijednost određenom svojstvu znaka ili glifa;
- <*mapping*> sadrži jedan ili više znakova vezanih uz određeni znak ili glif;
- <*figure*> grupira elemente koji sadrže ili predstavljaju grafičke informacije;

[25] Unicode: Where is my charecter? URL: http://unicode.org/standard/where/ (2015-03-10)
[26] Proposed New Characters: Pipeline Table URL: http://unicode.org/alloc/Pipeline.html (2015-03-10)
[27] Unicode Email Distribution Lists. URL: http://www.unicode.org/consortium/distlist.html (2015-03-10)
[28] Usp. Bosančić, Boris. Nav. Dj.

- <*note*> sadrži bilješke i citate.[29]

Elementi <*gloss*>, <*desc*>, <*figure*> i <note> su definirani u sklopu drugih TEI modula, a njihova funkcija u sklopu *gaiji* modula je jednaka. S druge strane, element <*figure*> se koristi isključivo kao poveznica na sliku znaka ili glifa. Također, element <*figure*> može sadržavati više <*graphic*> elemenata koji upućuju na slike različitih rezolucija i formata ili duplikat slike znaka.

Element <*mapping*> se koristi za izražavanje bilo kakve povezanosti između znaka ili glifa o kojem se raspravlja i znakova i glifova koji su spomenuti negdje u dokumentu. Može sadržavati bilo koji *Unicode* znak ili <*g*> element povezan s određenim <*char*> ili <*glyph*> elementom kako bi se, primjerice, pokazala povezanost između dva nestandardna znaka. Vrsta povezanosti je definirana atributom *type*, koji može imati neke od slijedećih vrijednosti:

- *exact*
- *standard*
- *uppercase*
- *lowercase*
- *simplified*
- *composed*
- *Unicode*
- *PUA*
- *Modern.*[30]

Na primjeru označavanja malog latinskog slova 'a' u dokumentu, to izgleda ovako:

```
      <charDecl>
      <char xml:id="aenl">
      <charName>LATIN LETTER ENLARGED SMALL A</charName>
<charProp>
      <localName>entity</localName>
      <value>aenl</value>
      </charProp>
       <mapping type="standard">a</mapping>
      </char>
```

[29] Usp. Isto
[30] Usp. Isto

```
        </charDecl>
```

Nadalje, <*mapping*> element može se koristiti za mapiranje znakova ili glifova iz *Privatnog područja korištenja* (PUA), prostora slobodnih kodnih vrijednosti u okviru *Unicode*-a koje su slobodne za korištenje za privatne svrhe u lokalnom okruženju:

```
        <charDecl>
        <glyph xml:id="z103">
        <glyphName>LATIN LETTER Z WITH TWO STROKES</glyphName>
<mapping type="standard">Z</mapping>
        <mapping type="PUA">U+E304</mapping>
        </glyph>
         </charDecl>
```

Nešto preciznija dokumentacija svojstva znakova i glifova se može prikazati putem <*charProp*> elementa, a nešto više o tome izložit će se u narednom poglavlju.[31]

3.3. Svojstva znakova i glifova

Kao što je prethodno spomenuto, element <*charProp*> omogućava korisniku da osigura informacije o znakovima i glifovima. Elementi koji se koriste za zapisivanje svojstava znakova i glifova su:

- <*unicodeName*> sadrži naziv *Unicode* svojstva;
- <*localName*> sadrži naziv svojstva koji je lokalno definiran;
- <*value*> sadrži jedinstvenu vrijednost određenog svojstva, atributa ili analize.[32]

3.4. Anotiranje (označavanje) znakova i glifova

Označavanje je izuzetno važno radi potrebe razlikovanja određenih aspekata znakova i glifova. Primjerice, kod starih knjiga i rukopisa jasno se mogu prepoznati različiti oblici slova *r*, što ostavlja mogućnost razlikovanja znakova u analitičke svrhe. Digitalni faksimil, posebno verzije povezane s transkripcijom i označenim tekstom, će uvijek pružiti kvalitetniji vizualni

[31] Usp. Isto
[32] Usp. Isto

prikaz, ali se on ne može koristiti da podrži argument zasnovan na distribuciji različitih oblika. Tu je rješenje u vidu označavanja i komentiranja znakova. Kako bi se mogli razlikovati drugačiji oblici glifova od standardnog prikazivanja znakova, potrebno je precizno definirati <*glyph*> elemente kao u sljedećem primjeru:

```
<charDecl>
      <glyph xml:id="r1">
            <glyphName>LATIN SMALL LETTER R WITH ONE FUNNY
      STROKE</glyphName>
            <charProp>
                  <localName>entity</localName>
                  <value>r1</value>
            </charProp>
            <figure>
                  <graphic url="r1img.png"/>
            </figure>
      </glyph>
      <glyph xml:id="r2">
            <glyphName>LATIN SMALL LETTER R WITH TWO FUNNY
      STROKES</glyphName>
            <charProp>
                  <localName>entity</localName>
                  <value>r2</value>
            </charProp>
            <figure>
                  <graphic url="r2img.png"/>
            </figure>
      </glyph>
</charDecl>
```

Također, različiti oblici glifova se mogu prikazati pomoću elementa <*g*>:

```
<p>Wo<g ref="#r1">r</g>ds in this
manusc<g ref="#r2">r</g>ipt are
sometimes written in a funny way.</p>
```
[33]

[33] Usp. Isto

Kao što se može vidjeti u primjeru, element <*glyph*> na koji se poziva element <*g*> interpretira se kao anotacija na sadržaj elementa <*g*>. Spomenuti mehanizam se može koristiti za prikaz učestalih rukopisnih skraćenica ili ligatura:

```
<p> ... <g ref="#Filig">Fi</g>lthy riches...</p>
<!-- in the charDecl -->
<glyph xml:id="Filig">
      <glyphName>LATIN UPPER F AND LATIN LOWER I LIGATURE</glyphName>
      <figure>
            <graphic url="Filig.png"/>
      </figure>
</glyph>

<p> ... <abbr>
      <g ref="#per">per</g>
      </abbr> ardua</p>
<!-- in the charDecl -->
<glyph xml:id="per">
      <glyphName>LATIN ABBREVIATION PER</glyphName>
      <figure>
            <graphic url="per.png"/>
      </figure>
</glyph>
```

Ligatura Fi nije zastupljena u *Unicode* standardu, a korisnicima koji se bave jezičnim označavanjem se preporučuje da ju ne koriste, prije svega radi jednostavnije jezične i tekstualne obrade, poput indeksiranja. S ovakvim označiteljskim mogućnostima moguće je pisanje i stvaranje programa za analizu distribucije različitih slova r, kao i stvaranja što vjernijih prikaza originala. Nadalje, moguće je stvaranje normaliziranih verzija na način da se ignorira komentare koji povezuju ili pokazuju na element <g>:

```
<!ENTITY r1 '<g ref="#r1">r</g>' >
<!ENTITY r2 '<g ref="#r2">r</g>' >
```

Time se stvara mogućnost označavanja istog materijala na slijedeći način:

```
<p>Wo&r1;ds in this manusc&r2;ipt are
```

```
sometimes written in a funny way.</p>
```[34]

Ista se tehnika može koristiti za prikaz označiteljskih objekata koji pružaju poveznicu između znaka u dokumentu i komentara na taj znak. To znači da se ne može koristiti na mjestima gdje takva metoda označavanja nije dozvoljena, primjerice kod vrijednosti atributa. Budući da je potreba za korištenjem spomenutih objekata u svrhu anotiranja učestala u dokumentima na kineskom, japanskom i korejskom jeziku, jasno je da postoje određeni problemi. Dva su osnovna problema. Prvi problem su česti slučajevi postojanja više glifova koji opisuju jedan znak, a razlog tomu je način na koji je *Unicode* zapisan. U tom slučaju je moguće zadržati znak u svom osnovnom obliku, te pridodati informacije na način da ih normalizator (za operacije pretraživanja i indeksiranja) može što bolje iskoristiti. To se može postići na slijedeći način:

```
<charDecl>
<glyph xml:id="u8aaa">
<mapping type="Unicode">說</mapping>
<mapping type="standard">説</mapping>
</glyph>
</charDecl>
```

Kao što se može primijetiti, osnovna struktura se sastoji od elemenata *<charDecl>*, *<glyph>* i dva *<mapping>* elementa. Prvi *<mapping>* prostorno definira glif putem mjesta na kojem je zastupljen u *Unicode*-u. Drugi *<mapping>* označava glif na osnovu konteksta, odnosno kako korisnici percipiraju taj glif (varijacija standardnog znaka). Na identifikator u8aaa se može pozvati iz samog teksta:

```
<g ref="#u8aaa">說</g>
```[35]

Drugi problem može nastati u slučaju kada postoji više drugačijih oblika (varijanti) znakova i glifova, a niti jedan nije definiran u *Unicode*-u. Tada je potrebno definirati jedan oblik kao novi znak koristeći element *<char>*, a druge oblike kao glifove koristeći element *<glyph>*:

[34] Usp. Isto
[35] Usp. Isto

```
<charDecl>
      <char xml:id="newchar1">
<!-- more properties here -->
      </char>
      <glyph xml:id="varofnewchar1">
<!-- more properties here -->
            <mapping type="standard">
                  <g ref="#newchar1"/>
            </mapping>
      </glyph>
</charDecl>
```
[36]

3.5. Dodavanje novih znakova

Stvaranje novih znakova je slično postupku anotiranja. Prije svega, koristi se isti element *<g>* za povezivanje znakova u samom tekstu s informacijama i detaljima o znaku unutar *<charDecl>* elementa. Element *<g>* je često prazan, no može sadržavati kodne vrijednosti iz PUA-e. Moguće je prethodno sastaviti jedinstveni oblik znaka koji je u *Unicode* standardu označen kao niz kodnih vrijednosti. Primjer tomu može biti znak iz starog nordijskog jezika (y s točkom i preglasom iznad njega) koji se često javlja u starom nordijskom tekstu, omogućujući korisniku da postupa s njim kao jedinstvenim prethodnom sastavljenim znakom koji ima samo jednu kodnu vrijednost. Za prikaz se preporuča korištenje elementa *<g>*.

```
      <g ref=''#ydotacute'' />
```
[37]

Nadalje, korisnik može izraditi kvalitetnu dokumentaciju za određeni znak ili glif kao u sljedećem primjeru.

```
<char xml:id="ydotacute">
      <charName>LATIN SMALL LETTER Y WITH DOT ABOVE AND
            ACUTE</charName>
      <charProp>
            <localName>entity</localName>
            <value>ydotacute</value>
      </charProp>
```

[36] Usp. Isto
[37] Usp. Isto

```
      <mapping type="composed">&#x0079;&#x0307;&#x0301;</mapping>
      <mapping type="PUA">U+E0A4</mapping>
</char>
```
[38]

3.6. Korištenje kodnih vrijednosti iz PUA-e

Pri objavi nove verzije *Unicode*-a (*Unicode* 5.0), ukupan broj dostupnih kodnih vrijednosti bio je oko 137 000, što je i više nego dovoljno za većinu potreba prosječnog i naprednog korisnika. Unošenje te pridodavanje novih kodnih vrijednosti postojećima dolazi isključivo od strane korisnika, a to znači da standardna tijela ne sudjeluju u tom procesu. Stoga, kodne vrijednosti iz PUA-e se ne preporučaju za direktno korištenje u dokumentima namijenjenim za nasumično izmjenjivanje (engl. *blind interchange*).[39]

3.7. Načini (stilovi) pisanja

Pisma koja se koriste za pisanje se razlikuju ne samo po znakovima i glifovima, već i po načinu, odnosno smjeru pisanja znakova i glifova. Termin 'način pisanja' se odnosi na orijentaciju određenih glifova unutar linije, odnosno niza te na smjer čitanja glifova i linija. Što se tiče većine svjetskih jezika i pisama, pisanje je uređeno na način da se nizovi slova i znakova čitaju od gore prema dolje. Unutar svakog pojedinog niza, znakovi su najčešće prikazani s lijeva na desno (engleski, ruski, grčki), a postoje i pisma koja prikazuju znakove s desna na lijevo (arapsko, hebrejsko). Nadalje, moguće je često se susresti s pismima koja uključuju vertikalno prikazivanje glifova, a čitaju se s desna na lijevo. Primjer tome su stara pisma istočne Azije (japanska Kana, korejski Hangul). U većini slučajeva jedan jezik koristi isti način pisanja, ali postoje i iznimke u kojima se isti jezik pojavljuje u više načina pisanja (*writing modes*), primjerice vertikalno ili horizontalno. Brojna pisma s područja istočne Azije se tradicionalno zapisuju od gore prema dolje unutar niza, smjerom čitanja s desna na lijevo. Iako moderni japanski, kineski i korejski jezik se često zapisuju horizontalno, česta je i upotreba tradicionalnog vertikalnog zapisivanja. Također, postoje i uglavnom rijetki slučajevi starih pisma zapisanih s lijeva na desno, čitanih u smjeru gore prema dolje, a primjer tome su staro *Uighur* pismo, klasično mongolsko i *Manchu* pismo, te *Ogham* pismo gdje smjer pisanja može započeti s donje lijeve strane i kretati se oko rubova ucrtanog predmeta.[40]

[38] Usp. Isto
[39] Usp. Isto
[40] Usp. Isto

Moguće je i kombiniranje različitih jezika koji imaju drugačije stilove pisanja. Kao primjer se može uzeti tekst na hebrejskom ili arapskom (s desna na lijevo) koji sadrži brojeve na latinskom jeziku (s lijeva na desno). U TEI dokumentu, jezik i pismo su definirani u označitelju atributom *xml:lang*, što određuje stil pisanja. Na mjestima gdje atribut nije korišten, dijelovi teksta na različitim pismima će koristiti različite *Unicode* znakove, te će time označavati zadani stil pisanja.[41]

```
The Arabic term قلم رصاص means "pencil".

<s xml:lang="en">The Arabic term

<term xml:lang="ar"> قلم رصاص </term> means "pencil".</s>
```
[42]

3.8. Primjeri stilova pisanja

Vrijednosti koje se mogu koristiti za stilove pisanja su *direction* (*ltr or rtl*); *writingmode* (*horizontal-tb, vertical-rl,* ili *vertical-lr*); i *text-orientation* (*mixed, upright, sideways,...*).[43]

3.8.1.Vertikalni stil pisanja

Korištenje svojstva *writing-mode* je praktično kod jezika koji se mogu prikazati u više stilova pisanja, poput kineskog i japanskog. Vrijednosti kojima se može opisati *su horizontal-tb, vertical-rl i vertical-lr*. Spomenute vrijednosti označavaju smjer kretanja teksta (horizontalno ili vertikalno) te smjer prikazivanja znakova unutar linije, kao i samih linija (od gore prema dolje, s desna na lijevo, s lijeva na desno). Primjer ispod prikazuje tri verzija iste pjesme. Prva verzija je na japanskom, zapisana od gore prema dolje. Druga verzija je *romaji* verzija (japanski jezik na latinici), a treća verzija je prijevod na engleskom.[44]

```
<div>
<lg xml:lang="ja"
style="writing-mode: vertical-rl">
<l>古池や</l>
```

[41] Usp. Isto
[42] Usp. Isto
[43] Usp. Isto
[44] Usp. Isto

```
<l>蛙</l>

<l>飛び込む</l>

<l>水の音</l>

</lg>
<lg xml:lang="ja-Latn"
style="writing-mode: horizontal-tb">
<l>furu ike ya</l>
<l>kawazu tobikomu</l>
<l>mizu no oto</l>
</lg>
<lg xml:lang="en">
<l>Old pond,</l>
<l>and a frog dives in—</l>
<l>"Splash"!</l>
</lg>
</div>
```
[45]

Japanska verzija sadrži svojstvo *writing-mode: vertical-rl*, koje je obavezno jer se japanski može pisati vertikalno i horizontalno. Nadalje, *romaji* verzija koristi atribut *xml:lang* koja omogućava korištenje vrijednost *ja-latn*, prikazujući japanski jezik napisan na latinici. Važno je označavanje atributa *style* sa horizontalnim stilom pisanja, iz razloga što postoje vertikalno zapisani *romaji* tekstovi.[46]

3.8.2. Vertikalni tekst s umetnutim horizontalnim tekstom

Kada se japanski jezik zapisuje vertikalno, orijentacija glifova i znakova unutar linije ostaje ista kao i kada se zapiše horizontalno. To znači za glifovi nisu rotirani, no često je slučaj da jezici koji su zapisani vertikalno imaju umetnute nizove teksta jezika koji se načelno zapisuju horizontalno. Tu se postavlja pitanje orijentacije glifova horizontalnih jezika, odnosno, zapisuju li se oni uspravno, kao što bi se zapisivali u horizontalnom tekstu, ili su oni rotirani? Na primjeru članka na japanskom o indonezijskom jeziku svojstvo *text-orientation* omogućava korisniku da odluči jesu li glifovi rotirani ili ne.[47]

[45] Usp. Isto
[46] Usp. Isto
[47] Usp. Isto

```
<list type="gloss" xml:lang="ja"
style="writing-mode: vertical-rl; text-orientation: mixed">
<label xml:lang="id">hampir</label>
<item>「近い、ほとんど」</item>
<label xml:lang="id">baru</label>
<item>「新しい、ばかい」</item>
<!--... -->
</list>
```

Text-orientation: mixed omogućava pozicioniranje znakova horizontalnih zapisa za 90 stupnjeva u smjeru kazaljke na satu u usporedbi sa standardnom orijentacijom u horizontalnom tekstu.[48]

```
<list type="gloss" xml:lang="ja"
style="writing-mode: vertical-rl; text-orientation: upright">
<label xml:lang="id">hampir</label>

<item>「近い、ほとんど」</item>
<label xml:lang="id">baru</label>
<item>「新しい、ばかい」</item>
<!--... -->
</list>
```

Text-orientation: upright omogućava uspravno pozicioniranje znakova horizontalnih zapisa, odnosno u njihovoj standardnoj horizontalnoj orijentaciji.[49]

3.8.3. Vertikalna orijentacija u horizontalnim pismima

Moguće je pronaći tekstove koji se inače zapisuju u horizontalnom obliku, ali se iz određenih razloga zapisuju vertikalno u sklopu cjeline koja je bazirana na horizontalnom obliku. Primjer tome je slijedeća slika:

[48] Usp. Isto
[49] Usp. Isto

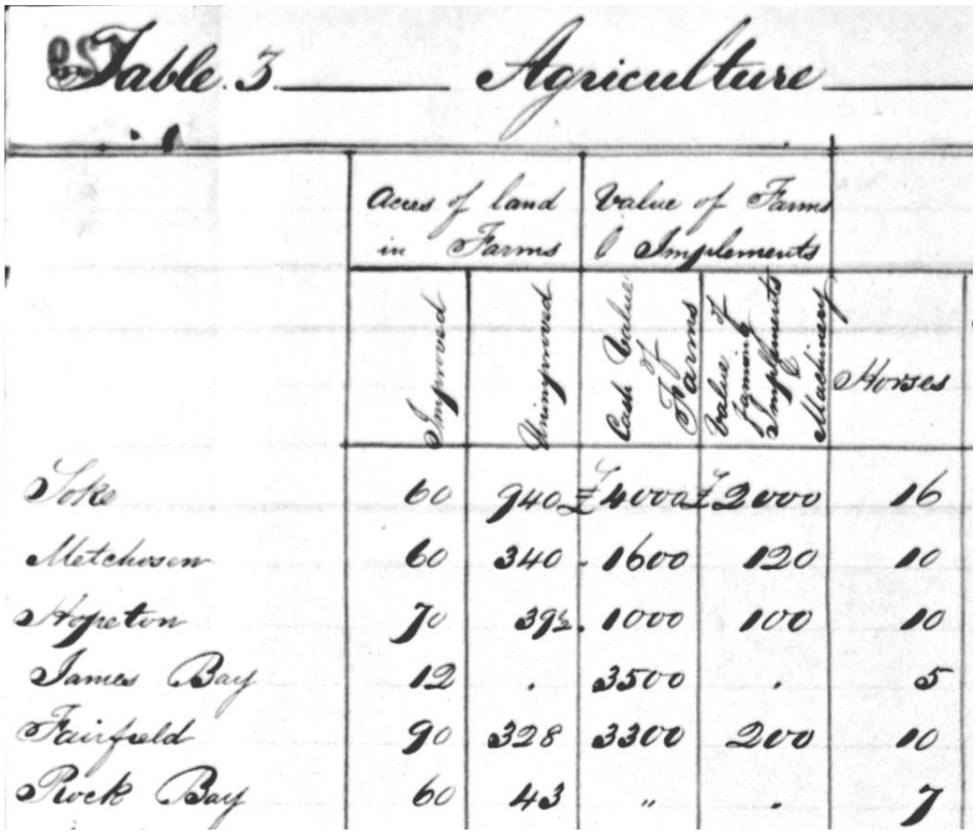


Table 3 — Agriculture

	Acres of land in Farms		Value of Farms & Implements		
	Improved	Unimproved	Cash Value of Farms	Value of Farming Implements & Machinery	Horses
Yale	60	340	£4000	2000	16
Metchosen	60	340	1600	120	10
Hopeton	70	392	1000	100	10
James Bay	12	.	3500	.	5
Fairfield	90	328	3300	200	10
Rock Bay	60	43	"	.	7

Slika 1. Isječak iz tablice s informacijama o razvoju poljoprivrede na području Vancouvera iz 1855. godine.[50]

Za potrebe označavanja koristi se svojstvo *text-orientation* s mogućim oblicima *mixed, upright, sideways-right, sideways-left, side-ways* i *use-glyph-orientation*. Prikaz treće od četiri vertikalne ćelije bi izgledao ovako:

```
<cell style="writing-mode: vertical-lr; text-orientation: sideways-left">
<lb/>Cash Value
<lb/>of
<lb/>Farms
</cell>
```

Kao što je vidljivo, tekst je zapisan vertikalno, a čita se s lijeva na desno. Svojstvo *text-orientation* označava orijentaciju teksta (90 stupnjeva suprotno smjeru kazaljke na satu).[51]

[50] Isto

[51] Usp. Isto

3.8.4. Pisanje od dolje prema gore

Najbolji primjer spomenutog stila je Ogham, abeceda koja se koristila za zapisivanje starog irskog jezika. Ogham se tradicionalno zapisivao uz rubove uspravnog kamena, počevši od njegovog uporišta (temelja). Za označavanje se koristi svojstvo *writing-mode*, a oblik je identičan prethodnom primjeru (*writing-mode: vertical-lr; text-orientation: sideways-left*).[52]

3.8.5. Mješoviti horizontalni smjer

Na primjeru arapskog jezika svojstvo *direction* se može koristiti kako bi se deklarirao obavezni smjer kretanja znakova.

```
The Arabic term رصاص قلم means "pencil".
<s xml:lang="en" style="direction: ltr">The Arabic term
<term xml:lang="ar"
style="direction: rtl; unicode-bidi: embed"> رصاص</ قلم term> means
"pencil".</s>
```

Svojstvo *unicode-bidi* načelno ignorira vrijednosti koje osigurava svojsto *direction*. Mješoviti horizontalni smjer je čest u arapskom i hebrejskom, posebno kada su umetnuti brojevi i fraze.[53]

3.8.6. Primjeri rotacije teksta



Slika 2. Primjer rotiranog teksta

[52] Usp. Isto
[53] Usp. Isto

```
<ab style="transform:rotateZ(-45deg)">TEI-C.ORG</ab>
```



Slika 3. Primjer rotiranog teksta

```
<ab style="transform:rotateY(45deg)">TEI-C.ORG</ab>
```

4. PRIMJER OZNAČAVANJA GLAGOLJIČKOG TEKSTA POMOĆU TEI STANDARDA

4.1. Općenito o glagoljici, vrstama glagoljice i posebno o uglatoj glagoljici

Glagoljica je pismo starih Slavena nastalo u 9. stoljeću. Njezin autor je Sv. Ćiril (Konstantin). Naziv je dobila prema glagolu glagoljati (nesvrš. prez. -ām, pril. sad. -ajūći, gl. im. –ānje), koji ima dva značenja; „pov. služiti misu na staroslavenskom jeziku“ te „(u šali) pričati, govoriti, divaniti (u smislu duže pričati, raspričati se); glagoljiti.“[54]

Glagoljičko pismo u Hrvata ima dugu povijest kao jedno od tri pisma koja su se koristila na ovim prostorima, uz ćirilično (bosančica) i latinično pismo. Hrvati su bili jedini narod u Europi koji su koristili vlastiti jezik i pismo u liturgiji, prema odredbi pape Inocenta IV.[55]

Uspoređujući ju s latinicom, glagoljica ima drugačiji raspored slova. Naime, ona započinje slovima a, b, v, g, ..., dok latinica započinje s a, b, c, d, i td. Kod glagoljice je važno spomenuti da svako slovo ima i svoj naziv. Primjerice, a je az, a b je buky, te od njih proizlazi naziv za grafemski sustav glagoljice koji se zove azbuka. Ona se sastoji od 38 znakova.[56]

Postoje dva oblika glagoljice, a to su obla i uglata. Obla je stariji oblik glagoljice, korištena u svim slavenskim zemljama. S druge strane, uglata je korištena samo u Hrvatskoj. Razlog tome je njezin kasniji vremenski okvir nastajanja (13. st.), u razdoblju kada se većina slavenskih zemalja okrenula ćirilici (12. st.). Glagoljica se dugo zadržala u hrvatskoj, čak do početka 20. st. Uz liturgiju, korištena je i u administrativne svrhe.[57]

Najpoznatija djela na glagoljici su Baščanska ploča, nastala u 11. stoljeću, vinodolski zakonik iz 1288. godine, te Misal po zakonu rimskoga dvora iz 1483. godine.[58]

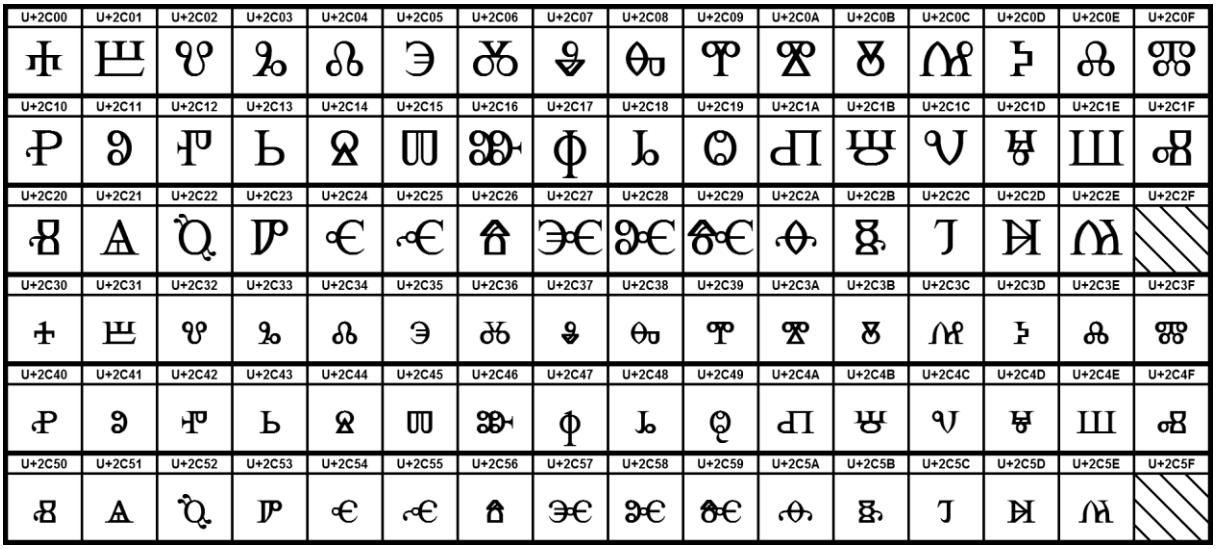
4.2. Svrha, ciljevi i metodologija rada

Svrha praktičnog dijela rada je označavanje nestandardnih znakova pomoću TEI standarda. Naglasak je pritom stavljen na znakove uglate glagoljice, koja je specifična po tome što je, za razliku od oble glagoljice, pismo koje se koristilo isključivo na prostoru Hrvatske, dok se obla

---

[54] Hrvatski jezični portal. URL: http://hjp.novi-liber.hr/index.php?show=search_by_id&id=fFdgXxY%3D (2015-06-02)
[55] Usp. Croatian Glagolitic Script. URL: http://www.croatianhistory.net/etf/et03.html (2015-05-10)
[56] Usp. Čunčić, M. Granice geometrije i simbolike u glagoljskoj paleografiji, URL: http://bib.irb.hr/datoteka/335643.Cuncic_Granice_geometrije_i_simbolike_u_glagoljskoj_paleografiji.pdf (2015-06-08)
[57] Usp. Čunčić, M. Jagićev srednji oblik glagoljice u svjetlu suvremene paleografije. // Slovo, 58 (2008), str. 267 - 283. URL: 267_283.pdf (2015-06-04)
[58] Usp. Moguš, Milan. Povijest hrvatskoga književnoga jezika. Zagreb: Nakladni zavod Globus, 1993.

glagoljica koristila u svim slavenskim zemljama od 9. do 12. stoljeća. Također, obla glagoljica se nalazi u *Unicode*-u, u rasponu kodnih točaka od U+2C00 do U+2C5E.



Slika 4. Prikaz oble glagoljice u sklopu *Unicode*-a

Označavanjem uglate glagoljice nastoji se poticati daljnji rad i istraživanje koje bi moglo ukazati na potrebu standardiziranja i reprezentacije ovog pisma u sklopu *Unicode*-a. Nadalje, iz navedene svrhe rada proizlaze i sljedeći ciljevi:

1. Označavanje transliteriranog glagoljičkog teksta Bašćanske ploče putem TEI standarda.[59]
2. Označavanje glagoljičkog teksta bašćanske ploče čiji su znakovi mapirani s kodnim vrijednostima PUA-e *Unicode* standarda.
3. Razmotriti odgovarajuću metodu prikaza označenog glagoljičkog teksta u pregledniku pomoću XSL-a.

Metodologija koja se koristi u praktičnom dijelu rada zasnovana je na postupku označavanja nestandardnih znakova putem TEI standarda, konkretnije prema načelima koja proizlaze iz petog poglavlja TEI P5 smjernica. Program koji se koristi za provedbu praktičnog dijela rada je Notepad ++.

[59] Ovaj tekst je originalno napisan u prijelaznom obliku glagoljice, na latinici. Osim vidljivih znakova (najčešće slova) posebno će se označavati i znakovi koji nedostaju u originalnom tekstu, a koji se odnose na posebna slova i rečenične dijelove.

4.3. Opis postupka označavanja

U svrhu ove specifične vrste istraživanja odabran je tekst Bašćanske ploče, napisan prijelaznim oblikom glagoljice (iz oble u uglatu), a za potrebe rada bit će prikazan na uglatoj glagoljici. Tekst se, prema stručnjaku Branku Fučiću, sastoji od šest dijelova. Prvi dio (1. red) uključuje invokaciju, odnosno molitveni zaziv. Drugi dio (2. – 6. red) sadrži zapis svećenika, opata Držihe, koji spominje ime hrvatskog kralja Zvonimira u kontekstu darivanja Crkvi sv. Lucije zemljom (ledinom). Pri tome, opat nabraja i svjedoke darivanja. Treći dio (6. – 7. red) spominje kletvu koja će zapasti one koji negiraju darovanje zemlje. Četvrti dio (8. – 9. red) sadrži obavezu molitve redovnika za darovatelja. Peti dio (9. – 12. red) sadrži zapis jednog opata, opata Dobrovita, koji je uz pomoć devetero samostanskih redovnika, tj. svoje samostanske braće, sagradio crkvu u vrijeme kneza Kuzme. Posljednji, šesti dio (12. – 13. red) sadrži zapis koji spominje samostan sv. Nikole u Otočcu, koji je u to vrijeme bio u zajednici s Crkvom sv. Lucije.[60]

Tekst Bašćanske ploče postoji i u transliteriranom obliku[61], s posebno uvedenim oznakama:

- glagoljički "jus" označen je sa Ju,
- glagoljički "jat" označen je sa Ja ili Je,
- glagoljički "iže" označen je sa Ï,
- apostrof se koristi kako bi se mogao označiti glagoljički poluglas "jer",
- nedostajuća slova u riječima stavljena su u zagrade.[62]

Nadalje, u tekstu se javlja znak BÏ koji predstavlja broj 12, kao i slovo G koje prikazuje broj 4. Također, u tekstu se javlja posebni znak za slovo 'a' (az), a to je 'trokutasti a'. Znak se javlja i u drugim glagoljičkim natpisima u Hrvatskoj, a uobičajen je za prostor Bliskog Istoka.[63]

Nakon odabira teksta, slijedeći korak u praktičnom dijelu rada je označavanje teksta unutar TEI XML dokumenta. Kreirana su dva dokumenta, jedan s transliteriranim tekstom ploče na latiničnom pismu, a drugi s tekstom na glagoljici reprezentiran kodnim vrijednostima PUA-e. Prethodno spomenute oznake su posebno označene unutar teksta. Na kraju, okvirno je

[60] Usp. Isto
[61] Usp. Bašćanska ploča: dragi kamen hrvatske pismenosti. URL: http://www.croatianhistory.net/etf/basska.html
[62] Usp. Isto
[63] Usp. Isto

kreiran jednostavan XSL dokument putem kojeg su se razmotrile mogućnosti prikaza glagoljičkih znakova u pregledniku.

Transliterirani tekst Bašćanske ploče dan je u nastavku:

"AZ V' IME OTCA I S(I)NA I SVETAGO DUHA AZ'

OPAT' DR'ŽIHA PISAH SE O LEDINE JuŽE

DA Z'V'NIM(I)R KRAL' HR'VAT'SKI V'

DNI SVOE V' SVETUJu LUCIJu I SVEDO -

MI ŽUPAN' DESIMIRA KR'BAVE MARTIN' V L(I) -

CE PRB'NEBŽA S' POSL' VIN(0)DOLE JaK(O)V' V O-

TOCE DA IŽE TO POREČE KL'NI I BO(G) I *BÏ* AP(OSTO)LA I *G* E -

VANJELISTI I S(VE)TAJa LUCIJa AM(E)N' DA IŽE SDE ŽIVE -

T' MOLI ZA NE BOGA AZ OPAT' DBROVIT' Z' -

DAH' CREK'V' SIJu I SVOEJu BRATIJu S DEV -

ETIJu V' DNI K'NEZA KOS'M'TA OBLAD -

AJuĆAGO V'SU K'RAINU I BJeŠE V' T' DNI M -

IKULA V' OTOČ'CI S' SVETUJu LUCIJu V' EDINO"[64]

[64] Bašćanska ploča: dragi kamen hrvatske pismenosti. URL: http://www.croatianhistory.net/etf/basska.html

4.4. Provedba postupka označavanja glagoljičkog teksta putem TEI standarda

4.4.1. Označavanje transliteriranog glagoljičkog teksta Bašćanske ploče putem TEI standarda

Struktura prvog TEI XML dokumenta u kojoj se nalazi označen transliteriran glagoljički tekst slijedi uobičajenu strukturu koja se koristi u svim TEI dokumentima. Osnovni element kojim započinje dokument je <TEI> element uz koji se navodi atribut @xml:lang, oznaka kojom se definira jezik teksta, u ovom slučaju staroslavenski. Zatim, nakon osnovnog <TEI> elementa slijedi element <teiHeader> koji čini zaglavlje dokumenta, a sastoji se od dva osnovna elementa, <fileDesc> i <encodingDesc>. Element <fileDesc> se sastoji od nekoliko elemenata. Prije svega, tu je <titleStmt> koji se sastoji od elementa <title> kojim se označava naslov označenog teksta, a u ovom radu to je Bašćanska ploča. Zatim, tu se nalazi i element <respStmt>, koji sadrži podatke o odgovornoj osobi i odgovornosti (<name>,<forename>,<surname>,<resp>).

```
<respStmt>
      <name xml:id="dm">
            <forename>Dino</forename>
            <surname>Maganjić</surname>
      </name>
      <resp>enkoder</resp>
</respStmt>
```

Nakon <titleStmt> elementa dolazi element <publicationStmt>, koji sadrži element <date>, odnosno podatak o godini nastanka digitalne inačice teksta.

```
<publicationStmt>
     <date>2015</date>
</publicationStmt>
```

Slijedeći element je <sourceDesc> unutar kojeg se nalazi nekoliko elemenata, a prvi je <biblStruct> unutar kojeg su zapisani podatci o izvornom djelu čiji tekst je označen. Elementi <analytic>, <author>, <title>, <respStmt>, u sklopu kojeg se nalaze elementi

<resp> i <persName>, se koriste za opisivanje izvornika, odnosno zapisivanje podataka o vrsti izvornika, autoru, naslovu i odgovornosti.

```
<sourceDesc>
      <biblStruct>
            <analytic>
                  <author></author>
                  <title>Transliterirani tekst Bašćanske ploče</title>
                  <respStmt>
                        <resp>Transliterirano s originalne ploče</resp>
                        <persName>Branko Fučić</persName>
                  </respStmt>
            </analytic>
      </biblStruct>
 </sourceDesc>
```

S tim podatcima završava element <fileDesc>, a nakon njega dolazi element <encodingDesc> koji sadrži element <charDecl> za analizu znakova unutar teksta. U sklopu <charDecl> se nalazi element <glyph> pomoću kojeg se može pojasniti značenje određenih znakova u tekstu. U tekstu su odabrana tri znaka koja su detaljno analizirana. Prvi znak je 'trokutasti a':

```
<glyph xml:id="a1">
      <glyphName>TRIANGLE GLAGOLITIC LETTER A</glyphName>
      <charProp>
            <localName>entity</localName>
            <value>a1</value>
      </charProp>
           <graphic url="Glagolitic_Letter_Trokutasti_A.svg"/>
      </glyph>
```

Tekst Bašćanske ploče spominje dva broja, dvanaest i četiri. Brojevi su u tekstu prikazani na isti način kao i 'trokutasti a':

```
<glyph xml:id="a2">
      <glyphName>SQUARE GLAGOLITIC NUMBER 12</glyphName>
      <charProp>
            <localName>entity</localName>
            <value>a2</value>
      </charProp>
</glyph>

<glyph xml:id="a3">
      <glyphName>SQUARE GLAGOLITIC NUMBER 4</glyphName>
      <charProp>
            <localName>entity</localName>
            <value>a3</value>
      </charProp>
</glyph>
```

Time je zaključeno zaglavlje dokumenta, nakon kojega slijedi sâm tekst unutar elementa <text>. Osnovni elementi strukture TEI dokumenta su <body>, i <p>. Element <body> označava primarni tekst dokumenta, a u sklopu njega se nalazi element kojim se označavaju logičke cjeline u tekstu. Na kraju, elementom <p> se označavaju manje cjeline unutar teksta, poput odlomka ili rečenice. U sklopu elementa <p> korišten je atribut @n kojim su označene rečenice u tekstu. Budući da tekst Bašćanske ploče nema interpunkcijske znakove, a transliteracija koristi samo znakove za novi redak, atribut @n je korišten samo jednom:

```
<p n="1">
```

Redovi u tekstu su posebno označeni oznakom <lb/> koja tekst prelama u slijedeći red:

```
<lb/>AZ V' <w type="contraction" subtype="noun">IME</w> <w
type="contraction" subtype="noun">OTCA</w> I <w type="contraction"
subtype="noun"><expan>S<ex>I</ex>NA</expan></w> I SVETAGO <w
type="contraction" subtype="noun">DUHA</w> AZ'

<lb/><w type="contraction" subtype="noun">OPAT'</w> <w type="contraction"
subtype="noun"><name type="person">DR'ŽIH<g ref="#a1">A</g></name></w>
PISAH SE O <w type="contraction" subtype="noun">LEDINE</w> <choice>
<orig>jus</orig> <reg>Ju</reg> </choice>ŽE
```

Kao što je prethodno navedeno, u tekstu se na nekoliko lokacija (točnije 14) javlja posebni znak za slovo 'a', 'trokutasti a'. Spomenuti znak je unutar transliteriranog teksta označen na slijedeći način:

```
<g ref="#a1">A</g>
```

Također, brojevi dvanaest i četiri su posebno označeni u tekstu:

```
<g ref="#a2">*BÏ*</g>
```

```
<g ref="#a3">*G*</g>
```

U tekstu su na pojedinim mjestima posebno označene vrste riječi, točnije imenice, a to praktično izgleda ovako:

```
<w type="contraction" subtype="noun">IME</w>
```

Nadalje, u tekstu su označena slova i znakovi koji su ispušteni, a za potrebe njihova označavanja korišteni su elementi <expan>, kojim se označava puni oblik, te <ex> kojim su označeni ispušteni znakovi:

```
<expan>S<ex>I</ex>NA</expan>
```

Također, moguće je i označavanje osobnih imena, kao u sljedećem primjeru:

```
<name type="person">PRB'NEBŽA</name>
```

Nadalje, moguće je i ispravljanje određenih znakova. U slučaju glagoljice, radi se o znakovima 'jat' i 'jus', koji predstavljaju stare oblike te su zamijenjeni novim oblicima. Jat je zamijenjen s 'ja' i 'je', a 'jus' s 'ju'. Za potrebe ispravljanja korišten je element <choice> koji ima zadatak grupiranja alternativnih verzija za isti izvor u tekstu, a elementima <orig> i <reg> se prikazuje originalni zapis i normalizirani zapis:

```
<choice> <orig>jus</orig> <reg>Ju</reg> </choice>
```

Tekst ploče nema razmaka među riječima, što otežava čitanje, a prilikom transliteracije ubačeni su znakovi za prelazak u novi red (<lb/>). Kreiran TEI XML dokument s označenim transliteriranim tekstom Bašćanske ploče na latinici dan je u Prilogu 1.

4.4.2. Označavanje glagoljičkog teksta Bašćanske ploče s nestandardnim znakovima putem TEI standarda

Po pitanju označavanja glagoljičkog teksta Bašćanske ploče čiji znakovi su mapirani s kodnim vrijednostima PUA-e, važno je spomenuti da se razlike u odnosu na označavanje istog, transliteriranog teksta vidljive tek počev elementa <encodingDesc>.

Element <encodingDesc> sadrži identična tri znaka koja su detaljno analizirana.

```
<encodingDesc>
      <charDecl>
            <glyph xml:id="a1">
                  <glyphName>TRIANGLE GLAGOLITIC LETTER A</glyphName>
                  <charProp>
                        <localName>entity</localName>
                        <value>a1</value>
                  </charProp>
                  <mapping type="PUA">&#xE02D;</mapping>
            </glyph>

            <glyph xml:id="a2">
                  <glyphName>SQUARE GLAGOLITIC NUMBER 12</glyphName>
                  <charProp>
                        <localName>entity</localName>
                        <value>a2</value>
                  </charProp>
                        <mapping type="PUA">&#xE001;&#xE009;</mapping>
            </glyph>

            <glyph xml:id="a3">
                  <glyphName>SQUARE GLAGOLITIC NUMBER 4</glyphName>
```

```
                <charProp>
                      <localName>entity</localName>
                      <value>a3</value>
                </charProp>
                <mapping type="PUA">&#xE003;</mapping>
          </glyph>
    </charDecl>
</encodingDesc>
```

Kod označavanja transliteriranog teksta, pri analizi prvog znaka ('trokutasti a') korišten je element <graphic>, koji je sada zamijenjen elementom <mapping> u kojem je naveden atribut @type kako bi se jasno naznačilo da su korištene PUA kodne vrijednosti. Znakovi za brojeve dvanaest i četiri također sadrže element <mapping> s atributom @type:PUA. U sklopu teksta, spomenuti znakovi su označeni na slijedeći način:

```
<g ref="#a1">&#xE02D;</g>
<g ref="#a2">&#xE001;&#xE009;</g>
<g ref="#a3">&#xE003;</g>
```

Kreiran TEI XML dokument (Prilog 2) s PUA oznakama sadrži tekst Bašćanske ploče, te uključuje identične elemente poput prvog dokumenta. Prvi red kôda s PUA kodnim vrijednostima izgleda ovako:

```
<lb/> &#xE000;&#xE008; &#xE002;&#xE020;
<w type="contraction" subtype="noun">&#xE00B;&#xE00F;&#xE005;</w>
<w  type="contraction"  subtype="noun">&#xE011;&#xE015;&#xE01C;&#xE000;</w>
&#xE00B;
<w type="contraction" subtype="noun">
<expan>&#xE014;<ex>&#xE00B;</ex>&#xE010;&#xE000;</expan></w>
&#xE00B; &#xE014;&#xE002;&#xE005;&#xE015;&#xE000;&#xE003;&#xE011;
<w  type="contraction"  subtype="noun">&#xE004;&#xE016;&#xE018;&#xE000;</w>
&#xE000;&#xE008;&#xE020;
```

4.4.3. Prikaz znakova glagoljičkog teksta u pregledniku

Za potrebe prikazivanja TEI XML dokumenata s označenim glagoljičkim tekstom krajnjem korisniku u pregledniku, između ostalih, koristi se i XSL tehnologija. Sadržaj XML dokumenta procesuira se posebnim XSL označiteljskim jezikom koji zahtjeva zasebnu XSL datoteku. Bez dubljeg ulaženja u XSL tehnologiju, a za potrebe ovog rada samo će se spomenuti XSL element <xsl:value-of> s obveznim atributom @select koji je zadužen za prikaz sadržaja konkretnog TEI XML elementa u pregledniku. Pritom, kao vrijednost atributa @select navodi se apsolutna ili relativna putanja do TEI XML elementa u TEI XML datoteci čiji sadržaj se želi prikazati. Opisani mehanizam omogućuje i prikazivanje glagoljičkih znakova u pregledniku, ali samo onih koji su zastupljeni u *Unicode*-u; a to znači, samo znakova oble glagoljice. Na osnovi rečenog, vidi se koliko je važna *Unicode* reprezentacija bilo kojeg znaka, jer samo na taj način isti može biti pohranjen u XML dokumentu i prikazan u pregledniku. Što se tiče uglate glagoljice trenutno su dostupne samo grafičke reprezentacije njenih znakova u obliku odgovarajućeg fonta.

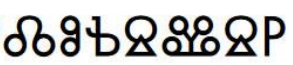
U sljedećem primjeru, u XML kôdu reprezentirano je ime hrvatskog kralja Zvonimira (bez ekstrakcija i poluglasova) putem kodnih vrijednosti oble glagoljice:

```
<name type="person">
&#x2C08;&#x2C02;&#x2C11;&#x2C10;&#x2C0B;&#x2C0F;&#x2C0B;&#x2C13;
</name>
```

Putem sljedećeg XSL kôda moguće je navedeni XML kôd prikazati u pregledniku:

```
<xsl:value-of select="name"/>
```

Prikaz imena hrvatskog kralja Zvonimira u pregledniku izgleda kao na slici 5. Prethodno, potrebno je na računalo instalirati odgovarajući font odnosno grafičku reprezentaciju znakova oble glagoljice.



Slika 5. Prikaz imena hrvatskog kralja Zvonimira na obloj glagoljici

Ovim primjerom želi se ukazati na općenitu metodu koja se može koristiti za prikaz znakova uglate glagoljice u pregledniku, ako ista bude zastupljena u Unicode standardu kao zasebna vrsta pisma, a što danas nije slučaj. Pritom, jasno je da se tekstovi uglate glagoljice

već sada mogu označavati putem TEI-ja na način da se pojedinim znakovima dodjeljuju *Unicode*-ove PUA kodne vrijednosti (što je ujedno bio i drugi cilj rada), ali da njihov prikaz u pregledniku mora sačekati ulazak uglate glagoljice u *Unicode*. Ako se to jednoga dana dogodi, jednostavnom zamjenom postojećih PUA kodnih vrijednosti sa prihvaćenim *Unicode* kodnim vrijednostima, omogućit će putem iste metode i prikaz znakova uglate glagoljice u pregledniku.

5. ZAKLJUČAK

U radu je prikazana metoda označavanja nestandardnih znakova pomoću TEI standarda, a tekst odabran za istraživanje je tekst Bašćanske ploče na uglatoj glagoljici (a prijelazom na oblu). Nestandardni znakovi koji su korišteni su znakovi uglate glagoljice, koja je specifična za Hrvatsku i hrvatski jezik. U praktičnom dijelu rada prikazani su elementi, atributi i mehanizmi potrebni za označavanje glagoljičkih tekstova. Sâm tekst Bašćanske ploče sadrži brojne stavke koje je moguće označiti, poput vrsta riječi, imena, posebnih znakova, ispuštenih znakova i interpunkcijskih znakova.

Što se tiče teorijskog dijela rada, prikazan je razvoj digitalne humanistike i TEI standarda, kao i XML-a i *Unicode*-a. Nadalje, opisani su nestandardni znakovi i metode unošenja novih znakova, kao i metode označavanja znakova i glifova. Također, prikazani su i stilovi, odnosno oblici i načini pisanja znakova i glifova, koji se razlikuju među narodima i kulturama svijeta. Na kraju, ukratko je obrađena glagoljice te je prikazano istraživanje koje se odnosi na označavanje glagoljičkog teksta pomoću TEI standarda.

Svrha rada odnosila se na označavanje znakova uglate glagoljice pomoću TEI standarda, kao i na ukazivanje na potrebu uvrštavanja uglate glagoljice u *Unicode* standard kako bi se propisi, zakoni i drugi povijesni izvori napisani spomenutim pismom mogli pravilno prikazati u mrežnim preglednicima, omogućujući stručnjacima i drugim korisnicima učinkovitiji pristup ovakvim tekstovima. Iako je pitanje uključivanja uglate glagoljice u *Unicode* ostalo izvan razmatranja ovog rada, ipak njime se opisala metoda koja se može koristiti za prikaz znakova uglate glagoljice u pregledniku, ako su isti pohranjeni u XML dokumentu.

Prvi cilj rada bio je označavanje transliteriranog teksta Bašćanske ploče, napisanog na glagoljici putem TEI standarda. U tekstu su označena ispuštena slova, redovi u tekstu, osobe, kao i imenice i ispravljeni znakovi. Drugi cilj rada se također odnosio na označavanje glagoljičkog teksta bašćanske ploče, a ovoga puta znakovi u tekstu su mapirani s kodnim vrijednostima iz Privatnog korisničkog područja (PUA) *Unicode* standarda. Struktura drugog dokumenta je identična prvom, a znakovi na latiničnom pismu su zamijenjeni odabranim kodnim vrijednostima. Na kraju, trećim ciljem rada se pokazalo na koji način ovako označen tekst prikazati u mrežnom pregledniku.

Osnovni problem u istraživanju odnosio se na odabir teksta za označavanje. Budući da je uglata glagoljica specifična za Hrvatsku, ali ne i za ostale zemlje u regiji i šire, jedan problem s odnosio na pronalazak glagoljičkog teksta koji posjeduje i detaljnu transliteriranu

verziju. Tekst je ujedno morao imati i brojne mogućnosti označavanja elemenata, bilo da se radi o ligaturama, vrstama riječi, ispuštenim slovima ili riječima, različitim pismima ili jezicima i sl. Odabrani primjerak teksta za istraživanje, tekst Bašćanske ploče, pokazao se gotovo idealnim za ovakav tip istraživanja jer nudi korisniku ili stručnjaku širok spektar mogućnosti uređivanja i označavanja teksta. Program korišten u istraživanju je Notepad ++, besplatan program dostupan svima, prilično jednostavan za korištenje, no ne nudi korisniku napredne funkcije, niti ne prikazuje pogreške u označavanju.

Usporedbom prethodnih verzija TEI standarda s današnjom (P5) u postupku označavanja teksta, kao i razvoja računalne tehnologije i potreba korisnika tijekom vremena, vidljivo je da se standardi, potrebe korisnika i obrasci korištenja tehnologije uvelike mijenjaju u relativno kratkom vremenskom roku, a stalne nadopune i provjere postojećih standarda postaju nužne.

LITERATURA:

1. Bašćanska ploča: dragi kamen hrvatske pismenosti. URL: http://www.croatianhistory.net/etf/basska.html

2. Bosančić, Boris. Osnovna terminologija i pojmovi označavanja teksta i povijest digitalne humanistike. Digitalna humanistika i označavanje teksta. Sveučilište J. J. Strossmayer, Filozofski fakultet, Odsjek za informacijske znanosti. Osijek, 2. 11.2012.

3. Bosančić, B. Uloga opisnih označiteljskih jezika u razvoju digitalne humanistike. // Libellarium: časopis za povijest pisane riječi, knjige i baštinskih ustanova. 4 (2001). Str. 73. URL: http://hrcak.srce.hr/92394?lang=en (2015-03-01)

4. Brown University Women Writers Project. URL: http://www.tei-c.org/Activities/Projects/br04.xml (2015-03-07)

5. Coombs, J. H.; Renear, A. H.; DeRose, S. J. Markup systems and the future of scholarly text processing // Communications of the ACM. 30, 11(1987), str. 933-947. URL: http://xml.coverpages.org/coombs.html

6. Croatian Glagolitic Script. URL: http://www.croatianhistory.net/etf/et03.html (2015-05-10)

7. Croatian Language Corpus (Hrvatski jezični korpus). URL: http://www.tei-c.org/Activities/Projects/cr01.xml (2015-03-03)

8. Čunčić, M. Granice geometrije i simbolike u glagoljskoj paleografiji, URL: http://bib.irb.hr/datoteka/335643.Cuncic_Granice_geometrije_i_simbolike_u_glagoljskoj_paleografiji.pdf (2015-06-08)

9. Čunčić, M. Jagićev srednji oblik glagoljice u svjetlu suvremene paleografije. // Slovo, 58 (2008), str. 267 - 283. URL: 267_283.pdf (2015-06-04)

10. Hockey, S. The History of Humanities Computing. // A companion to digital humanities / Susan Schreibman, Raymond George Siemens and John M. Unsworth. Wiley-Blackwell, 2004. URL: http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1&toc.depth=1&toc.id=ss1-2-1&brand=9781405103213_brand (2015-03-01)

11. Hrvatski jezični portal. URL: http://hjp.novi-liber.hr/index.php?show=search_by_id&id=fFdgXxY%3D (2015 -06-02)

12. Ide, N. Encoding standards for large text resources: the Text encoding Initiative. URL: http://clair.eecs.umich.edu/aan/paper.php?paper_id=C94-1094#pdf (2015-03-01)

13. Index Thomisticus. URL: http://www.corpusthomisticum.org/it/index.age (2015-09-03)

14. Journal of the Text Encoding Initiative. URL: http://journal.tei-c.org/journal/index (2015-03-05)

15. Moguš, Milan. Povijest hrvatskoga književnoga jezika. Zagreb: Nakladni zavod Globus, 1993.

16. Oxford Text Archive. URL: http://www.tei-c.org/Activities/Projects/ox01.xml (2015-03-04)

17. Perseus Project. URL: http://www.tei-c.org/Activities/Projects/pe01.xml (2015-03-03)

18. Proposed New Characters: Pipeline Table URL: http://unicode.org/alloc/Pipeline.html (2015-03-10)

19. TEI P5: Guidelines for Electronic Text Encoding and Interchange / TEI Consortium eds. URL: http://www.tei-c.org/Guidelines/P5/ (2015-03-01)

20. Unicode Email Distribution Lists. URL: http://www.unicode.org/consortium/distlist.html (2015-03-10)

21. Unicode: Summary Narrative. URL: http://www.unicode.org/history/summary.html (2015-03-10)

22. Unicode: What is Unicode? URL: http://www.unicode.org/standard/WhatIsUnicode.html (2015-03-10)

23. Unicode: Where is my character? URL: http://unicode.org/standard/where/ (2015-03-10)

24. Žubrinić, Darko. Hrvatska glagoljična slova Bašćanske ploče, 2010. URL: http://www.croatianhistory.net/glagoljica/basc.pdf (2015-05-25)

Prilozi

Prilog 1 - Označavanje transliteriranog glagoljičkog teksta Bašćanske ploče putem TEI standarda **(bascanska_ploca.xml)**

```
<?xml version="1.0" ?>
<TEI xml:lang="hr">

      <teiHeader>
            <fileDesc>
            <titleStmt>
                        <title xml:lang="hrv">Bašćanska ploča</title>
                              <respStmt>
                                    <name xml:id="dm">
                                          <forename>Dino</forename>
                                          <surname>Maganjić</surname>
                                    </name>
                                    <resp>enkoder</resp>
                              </respStmt>
            </titleStmt>
            <publicationStmt>
                  <date>2015</date>
            </publicationStmt>
            <sourceDesc>
                <biblStruct>
                    <analytic>
                                    <author></author>
                                    <title>Transliterirani tekst Bašćanske
ploče</title>
                                    <respStmt>
                                          <resp>Transliterirano s originalne
ploče</resp>
                                          <persName>Branko Fučić</persName>
                                    </respStmt>
                              </analytic>
                </biblStruct>
```

```
            </sourceDesc>
         </fileDesc>

            <encodingDesc>
                  <charDecl>
                        <glyph xml:id="a1">
                              <glyphName>TRIANGLE GLAGOLITIC LETTER
A</glyphName>
                     <charProp>
                         <localName>entity</localName>
                         <value>a1</value>
                     </charProp>
                     <graphic url="Glagolitic_Letter_Trokutasti_A.svg"/>
                 </glyph>

                         <glyph xml:id="a2">
                     <glyphName>SQUARE GLAGOLITIC NUMBER 12</glyphName>
                     <charProp>
                         <localName>entity</localName>
                         <value>a2</value>
                     </charProp>
                 </glyph>

                         <glyph xml:id="a3">
                     <glyphName>SQUARE GLAGOLITIC NUMBER 4</glyphName>
                     <charProp>
                         <localName>entity</localName>
                         <value>a3</value>
                     </charProp>
                 </glyph>
                   </charDecl>
            </encodingDesc>
      </teiHeader>
      <text>
            <body>
                  <div>
                        <p n="1">
                        <lb/>AZ V' <w type="contraction"
subtype="noun">IME</w> <w type="contraction" subtype="noun">OTCA</w> I <w
type="contraction" subtype="noun"><expan>S<ex>I</ex>NA</expan></w> I
SVETAGO <w type="contraction" subtype="noun">DUHA</w> AZ'
```

<lb/><w type="contraction" subtype="noun">OPAT'</w>
<w type="contraction" subtype="noun"><name type="person">DR'ŽIH<g
ref="#a1">A</g></name></w> PISAH SE O <w type="contraction"
subtype="noun">LEDINE</w> <choice> <orig>jus</orig> <reg>Ju</reg>
</choice>ŽE

<lb/>DA <w type="contraction" subtype="noun"><name
type="person"><expan>Z'V'NIM<ex>I</ex>R</expan></name></w> <w
type="contraction" subtype="noun">KRAL'</w> HR'V<g ref="#a1">A</g>T'SKI V'

<lb/><w type="contraction" subtype="noun">DNI</w>
SVOE V' SVETU<choice> <orig>jus</orig> <reg>Ju</reg> </choice> <w
type="contraction" subtype="noun"><name type="person"> LUCI<choice>
<orig>jus</orig> <reg>Ju</reg> </choice></name></w> I SVEDO <pc>-</pc>

<lb/>MI <w type="contraction" subtype="noun">ŽUP<g
ref="#a1">A</g>N'</w> <name type="person">DESIMIR<g ref="#a1">A</g></name>
<w type="contraction" subtype="noun">KR'B<g ref="#a1">A</g>VE</w> <name
type="person">MR<g ref="#a1">A</g>TIN'</name> V <expan>L<ex>I</ex></expan>
<pc>-</pc>

<lb/>CE <name type="person">PRB'NEBŽA</name> S'
POSL' <expan>VIN<ex>0</ex>DOLE</expan> <name type="person"><choice>
<orig>jat</orig> <reg>Ja</reg> </choice>
<expan>K<ex>O</ex>V'</expan></name> V O<pc>-</pc>

<lb/>TOCE DA IŽE TO POREČE KL'NI I
<expan>BO<ex>G</ex></expan> I <g ref="#a2">*BÏ*</g> <w type="contraction"
subtype="noun"><expan>AP<ex>OSTO</ex>LA</expan></w> I <g ref="#a3">*G*</g>
E <pc>-</pc>

<lb/>VANJELISTI I <expan>S<ex>VE</ex>TA</expan>
<orig>jat</orig> <reg>Ja</reg> <name type="person">LUCI<choice>
<orig>jat</orig> <reg>Ja</reg> </choice>
</name><expan>AM<ex>E</ex>N'</expan> DA IŽE SDE ŽIVE <pc>-</pc>

<lb/>T' MOLI Z<g ref="#a1">A</g> NE <w
type="contraction" subtype="noun">BOGA</w> AZ OP<g ref="#a1">A</g>T' <name
type="person">DBROVIT'</name> Z' <pc>-</pc>

<lb/>D<g ref="#a1">A</g>H' CREK'V' SI<choice>
<orig>jus</orig> <reg>Ju</reg> </choice> I SVOE<choice> <orig>jus</orig>
<reg>Ju</reg> </choice> BR<g ref="#a1">A</g>TI<choice> <orig>jus</orig>
<reg>Ju</reg> </choice> S DEV <pc>-</pc>

<lb/>ETI<choice> <orig>jus</orig> <reg>Ju</reg>
</choice> V' DNI K'NEZA <name type="person">KOS'M'T<g
ref="#a1">A</g></name> OBLAD <pc>-</pc>

<lb/>A<choice> <orig>jus</orig> <reg>Ju</reg>
</choice>Ć<g ref="#a1">A</g>GO V'SU K'R<g ref="#a1">A</g>INU I B<choice>
<orig>jat</orig> <reg>Je</reg></choice>ŠE V' T' DNI <name type="person">M
<pc>-</pc>

<lb/>IKUL<g ref="#a1">A</g></name> V' OTOČ'CI S'
SVETU<choice> <orig>jus</orig> <reg>Ju</reg> </choice> <name
type="person">LUCI<choice> <orig>jus</orig> <reg>Ju</reg> </choice>
</name>V' EDINO
</p>

</body>
</text>
</TEI>

Prilog 2 - Označavanje glagoljičkog teksta bašćanske ploče čiji su znakovi mapirani s kodnim vrijednostima PUA-e *Unicode* standarda **(bascanska_ploca_PUA.xml)**

```
<?xml version="1.0" ?>
<TEI xml:lang="hrv">

      <teiHeader>
            <fileDesc>
            <titleStmt>
               <title xml:lang="hrv">Bašćanska ploča PUA</title>
                        <respStmt>
                              <name xml:id="dm">
                                    <forename>Dino</forename>
                                    <surname>Maganjić</surname>
                              </name>
                              <resp>enkoder</resp>
                        </respStmt>

            </titleStmt>
            <publicationStmt>
                <date>2015</date>
            </publicationStmt>
            <sourceDesc>
                <biblStruct>
                    <analytic>
                                    <author></author>
                                    <title>Transliterirani tekst Bašćanske
ploče</title>
                                    <respStmt>
                                          <resp>Transliterirano s originalne
ploče</resp>
                                          <persName>Branko Fučić</persName>
                                    </respStmt>
                              </analytic>
                </biblStruct>
            </sourceDesc>
        </fileDesc>

            <encodingDesc>
```

```
                <charDecl>
                      <glyph xml:id="a1">
                            <glyphName>TRIANGLE GLAGOLITIC LETTER
A</glyphName>
                   <charProp>
                       <localName>entity</localName>
                       <value>a1</value>
                   </charProp>
                   <mapping type="PUA">&#xE02D;</mapping>
               </glyph>

                      <glyph xml:id="a2">
                   <glyphName>SQUARE GLAGOLITIC NUMBER 12</glyphName>
                   <charProp>
                       <localName>entity</localName>
                       <value>a2</value>
                   </charProp>
                <mapping type="PUA">&#xE001;&#xE009;</mapping>
               </glyph>

                      <glyph xml:id="a3">
                   <glyphName>SQUARE GLAGOLITIC NUMBER 4</glyphName>
                   <charProp>
                       <localName>entity</localName>
                       <value>a3</value>
                   </charProp>
                            <mapping type="PUA">&#xE003;</mapping>
               </glyph>
                </charDecl>
            </encodingDesc>
      </teiHeader>

      <text>
            <body>
                <div>
                      <p n="1">
                            <lb/> &#xE000;&#xE008; &#xE002;&#xE020; <w
type="contraction" subtype="noun">&#xE00B;&#xE00F;&#xE005;</w> <w
type="contraction" subtype="noun">&#xE011;&#xE015;&#xE01C;&#xE000;</w>
&#xE00B;
```

<w type="contraction"
subtype="noun"><expan><ex></ex></expan></w>
 
<w type="contraction"
subtype="noun"></w>


<lb/><w type="contraction"
subtype="noun"></w> <w
type="contraction" subtype="noun">
<name
type="person"><g
ref="#a1"></g></name></w> 
 
<w type="contraction"
subtype="noun"></w>
<choice> <orig></orig> <reg></reg>
</choice>

<lb/> <w type="contraction"
subtype="noun"><name
type="person"><expan>&#xE00
F;<ex></ex></expan></name></w>
<w type="contraction"
subtype="noun"></w>
<g
ref="#a1"></g>


<lb/><w type="contraction"
subtype="noun"></w>
 

<choice> <orig></orig>
<reg></reg> </choice> <w type="contraction" subtype="noun"><name
type="person"> <choice>
<orig></orig> <reg></reg> </choice></name></w>

 <pc>-</pc>

<lb/> <w type="contraction" subtype="noun"><g ref="#a1"></g></w>
<name type="person">M<g ref="#a1"></g></name> <w type="contraction" subtype="noun"><g ref="#a1"></g>
</w> <name type="person">M<g ref="#a1"></g></name>  <expan><ex></ex> <pc>-</pc>

<lb/></expan> <name type="person"></name> 


<expan><ex></ex></expan> <name type="person"><choice> <orig></orig> <reg></reg> </choice>

<expan><ex></ex></expan></name>  <pc>-</pc>

<lb/>
  O

 

<expan><ex></ex></expan>  <g ref="#a2"></g> <w type="contraction" subtype="noun">

<expan><ex>T</ex></expan></w>  <g ref="#a3"></g>  <pc>-</pc>

<lb/>
 
<expan><ex></ex></expan>
<choice> <orig></orig>
<reg></reg> </choice> <name

type="person"><choice>
<orig></orig> <reg></reg> </choice> </name>

<expan>M<ex></ex></expan>
  
 <pc>-</pc>

<lb/>
M <g ref="#a1"></g>
 <w type="contraction"
subtype="noun"></w>
 <g
ref="#a1"></g> <name
type="person">OI</name>
 <pc>-</pc>

<lb/><g
ref="#a1"></g>

<choice> <orig></orig> <reg></reg>
</choice>
 O<choice>
<orig></orig> <reg></reg> </choice> <g
ref="#a1"></g>
<choice> <orig></orig>
<reg></reg> </choice>   <pc>-</pc>

<lb/><choice>
<orig></orig> <reg></reg> </choice> 
 
<name
type="person">OM<g
ref="#a1"></g></name> O <pc>-</pc>

<lb/><choice> <orig></orig>
<reg></reg> </choice><g ref="#a1"></g>O
 
<g
ref="#a1"></g>  <choice>
<orig></orig> <reg></reg></choice>

```
&#xE002;&#xE020; &#xE015;&#xE01F; &#xE004;NI <name type="person">M <pc>-
</pc>

                              <lb/>I&#xE00D;&#xE016;&#xE00E;<g
ref="#a1">&#xE02D;</g></name> &#xE002;&#xE01F;
O&#xE015;O&#xE01D;&#xE01F;&#xE01C;&#xE00B; &#xE014;&#xE01F;
&#xE014;&#xE002;&#xE005;T&#xE016;
                              <choice> <orig>&#xE022;</orig>
<reg>&#xE022;</reg> </choice> <name
type="person">&#xE00E;&#xE016;&#xE01C;I<choice> <orig>&#xE022;</orig>
<reg>&#xE022;</reg> </choice> </name>&#xE002;&#xE01F;
&#xE005;&#xE004;&#xE00B;NO
                         </p>
                  </div>
            </body>
      </text>
</TEI>
```