Uloga opisnih označiteljskih jezika u razvoju digitalne humanistike

Bosančić, Boris

Source / Izvornik: **Libellarium : časopis za povijest pisane riječi, knjige i baštinskih ustanova, 2012, 4, 65 - 82**

Journal article, Published version
Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:142:178858

Rights / Prava: Attribution 3.0 Unported/Imenovanje 3.0

Download date / Datum preuzimanja: **2024-03-28**

FILOZOFSKI FAKULTET
SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU

Repository / Repozitorij:

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

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

Uloga opisnih označiteljskih jezika u razvoju digitalne humanistike

Boris Bosančić, bbosancic@ffos.hr
Sveučilište J. J. Strossmayera u Osijeku, Filozofski fakultet, Odsjek za informacijske znanosti

Libellarium, IV, 1 (2011): 65 - 82.
UDK: 168.522:004.43+004.9(091)
Pregledni znanstveni rad

Sažetak

Rad razmatra ulogu opisnih označiteljskih jezika u razvoju digitalne humanistike, novije znanstvene discipline u sklopu društvenih i humanističkih znanosti, koju karakterizira proučavanje mogućnosti uporabe računala u znanstvenoistraživačkom radu. Izložen je kronološki prikaz razvoja digitalne humanistike, a potom i opisnih označiteljskih jezika, u nekoliko razvojnih faza. Pokazano je da je razvoj digitalne humanistike od sredine 1980-ih i pojave SGML-a, označiteljskog jezika na kojem počiva TEI, ključni standard za označavanje i razmjenu humanističkih tekstova u digitalnom okruženju, neodvojiv od razvoja označiteljskih jezika. Posebna je pozornost pridana predstavljanju razvoja Inicijative za označavanje teksta (*Text Encoding Initiative - TEI*), ključne organizacije koja je razvila istoimeni standard, kako s organizacijskog tako i s označiteljskog aspekta. Do danas TEI standard objavljen je u pet inačica, a tijekom dvijetisućitih SGML je zamijenjen XML označiteljskim jezikom.

KLJUČNE RIJEČI: označiteljski jezici, digitalna humanistika, označavanje teksta, TEI, SGML, XML.

Uvod

Digitalna humanistika nastaje sredinom 20. stoljeća, ubrzo nakon pojave prvih računala, kao posljedica mogućnosti korištenja i proučavanja humanističkih tekstova u digitalnom okruženju. Pojava opisnih označiteljskih jezika (engl. *descriptive markup languages*) vezana je uz širu uporabu računala za obradu teksta sredinom 1960-ih. Na početku razdvojen, njihov razvojni put konvergira sredinom 1980-ih kada započinje šira primjena opisnih označiteljskih jezika u digitalnoj humanistici koja traje sve do danas. Svrha je ovog rada razmotriti ulogu opisnih označiteljskih jezika u razvoju digitalne humanistike, s posebnim osvrtom na ulogu TEI-a kao vodećeg standarda temeljenog na XML opisnom označiteljskom jeziku koji se rabi za označavanje humanističkih tekstova.

U skladu sa svrhom rada, bez dubljeg razmatranja samog predmeta digitalne humanistike i njenog odnosa sa srodnim disciplinama, može se reći da se, općenito, svaki vid korištenja računala u istraživačkom radu u području humanistike i srodnih područja smatra predmetom proučavanja novonastale discipline. S druge strane, opisni

označiteljski jezici predstavljaju učinkovite alate za obradu teksta od kojih najprije vrijedi izdvojiti SGML (*Standard General Markup Language*), opisni označiteljski jezik iz kojeg će se u 1990-im razviti njegova mnogo poznatija inačica - XML (*eXstensible Markup Languages*).

Za potrebe razumijevanja uloge opisnih označiteljskih jezika u razvoju digitalne humanistike u radu je najprije dan kronološki prikaz razvoja digitalne humanistike od početaka do sredine 1980-ih. Naglašene su razvojne faze discipline kao i uvjeti koji su na koncu i doveli do primjene označiteljskih jezika u području. Iza toga slijedi kronološki prikaz razvoja opisnih označiteljskih jezika koji se u početku odvijao u području grafičko-nakladničke industrije kulminiravši objavom SGML-a, opisnog označiteljskog standarda s ISO certifikatom. Kronološki prikaz razvoja digitalne humanistike nastavlja se prikazom osnutka Inicijative za označavanje teksta (*Text Encoding Initiative - TEI*), organizacije koja će proizvesti istoimeni standard za označavanje i razmjenu humanističkih tekstova u digitalnom okruženju, a koji je u prvo vrijeme počivao na SGML-u. Razvoj digitalne humanistike od sredine 1980-ih neodvojiv je od razvoja označiteljskih jezika, i on se u 1990-im prati kako kroz nastavak razvoja Inicijative za označavanje teksta i istoimenog označiteljskog standarda tako i kroz razvoj XML-a, opisnog označiteljskog jezika koji je uspješno zamijenio SGML. Rad završava prikazom razvoja Inicijative za označavanje teksta u dvijetisućitim te osvrtom na današnju ulogu opisnih označiteljskih jezika u digitalnoj humanistici.

Kronološki prikaz razvoja digitalne humanistike od početaka do druge polovine 1980-ih i pojave TEI-a

Za razliku od nekih srodnih interdisciplinarnih područja (poput informacijske znanosti), oko nastanka digitalne humanistike među stručnjacima nema dvojbi. Nastala je 1949. kada je talijanski isusovac Roberto Busa posjetio tvrtku IBM tražeći podršku za provedbu svog ambicioznog plana: stvoriti indeks djela Tome Akvinskog i drugih srodnih srednjovjekovnih latinista. Indeks će u literaturi kasnije postati poznat pod imenom *Index Thomisticus* i doživjet će buran razvoj. Međutim na samom početku cilj je bio proizvesti tiskano izdanje Indeksa, u čemu se konačno i uspjelo 1974. kada je tiskan prvi svezak.[1] Od tada je *Index Thomisticus* objavljen u trideset jednom tiskanom svesku s oko 36 000 stranica (Fraser 1996). Godine 1992. pojavila se i verzija na CD-ROM-u koja je uključivala i neke hipertekstualne mogućnosti povezivanja naziva.[2] Danas je *Index Thomisticus* dostupan i preko interneta.[3]

1 Busa, R. *Index Thomisticus*. Stuttgart: Frommann - Holzboog, 1974 - 1980.

2 Busa, R. *Thomae Aquinatis Opera Omnia Cum Hypertextibus in CD-ROM*. Milano: Editoria Elettronica Editel, 1992.

3 *Index Thomisticus*, http://www.corpusthomisticum.org/it/index.age

Zanimljivost pionirskog rada oca Buse ogleda se u činjenici da je za provedbu svog plana postavio visoke kriterije usprkos skromnoj podršci tadašnje tehnologije. Sva djela Tome Akvinskog i poznatih latinista prvo su prebačena na bušene kartice, da bi zatim bila obrađena odgovarajućim programom za stvaranje registra odnosno indeksa svih riječi spomenutih u tijelu teksta u njihovom kontekstu - konkordancija (Aarseth 1997). Međutim otac Busa nije se zadovoljio uobičajenim stvaranjem indeksa naziva koji bi bili poredani abecednim redom, već je odlučio lematizirati, tj. grupirati sve moguće oblike jednog leksema koji se javlja u tekstu. Takva mu je odvažnost, kad je posao konačno bio gotov, i donijela priznanje zajednice digitalne humanistike te je 1998. bio prvi koji je primio *Busa* nagradu za postignuća u digitalnoj humanistici koja je po njemu dobila naziv.

Metoda kojom se otac Busa koristio u slučaju *Indexa Thomisticusa* i kojom će se uskoro početi koristiti drugi stručnjaci digitalne humanistike za svoja istraživanja porijeklo vuče iz metoda koje datiraju davno prije računalnog doba. Još 1851. profesor matematike A. Morgan shvatio je da se mjerenjem duljine riječi u određenom tekstu može doći do značajnih otkrića u pogledu dokazivanja autorstva djela te je na taj način pokušao istražiti autorstvo Pavlovih poslanica (usp. Fraser 1996). T. C. Mendenhall je 1901. preko mjerenja frekvencije uporabe najčešće korištenih riječi u djelima W. Shakespearea i njegovih suvremenika (primjerice C. Marlowa i F. Bacona) želio saznati tko stoji iza imena W. Shakespearea izvodeći logičan zaključak da bi isti odnos frekvencije riječi u pojedinim djelima upućivao na to da su djela proizvod istog autora (Mendenhall 1901). Naravno, uz pomoć računala tom se metodom mogu dobiti mnogo precizniji podaci koji se mogu iskoristiti u istim ili sličnim istraživanjima.

To je shvatio i škotski svećenik A. Morton, koji je u suradnji s G. H. C. MacGregorom 1957. uz uporabu računala započeo stilističku analizu Novog zavjeta orijentirajući se ponovno na provjeru autorstva poslanica svetog Pavla. Osim što je uz pomoć računala mjerio frekvenciju uporabe najčešćih riječi, tomu je dodao i mjerenje dužine rečenice.[4] Rezultate istraživanja objavio je 1963. u članku u kojem je, na iznenađenje mnogih, ustvrdio kako je sintaktička analiza, koju je u istraživanju proveo pomoću računala, pokazala kako je sveti Pavao autor svega četiriju poslanica od ukupno trinaest za koje se dotad vjerovalo da ih je doista i napisao (Morton 1965).

Prvi broj časopisa „Computers and the Humanities" koji je utemeljio američki znanstvenik J. Raben objavljen je 1966., iste godine kada su A. Ellis i A. Favat ustvrdili da za analizu humanističkih tekstova računala predstavljaju isto što i teleskop za uvećanje slike svijeta (usp. Fraser 1996). Već u svom prvom broju časopis je uspio odraziti entuzijazam koji je zahvatio istraživače u područjima lingvistike, povijesti umjetnosti, književnih studija, antropologije, muzikologije i sl. od Sjeverne Amerike do Europe, a koji su se u to vrijeme već koristili računalom u svom radu za provođenje nekih od

4 Ove tehnike istraživanja danas pripadaju stilometriji, dijelu lingvistike.

prethodno spomenutih metoda za analizu teksta (usp. McCarty 2003: 1226). Program COCOA, koji se koristio za brojenje riječi i stvaranje konkordancija, pojavio se 1967. i u tom je trenutku predstavljao najnapredniji vid obrade humanističkih tekstova u računalom okruženju. Program COCOA osmislio je D. Russell u Računalnom laboratoriju Atlas (*Atlas Computer Laboratory*) u Chiltonu u Sjedinjenim Američkim Državama (Russell 1965). Riječ je o prvom programu uopće koji se koristio nekim vidom označavanja teksta. Na taj način, putem programa COCOA, mogla se kreirati specifikacija strukture dokumenta, a zbog rezerviranja triju znakova za reprezentaciju jednog slova pisma moguća je bila i uporaba slova neengleske abecede (Russell 1965).

Uvodno razdoblje razvoja digitalne humanistike, tako, karakterizira istraživački rad na konkretnim projektima u kojima se računalo koristi za procesuiranje humanističkih tekstova na različite načine, od stvaranja konkordancija do naprednih metoda stilističke analize koje se temelje na izračunu učestalosti pojavljivanja najzastupljenijih riječi u tekstu, bez upliva teorijskih promišljanja o predmetu i metodama digitalne humanistike kao takve.

U razdoblju između 1970-ih i sredine 1980-ih nastupio je proces koji S. Hockey naziva „konsolidacijom područja" (Hockey 2004: 7). U Velikoj Britaniji, u Cambridgeu se 1970. održava prvi u nizu simpozija literarnog i lingvističkog računalstva (*Literary and Linguistic Computing*) koji obilježava znanstvenu aktivnost istraživača područja u 1970-ima. Simpozij se održavao redom u Edinburghu (1972.), Cardiffu (1974.), Oxfordu (1976.), Birminghamu (1978). i, konačno, opet u Cambridgeu (1980.), svaki put rezultirajući kvalitetnim radovima iz navedenog područja. Godine 1973. osnovano je Udruženje za literarno i jezično računarstvo (*ALLC - The Association for Literary and Linguistic Computing*[5]) koje preuzima organizaciju simpozija, a koji stoga mijenja naziv u Simpozij ALLC. ALLC započinje i s izdavanjem vlastitog biltena triput na godinu, a taj će bilten 1986. prerasti u časopis istog naziva („Literary and Linguistic Computing"[6]). Godine 1976. S. Hockey, jedna od najutjecajnijih među istraživačima digitalne humanistike uopće, održava na Sveučilištu Oxford niz predavanja, uključujući i dva najpoznatija - *Concordances, word indexes and dictionaries* i *Stylistic analysis and authorship studies* - u kojima se osvrće na pionirske radove oca Buse, A. Mortona i drugih istraživača ranije spomenutih u ovom radu (Fraser 1996). Među prvima u humanističkoj zajednici u Velikoj Britaniji nastoji potaknuti zanimanje za programske jezike održavajući kratke tečajeve programskog jezika SNOBOL (*String Oriented Symbolic Language*) i nazivajući ga „očitim izborom programskog jezika za istraživače iz humanističkih znanosti" (Fraser 1996).

U godini u kojoj je S. Hockey održala svoja poznata predavanja, 1976., stručnjak za baze podataka s diplomom iz engleske književnosti L. Burnard uspostavlja Arhiv tek-

5 *The Association for Literary and Linguistic Computing*, http://www.allc.org/.

6 *Oxford journals: Literary and linguistic computing*, http://llc.oxfordjournals.org/.

stova Sveučilišta Oxford (*Oxford Text Archives - OTA*)[7], projekt koji će u godinama koje slijede ponijeti epitet jednog od najvećih projekata zajednice stručnjaka koji se bave označavanjem teksta uopće. OTA je prvotno zamišljena kao obični repozitorij u koji će se pohranjivati tekstovi koje su pojedini istraživači iz humanističkih znanosti proučavali kako ne bi bili izgubljeni (Hockey 2004: 8). OTA 1995. postaje sastavni dio upravo osnovanog *UK Arts and Humanities Data Servicea* s pridodanim zadaćama kreiranja upotrebljivijih metapodataka te pripremanja učinkovitije diseminacije tekstova (Hockey 2000: 14).

Od ostalih projekata vrijedi izdvojiti Tezaurus starogrčkog jezika (*Thesaurus Linguae Graecae* - TLG), projekt koji je inicijalno financirala Marianne McDonald i koji je pokrenut 1971. na Sveučilištu Kalifornija u Irvinu, a koji je mnogo godina održavao T. Brunner. T. Brunner je TLG zamislio kao veliku bazu podataka starogrčkih tekstova od Homera do 600. godine pr. Krista (Brunner 1993). Softversku i hardversku podršku na projektu osiguravao je David W. Packard. Do 1997. u okviru tezaurusa prikupljeno je oko 70 milijuna riječi, a projekt je specifičan i po tome što se koristi vlastitim sustavom označavanja teksta („beta code") (Hockey 2000: 14).

U okviru Simpozija ALLC 1978. godine osniva se i Udruženje za računala i humanistiku (*ACH* Đ *Association for Computers and the Humanties*[8]) koje također započinje s održavanjem svojih skupova jednom na godinu. Godine 1989., nakon šesnaest održanih ALLC-ovih i devet ACH-ovih skupova, održana je i prva zajednička konferencija ALLC-a i ACH-a koja nastavlja s održavanjem jednom godišnje i koja je tijekom 2000-ih promijenila naziv u *Digital Humanities*.

Karakteristika ovog razdoblja razvoja digitalne humanistike može se sažeti u tvrdnji da je naglasak stavljen više na diseminaciju radova iz područja nego na provedbi praktičnih projekata kao u 1960-ima, što je dovelo do održavanja skupova i pokretanja izdavačke djelatnosti.

Razvoj opisnih označiteljskih jezika do druge polovine 1980-ih i pojave SGML-a

Šira uporaba računala za računalnu obradu teksta započela je sredinom 1960-ih. Postupak formalnog oblikovanja teksta za tiskanje ili neku drugu namjenu ostao je manje-više isti sve do danas. Pritom, formalno se oblikovanje teksta u digitalnom obliku tada sastojalo od tri različite vrste znakova:

- znakovi koji predstavljaju sam tekstualni sadržaj
- znakovi koji predstavljaju naredbe za formatiranje tog tekstualnog sadržaja

7 *The Oxford Text Archive*, http://ota.ahds.ac.uk/.

8 *Association for computers and the humanities*, http://www.ach.org/.

- graničnici (engl. *delimiters*), posebna vrsta znakova koja služi za razdvajanje prvih dviju vrsta spomenutih znakova (usp. Renear 2004: 220).

Takva bi datoteka potom bila računalno obrađena prema ugrađenim naredbama formatiranja te bi kreirala odgovarajući prikaz na ekranu. Prvi računalni program za formatiranje teksta koji je koristio opisani označiteljski princip bio je RUNOFF koji je 1964. razvio J. H. Saltzer[9] (usp. Saltzer 1964). Opisani sustav koji je na ovaj način funkcionirao poznat je pod nazivom sustav označavanja dokumenata (engl. *document markup system*).

Sam termin *markup* prvo se započeo koristiti u izdavaštvu i vuče svoje porijeklo od engleskog izraza *mark up* koji se na hrvatski jezik također može prevesti kao 'označavanje'. Naime, urednici rukopisa za objavljivanje su 'označavali' (engl. *marks up*) rukopise na način dodavanja anotacija ili simbola uz sam tekst (npr. „centrirati tekst" ili „naslov") kako bi naznačili njegov izgled u tiskanom obliku (usp. Spring 1989: 111).

Već krajem 1960-ih i početkom 1970-ih glomazne naredbe formatiranja koje su bile ugrađene u prikaz teksta u digitalnom obliku polako se počinju zamjenjivati njihovim skraćenicama koje dobivaju naziv makro naredbe (engl. *macros*) prema terminologiji posuđenoj iz područja strojnih (asemblerskih) jezika, a u kojima se tim nazivom označavala viša razina naredbi, za razliku od onih na nižoj, primitivnijoj razini. Već u tom ranom stadiju razvoja označiteljskih jezika nije bilo potpuno jasno mogu li se makro naredbe shvatiti samo kao skraćenice za složenije naredbe formatiranja ili čak kao posebne oznake za identificiranje tekstualnih komponenti (npr. naslova, velikih slova) (Renear 2004: 220). U tom se smislu govori o začetku razvoja deskriptivnih odnosno opisnih označiteljskih jezika koji se pojavljuju u trenutku kada se makro naredbe počinju koristiti kao posebne oznake komponenata sadržaja, a ne kao skraćenice naredbi formatiranja.[10]

Ovaj novi pristup u označavanju teksta donio je niz prednosti. Kao jednu od ključnih možemo navesti onu koja na globalnoj razini omogućuje mijenjanje naredbi za formatiranje pojedinih komponenti teksta. Pa tako, npr., ukoliko želimo da 'naslov druge razine' (*Heading 2*) u tekstu bude podcrtan dovoljno je samo na jednom mjestu izmijeniti naredbu formatiranja (dodati naredbu podcrtavanja), a promjena će se odraziti na sve 'naslove druge razine' zastupljene u samom tekstu. Konačnu promociju opisnih označiteljskih jezika u računalnoj obradi teksta kao i u području digitalnog izdavaštva prema A. Renearu obilježila su tri događaja:

9 Zapravo, riječ je o dva programa: TYPSET-u koji se koristio kao editor dokumenta i RUNOFF-u koji je uređivao prikaz na ekranu.

10 Drugu vrstu označiteljskih jezika predstavljaju tzv. proceduralni označiteljski jezici koji se nastavljaju usporedo razvijati, zasnivajući se na razumijevanju oznaka u tekstu kao naredbi formatiranja. Kao primjer široko korištenog proceduralnog označiteljskog jezika može se navesti LaTeX (vidi http://www.latex-project.org/).

- prezentacija W. Tunnicliffea, tadašnjeg predsjednika *Graphic Communications Association's Composition Committee*, u rujnu 1967. u kojoj on o postupku označavanja teksta govori kao o „*generic codingu*"
- razvoj univerzalnog kataloga elemenata „uredničke strukture" teksta u sklopu projekta S. Ricea koji će pojednostaviti postupak pripreme i proizvodnje knjiga te
- u kasnim 1960-im rad C. Goldfarba zajedno sa suradnicima E. Mosherom i R. Lorie na makro-označiteljskom jeziku GML (*Generalized Markup Language*).

Povrh tih događaja, N. Scharpf, ravnatelj Udruge za grafičku komunikaciju (*Graphics Communication Association*), inicirao je pokretanje projekta *GenCode* koji je imao za cilj razvoj standardne inačice označiteljskog jezika za potrebe izdavaštva.

S druge strane, C. Goldfarb nastavlja razvoj prve inačice GML-a u tvrtci IBM kojeg prilagođava komercijalnoj uporabi. Iako je stvoren 1969. zbog straha tvrtke IBM od krađe poslovne ideje, GML je prvi put objavljen tek 1973. (usp. Goldfarb 1996). GML je dobio takav naziv jer elemente korištene za označavanje jednog dokumenta ne svodi na jednu aplikaciju ili stil formatiranja već se oni mogu odnositi generalno (*general*) na sve dokumente. GML se temeljio na dva postulata:

- označavanje se koristi za opis strukture dokumenta
- označavanje podliježe strogim sintaktičkim pravilima zbog pretpostavljene naknadne obrade dokumenata u drugim programima (usp. Goldfarb 1981: 68).

Krajem 1970-ih aktivnosti oko projekta *GenCode* i GML-a konvergiraju (Goldfarb 1996) i napokon se ujedinjuju, a u zajedničkim naporima pridružuje im se i ANSI (*American National Standards Institute*), na čiju inicijativu C. Goldfarb preuzima mjesto voditelja cijelog projekta. Konačno, u zajednički projekt *GenCode*-a i GML-a uključuje se i ISO (*International Organization for Standardization*), koji 1986. objavljuje prvu službenu inačicu SGML-a pod brojem ISO 8879 i s punim naslovom *Information Processing - Text and Office Systems - Standard Generalized Markup Language (SGML)*.[11]

SGML na kraju nije poprimio oblik tipičnog označiteljskog jezika s propisanim skupom elemenata koji se koristi za potrebe označavanja komponenti teksta, već postaje *metajezik* - jezik za definiranje opisnih označiteljskih jezika. Drugim riječima, SGML je ponudio samo stroga sintaktička pravila za kreiranje vlastitog označiteljskog jezika. Zato je Američko udruženje izdavača, također usko povezano s razvojem SGML-a, 1988. kreiralo prvu SGML aplikaciju, što je prvi konkretan opisni označiteljski jezik temeljen na SGML-u koji se sastojao od propisanog skupa oznaka kojima bi se trebali služiti članovi Udruženja za potrebe pripreme dokumenata za objavljivanje. U to je doba bilo

11 *International organization for standardization: information processing - text and office systems - Standard Generalized Markup Language (SGML)*, http://www.iso.org/iso/catalogue_detail.htm?csnumber=16387.

za očekivati da će se SGML-om početi koristiti i mnoge druge zajednice za potrebe računalne obrade teksta prilagođujući ga svojim potrebama kroz definiranje vlastitog skupa oznaka koje podliježu njegovim strogim sintaktičkim pravilima. Međutim, iako se činilo da se pred SGML-om u to doba pružala lijepa budućnost, njegov ga je razvoj, kako će biti prikazano u sljedećim poglavljima, na kraju odveo u sasvim drugom smjeru.

Uloga opisnih označiteljskih jezika u kriznom periodu razvoja digitalne humanistike u 1980-im i 1990-im

Sredinom 1980-ih, u široj zajednici stručnjaka digitalne humanistike mogla se osjetiti potreba za standardnim označiteljskim rješenjem koje bi unaprijedilo postupke arhiviranja i uređivanja tekstova i ponovno usmjerilo aktivnost zajednice u praktičnom smjeru. Mnogi stručnjaci slažu se da je vrijeme sredine 1980-ih u području digitalne humanistike obilježila svojevrsna kriza razvoja koja se ogledala u udaljavanju od pravih problema u području i nedostatku koordinacije u provođenju projekata.

Štoviše, J. H. Coombs, A. H. Renear i S. J. DeRose započinju svoj vrlo utjecajan rad iz 1987. tvrdnjom kako je računalna obrada humanističkih tekstova sredinom 1980-ih sve više nazadovala. Razlog takvom stanju vide u tri problema koji su se u području digitalne humanistike u međuvremenu pojavili:

- pažnja se sve više poklanja razvoju onih dijelova računalnih sustava koji su manje bitni za daljnji razvoj obrade humanističkih tekstova poput fontova, boja, printera, tipkovnice i sl.
- izgubljena je svijest o postojanju dviju strana jednog te istog digitalnog dokumenta: jedna je njegov prikaz na ekranu, a druga izvornik, datoteka u kojoj je pohranjen na računalu
- znanstvenici sve više svog vremena i energije troše na formatiranje dokumenta za završnu prezentaciju, a manje na bavljenje samim tekstom (Coombs, Renear i DeRose 1987: 933).

Iako nema jednostavnih rješenja za navedene probleme, autori se slažu da bi primjena opisnih označiteljskih jezika, koji su u to doba zaživjeli u praksi, umnogome pomogla njihovu rješavanju (Coombs, Renear i DeRose 1987: 946).

Možda je najveća zasluga ovog rada sadržana u eksplicitno izraženoj potrebi za dijeljenjem humanističkih tekstova u digitalnom okruženju jer se otvoreno govori o „nedostatku označiteljskog standarda" koji bi jamčio uspješnost te razmjene. Doista, u to su doba na uspostavljenoj označiteljskoj sceni već postojale neke označiteljske sheme (*markup scheme*) odnosno jezici koji su se koristili u svrhu označavanja humanističkih tekstova u digitalnom obliku, a koje smo spomenuli prije u ovom radu (COCOA, beta code projekta TLG i sl.). Međutim često se situacija tog doba, vezana uz označavanje teksta, opisuje kao kaotična (Burnard 1988: 131). Iako je prevladavalo shvaćanje da

je postupak označavanja teksta od esencijalne važnosti, ipak je nedostajala opća označiteljska shema koja će biti prihvaćena od strane većine stručnjaka u zajednici. Mnoge označiteljske sheme bile su razvijene za potrebe samo jednog projekta, s nedostatnom dokumentacijom, a pravi problem nastajao je kad je trebalo konvertirati tekstove označene pomoću jedne označiteljske sheme u tekstove označene drugom shemom. Drugim riječima, iskazana potreba za standardizacijom postupka označavanja teksta morala je uskoro rezultirati i svojim razrješenjem.

Konačno, 1986. nastupa prekretnica u razvoju digitalne humanistike. Te je godine, kako je ranije spomenuto, na scenu stupio novi standard za označavanje teksta - SGML (*Standard General Markup Language*) - nudeći mehanizam definiranja vlastitog skupa elemenata koji bi se koristio za označavanje teksta, a kojim bi se mogli riješiti tadašnji problemi u postupku označavanja teksta poput upravljanja heterogenim vrstama tekstova i pridruženim metapodacima, kao i potreba za strožim pristupom u definiranju strukture dokumenata.

Osim toga, SGML je predstavljao označiteljski jezik iz kojeg će se kasnije razviti i njegova mnogo poznatija inačica - XML (*eXstensible Markup Languages*).

Osnutak TEI-a

Iako je ranije spomenuti rad J. H. Coombsa, A. H. Reneara i S. J. DeRosea, koji se pojavio u kratkom međurazdoblju između pojave SGML-a i TEI-a, na nekoliko mjesta već spominjao SGML, autori su bili svjesni da SGML još uvijek ne predstavlja traženo standardizirano rješenje za razmjenu humanističkih tekstova, već samo alat koji bi to rješenje trebao kreirati (Coombs, Renear i DeRose 1987: 941). Nedugo potom, kad je objavljen spomenuti rad, na Sveučilištu Vassar osnovana je Inicijativa za označavanje teksta (TEI) sa svrhom razvoja standarda za označavanje teksta baziranog na SGML-u. U tom smislu može se reći da je osnutak TEI-a bio potaknut reakcijom istraživača iz područja digitalne humanistike na objavljivanje SGML-a. SGML je bio taj traženi alat koji je zajednica u godinama koje su prethodile toliko željno očekivala kako bi omogućila standardizaciju označavanja teksta u svom području. U osnivanju TEI-a, jednom od najznačajnijih događaja u povijesti digitalne humanistike, organizacijama već spomenutim u ovom radu - Udruženju za računala i humanistiku te Udruženju za literarno i jezično računalstvo - pridružilo se i Udruženje za računalnu lingvistiku (*Association for Computational Linguistics - ACL*[12]). Cilj TEI-a bio je dvojak: osigurati uspješnu razmjenu humanističkih tekstova u istraživanjima i sugerirati principe označavanja teksta temeljem korištenja zasebne TEI SGML aplikacije. Prema W. McCartyu, TEI predstavlja najvažniji projekt u sklopu kojeg se ostvaruje suradnja istraživača u području digitalne humanistike (McCarty 2003: 1225). S takvom ocjenom slaže se većina stručnjaka i istraživača (Hockey 2004: 12). Projekt se mogao pohvaliti izdašnom

12 *Association for Computational Linguistics*, http://www.aclweb.org/ .

financijskom podrškom od milijun dolara od strane institucija u Sjevernoj Americi i Komisije Europske zajednice (*Commission of the European Communities*), već u svom početnom stadiju (Hockey 1994: 686). Tijekom 1990-ih TEI je uspio realizirati tri inačice svojih TEI smjernica (*TEI Guidelines*) za označavanje teksta koje su opisivale TEI SGML aplikaciju sastavljenu od propisanog skupa elemenata i atributa za označavanje teksta. U lipnju 1990., tri godine nakon osnutka TEI zajednice, realizirana je prva inačica (P1) TEI smjernica za označavanje teksta sa specifikacijom temeljenom upravo na SGML-u. U razdoblju između 1990. i 1993. petnaest različitih radnih skupina radilo je na reviziji TEI smjernica (P1-inačicu standarda 1992. su za kratko vrijeme zamijenili novom inačicom P2) da bi u svibnju 1994. konačno bila objavljena prva službena verzija TEI smjernica za označavanje teksta u P3-inačici.[13] Iza P3-inačice TEI smjernica kao glavni urednici stoje dva ugledna stručnjaka zajednice TEI i dobra poznavatelja SGML-a i označiteljske teorije - L. Burnard i M. Sperberg-McQueen. Potonji će u drugoj polovini 1990-ih dobiti i prihvatiti poziv W3 konzorcija (*World Wide Web Consortium*) i T. Berners-Leeja za sudjelovanje u razvoju novog označiteljskog standarda u mrežnom okruženju - XML-a.

Projekti označavanja teksta pomoću TEI-a

Jedan od prvih projekata koji se koristio SGML-om ne samo za označavanje teksta već i za strukturiranje dodanih informacija o samom tekstu jest projekt *Orlando* Sveučilišta Alberta i Guelph. Projekt *Orlando* koristi se SGML-om kako bi kreirao „integriranu digitalnu povijest tekstova britanskih spisateljica na engleskom jeziku" (Brown i ostali 1998: 271). Primarnim tekstovima pridružene su i druge vrste tekstova poput biografija spisateljica, književnih i povijesnih interpretacija i sadržaja od velikog povijesnog značenja, koji su također bili označeni putem sofisticiranog mehanizma SGML-a.

Godine 1987. na Sveučilištu Harvard započinje i projekt *Perseus Digital Library*, koji je u početku imao za cilj okupiti „kritičnu masu heterogenih sadržaja vezanih uz starogrčku civilizaciju" (Crane 1998). S vremenom je projekt počeo okupljati i druge vrste sadržaja da bi danas uspostavio zbirke renesansne građe, američke književnosti 19. stoljeća, zbirke dokumenata na arapskom jeziku i dr. *Perseus Digital Library* pokrenut je s idejom zamisli funkcioniranja knjižnica u digitalnom okruženju. Projekt je razvio zasebno programsko rješenje koje danas uključuje mogućnost označavanja teksta pomoću TEI-a. Danas projekt održava G. Crane sa Sveučilišta Tufts.[14]

Na Sveučilištu Brown 1988. započinje Projekt spisateljica (*The Women Writers Project - WWP*)[15] kojemu je cilj uspostava baze podataka tekstova na engleskom jeziku koje

13 *TEI: Text Encoding Initiative*, http://www.tei-c.org/index.xml.

14 *Perseus Digital Library*, http://www.perseus.tufts.edu/hopper/.

15 *Women Writers Project*, http://www.wwp.brown.edu/.

su u razdoblju između 1330. i 1830. godine napisale žene. Projekt je na neki način i proslavio uporabu označavanja teksta putem TEI-a.

Mnogi drugi srodni projekti koji se koriste nekim vidom označavanja teksta započinju u istom razdoblju od realizacije SGML-a i pojave TEI standarda: *The Prosopography of the Byzantine Empire*[16] (1988 -), *EpiDoc*[17] (1990 -), *British National Corpus*[18] (1991 -), *Early American Fiction Collection* (1996 -)[19] i dr. Popis većine projekata koji danas upotrebljavaju TEI za neki vid označavanja teksta nalazi se u sklopu mrežnog mjesta TEI.[20]

Uloga opisnih označiteljskih jezika u razvoju digitalne humanistike u 1990-ima

Kako je kraj 80-ih godina prošlog stoljeća obilježila pojava *World Wide Weba*, a slijedom toga i rapidan porast uporabe interneta, mogućnost hipertekstualnog povezivanja te manipuliranja multimedijom (slike, audio, video i sl.) u mrežnom okruženju predstavljala je novi izazov za zajednicu stručnjaka digitalne humanistike, pogotovo za onaj njezin dio koji je sudjelovao u razvoju TEI-a. Međutim upravo mogućnost upravljanja multimedijom u mrežnom okruženju na početku je dovela do rasprava više tehničke prirode, o formatima slikovnih datoteka, dubini boje i rezoluciji, a manje, primjerice, o tome što će se sa slikama moći činiti, osim što će ih se odsad moći gledati i u digitalnom okruženju.

S druge strane, pojava SGML-a, osim u nekim stručnim zajednicama poput TEI-a, nažalost, nije pobudila veće zanimanje, pogotovo ne onih stručnjaka od kojih se to možda najviše očekivalo - razvojnih timova programa za obradu teksta, koji su na dolazak SGML-a ostali potpuno ravnodušni. Iako su još 1960-ih istraživači preporučivali računalnu obradu orijentiranu prema sadržaju teksta, računalni programi za obradu teksta koji su se pojavili na tržištu 1980-ih uglavnom su ignorirali upućene preporuke.[21]

Situacija se drastično izmijenila dolaskom weba i pojavom HTML-a, opisnog označiteljskog jezika baziranog na SGML-u koji se počeo koristiti kao temeljni format za izradu mrežnih stranica: jedna je SGML aplikacija napokon šire prihvaćena u svijetu, čak i izvan industrije računalnih programa za obradu teksta. Međutim veza između SGML-a

16 *The Prosopography of the Byzantine Empire*, http://www.pbw.kcl.ac.uk/.

17 *EpiDoc*, http://epidoc.sourceforge.net/index.shtml.

18 *British National Corpus*, http://www.natcorp.ox.ac.uk/.

19 *Early American Fiction Collection*, http://etext.lib.virginia.edu/eaf/.

20 *Text Encoding Initiative: Projects*, http://www.tei-c.org/Activities/Projects/.

21 Kao dobar primjer takve prakse možemo izdvojiti danas u svijetu možda i najpoznatije programsko rješenje za obradu teksta - *Microsoft Word* - koji se od svog početka bazirao na tzv. *page layout* modelu, a to je značilo modelu teksta koji posvećuje punu pažnju smještanju elemenata na stranici i izgledu same stranice, a manje njenoj sadržajnoj strani. Tek 2007. *Microsoft Word* je napokon prešao na XML format pohrane svojih dokumenata zadržavajući sučelje u obliku WYSIWYG (*What You See Is What You Get*) programa za obradu teksta.

i HTML-a u startu je bila labava i nepotpuna (Renear 2004: 226). Primjerice, HTML se počeo koristiti i prije nego što je službeno objavljen njegov DTD (*Document Type Definition*) odnosno dokument koji je propisivao njegovu strukturu, kao i sastav elemenata i atributa od kojih će se sastojati. Odluka njegovih tvoraca da neselektivno uključuju kako tipove elemenata opisnog tako i proceduralnog označavanja (npr. centriranje teksta, boju teksta, font slova i sl.) također je primjer nepromišljene odluke koja se temeljila na zanemarivanju prednosti odvajanja forme od sadržaja teksta (Renear 2004: 229). Isto tako, ni danas ne postoji poseban softver koji bi procesuirao HTML dokument (kao što to imamo u primjeru XML-a, gdje navedenu zadaću obavlja XML Parser), već je konačni prikaz HTML dokumenata na ekranu računala ostavljen na volju mrežnim preglednicima (engl. *web browsers*) koji, svaki na svoj način, tumače HTML kôd. Posljedica su česte razlike u prikazu istog HTML koda, a *web* dizajneri mrežnih stranica imaju mnogo problema s uspostavljanjem kompatibilnosti HTML koda s pojedinim mrežnim preglednicima. DTD se počeo činiti suvišnim jer HTML dokumenti više nisu morali slijediti pravila DTD-a, već ona mrežnih preglednika (Renear 2004: 229). Dakle, mnogi stručnjaci za TEI brzo su primijetili slabost HTML-a, opisnog označiteljskog jezika za izradu mrežnih stranica kao opisnog jezika za opis sadržaja. Međutim tvorci HTML-a namijenili su HTML u prvom redu za uređivanje prikaza sadržaja u mrežnim preglednicima, a ne za označavanje i opis istog. Stoga se HTML u daljnjem razdoblju i nije smatrao ozbiljnijim alatom koji bi se mogao koristiti za označavanje tekstova u humanističkim istraživanjima.

Imajući u vidu navedene probleme, bilo je očigledno da će tvorci *weba* u dogledno vrijeme nastojati stvoriti inačicu označiteljskog jezika koja bi bila sofisticiranija i lakša za upravljanje u skladu sa zahtjevima mrežnog okruženja. Sukladno tomu osobito se zanimljivom čini odluka da se ne ide u modifikaciju HTML-a, generatora svih problema, već SGML-a, kompleksnog metajezika na kojem se bazirao HTML. Cilj je ovaj put bio stvoriti jezik koji bi kombinirao jednostavnost HTML-a te fleksibilnost i izražajnu snagu SGML-a (Huitfeldt: 2004). Tako je nastao XML, pojednostavljena inačica SGML-a prilagođena mrežnom okruženju. Dovoljno je reći da SGML-specifikacija obuhvaća 155 stranica, dok XML-specifikacija u svojoj prvoj inačici obuhvaća tek 25, pa da se razumije zašto se W3 konzorcij odlučio na takav korak. Osnovna razlika u odnosu na SGML sastojala se u tome što su se XML dokumenti mogli *parsirati* (obraditi) bez da su morali imati pripadni DTD.

XML je u prvom nacrtu realiziran još 1996., ali je kao službena preporuka W3 konzorcija (*W3 Recommendation*) obznanjen 1998. Iako je radna skupina brojala 100 do 150 ljudi, završni rad na XML-specifikaciji verzije 1.0 bio je povjeren jedanaestorici članova uredničkog odbora među kojima se našao i M. Sperberg-McQueen, jedan od urednika TEI smjernica za označavanje teksta. XML-specifikacija dosad je objavljena u dvije verzije:

- XML 1.0 verzija prvi je put objavljena 1998., a 2008. zabilježila je svoje peto izdanje

- XML 1.1 verzija prvi je put objavljena 2004., a 2006. objavljeno joj je drugo i trenutno posljednje izdanje.

Najveća razlika između XML verzija 1.0 i 1.1. jest u tome što potonja više ne ovisi o trenutnoj inačici Unicodea, standarda za predstavljanje znakova i simbola većine pisama svijeta u digitalnom okruženju, već se uvijek koristi najnovijom inačicom.[22]

Dakle, kao ni SGML, niti XML u svom temeljnom smislu nije bio predviđen za to da bude označiteljski jezik, već označiteljski standard koji preko vlastitih pravila propisuje kako kreirati pojedini označiteljski jezik. Drugim riječima, pojednostavljena XML-sintaksa propisuje kako možemo kreirati vlastiti označiteljski jezik.

Ono što se s XML-om u startu željelo postići jest da prethodna komunikacija između stručnjaka koji kreiraju programe i samih kreatora sadržaja ne bude više neophodna. Za razliku od HTML-a, XML dokument, kako je prije navedeno, obrađuje posebno napisana aplikacija za tu namjenu - XML Parser. Pa, iako se i XML Parseri mogu razlikovati od slučaja do slučaja, svaki se oslanja na stroga sintaktička pravila kojima XML podliježe. Međutim, ono što je ovdje bitno naglasiti, XML Parseri mogu obrađivati XML dokument i bez pripadnog DTD-a. Naravno da je to, onda, rezultiralo nužnim i veoma strogim sintaktičkim pravilima prema kojima u XML dokumentima nije dozvoljeno izostaviti nijednu oznaku, a tu se prvenstveno misli na završnu oznaku elementa. Osim toga, u XML dokumentu elementi se ne smiju preklapati, vrijednosti atributa moraju biti navedeni u navodnim znakovima, a velika i mala slova se međusobno razlikuju. U navedenim pravilima sadržane su i najveće razlike između parova označiteljskih metajezika i jezika - SGML-a i HTML-a s jedne strane i XML-a i XHTML-a (XML aplikacije HTML-a) s druge strane: HTML i SGML za razliku od XML-a i XHTML-a, posjeduju daleko blaža sintaktička pravila (primjerice, dopuštaju izostavljanje završnih oznaka elemenata).

XML je odmah izazvao zanimanje i simpatije na *webu* jer je obećavao hardverski i softverski neovisno rješenje za pohranu i prijenos teksta u mrežnom okruženju. To svojstvo XML-a pokazalo se ključnim za njegovu široku primjenu.

Razvoj TEI-a u dvijetisućitima i uloga opisnih označiteljskih jezika u digitalnoj humanistici u današnje vrijeme

Godine 1999. Sveučilište Virginia i Sveučilište Bergen u Norveškoj predlažu osnivanje međunarodne udruge danas poznate pod nazivom TEI konzorcij (*TEI Consortium*), koja bi se brinula za održavanje, razvoj i promociju TEI-a u svijetu. Pored Sveučilišta Virginia i Sveučilišta Bergen, stalno domaćinstvo i skrb za TEI konzorcij dodijeljeni su

22 *W3 Consortium: XML core working group public page*, http://www.w3.org/XML/Core/#Publications.

i Sveučilištu Brown i Sveučilištu Oxford zbog dugogodišnjih veza s TEI-em. Prvi članovi odbora TEI konzorcija izabrani su početkom 2001.

U dvijetisućitima TEI objavljuje još dvije inačice svog standarda: inačicu P4 u lipnju 2002., a inačicu P5 krajem 2007. Dok su se prve tri inačice TEI smjernica bazirale na sintaktičkim pravilima SGML-a, inačica P4 omogućavala je izbor između SGML-a i XML-a, da bi se naposljetku P5 koristio isključivo XML-om kao sintaksom za označavanje teksta, što je u duhu mrežnog okruženja u kojem se većina tekstova u digitalnom obliku danas i nalazi.

Unutar TEI zajednice tijekom 2000-ih započinje djelovanje cijelog niza interesnih skupina zainteresiranih za razvoj postupka označavanja teksta putem TEI standarda koji se tiče ili specifičnih vrsta sadržaja (npr. rukopisi, ontologije[23] i sl.) ili pojedinih tematskih područja poput glazbe, korespondencije, izobrazbe i sl. ili su namijenjene razvoju TEI-a u pojedinim zajednicama (poput knjižničarstva, za potrebe projekata digitalizacije i digitalnih knjižnica).

U okviru TEI zajednice razvijeni su i mnogi alati namijenjeni olakšanom označavanju i obradi tekstova u digitalnom okruženju. Izdvaja se *Roma* - alat za generiranje *shema* dokumenata (DTD, XML Schema, RELAX NG) koji opisuje strukturu TEI datoteka za potrebe provjere njihove valjanosti.[24] Tu je i XSLT-specifikacija za transformaciju TEI XML dokumenata u druge formate (XHTML, LaTeX, ePub i dr.).[25] U okviru TEI označavanja teksta osigurana je podrška postupcima internacionalizacije (engl. *internationalization*) i lokalizacije (engl. *localization*) odnosno mogućnosti prijevoda TEI elemenata na bilo koji drugi jezik. Postoji iznimno bogat TEI Wiki,[26] izvor informacija o praksama TEI označavanja teksta, a u najnovije vrijeme pokrenut je i „Journal of Text Encoding Initiative", časopis o TEI-u.[27] Isto tako, TEI mrežno mjesto predstavlja ključni izvor informacija za svakog stručnjaka zainteresiranog za postupak označavanja teksta.

U okviru šire zajednice digitalne humanistike, određen broj stručnjaka bavi se i predviđanjem mogućih scenarija razvoja XML-a, označiteljskog jezika na kojem TEI danas počiva. C. Huitfeldt, supsumirajući te napore, prognozira: ili će XML ostati vodeći standard za obradu teksta u „digitalnoj okolini drugih srodnih tehnologija koje će iz dana u dana postajati sve kompleksnije (PDF, RTF, PostScript i dr.)" ili će XML „evoluirati u smjeru uključivanja tih drugih tehnologija u svoju specifikaciju" (Huitfeldt 2009). Oba scenarija posjeduju nedostatke. U prvom slučaju može se dogoditi da se srodne druge

23 Ontologije su, u najjednostavnijem smislu, „specifikacije konceptualizacije" odnosno formalni oblik predstavljanja znanja nekog znanstvenog područja. Vidi: Gruber, T. R. *What is an ontology*?, http://www-ksl.stanford.edu/kst/what-is-an-ontology.html.

24 *Roma: generating validators for the TEI*, http://www.tei-c.org/Roma/.

25 *TEI XSL customization handbook*, http://www.tei-c.org/Tools/Stylesheets/customize.xml.

26 *TEI Wiki*, http://wiki.tei-c.org/index.php/Main_Page.

27 „Journal of Text Encoding Initiative", http://journal.tei-c.org/journal/index.

tehnologije razviju u smjeru veće inkompatibilnosti s XML-om, a u drugom da sam XML izgubi svoju jednostavnost i postane izrazito kompleksna tehnologija (Huitfeldt 2004).

Konferencija *Digital Humanities* koja se održava svake godine i koju danas sponzorira Alijansa organizacija digitalne humanistike (*The Aliance of Digital Humanities Organizations - ADHO*),[28] a koja je nastala konvergencijom ranije spomenutih organizacija - ALLC-a i ACH-a te Društva za digitalnu humanistiku (*The Society for Digital Humanities / Société pour l'étude des médias interactifs - SDH-SEMI*), naglašeno tematizira tekstualne analize i postupak označavanja teksta (usp. Svensson 2009).

Općenito, na postupak označavanja teksta mnogi stručnjaci u području danas gledaju kao na jezgru digitalne humanistike (usp. Svensson 2009). Štoviše, prema K. de Smedtu postupak označavanje teksta kreira temelj za gotovo svaku uporabu računala u humanistici (usp. De Smedt 2002).

Zaključak

Opisni označiteljski jezici kao učinkoviti alati za obradu teksta danas predstavljaju nezaobilazno rješenje u humanističkom istraživačkom radu koji se provodi u digitalnom okruženju. Svoju najveću primjenu pronašli su u stvaranju TEI-a, danas vodećeg standarda za označavanje i razmjenu humanističkih tekstova u digitalnom okruženju iza kojeg stoji i istoimeni konzorcij.

U radu su najprije zasebno izloženi kronološki prikazi razvoja digitalne humanistike i opisnih označiteljskih jezika do sredine 1980-ih. U 1950-im i 1960-im razvoj digitalne humanistike obilježavaju praktični projekti uporabe računala u stilističkoj analizi tekstova. U 1970-ima dolazi do „konsolidacije područja" koju obilježava okrenutost stručnjaka iz tog područja objavljivanju sve većeg broja radova u novopokrenutim časopisima kao što je, primjerice, „Computer and Humanties". U 1970-ima osnivaju se i ključne organizacije (ALLC i ACH) koje preuzimaju skrb u vezi s organizacijom održavanja znanstvenih skupova. S druge strane, razvoj opisnih označiteljskih jezika započinje sredinom 1960-ih u grafičko-nakladničkoj industriji. Usporedo, razvijaju se dva opisna označiteljska jezika, GenCode i GML koji se krajem 1970-ih ujedinjuju te započinju s pripremama za objavu prvog standardiziranog opisnog označiteljskog jezika - SGML-a. U 1980-ima, razvoj digitalne humanistike pogađa kriza koja se ogleda u nedostatku standardiziranog rješenja za potrebe razmjene tekstova u znanstveno-istraživačkom radu. Konačno, godina 1986. uzima se kao prekretnica u razvoju digitalne humanistike jer je upravo te godine objavljena prva službena inačica SGML-a, opisnog označiteljskog jezika koji će ubrzo poslužiti za kreiranje nedostajućeg standarda za označavanje i razmjenu tekstova u okviru društvenih i humanističkih znanosti kroz osnutak Inicijative za označavanje teksta - TEI-a.

28 *ADHO: The aliance of digital humanities organizations*, http://www.digitalhumanities.org/.

Ubrzo nakon osnutka, TEI postaje vodeći standard za označavanje teksta i nezaobilazan sudionik svih projekata koji su uključivali naprednu računalnu obradu humanističkih tekstova. TEI standard propisuje skup elemenata i atributa za potrebe označavanja humanističkih tekstova uz odgovarajuće upute o provedbi postupka označavanja teksta u duhu označiteljske prakse TEI. Do danas objavljen je u pet inačica. Prve tri inačice temeljile su se na SGML opisnom označiteljskom jeziku, četvrta je dopuštala uporabu SGML-a i XML-a, dok se peta, trenutno važeća inačica, u potpunosti temelji na XML-u.

Može se zaključiti da je uloga opisnih označiteljskih jezika u razvoju digitalne humanistike do sada imala zamjetan značaj, ali i da se ona najprije ogledala u podršci razvoju TEI-a, označiteljskog standarda baziranog na XML opisnom označiteljskom jeziku koji se široko primjenjuje u zajednici digitalnih humanista. Što će se dogoditi u budućnosti, teško je predvidjeti. Može se dogoditi da razvoj XML-a usmjere potrebe humanističkih znanstvenika koji ga danas koriste za postupak označavanja teksta. Potreba za izmjenama TEI-a može dovesti do potreba za izmjenama označiteljskog jezika na kojem počiva. U svakom slučaju, pretpostavka je da će ocjena razvoja TEI-a ostati ključan kriterij za svaku ocjenu uloge opisnih označiteljskih jezika u daljnjem razvoju područja digitalne humanistike.

Literatura

Aarseth, E. 1997. *The field of humanistic informatics and its relation to the humanities*, http://www.hf.uib.no/hi/espen/HI.html.

Brown, S. i ostali. 1998. SGML and the Orlando project: descriptive markup for an electronic history of women's writing. *Computers and the humanities* 31: 271 - 284.

Brunner, T. F. 1993. Classics and the computer: the history of a relationship. U *Accessing antiquity: the computerization of classical studies*, uredio J. Solomon, Tucson: University of Arizona Press, 10 - 33.

Burnard, L. 1988. Report of workshop on text encoding guidelines. *Literary & Linguistic Computing* 3 (2): 131 - 133.

Coombs, J. H., Renear, A. H., i S. J. DeRose. 1987. Markup systems and the future of scholarly text processing. *Communications of the ACM* 30 (11): 933 - 947.

Crane, G. R. 1998. The Perseus project and beyond: how building a digital library challenges the humanities and technology. *D-Lib Magazine* 4 (1), http://www.dlib.org/dlib/january98/01crane.html.

De Smedt, K. 2002. Some reflections on studies in humanities computing. *Journal of Linguistic and Literary Computing* 17: 89-101.

Fraser, M. 1996. *History of humanities computing*. http://info.ox.ac.uk/ctitext/history/index.html.

Goldfarb, C. F. 1981. A generalized approach to document markup. U *Proceedings of the ACM SIGPLAN-SIGOA Symposium on Text Manipulation*: 68 – 73

Goldfarb, C. F. 1996. *The SGML history niche: the roots of SGML – a personal recollection*, http://www.sgmlsource.com/history/roots.htm.

Hockey, S. 1978. *Computing in the arts at Oxford University : developments during the academic year 1976-77*. Oxford: Oxford University Computing Service.

Hockey, S. 2000. Electronic texts in the humanities: principles and practice. Oxford: Oxford University Press.

Hockey, S. 1994. Evaluating electronic texts in the humanities. *Library trends* 42 (4): 676 – 693.

Hockey, S. 2004. The history of humanities computing. U *A companion to digital humanities*, uredili Susan Schreibman, Raymond George Siemens i John M. Unsworth. Wiley – Blackwell, 3 – 20.

Huitfeldt, C. 2004. *Scholarly text processing and future markup systems*, http://computerphilologie.uni-muenchen.de/jg03/huitfeldt.html.

McCarty, W. 2003. Humanities Computing. U *Encyclopedia of Library and Information Science*, uredila Miriam A. Drake. New York: Marcel Dekker, 1224 – 1236.

Mendenhall, T. C. 1901. A mechanical solution of a literary problem. *Science Monthly* 60: 97 – 105.

Morton, A. Q. 1965. The authorship of the Pauline Epistles: a scientific solution. Saskatoon: University of Saskatchewan.

Renear, A. H. 2004. Text encoding. U *A companion to digital humanities*, uredili Susan Schreibman, Raymond George Siemens and John M. Unsworth. Wiley – Blackwell, 218 – 240.

Russell, D. B. 1965. COCOA – a word count and concordance generator, http://www.chilton-computing.org.uk/acl/applications/cocoa/p001.htm.

Saltzer, J. H. TYPSET and RUNOFF, memorandum editor and type-out commands. 1964. URL: http://mit.edu/Saltzer/www/publications/CC-244.html (2010-09-01)

Spring, M. B. 1989. The origin and use of copymarks in electronic publishing. *Journal of Documentation* 45: 110-123.

Svensson, P. 2009. Humanities Computing as Digital Humanities. *Digital humanities quaterly* 3 (3), http://www.digitalhumanities.org/dhq/vol/3/3/000065/000065.html

TEI: Text Encoding Initiative, http://www.tei-c.org/index.xml.

TEI XSL customization handbook, http://www.tei-c.org/Tools/Stylesheets/customize.xml.

TEI Wiki, http://wiki.tei-c.org/index.php/Main_Page.

W3 Consortium: XML core working group public page. URL: http://www.w3.org/XML/Core/#Publications.

Summary

Descriptive markup languages and the development of digital humanities

The paper discusses the role of descriptive markup languages in the development of digital humanities, a new research discipline that is part of social sciences and humanities, which focuses on the use of computers in research. A chronological review of the development of digital humanities, and then descriptive markup languages is exposed, through several developmental stages. It is shown that the development of digital humanities since the mid-1980s and the appearance of SGML, markup language that was the foundation of TEI, a key standard for the encoding and exchange of humanities texts in the digital environment, is inseparable from the development of markup languages. Special attention is dedicated to the presentation of the *Text Encoding Initiative - TEI* development, a key organization that developed the titled standard, both from organizational and markup perspectives. By this time, TEI standard is published in five versions, and during 2000s SGML is replaced by XML markup language.

KEY WORDS: markup languages, digital humanities, text encoding, TEI, SGML, XML.