

Organiziranje projekta vizualizacije podatkov

Narvika Bovcon, * Jure Demšar, * Aleš Vaupotič†

* Fakulteta za računalništvo in informatiko, Univerza v Ljubljani
Večna pot 113, SI-1000 Ljubljana

narvika.bovcon@fri.uni-lj.si

* Fakulteta za računalništvo in informatiko, Univerza v Ljubljani
Večna pot 113, SI-1000 Ljubljana

jure.demsar@fri.uni-lj.si

† Raziskovalni center za humanistiko, Univerza v Novi Gorici
ales.vaupotic@ung.si

Povzetek

Članek bo predstavil vizualizacije zbirke literarnovednih podatkov, projekt, ki je potekal v okviru študijskega procesa na Fakulteti za računalništvo in informatiko Univerze v Ljubljani, v povezavi sodelavcev iz treh disciplin: humanistike, računalništva in informatike ter grafičnega oblikovanja.

1 Uvod

Področje digitalne humanistike je izoblikovalo specifične žanre in metode, s katerimi se loteva urejanja in raziskovanja digitaliziranih kulturnih zbirk podatkov. Osrednji poudarki in raziskovalni pristopi so opisani v petnajstih poglavjih zbornika *Digital Humanities* (Burdick et al., 2012), kot teoretski okvir področja jih poskusimo strniti v naslednji seznam: (1) kuriranje večpredstavnostnih zbirk, (2) označevanje besedil, (3) oddaljeno branje in strojno učenje, (4) različni pogledi na iste podatke, (5) kulturna analitika in podatkovno rudarjenje, (6) vizualizacija podatkov in oblikovanje informacij, (7) platenje podatkov na geolokacijah, (8) digitalne skupnosti, (9) porazdeljeno proizvajanje znanja in množičenje, (10) resne igre, (11) refleksija programja, (12) narativizacija podatkovnih zbirk, (13) kultura remixa in ponovne uporabe, (14) vseprisotna infrastruktura, (15) javna dostopnost vsebin. Gre za dejavnosti in problemske sklope, ki se med seboj povezujejo in nadgrajujejo, zato jih lahko predstavimo tudi v bolj skrčenem naboru temeljnih pojmov. Teorija novomedijskega objekta Leva Manovicha (2001) poveže podatkovno zbirko in množico vmesnikov, prek katerih dostopamo do podatkov v zbirki. Ta koncept združuje točke 1, 2, 3, 4, 5 in 12. Naslednja pomembna enota se nanaša na samo zasnovo in jo zato lahko prepoznamo kot osnovni element digitalne humanistike, to so digitalni modeli kulturnih artefaktov, "oblike argumenta, izražene v informacijskih strukturah – vmesnikih, podatkovnih zbirkah, orodjih, platformah" (Burdick et al., 2012), ta koncept poveže novomedijski objekt s točkama 11 in 13, z refleksijo lastnega obstoja in učinkovanja ter načini produkcije in uporabe. Kot poseben poudarek raziskovanja digitalne humanistike ostaja točka 6, vizualizacije, ki prenašajo podatke v vidno razumljivo obliko prek likovnega jezika. Točke 7, 14 in 15 se povežejo v enoto, ki omogoča obstoj in delovanje v mešani resničnosti, kjer je poudarek lahko tudi na telesnosti. Točki 8 in 10 se nanašata na pravila, ki so ključna tako za resne igre kot za delovanje digitalnih skupnosti; s tem se ukvarja teorija kiberteksta (Aarseth, 1997). Točka 9 izpostavi družbo in družbeno raznolikost kot enega ključnih akterjev, udeleženih v pristopih digitalne humanistike.

2 Cilji

V projektu vizualizacije literarnovedne podatkovne zbirke smo uporabljali predvsem metode, povezane z novomedijskim objektom, torej v relaciji med podatkovno zbirko in njenimi vmesniki, ter metode vizualizacije podatkov in oblikovanja informacij (Manovich, 2011). Sama izgradnja podatkovne zbirke ni bila v domeni našega projekta, saj smo izdelovali zgolj vmesnike za obstoječo podatkovno zbirko *Women Writers* (<http://neww.huygens.knaw.nl>), ki je nastala v okviru evropskega projekta *Women Writers In History* (COST Action, 2009–2013). Z našim pristopom, ki sodi med generativne humanistične metode (Burdick et al., 2012), in množico konkretnih partikularnih rešitev smo dejavno prispevali k oblikovanju digitalnega modela kulturnega artefakta podatkovne zbirke *Women Writers*: sama podatkovna zbirka namreč še ni tematizirala možnosti vizualizacij kot svojega integralnega dela, ki skupaj z zbirko tvori argument, izražen v informacijski strukturi. Pravzaprav vsaka posamezna prototipna vizualizacija podatkovne zbirke predstavlja specifično obliko tega digitalnega kulturnega artefakta. Uredniki podatkovne zbirke so izdelali drugačne oblike argumentov, s tem ko so zbirko predstavili s spletno stranjo, ki ob iskanju izpiše seznam zadetkov, ter jo povezali s posebej zanjo načrtovanim in izdelanim virtualnim raziskovalnim okoljem, ki bo v okviru projekta *Travelling Texts 1790–1914* (HERA CRP, 2013–2016) dokončano letos (<http://resources.huygens.knaw.nl/womenwriters>).

Drugačen okvir za pregledovanje in sistematizacijo področja digitalne humanistike ponuja npr. pregled vsakoletnih nagrajenih projektov na spletišču *Digital Humanities Awards* (<http://dhawards.org>). Primerjava s podobnimi projekti omogoča natančnejše obravnavanje posebnosti, povezanih z vizualizacijami podatkovne zbirke *Women Writers*.

Poleg teoretskega ozadja in primerjave z referenčnimi sorodnimi projekti bo članek predstavil tudi samo organizacijo tovrstnega interdisciplinarnega ter v pedagoški proces integriranega projekta ter s tem povezane omejitve in priložnosti. V prvem podpoglavju predstavimo metodo, ki je v svojem temelju povezana z organizacijo projekta.

2.1 Metode in organizacija projekta

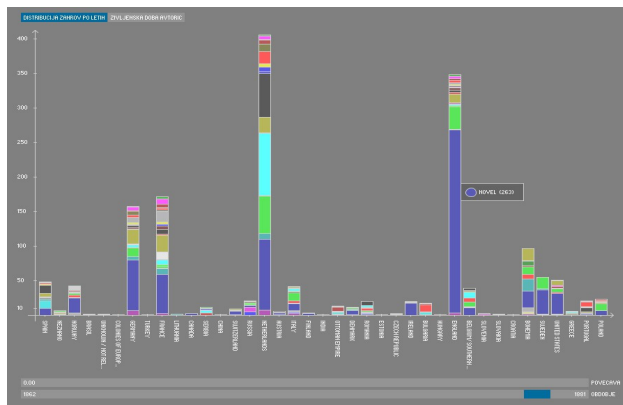
Projekt je potekal v treh organizacijskih oblikah, ki so vsaka po svoje strukturirale odnose med sodelavci z različnih raziskovalnih področij. Prva organizacijska oblika je študente računalništva in informatike soočila z literarnovedno podatkovno zbirko in mentorico, grafično oblikovalko. Druga organizacijska oblika je vpeljala literarnovedne raziskovalke, ki so soustvarjale zbirko Women Writers, da so zastavile znanstvena vprašanja, na katera so bili odgovori pridobljeni skozi vizualizacije. Tretja oblika je namesto profesionalnih raziskovalcev vključila študente humanistike in s tem razširila pedagoški proces še na drugo disciplino. V vseh organizacijskih oblikah so vizualizacije izdelali študenti Fakultete za računalništvo in informatiko Univerze v Ljubljani pod mentorstvom izr. prof. dr. Narvike Bovcon in asistentov Tadeja Zupančiča in Jureta Demšarja. Koordinacija projekta s sodelavkami zbirke Women Writers je potekala prek doc. dr. Aleša Vaupotiča, ki je bil tudi mentor študentom humanistike na Univerzi v Novi Gorici.

2.1.1 Množica vmesnikov do podatkovne zbirke

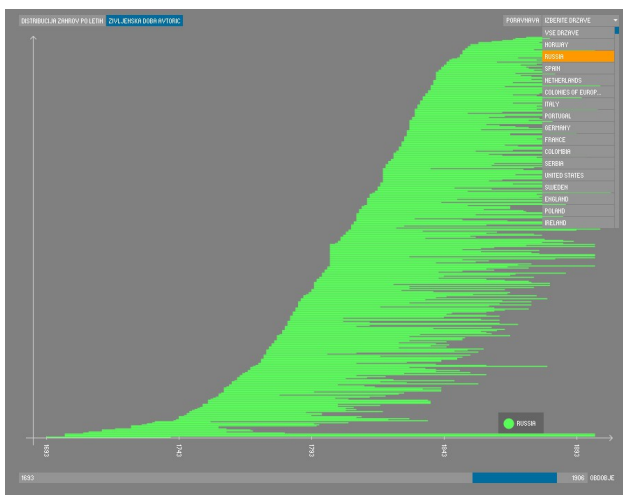
Metoda generativnih humanističnih raziskav izdeluje argumente s pomočjo prototipov, ki jih postopoma izboljšuje. Gre za pristop, ki je značilen za oblikovalske projekte, kjer je poudarek na izdelavi konkretnih rešitev, množice prototipov, ter njihovo ovrednotenje in nadgrajevanje (Kräutli, 2016). Prednost umestitve izdelovanja prototipov v študijski proces je ta, da razred stotih študentov, ki izdela v enem mesecu nalogo, izdela dejansko sto povsem različnih prototipov. Takemu obsegu produkcije je v okvirih industrije oz. komercialno zasnovanega projekta zelo težko slediti, saj se finančno in terminsko ne izide, z bistveno večjo težavo tudi zagotovi toliko človeških virov. Pri tem želimo poudariti, da se je v našem pristopu individualnost študentov jasno odražala v veliki raznolikosti njihovih izdelkov, medtem ko izdelki istega avtorja bistveno težje ali sploh ne morejo doseči tolikšne raznolikosti izjav. Izdelava naloge v več zaporednih študijskih semestrih (naš projekt je potekal od leta 2012 do leta 2016) prinese na stotine rešitev, med katerimi je mogoče izbrati najboljše: tiste, ki so kvalitetne na ravni vsebine, torej smiselne in zanimive glede samega izbora prikazanih odnosov med informacijami iz podatkovne zbirke, in na ravni vizualne predstavitve teh odnosov in informacij, torej grafičnega oblikovanja vmesnika. Za študente je tak projekt zanimiv (kljub težavam in predsodkom, pogosto povezanim z interdisciplinarnim sodelovanjem), ker se vključujejo v raziskovalno prakso.

Raznolikost prototipnih rešitev izvira v veliki meri tudi iz našega načrtno fragmentarnega pristopa k vizualizaciji. Namen ni bil, vizualizirati celotno zbirko oz. izbrati nekaj ustreznih grafov za prikaz vseh povezav med deli zbirke. Temu smo se odrekli zato, ker je zbirka preveč obsežna in bi že zaradi tega tovrstna naloga ne bila več izvedljiva v okvirih študentskega seminarskega projekta. Drugi, pravzaprav pglavitni razlog je bil počasno delovanje prikaza, ko se vizualizirajo vsi podatki iz zbirke, kar naredi vizualizacijo praktično nefunkcionalno za razbiranje prikazanih podatkov in odnosov med njimi (tovrsten prikaz smo poskusili izdelati v obliki različnih grafov in z različnimi tehnologijami, vendar je vsakokrat

deloval prepočasi in bil zato neuporaben). Tretji razlog za izbiro parcialnih pogledov na zbirko je bila večja fleksibilnost in inovativnost pri formuliranju posameznih smiselnih znanstvenih vprašanj v zvezi s podatki iz zbirke (osnova za to so bile kategorije podatkov), saj so bili študenti motivirani, da so s pomočjo vizualizacije poiskali odgovor na konkretno vprašanje, ki se jim je porodilo ob pregledovanju in razbiranju ustroja zbirke, brez vizualizacije pa odgovora nanj ni ali pa ni očitno (Slika 1). Na ta način so bili tematizirani različni vstopi v zbirko, ki so temeljili na konkretnih vprašanjih v zvezi s podatki, torej so zbirko skozi vizualizacijo narativizirali.



Slika 1: Dva pogleda na zbirko, izbira prek gumbov v zgornji vrsti – popularnost žanrov in življenjska doba avtoric. (Študenta: Aleksandar Kojić, Dejan Grbec.)



Slika 2: Vizualizacija pokaže napako pri vnosu življenjske dobe avtorice.

Tovrstni fragmentarni, skicozni, partikularni pristop je najustreznejši način za izdelavo eksperimentalnih vizualizacij, ki služijo mdr. namenu promocije zbirke, javnemu predstavljanju projekta za različna občinstva ter tudi odkrivanju napak in iregularnosti v zvezi z vnosi v zbirko (Slika 2). Zbirka Women Writers nujno potrebuje uspešno predstavitev in promocijo, saj gre za raziskave in predstavljanje podatkov o do sedaj zapostavljenem in skorajda nevidnem ustvarjalnem delu žensk vse do vključno t. i. dolgega 19. stoletja, ki se danes mora družbeno afirmirati in uvrstiti v šolske učne programe.

Posamezni prototipi vmesnikov so bili izdelani skozi stopnje korektur v zvezi z grafičnim oblikovanjem in strukturiranim posredovanjem informacij, ki so potekale pod mentorstvom oblikovalke, zato so nekateri med njimi – skozi postopno odstranjevanje napak in v kombinaciji s softverskimi predlogami za različne diagramске prikaze iz prosto dostopnih knjižnic, ki zagotavljajo dokaj visoko stopnjo grafične urejenosti – dosegli raven ustrezne berljivosti, smiselne razporeditve informacij po ekranski plošvi ter za uporabnika intuitivne interakcije. Študenti računalništva in informatike so na ta način, tj. skozi interdisciplinarno sodelovanje, oblikovali čitljive in likovno urejene vmesnike za vizualizacije odnosov med podatki. Če razmislimo, ali bi bilo bolje, da bi vmesnike oblikovali študenti oblikovanja vizualnih komunikacij ali pa celo profesionalni oblikovalci, hitro ugotovimo, da obstaja ovira, ki je tehnološke narave: oblikovalci bi se morali naučiti nekaj programiranja, kar v Sloveniji danes še ni prav pogosta praksa, čeprav je bil npr. Processing (<http://processing.org>) razvit za oblikovalce vizualnih komunikacij v novih medijih in v tujini obstajajo številni interdisciplinarni novomedijski študiji, ki temeljijo na oblikovanju. V Sloveniji je organizacija te vrste interdisciplinarnega projekta danes lažje izvedljiva v okviru študija računalništva in informatike.

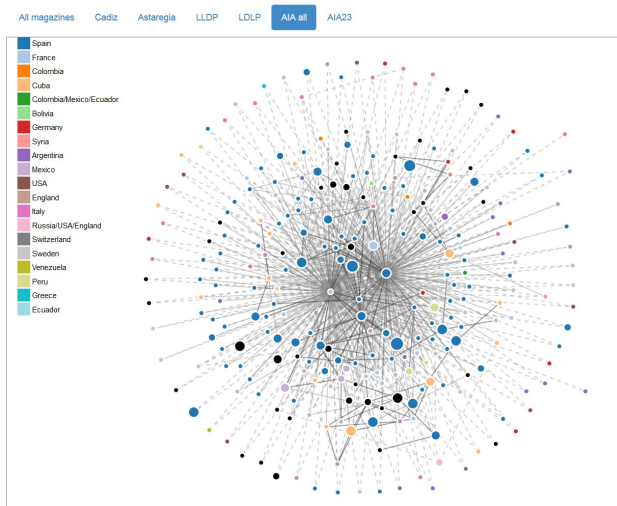
Pregled najboljših vizualizacij, ki so bile izdelane v letih 2012–2013 – te so bile dveh vrst: interaktivne, izdelane s pomočjo platforme Processing, za katere je bilo značilno približevanje in oddaljevanje pogleda na podatke prek animacij, interakcije in postopnih prikazov podrobnosti, ter na drugi strani videi in gibljive informacijske grafike, izdelane s pomočjo programa Adobe After Effects, ki so s svojo linearno obliko izrazito narativizirale podatkovno zbirko – so podrobneje opisane in dokumentirane z ekranskimi slikami v članku Narvike Bovcon (2014). Vizualizacije so jasno pokazale, da je vsakršna statistična, tj. količinska primerjava med podatki v bazi, povezana predvsem z vnosi v bazo in ne odslikuje v celoti dejavnosti avtoric, npr. količine izdanih del ali prejetih recepcij po Evropi, saj različne države in jezikovni teritoriji bolj ali manj celovito vnašajo gradivo v bazo: Holandija ima daleč največ vnosov, saj je baza nizozemski projekt. Torej pri poskusu oddaljenega branja te podatkovne zbirke ne gre za dilemo o redukciji kompleksnosti humanističnih podatkov na nekaj ali eno njihovo lastnost, prek katere se jih primerja (Drucker, 2011), ampak za strukturno razliko nezaključene baze, kakršna je Women Writers, ki je projekt v nastajanju (vizualizacije se povežejo z bazo in odražajo trenutno stanje zbirke), v primerjavi z zaključeno bazo, kakršna je neko dramsko besedilo, ki je napisano in vsebuje določeno število junakov, vizualizacija pa lahko riše razmerja med njimi, soprisotnost junakov v dramskih prizorih ali pa količino izgovorjenega teksta (Grandjean, 2015). Zaradi velikih teritorialnih in časovnih razdalj ter neenakomerne izpolnjenosti zbirke izrisovanje mrež povezav v zbirki Women Writers zahteva omejitve pri prikazih (npr. omejitev na eno državo), če želimo dobiti dokaj primerljive nabore podatkov, kot npr. v projektu, ki izrisuje mreže povezav med člani Belfastske skupine (<http://belfastgroup.digitalscholarship.emory.edu/network>), kjer gre za bistveno manjši in bolj povezan nabor možnih podatkov, čeprav še niso vsi zavedeni in

upoštevani. V tej smeri so šle nadaljnje raziskave, v katerih so bila znanstvena vprašanja izrazito fokusirana ter podatki posebej urejeni s strani strokovnjakinj.

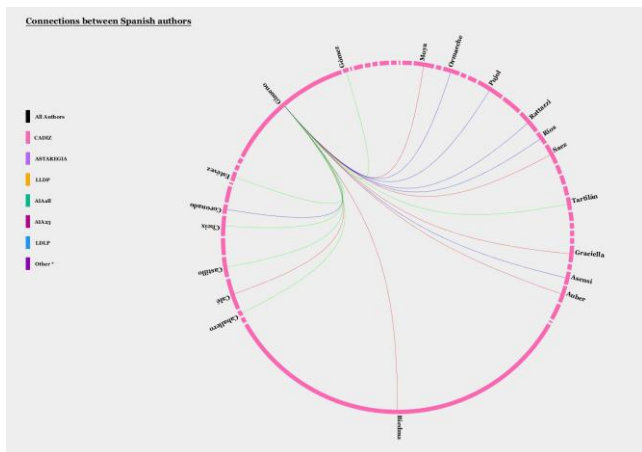
2.1.2 Nove raziskave – vizualizacije strokovno oblikovanih znanstvenih vprašanj v zvezi z izbranimi podatki

Druga stopnja organizacije projekta je testirala vključitev profesionalnih humanistov, torej profesorjev in raziskovalcev podatkov, vključenih v gradnjo zbirke, pri zastavitvi znanstvenega vprašanja (to je bilo v prvi stopnji projekta prepuščeno samim študentom računalništva in informatike in je bilo v več primerih zastavljeno premalo kompleksno ali pa sploh narobe – povezovalo je podatke, ki v resničnosti niso bili soodvisni). Na ta način smo zagotovili najvišjo stopnjo zahtevnosti in doslednosti v zvezi z znanstvenim vprašanjem, naloga študentov pa je bila, to vprašanje čim bolje vizualizirati: fokus naloge je bil prestavljen z načrtovanja vsebine na načrtovanje oblike vizualizacije. Poseben, skrbno pripravljen sklop podatkov, povezanih z njunima raziskovalnima znanstvenima vprašanjema, sta posredovali dr. Suzan van Dijk s Huygens ING, o povezavah med nizozemskimi avtoricami, in Judith Rideout, doktorska študentka z Univerze v Glasgowu, o moči povezav med španskimi avtoricami, ki so objavljale pri različnih literarnih revijah; avtorice so se bodisi poznale enostransko, vzajemno, ali pa so delale na skupnih projektih, kar pomeni najmočnejšo povezavo. Izdelana sta bila dva, povsem različna prikaza moči povezav med španskimi avtoricami.

Visualization of Spanish Authors



Slika 3: Tri debeline črt predstavljajo tri stopnje moči povezav med španskimi avtoricami (avtorice so predstavljene s točkami na mreži), ki so pisale za konkretno revijo (izbrana v zgornjem meniju). Barve označujejo države. Ob kliku na posamezno točko na grafu se izpišejo podrobnejši podatki o izbrani avtorici. Mreža je interaktivna in jo je mogoče premikati, razplesti, povezave delujejo kot elastike – interakcija omogoča postopno razbiranje povezav, ki zaradi številnosti niso razvidne prek pogleda na celoten graf. (Študentki: Mojca Komavec, Viki Petrovič.)



Slika 4: Moč povezave med avtoricami, ki so razporejene po krogu, je označena s tremi različnimi barvami črt. Barva krožnice predstavlja izbrano revijo (izbranih je lahko več revij hkrati) in se povezuje z legendo na levi. (Študentka: Marija Djurdjević.)

Posebej velja omeniti primer t. i. resne igre, ki recepcije med nizozemskimi avtoricami postavi v interaktivno virtualno resničnost, narejeno s pomočjo pogona Unity, v kateri se uporabnik sprehaja po tipični nizozemski pokrajini med hišami posameznih avtoric, vzdolž poti pa bere citat iz recepcije, ki jo je napisala avtorica, katere hišo zapušča, o avtorici, proti hiši katere hodi. V tem primeru je pomembno izumljanje vmesnika, ki omogoča uprostorjeno branje, s tem ko poveže računalniško generirani 3D prostor z zbirko citatov in mrežo povezav (Grosar in Demšar in Bovcon, 2015).

V tretjem modelu sodelovanja je bila naloga študentov humanistike, postaviti znanstveno vprašanje v zvezi s

Literatura

- Espren Aarseth. 1997. *Cybertext: Perspectives on Ergodic Literature*. Baltimore: The Johns Hopkins Univ. Press.
- Narvika Bovcon. 2014. Jezik gibljivih slik v računalniških vizualizacijah literarnozgodovinske podatkovne zbirke. *Primerjalna književnost*, 37(2):119–133, 235–242.
- Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner in Jeffrey Schnapp. 2012. *Digital Humanities*. Cambridge, Mass.: MIT Press.
- Digital Humanities Awards. 2012–2015*. <http://dhawards.org>.
- Johanna Drucker. 2011. Humanities Approaches to Graphical Display. *Digital Humanities Quarterly*, 5(1). <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>.
- Johanna Drucker. 2014. *Graphesis: Visual Forms of Knowledge Production*. Harvard UP.
- Belfast Group Poetry: Networks*. Emory University's Manuscript, Archives, and Rare Book Library. <http://belfastgroup.digitalscholarship.emory.edu/network>.

podatki iz zbirke in ga formulirati v obliki jasnega navodila, na podlagi katerega študenti računalništva in informatike izdelajo vizualizacijo. Prototipe rešitev so študenti humanistike v zaključnem eseju tudi ovrednotili, razčlenili, kako in ali so podatki prikazani na razumljiv način, kako bi bili posamezni deli lahko prikazani bolj nazorno in razumljivo, kakšno vlogo imajo sredstva grafičnega oblikovanja pri podajanju informacij. S tem so aktivno vstopili v študij digitalne humanistike.



Slika 5: Navodilo na dnu ekrana usmerja uporabnika k hiši naslednje avtorice, omenjene v citatu, ki ga bere na zgornjem delu ekrana. (Študent: Jernej Grosar.)

3 Sklep

Podatkovna zbirka Women Writers se je izkazala kot ustrezna, dovolj kompleksna zbirka podatkov, ob kateri se študenti naučijo kritično razmišljati tako o različnih smiselnih povezavah med kategorijami podatkov kot tudi o načinih grafičnega prikaza teh odnosov na interaktivnih vmesnikih. V prihajajočem mesecu bomo izdelali predstavitevno spletno stran, na kateri bodo dostopne izbrane eksperimentalne vizualizacije te zbirke.

Martin Grandjean. 2015. *Network visualization: mapping Shakespeare's tragedies*.

<http://www.martingrandjean.ch/network-visualization-shakespeare>.

Jernej Grosar, Jure Demšar, Narvika Bovcon. 2015. 3D walk through the references of Dutch women writers. V: *StuCoSReC: proceedings of the 2015 2nd Student Computer Science Research Conference*. Koper: University of Primorska: 39-42, ilustr.

Florian Kräutli. 2016. *Visualising Cultural Data: Exploring Digital Collections Through Timeline Visualizations*. Ph.D. thesis, Royal College of Art.

Lev Manovich. 2001. *The Language of New Media*. Cambridge, Mass.: MIT Press.

Lev Manovich. 2011. What is visualization? *Visual Studies*, 26(1):36–49.

Processing. <http://processing.org>.

Aleš Vaupotič. 2007. Literarno-estetski doživljaj in novi mediji - prihodnost literature? *Primerjalna književnost* 30(1):203–216.

WomenWriters. 2012–2016. Ur. Suzan van Dijk. <http://neww.huygens.knaw.nl> i n <http://resources.huygens.knaw.nl/womenwriters>.