

PRIČA IZMEĐU REDAKA: ISTRAŽIVANJE ZBIRKI DOKUMENATA O KULTURNOJ BAŠTINI DISTRIBUIRANIH *ONLINE* KORIŠTENJEM METODA NA BAZI ONTOLOGIJA

**EFTHIMIOS C. MAVRIKAS,
EVANGELIA KAVAKLI,
NICOLAS NICOLOYANNIS**
Mitilena, otok Lezbos, Grčka &
Bron Cedex, Francuska

PISANJE ISKORISTIVE PROŠLOSTI

Kolektivno je sjećanje utemeljeno na usmenoj tradiciji definiranoj u smislu shematskih iskaza [19]. Diskurs kulturne baštine koristi se činjenicama iz tih kognitivnih instrumenata [13] kako bi shvatio podatke o predmetima / artefaktima, ljudima, događajima, motivacijama i njihovim okruženjima radi aktivnog stvaranja prošlosti u prostoru modernističkog entiteta [18]. Povijesni, arheološki, muzeološki i antropološki diskurzivni kodovi deklarativni su, potvrdni iskazi događaja i pojava [2] odražavajući rigoroznu analizu dokaza, pri čemu ostaju oblikovani implicitnim teorijama i neobjektivnošću, a odnose se na materijalne dokaze i pronalaženje znanja, zauzimanje stavova i stvaranje znanja. Socijalni i akademski teret diskursa kulturne baštine predočen je idejama koje su prenesene riječima [1]. Riječi i pi-

sanje stvaraju interpretacije prošlosti svojstvene tekstualnome mediju, intelektualnom proizvodu koji podjednako obitava u ideji teksta i u lingvističkome materijalu.

Lingvistički materijal teksta obuhvaća sva eksplicitno očitovana lingvistička sredstva signaliziranja, uključujući ekološke, morfološke, sintaktičke, semantičke i pragmatičke konstrukcije. Redosljed dijelova teksta, lokacijskih znakova i znakovnih fraza, korištenje zamjenica i referenci, vremena i aspekti glagola sredstva su koja često signaliraju specifičnu komunikacijsku namjeru. Međutim, ukupna predviđena struktura [10] teksta obuhvaća i ono što nije napisano, koliko i ono što jest [18]. Ideja teksta mora biti potražena u skrivenim konceptima i značenjima, kao i u sukobu s eksplicitnim signalnim sredstvima. Tilly, citirajući Machereya, koji pak citira Nietzschea, podsjeća nas da si postavimo pitanje o tekstu:

Kada smo suočeni s bilo kojom manifestacijom koju je netko dopustio da vidimo, možemo upitati: što ona treba sakriti? Od čega nam treba odvratiti pozornost? Koje predrasude želi stvoriti? I, opet, koliko daleko suptilnost maske mora ići? [15]

Zatim vješto sažimlje problem:

Adekvatno pročitati tekst znači ponovno ga napisati, ispuniti sve one nedostatke pronađene u marginama teksta i u prazninama između redaka i riječi. Znanost teksta ne ostavlja ga onakvim kakav jest, već ga mijenja. [18]

RASPON *ONLINE* PRISUTNOSTI

Proizvodi li drvo koje padne u šumi zvuk ako nema nikoga da ga čuje?

Dokument kulturne baštine - pisani iskaz u smislu informacije-kao-stvari [5] - može prenijeti znanje samo ako komunicira s publikom. Čitatelj je uvodničar u uspostavljanju veza [3] među točkama u diskurzivnoj mreži stvorenoj namjernim strukturama nađenim unutar zbirke dokumenata. Zbirke dokumenata stvorene su od informacijskih institucija kao fizičkih mjesta gdje su resursi izabrani, organizirani, očuvani i dostupni [4] kao potpora zajednici čitatelja. *Online* zbirke dokumenata šire podržani doseg zajednice uključivanjem tradicionalnih informacijskih institucija kao što su knjižnice, muzeji, arhivi i škole, zajedno s učionicama, uredima, laboratorijima, domovima i javnim mjestima. Proširenoj zajednici čitatelja pružena su sredstva za pristupanje, stvaranje, spremanje, traženje, prikaz i istraživanje *online* zbirki dokumenata unutar prostora zapisa kulturne baštine.

Prostor zapisa kulturne baštine ima hipertekstualni dizajn i angažira zajednicu čitatelja u nelinearnom i transformativnom načinu interakcije sa zbirkama dokumenata kulturne baštine. Prostor zapisa kulturne baštine jest primjena tehnologije semantičke mreže i računalne semantike orijentirane prema pristupu smislu koji se temelji na očekivanju.

POSTUPANJE S ANALIZOM I STVARANJE TEKSTA

Arhitektura prostora zapisa kulturne baštine oslanja se na teoriju i metodologiju procesiranja ontološke semantike prirodnih jezika koje su uveli Nirenburg i Raskin, smještajući model konstruiranog znanja - ontologiju - u jezgru metode dohvata informacija i metode

zaključivanja o znanju izvučenom iz teksta te interaktivnošću čitatelja; ontologija je dopunjena epizodnim repozitorijem, leksikonom, onomastikom i skupom modula za procesiranje za željenu strukturu teksta.

Epizodni repozitorij čuva epizodno sjećanje [16] na znanje o primjerima predmeta / artefakata, ljudi, događaja, motivacija i njihovih okruženja, indeksiranih odgovarajućim konceptima ontologije i međusobno povezanim vremenskim, uzročnim i drugim svojstvima. Sustavno indeksiranje epizodnog znanja omogućuje zaključivanje na osnovi slučaja [17] i analogno zaključivanje [6].

Leksikon popisuje uobičajene izraze tipično objašnjene u smislu koncepta ontologije i određenoga fragmentarnog znanja, pretraživog prema formi citata i povezanoga sinonimima i izrazima u tezaursima. Onomastikon popisuje vlastita imena, obično objašnjena u smislu koncepta ontologije, uobičajene izraze, popisane u leksikonu i određeno fragmentarno znanje, indeksirano u obliku citata i povezano s jedinstvenim izrazima u tezaursima. Procesiranje planirane strukture teksta obuhvaća modularizaciju analize lingvističkih signalnih sredstava i generiranje u skup modula koji tvore protočnu strukturu (engl. pipeline).

Raščlanjivač teksta sastoji se od:

- znakovnog raščlanjivača (engl. tokeniser) koji se bavi prirodnim pitanjima kao što su posebno znakovlje i nizovi znakova, brojevi, simboli, granice riječi i razlike u fontovima, alfabeta i kodnim stranicama;
- morfološkog raščlanjivača koji se bavi odvajanjem leksičkih i gramatičkih morfema i utvrđivanjem značenja u gramatičkim morfemima;

- semantičkog raščlanjivača koji sadržava skup podmodula:
 - jezični razlikovnik (engl. lexical disambiguator), koji iz popisa značenja navedenih u leksikonu odabire prikladni smisao riječi;
 - skriptni graditelj semantičkih ovisnosti koji utvrđuje značenje u rečenicama;
 - skriptni graditelj ovisnosti na razini diskursa koji utvrđuje značenje u cijelom dokumentu;
 - skriptni modul koji utvrđuje stil pisanja;
 - skriptni modul koji upravlja pozadinskim znanjem potrebnim za razumijevanje sadržaja dokumenta, pogotovo referenci i koreferenci;
 - modul koji konstruira profil čitatelja određivanjem parametara situacije čitanja - vremenske, uzročne i druge - prateći odnos čitatelja prema sadržaju dokumenta i utvrđujući namjere čitatelja.

Ulazni podaci analize su dokumenti dodani *online* zbirci dokumenata. Izlazni podaci analize su skupovi struktura znanja - fragmentarno znanje koje se odnosi na unose u leksikonu i onomastikonu - dobiveni iz teksta i profila čitatelja. Izlazni podaci analize ulazni su podaci generiranja.

Sveobuhvatni se generator sastoji od:

- modula specifikacije sadržaja koji utvrđuje značenja u tekstu što će se generirati i sadržava skup podmodula;
- modul specifikacije interaktivne funkcije koji odlučuje o uključivanju ili isključivanju informacija na osnovi interaktivnosti čitatelja;
- modul specifikacije funkcije profila koji odlučuje o uključivanju ili isklju-

čivanju informacija na osnovi pretpostavljenog predznanja čitatelja;

- modul tekstualne strukture koji organizira značenja u rečenice i zavisne rečenice te ih dovodi u pravilan redoslijed;
- modul leksičke selekcije koji riješava semantičke ovisnosti i idiosinkratičke odnose kao što su kolokacije;
- modul odabira sintaktičke strukture;
- morfološki realizator riječi;
- modul koji ujednačava zavisne rečenice i riječi.

Izlazni podaci generiranja su sažeci dodanih dokumenata dobiveni iz procjene skupova strukture znanja nastalih tijekom analize i interaktivnosti čitatelja. Izlazni podaci generiranja odgovaraju stanju prostora zapisa kulturne baštine za specifične čitatelje s obzirom na njihovu interaktivnost i pretpostavljeno predznanje.

Cilj transformativnog procesiranja željene strukture teksta utemeljene na interaktivnosti čitatelja i pretpostavljenog predznanja jest oponašanje čitanja za pisanje sažetaka *online* dokumenata - čitanje istraživanje-do-vraćanja, nakon kojega slijedi čitanje razumijevanje-do-inventivnoga, nakon čega pak slijedi konektivno čitanje (vrijednost-do-značenja) [7] - pomoću:

1. izgradnje semantičke reprezentacije svakog dokumenta;
2. provedbe selekcije, agregacije i generalizacije funkcija na svakoj semantičkoj reprezentaciji tako da se stvore nove reprezentacije, kako utječu na akcije i profile čitatelja;
3. predstavljanja svake nove reprezentacije u prirodnom jeziku. [12]

Općenito, transformativno procesiranje željene strukture teksta aktivno je čitanje za sadržaj informacije i pasivno čitanje za razumijevanje kojim se otkriva diskurzivna mreža *online* zbirke dokumenata i zajednica čitatelja potiče na istraživanje aktivno čitajući za razumijevanje i pasivno čitajući za sadržaj informacije [7].

POSTUPANJE S ISKAZIMA I ČITATELJIMA

Pristup značenju utemeljen na očekivanju, pokušao je u prostoru zapisa o kulturnoj baštini pribaviti i definirati uzorke shematskih iskaza, objašnjavajući njihove kulturne i kognitivne konstrukcije s obzirom na koncept ontologije i pripadnoga fragmentarnog znanja. Uzorci shematskog iskaza proizlaze iz opisnih struktura nanovo identificiranih u diskursu kulturne baštine i daju strukturu osnovne čestice raspona objekata / artefakata, ljudi, događaja, motivacija i njihovih okoliša.

Uzorak shematskog iskaza za povijesni diskurs: *Trijumf nad stranim silama* [19] odigrava se ovako:

1. početna situacija, u kojoj narod živi u miru i nije nikakva prijetnja drugima; ta je situacija prekinuta:
2. početkom nevolja ili agresije stranih sila, što dovodi do:

3. razdoblja krize i velike patnje, koje je:

4. nadjačano trijumfom naroda što je djelovao herojski i samostalno prema stranim silama.

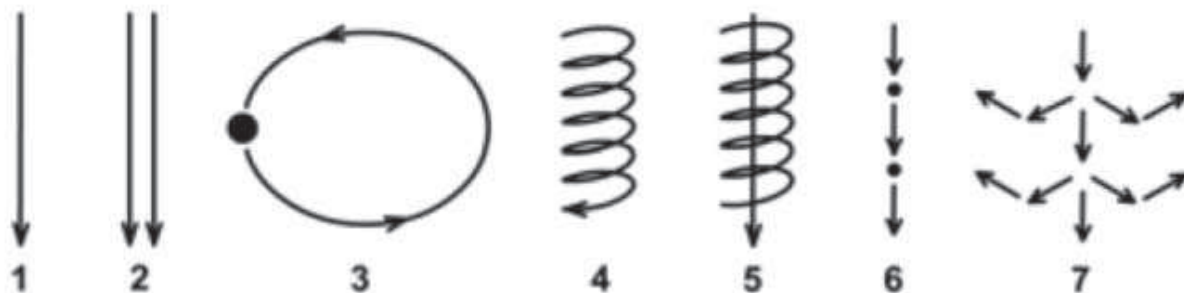
Drugi ustaljeni uzorci shematskog iskaza za povijesni diskurs obuhvaćaju: *Teški izbor*, *Traganje za slobodom* [19] i *Mističnost manifestne sudbine* [11].

Uzorak shematskog iskaza za arheološki diskurs: *Detektivska priča* [18] odigrava se kao:

1. prikaz problema koji se treba riješiti, nakon čega slijedi:
2. razrješenje i rasprava o tzv. tragovima, uz izgradnju napetosti dok se:
3. ne razbije napetost i otkrije se pravo značenje dokaza.

Zapisi koji definiraju uzorke shematskog iskaza kao što je *Trijumf nad stranim silama* ili *Detektivska priča*, koji proizlazi iz linearnih iskaznih struktura (ilustracija 1), donose očekivanja za procesiranje značenja semantičkih i pragmatičnih konstrukcija, procesiranje stila i procesiranje referenci i koreferenci.

Zapisi definiraju generaciju hipertekstualnog pogleda na *online* zbirku dokumenata unutar prostora zapisa kulturne baštine, kompilirajući sažetke dokumenata u alternativne strukture iskaza, ili neinteraktivne strukture kao što su:



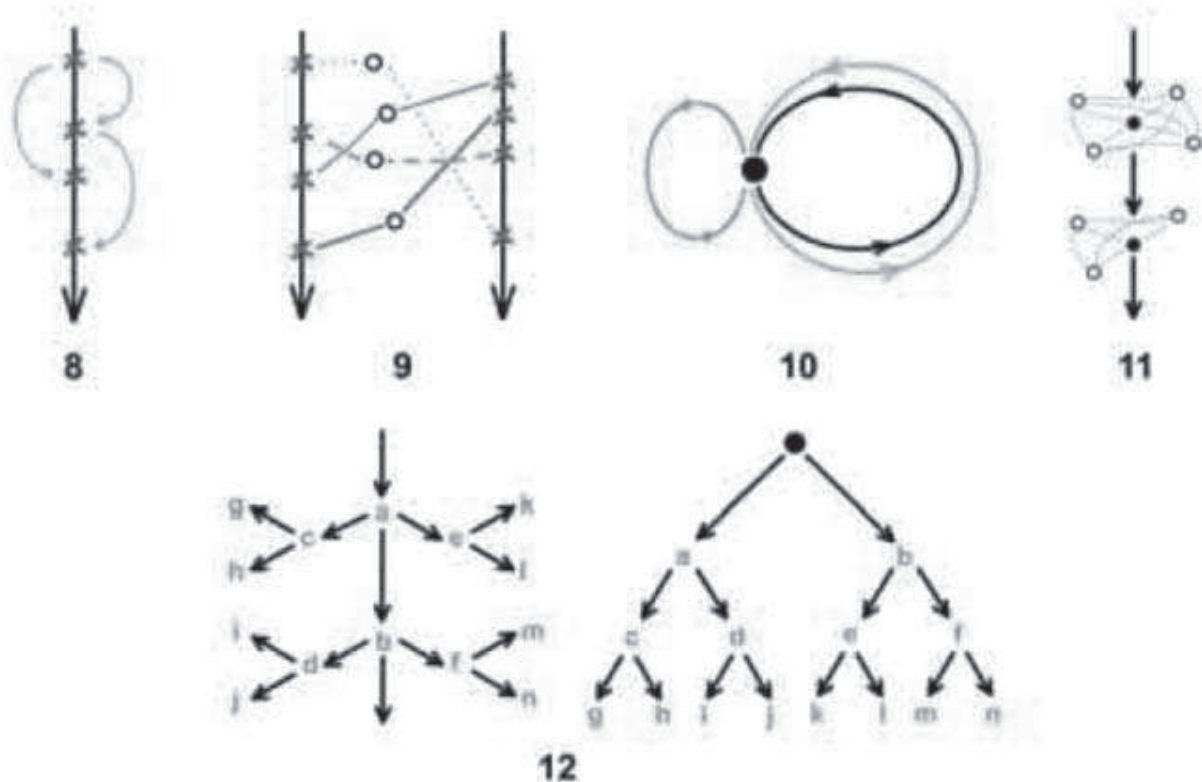
- *paralelne* (ilustracija 2), koje uspoređuju drugačije poglede na uobičajenu temu;
- *kružne* (ilustracija 3), koje postavljaju središnju točku kao početak i kraj u istraživanju teme;
- *spiralne* (ilustracija 4);
- *linearne*, sa *spiralnom medijacijom* (ilustracija 5);
- *aforistične* (ilustracija 6), koje kreću od točke polaska i istražuju temu, zaustavljajući se i odatle započinjući potpuno drugu temu, s tim da se nikad ne vraćaju na prvu;
- *tangencijalne* (ilustracija 7), koje kreću od točke polaska i istražuju temu, često proučavajući povezane teme na tangenti, periodično se vraćajući na prvotnu temu i istražujući je u drugim smjerovima. [18]

Strukture mogu biti i interaktivne, npr.:
 – *usmjerene* (ilustracija 8); propisuju tematski itinerer prema glavnoj temi

- koju definira čitatelj, s mnogim izlaznim i ulaznim točkama;
- *pletene* (ilustracija 9); propisuju različite poglede na uobičajenu temu koju čitatelj definira kao skup čestih koncepata;
- *orbitalne* (ilustracija 10); propisuju skup tematskih istraživanja koja dijele središnju točku onako kako je definira čitatelj;
- *izmiješane* (ilustracija 11); propisuju sekvenciju tematskih istraživanja s međupromjenjivim polaznim točkama koje definira čitatelj;
- *razgranate* (ilustracija 12); propisuju glavnu temu s tangencijskim istraživanjima koje čitatelj određuje kao hijerarhiju točaka odluke.

POKUSNI PRIMJER

Čitatelj želi istražiti koncept fragmentacije identiteta i izabire orijentalizam kao polazišnu točku istraživanja. Proizvede-



ni su nizovi međusobno povezanih sažetaka. Prvi stvara definiciju orijentalizma Edwarda Saida koja datira iz 1978., kako ga citira Močnik. Drugi je sažetak novi Močnikov odlomak koji uspoređuje orijentalizam s balkanizmom. Treći je sažetak odlomak Ditcheva, koji se pomalo odnosi na balkanizam.

Prema Edwardu Saidu, orijentalizam je konceptualna shema koju ideološki prenosi dominacija Zapada nad Orijentom. (...)

Suprotno orijentalizmu, gdje je logika dominacije uvedena kolonijalnom vladavinom, u balkanizmu je ta stalna logika logika samoustrojstva koja sama po sebi stvara nemogućnost shvaćanja sebe u bilo kojem drugom okviru osim u onome u kojemu dominira drugi. [14]

U slučaju poluneovisnih država kao što su one na Balkanu, narodni su glumci kon-

stantno rastrgani, s jedne strane, između potrebe da budu dio planova geopolitičkih sponzora, koji se pridržavaju općih ključnih riječi, normi i narativa, i s druge strane, potrebe da se međusobno razlikuju i da dobiju neovisno postojanje u jedinstvenoj izmišljotini modernosti. [9]

Legenda:

- **izražajna fraza**
- **autor**, natuknica u onomastikonu
- **prikaz**, natuknica u onomastikonu
- **fraza**, epizoda koja se odnosi na **prikaz**; ključni su sastavni izrazi natuknice u leksikonu
- **prikaz**, natuknica u onomastikonu
- **fraza**, epizoda koja se odnosi na **prikaz**, ključni su sastavni izrazi natuknice u leksikonu
- **fraza**, epizoda, ključni su sastavni izrazi natuknice u leksikonu

Sljedeći pseudokod djelomični je primjer korištenja konceptualnoga referentnog modela CIDOC-a [8] kao modela znanja prostora zapisa o kulturnoj baštini.

Natuknica u onomastikonu

|orijentalizam|

Instanca dizajna ili procesa E29

P69 povezuje se s:

Natuknica u onomastikonu

|balkanizam|

Instanca dizajna ili procesa E29

P103 bio je namijenjen za:

Natuknica u onomastikonu

|prikaz|

Instanca E55 tipa

P137 dokazan je:

Epizoda

|konceptualna shema koja ideološki prenosi dominaciju Zapada nad Orijentom|

Instanca E33 lingvističkog objekta

P2 je tipa:

Natuknica u onomastikonu

lfrazal
 Instanca E55 tipa
 P72 ima jezik:
 Natuknica u onomastikonu
 lengleskil
 Instanca jezika E65
 P94 stvorio je:
 Natuknica u leksikonu
 ldefinicijal
 Instanca događaja stvaranja E65
 P2 je tipa:
 Natuknica u onomastikonu
 lriječl
 Instanca E55 tipa
 P4 ima vremenski raspon:
 Natuknica u leksikonu
 lpublikacijal
 Instanca E52 vremenskog raspona
 P2 ima tip:
 Natuknica u onomastikonu
 lriječl
 Instanca E55 tipa
 P78 identificira se prema:
 Epizoda
 l1978l
 Instanca E50 datuma
 P14 izvodi:
 Natuknica u onomastikonu
 lEdward Saidl
 Instanca E21 osobe
 P106 čini:
 Natuknica u leksikonu
 lkonceptl
 Instanca E33 lingvistički objekt
 P2 je tipa:
 Natuknica u onomastikonu
 lriječl
 Instanca E55 tipa
 Natuknica u leksikonu
 lshemal
 Instanca E33 lingvistički objekt
 P2 je tipa:
 Natuknica u onomastikonu

riječl
 Instanca E55 tipa
 Natuknica u leksikonu
 lideologijal
 Instanca E33 lingvistički objekt
 P2 je tipa:
 Natuknica u onomastikonu
 riječl
 Instanca E55 tipa
 Natuknica u leksikonu
 lprenošenjel
 Instanca E33 lingvistički objekt
 P2 je tipa:
 Natuknica u onomastikonu
 riječl
 Instanca E55 tipa
 Natuknica u leksikonu
 ldominacijal
 Instanca E33 lingvistički objekt
 P2 je tipa:
 Natuknica u onomastikonu
 riječl
 Instanca E55 tipa
 P129, riječ je o:
 Natuknica u leksikonu
 lkomandal
 Instanca E33 lingvistički objekt
 P2 ima tip:
 Natuknica u onomastikonu
 riječl
 Instanca E55 tip
 Natuknica u leksikonu
 lkontrolal
 Instanca E33 lingvistički objekt
 P2 ima tip:
 Natuknica u onomastikonu
 riječl
 Instanca E55 tip
 Natuknica u onomastikonu
 lZapadl
 Instanca E53 mjesto
 P87 identificira se prema:
 Natuknica u onomastikonu

|Zapadna Europa|

Instanca E44: naziv mjesta

P70 dokumentira se u:

Natuknica u onomastikonu

|UNESCO-ov tezaurs|

Instanca E32, autoritativni dokument

P67 odnosi se na:

Epizoda

|databases.unesco.org/thesaurus/|

Instanca E31 dokument

Natuknica u onomastikonu

|Orijent|

Instanca E53 mjesto

P87 identificira se prema:

Natuknica u onomastikonu

|Bliski istok|

Instanca E44: naziv mjesta

P70 dokumentira se u:

Natuknica u onomastikonu

|Gettyjev tezaurs geografskih imena|

Instanca E32: autoritativni dokument

P67 odnosi se na:

Epizoda

|www.getty.edu/research/conducting_research/vocabularies/tgn/|

Instanca E31: dokument

P102 ima naslov:

Epizoda

|Bliski istok (opća regija)|

Instanca E35: naslov

P1 identificira se prema:

Epizoda

|7001526|

Instanca E42

identifikatora objekta

P106 čini dio:

Epizoda

|Balkan kao element u ideološkim mehanizmima|

Instanca E73: informacijski objekt

P128 znači:

Epizoda

|Balkan kao metafora|

Instanca E84: nosilac informacije

ZAHVALA

Istraživanje je proveo prvi autor djelomično ispunjujući zahtjev za doktoratom na Egejskom sveučilištu i Université Lumière-Lyon 2, slijedeći zajedničku tezu između dvije institucije (*cotutelle de thèse*). Ovaj su istraživački rad financirali Ministarstvo nacionalnog obrazovanja i religijskih poslova Grčke i Europska unija.

LITERATURA

- Althusser, L. (1971), *Lenin and Philosophy and Other Essays*, London: New Left Books.
- Barthes, R. (1977), *Image, Music, Text*, New York NY: Hill & Wang.
- De Beer, C. S. (1996), *Ontology of the Book: Text, Intertext, Hypertext*, in *South African Journal of Library and Information Science* 15 (6): 124-129.
- Borgman, C. (2003), *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World*, Cambridge MA: MIT Press.
- Buckland, M. K. (1991), *Information as Thing*, in *Journal of the American Society for Information Science* 42 (5): 351-360.
- Carbonell, J. G. (1983), *Learning by Analogy: Formulating and Generalizing Plans from Past Experience*, in *Machine Learning: An Artificial Intelligence Approach*, Michalski, R. S., Carbonell, J.G. and Mitchell, T.M. (eds.), volume 1: 137-161.
- Cremmins, E. T. (1996), *The Art of Abstracting*, Arlington: Information Resources Press.
- Crofts, N., Doerr, M., Gill, T., Stead, S. and Stiff, M. (eds.) (2004), *Definition of the CIDOC Conceptual Reference Model and Crossreference Manual v4.0*, ICOM/CIDOC CRM Special Interest Group.
- Ditchev, I. (2002), *The Eros of Identity*, in *Balkan as Metaphor: Between Globalization and Fragmentation*, Bjelić, D. and Savić, O. (eds.), Cambridge MA: MIT Press: 235-250.
- Kintsch, W. and Van Dijk, T. A. (1978), *Toward a Model of Text Comprehension and Production*, in *Psychological Review* 85 (5): 363-394.
- Lowenthal, D. (1985), *The Past is Another Country*, Cambridge: Cambridge University Press.

- Mani, I. (2001), *Automatic Summarization*, Amsterdam: John Benjamins.
- Mink, L.O. (1978), *Narrative Form as a Cognitive Instrument*, in *The Writing of History: Literary Form and Historical Understanding*, Canary, R.H. and Kozicki, H. (eds.), Madison WI: University of Wisconsin Press: 129-149.
- Močnik, R. (2002), *The Balkans as an Element in Ideological Mechanisms*, in *Balkan as Metaphor: Between Globalization and Fragmentation*, Bjelić, D. and Savić, O. (eds.), Cambridge MA: MIT Press: 79-115.
- Nietzsche, F. (1881), *Daybreak*, Cambridge: Cambridge University Press.
- Nirenburg, S. and Raskin, V. (2004), *Ontological Semantics*, Cambridge MA: MIT Press.
- Riesbeck, C. and Schank, R. (1989), *Inside Case-based Reasoning*, Hillsdale NJ: Lawrence Erlbaum.
- Tilley, C. (1990), *On Modernity and Archaeological Discourse*, in *Archaeology After Structuralism*, Bapty, I. and Yates, T. (eds.), London: Routledge. Also published online at archaeology.kiev.ua/meta/tilley.html
- Wertsch, J. V. (1998), *Mind as Action*, New York NY: Oxford University Press.

ONTOLOGY DRIVEN ACCESS TO MUSEUM INFORMATION

Cultural Heritage documents deal with objects/artifacts and the people that created, owned, used, or (re)discovered them. Their fates are intertwined in unique and complex stories forming a cumulative body of knowledge, often fragmented across large online document collections. While our collective memory has explicitly documented these stories, the heterogeneity and distribution of the available sources creates islands of information that can only be implicitly connected by a limited, expert audience. This paper presents a semantically consistent framework for the online presence of Cultural Heritage document collections, set upon a participatory centre stage and supported by a shared knowledge model, the CIDOC CRM ontology. In this framework, Cultural Heritage document contributors are peer-to-peer network nodes that benefit from: a schema-based network topology; a transparent, self-organised, self-optimised network infrastructure; knowledge-rich document processing modu-

les which analyse and classify each contribution, capture the notion of time and the unfolding of events spanning a single or multiple documents, and establish meaning connectivity over the entire collection. Overall, this framework assists a scholarly audience with the exploration of online distributed Cultural Heritage document collections, and offers an informed tap into the collective memory scattered therein.

Cultural Heritage documents deal with objects/artifacts and the people that created, owned, used, or (re)discovered them. Their fates are intertwined in unique and complex stories forming a cumulative body of knowledge, often fragmented across online document collections. While our collective memory has explicitly documented these stories, the heterogeneity of the available sources creates

islands of information that can only be implicitly connected by a limited, expert audience.

This paper presents a semantically consistent framework for the online presence of Cultural Heritage document collections, set upon a participatory centre stage and supported by a shared knowledge model. In this framework, Cultural Heritage document contributors benefit from knowledge-rich document processing modules which analyse and classify each contribution, capture the notion of time and the unfolding of events spanning a single or multiple documents, and establish meaning connectivity over the entire collection. Overall, this framework assists a scholarly audience with the exploration of online Cultural Heritage document collections, and offers an informed tap into the collective memory scattered therein.