

Mediji i demokracija: baza podataka, dubinsko pretraživanje, izravni pristup – digitalizacija novinskog arhiva i organizacija znanja u bazu podataka

Željko Hodonj

Hrvatska izvještajna novinska agencija HINA
HINA informacijska baza / HIB
Informacijsko dokumentacijski centar / IDC
Marulićev trg 16, Zagreb, Hrvatska
zeljko.hodonj@hina.hr

Lana Žaja

Hrvatska izvještajna novinska agencija HINA
HINA informacijska baza / HIB
Informacijsko dokumentacijski centar / IDC
Marulićev trg 16, Zagreb, Hrvatska
lana.zaja@hina.hr

Sažetak

HINA informacijska baza jedan je od alata kojim se služi HINA u vlastitom organiziranju, HIB Informacijsko dokumentacijski centar povezuje interes korisnika sadržaja što ih HINA generira kao svoj proizvod osiguravajući dostupnost javnog znanja. HIB/IDC suvremenim načinom osigurava zaštitu i javnu upotrebu hrvatske kulturne baštine zabilježene u tiskanim i AV medijima. HIB IDC podržava umreženje s bazama u okruženju. Godišnje će obraditi milijun dokumenata, tijekom 2008. godine razvit će automatsku kategorizaciju sustava.

Glavne riječi: Hina, baština, digitalizacija, Hrvatski sabor, baza podataka

Uvod

Hrvatsku izvještajnu novinsku agenciju osnovala je Republika Hrvatska s ciljem da prikupi i pošalje informacije utvrđene na činjenicama o zbivanjima u RH i u svijetu za potrebe medija i drugih sudionika javnog života. HINA djeluje po načelima neovisnog, nepristranog i profesionalnog novinsko-agencijskog izvještavanja. Svakodnevno Hina priprema dvadesetak općih i specijaliziranih servisa vijesti u kojima se dnevno na hrvatskom i engleskom jeziku objavljuje oko 400 tekstova s više od 150.000 riječi. Hina pruža i multimedijske usluge, ima vlastiti

foto (FaH) i audio servis te elektroničku bazu podataka (EVA). Tijekom 24 sata Hinine vijesti prima oko 500 korisnika u zemlji i svijetu. U agenciji je ukupno 180 stalno zaposlenih, od toga 130 novinara i urednika. Povremeno za Hina-u izvještava još šezdesetak vanjskih suradnika-dopisnika iz zemlje i svijeta. Hina je od 1993. članica Europske udruge novinskih agencija EANA čije su članice su nacionalne novinske agencije iz 30 europskih zemalja.

Hrvatska izvještajna novinska agencija – HINA, od osnutka 1990. godine, opredijelila se da će djelovati pomoću opće pristupnog interaktivnog sustava baze podataka. U samom početku to je značilo:

1. Stručno, sa stajališta agencijskog novinarsko uređivačkog posla, trebalo je svoju svakodnevnu aktivnost podvrgnuti normama obilježavanja / atribucije svakog novonastalog sadržaja: tekstualnog, vizualnog ili auditivnog.
2. Stvorena je EVA Hinina baza podataka s današnjih 1,3 milijuna dokumenata – pokazana je dimenzija zadatka pohrane dokumenata već na razini tekstualnih formi, njihove organizacije i pretraživanja. Trebalo je odgovoriti na pitanja dostupnosti sadržaja za krajnje korisnike. Bio je to veliki izazov novinarima, urednicima, IT timu.

Bilo je potrebno primijeniti opće prihvaćene standarde za taj posao – IPTC kategorizaciju¹.

EVA je postala sa stajališta zaštite baštine, neprocjenjiv izvor podataka / dokumenata o stvaranju samostalne hrvatske države, o događajima i osobama koji u svakodnevnom agencijskom djelovanju predstavljaju sadržajne silnice novinarskih tekstova.

U ovom ćemo tekstu predočiti zahtjeve koji su utjecali da se na jednom mjestu dogodi sinergija odgovora na pitanja o zaštiti baštine, podržavanju demokracije i dostupnosti organiziranog javnog znanja.

Na prijedlog Vlade RH 2006. godine, Hrvatski sabor je potvrdio zaštitu najveće novinske baze podataka u papirnatom i elektronskom formatu – Vjesnikovu novinsku dokumentaciju / VND online. Tako su stvoreni preduvjeti da se VND

¹ IPTC International Press Telecommunications Council – IPTC međunarodna organizacija novinskih agencija koja je osnovana sredinom prošlog stoljeća s ciljem unapređenja i profesionalizacije međusobne razmjene (tehničkih i novinskih) podataka. U tom kontekstu započelo je korištenje IPTC dokumenata koji su pomagali jednostavnijoj klasifikaciji i organizaciji. Prvenstveno je taj oblik metapodataka bio rješenje za upravljanje slikovnim dokumentima. Kao što je Eurovoc organiziran prema potrebama rukovanja spisima Europske unije, i to mu je odredilo strukturu i metodu označavanja, tako je i IPTC imao svoju prvotnu namjenu. Takav oblik označavanja metapodatcima zahtijevao je točnost i jednoznačnost područja. Među pojmovima međuodnos je bio određen širi/uži pojam. No, gledajući prema osnovnoj namjeni IPTC kategorizacija i njezina osnovna "područja" bliža su klasifikaciji HIB/IDC (odnosi se na razdoblje prije povezivanja s HINA-om kada je to još bila baza VND online) nego klasifikaciji Eurovoca jer je polazišna građa za koju se klasifikacija organizira sadržaj javnih glasila, novinski tekst. (Danijela Miklec)

proglasi razvojnim projektom HINA-e; te da se financira zaštita i razvoj baze podataka sredstvima državnog proračuna.²

Tako je projekt digitalizacije zaštićenog novinskog arhiva Vjesnikove novinske dokumentacije s dvanaest milijuna dokumenata i njegova organizacija u javno dostupnu bazu znanja postao je 2006./2007. godine nacionalni projekt i započela je njegova integracija u sustav Hrvatske izvještajne novinske agencije.

Postupajući u skladu s takvim opredjeljenjima Hrvatska izvještajna novinska agencija odmah je uputila Ministarstvu kulture RH projektnu ideju o povezivanju Hinine informacijske baze EVA i Vjesnikove novinske dokumentacije VND.

Osnova za pokretanje projekta bila je kako sljedeća:

Članovi Odbora za medije Hrvatskog sabora inicirali su u lipnju 2006. objedinjavanje Hinine baze podataka Eva i Vjesnikove novinske dokumentacije kako bi ta vrijedna arhiva mogla biti na korist državi i drugim korisnicima.

U saborskoj raspravi povodom Izvješća o radu Upravnog vijeća Hine za 2005/2006. godinu zatraženo je da Vlada razmotri mogućnost povezivanja HINA baza EVA i VND.

Mišljenje je Hrvatskog državnog arhiva iz siječnja 2006. g. da se VND, budući da Vjesnik d.d. to nije u mogućnosti, povjeri na skrb i korištenje institucijama poput HINA-e, HIDRA-e ili druge srodne ustanove.

Kao ciljevi projekta određeni su očuvanje primarne, recentne dokumentacijsko-informacijske vrijednosti gradiva Hinine baze EVA i VND, kao baze podataka i kao povijesno-kulturnog dobra; potpuna digitalizacija baze podataka; stvaranje i razvoj novih, samostalnih i međusobno povezanih informacijskih usluga za različite skupine korisnika (media monitoring, on line novinska dokumentacija i posebni projekti); komercijalno iskorištavanje gradiva kako bi se budući razvoj i održavanje u što većoj mjeri samofinancirali.

Takav kontekst djelatne zaštite kulturnog dobra predstavlja izuzetno kvalitetan i odlučan odgovor na pitanja o modelima zaštite baštine.³ Suvremene okolnosti zajedničkog života svojim su posebnostima tražile da zajedno s drugima odgovorimo na izazove vremena u kojem živimo.

² Krajem 2006. godine u Zagrebu je boravio prof. emeritus Robert Hayes i s nama je proveo desetak sati svog dragocjenog vremena. Potpuno besplatno je vrednovao naše rezultate dajući nam snažnu podršku za nastavak rada i integraciju u sustav HINA-e. O tome je iscrpno razgovarao s Fjedorom Polojcem, pomoćnikom ravnatelja HINA-e, koji je bio zadužen za provođenje projekta povezivanja i s Darkom Gulijom, voditeljem IT tima HINA-e. Vjerujemo da ćemo se prof. Robertu Hayesu uskoro moći dolično zahvaliti.

³ Državnim proračunom RH za 2007. godinu osigurana su sredstava za razvoj projekta ali je zbog administrativne procedure tijekom ustrojavanja uprave HINA-e, postupak provedbe projekta HIB/IDC započeo tek 1. lipnja 2007. umjesto 1. siječnja 2007. godine.

Historijski je trend⁴ da se dominantne funkcije i procesi u informacijskom dobu sve više organiziraju oko mreža. Kultura se dubinski transformira; nakon tisućljeća dominacije prirode nad kulturom i stoljeća dominacije kulture nad prirodom, ulazimo u doba dominacije kulture nad kulturom, gdje priroda postaje kulturni (re) konstrukt. Teoretičari poput Manuela Castellsa, sugeriraju da su čimbenici zaduženi za upravljanje našom tranzicijom u informacijsko doba državne institucije te da nam je u toj tranziciji potrebno političko zastupanje, demokracija u kojoj svi sudjeluju, procesi iznalaženja konsenzusa...

Trebamo imati na umu i to da je razvoj tehnologije rezultat triju društvenih procesa: potrebe ekonomije za većom fleksibilnošću, koja je u konačnici dovela i do uspona elektroničkog poslovanja; potrebe društva u kojem vladaju vrijednosti individualnih sloboda koje su omogućile trijumf umreženog individualizma – stvaranje mrežnih zajednica prema individualnim afinitetima; a (tek tada) i napretka u računarstvu i telekomunikacijama koji se dogodio krajem milenija.

Znanje – novac i moć

U takvom kontekstu trebalo je izabrati način pomoću kojeg će nam vlastiti dokumenti biti dostupni, organizirani u baze znanja, dostupni u svakom trenutku nama i drugima. Osobito nam je bilo važno omogućiti uvid u sadržaje koji su "prekriveni" prvom razinom čitanja.

Rečeni navod potkrijepit ćemo primjerom svoje prakse organizacije baze podataka. Primjer je krajnje jednostavan: pomoću javno dostupnih izvora / baza podataka možemo otkrivati podatke i sadržaje, odnose i procese koji su izvan standardne pozornosti. Primjer pokazuje potencijalni smjer za provjeru kolika je dostupnost i organiziranost baštine, odnosno koliko su nam dostupni kulturni artefakti vremena i to oni koji su potrebni znanju, sudjelovanju i odlučivanju. Izabrali smo skup podataka koji upozoravaju na informacijski paradoks što se u opasnoj mjeri može preobraziti u grotesku. Riječ je o projektu koji je od VND naručio Ured glasnogovornice Ministarstva pravosuđa RH 2004. godine s namjerom da dobije sliku o recepciji i percepciji reforme pravosuđa u Hrvatskoj.

Od tada do danas ta posebno projektirana baza podataka broji više od šezdeset tisuća dokumenata/članaka iz novina. Početkom 2006. godine projekt je predefiniran u *custom made product*, a trebalo je provjeriti koliko i kako hrvatska tiskana glasila pišu o korupciji, sukobu interesa, odnosno kako tiskana glasila prate i interpretiraju nastojanja hrvatske izvršne vlasti da obračuna s korupcijom na svim razinama hrvatskog društva.

Od 1. siječnja 2006. do 31. prosinca 2006. godine u bazi podataka VND online, područje Pravosuđe, sudstvo i kriminal; obrađen je 22.791 dokument/tekst iz 16 hrvatskih novina (Jutarnji list, Novi list, Večernji list, Slobodna Dalmacija, Glas

⁴ Manuel Castells, *Informacijsko doba*, Golden Marketing, Zagreb 2000/2002, Internet galaksija, Jesenski i Turk, Zagreb; 2003.

Slavonije, Vjesnik, Poslovni dnevnik, Globus, Feral, Nacional, Arena, Sportske novosti, Banka, Privredni vjesnik, Glorija, Vijenac).

Obrađeni su svi članci koji su imali relaciju prema temama: mito i korupcija, zloraba položaja i ovlasti, sukob interesa.

Usred opće kampanje za borbu protiv korupcije, za tzv. nultu točku tolerantnosti na korupciju, obrada izabranih glasila pokazala je da od ukupnog broja objavljenih tekstova koji su imali sadržajne veze s područjem Pravosuđe, sudstvo i kriminal, samo je 7 posto, odnosno u apsolutnom broju 1633 članka, imalo dodirnih sadržaja s pojmom ili konotaciju prema korupciji. Usporedbe radi, broj tekstova o korupciji od početka devedesetih do 2006. godine, s razine stotinjak napisa godišnje, porastao je do broja koji sada spominjemo. Tekstovi koji utječu na formiranje stavova su: kolumne (33), komentari (44), tzv. pisma (17); naglašavaju moralne principe i opisuju desetak istih događaja ili osoba. U toj godini novinari istraživači nisu objavili toliko tekstova koliko kolumnisti i komentatori. Možda je lakše pisati i objavljivati tekstove o principima nego istraživati stvarnost. Izabrana tiskana glasila zauzimaju približno 90 posto novinskih naklada. Informacije o korupciji najčešće su iz “druge ruke”.

U to je doba Hrvatska dobila ozbiljna upozorenja Europske komisije o korupciji u hrvatskom društvu, a poruke su bile potkrijepljene: anketama građana, istraživanja stavova građana i tome slično.

Za sada su bez odgovora pitanja o tome kako je moguće da ocjene o općoj korumpiranosti društva, što izravno dovodi u pitanje demokraciju *per se*, nemaju uporište u sadržajima hrvatskih tiskanih glasila. Hipotetički, možda je riječ o tome da – korupcija nema takvog maha i utjecaja koji joj se pripisuje?

Ili – mediji su korumpirani i zato su takve teme izvan njihova interesa?

Ili nešto treće?

Odgovor na takva pitanja moguć je temeljitim analitičkim i vrijednosnim uvidom u baze podataka, dubinskim pretraživanjem brojnih relacija.

Sliku jednog društva s njegovim kulturnim artefaktima, među kojima je i korupcija, treba promotriti što cjelovitije.

Klasične teorije demokracije pretpostavljaju da je društvena aktivnost građanska dužnost, a biračko tijelo ne bi trebalo biti samo “informirano”, nego bi trebalo kontinuirano procjenjivati i tako kreirati pučki interes, stvarati svojevrsnu samoinicijaciju za participaciju u društvenim poslovima. Biračko bi tijelo, s te točke gledišta, trebalo sudjelovati promišljeno i tolerantno prema različitim gledištima držeći se norme da je javni interes povrh pojedinačnog.

Praksa je pokazala da je takva omnikompetencija iznimka, a ne pravilo. Većinom je građanska aktivnost posljedica incidentnih poticaja ili ih – mobiliziraju mediji.

Javna glasila, mediji koji se bave političkim / javnim temama, ne samo da ocrtavaju / opisuju političko okruženje; oni u sve većoj mjeri sami upravo predstavljaju / jesu – političko okruženje.

Primijenjeno na pojedinačnog građanina – rastući dio javnog djelovanja zbiva se pod utjecajem medija.

U tom je kontekstu moćni utjecaj novih medijskih formata: Interneta i World Wide Weba; radija (govorni) i televizijskih kontakt programa, ali i raznovrsnih zabavnih programa koji pozivaju (političke i nepolitičke) goste u zabavnu raspravu o političkim pitanjima (poput zagrebačke PIRAMIDE, by Merlić & Ogresta)

Kroz te elektronske sustave građani su izloženi političkim informacijama, politički im se lideri čine dohvatnima, uključuje ih se u javne rasprave o važnim pitanjima, a političari su izloženi izravnom javnom propitivanju.

Sumnjičavci upozoravaju da novi mediji tako bespovratno utječu na žrtvovanje izravne, osobne komunikacije. Promatrano s te točke gledišta – tako se stvaraju uvjeti za manipulaciju i podržava se društvena izoliranosti. Ta skepsa nije neuvjerljiva ali uvjerljivo je i to da interaktivni mediji u najmanju ruku mogu potaknuti pojedinca / građanina na samoaktivnost.

Zato je HINA u svom strateškom opredjeljenju odlučila da sve svoje sadržaje i sadržaje drugih; i u dogovoru s njima; generira kao baze podataka "podržavajući umreženje svakog pojedinog građanina Hrvatske. Glavni je cilj osigurati dostupnost podataka iz javnih glasila pomažući osviještenu obaviještenost o izabranim razinama javnog znanja u hrvatskom društvu".

Zahvaljujući dosljednoj podršci Vlade RH prema poštivanju takvih kriterija djelovanja HINA-e, stvoreni su uvjeti za razvoj HINA informacijske baze / HIB Informacijsko dokumentacijskog centra.

Stoga je važno znati što je to motiviralo odluke Vlade RH i Hrvatskog sabora da 2006. godine donesu odluku kojom je najveća baza novinskih podataka u Hrvatskoj poznata kao Vjesnikova novinska dokumentacija, VND on line, zaštićena kao prvorazredno kulturno dobro.

Citiramo iz pismohrane Hrvatskog sabora⁵:

*"Vjesnikova novinska dokumentacija – VND online" novinski je dokumentacijski centar organiziran u računalnom mediju i jedini je u Hrvatskoj. Nastao je na temeljima Vjesnikove novinske dokumentacije koja je ustrojena 1964. godine.*⁶

Od tada do 2005. godine kreirano je više od 6.000 zbirki dokumenata / izrezaka, članaka / tekstova iz novina s približno 12 milijuna pojedinačnih zapisa, a

⁵ Interpelacija o odgovornosti Vlade RH za stanje društva Vjesnik d.d., 20. sjednica Hrvatskog sabora 1. lipnja 2006. godine.

⁶ Osnutak novinskog arhiva nekadašnje Vjesnikove kuće izravno se vezuje uz ime Bože Novaka, glavnog direktora NIŠPRO "Vjesnik" 1964. godine. Od tada pa sve do osnutka HINA informacijske baze / HIB Informacijsko dokumentacijski centar, bazu su održali dokumentalisti: Anđelka Grubišić Čabo, Vlado Kvitek, Deša Kolar, Jasminka Cesar Miličević, Božica Jurčić, Luca Jelić; i dosljedno inzistiranje Hrvatskog državnog arhiva da građa VND bude zaštićena kao baština iznimne spomeničke vrijednosti.

uza sve to je i posebna zbirka članaka za približno 100.000 domaćih i inozemnih javnih osoba. Baza podataka u izrascima članaka, mjereno smještajem na policama, iznosi približno 480 dužnih metara. Baza podataka u računalnom mediju⁷ broji 600.000 dokumenata.

VND je projektirana za pohranu i organizaciju poveznica starijih i novih podataka..

Tako su ostvarene, kao jedine u Hrvatskoj, posebne eksportne baze podataka tiskanih glasila na DVD-nosačima.

Baza podataka VND online u svom je papirnatom dijelu zaštićena norma-nadzorom Državnog arhiva. (...) Baze podataka su dio organizacije sustava i iznimno su važne u nacionalnoj mreži kulture. U svom općem servisu digitalne baze podataka razvijaju i poslovnu djelatnost povezivanjem agencijskih sadržaj i media monitoringa, a ustrojavanjem posebnih baza podataka izravna su podrška razvoju i postojanju novih medija.

Vlada Republike Hrvatske izrazila je Nadzornom odboru i Upravi Društva izričito protivljenje da se građa i djelatnosti Vjesnikove novinske dokumentacije (dalje VND) prodju (privatiziraju) jer je riječ o javnom arhivskom građivu. Istog mišljenja je Hrvatski državni arhiv koji je 17. siječnja 2006. godine uputio pismo sličnog sadržaja Ministarstvu kulture.

Istodobno, ako Vjesnik d.d. ne može dalje skrbiti o djelatnosti i građi VND, primjereno je da o navedenom skrbe i da ga koriste institucije koje se primarno bave informacijsko-dokumentacijskim poslovima. Primjerice: Hrvatska izvještajna novinska agencija (HINA)... (...)

Vjesnikova novinska dokumentacija od 26. studenog 1991. godine u režimu je zaštite Državnog arhiva u Zagrebu (Rješenje Historijski arhiv u Zagrebu, Opatička 29. od 26. studenog 1991.): "Građa je od izuzetne vrijednosti u prvom redu za praćenje i analizu događaja kako za period RAT U HRVATSKOJ tako i za sva zbivanja od 1945. do danas u našoj Republici i šire."

Povijesni arhiv u Zagrebu svojim rješenjem 22. srpnja 1994. godine na temelju Zakona o zaštiti arhivske građe donio je rješenje kojim se nalaže: "Sačuvati cjelovitu dokumentaciju te se pojedini dijelovi ne smiju izdvajati."

Istodobno se ističe: "Zbog izuzetne vrijednosti građe potrebno je što hitnije riješiti status stalnog financiranja ove službe." Državni arhiv u Zagrebu svojim Zapisnikom od 11. lipnja 2001. i 8. svibnja 2002. ističe: "Budući da ova služba

⁷ Željko Hodonj je autor prijedloga koncepta organizacije baze podataka i njegovo prihvaćanje od strane Uprave Vjesnik d.d. bio je preduvjet za dolazak (Ž. H.) na mjesto zamjenika glavnog urednika dnevnog lista "Vjesnik" od 1. siječnja 2001. godine. Nakon ostavke na mjesto zamjenika glavnog urednika, Željko Hodonj je Upravi "Vjesnik" d.d. 10. listopada 2002. godine predložio "Projekt pokretanja poduzeća – Dokumentacija Vjesnik; poduzeća za proizvodnju i trgovinu informacijama, dokumentaciju i analizu, stvaranje baza podataka za javnu i posebnu upotrebu". Projekt je prihvaćen u travnju 2003., 1. svibnja 2003. Projekt je formaliziran, a nakon parlamentarnih izbora 2003. godine Uprava Vjesnik d.d. donijela odluku da se nastavi s razvojem Projekta.

vrši vrlo značajan posao, nužno bi bilo da se organizacija i tehnološka podrška modernizira."

"Odgovarajući na zahtjeve Državnog arhiva, Vjesnikova novinska dokumentacija/VND provela je naloge iz Rješenja i Zapisnika. Sadašnji ustroj i organizacija obrade građe nastali su tijekom 2004. i 2005. godine. (...) Baza je kreirana za pohranu, organizaciju i pretraživanje 360.000 članaka godišnje. Sustav je ustrojen kao pouzdana informacijska paradigma za pretraživanje i razvrstavanje dokumenata na zahtjev korisnika dokumentacijske građe. (...) U 2006. Uprava Društva ukinula je sva investiranja u daljnji razvoj te djelatnosti i trenutačno provodi postupak izrade popisa građe prema zahtjevu Vlade Republike Hrvatske i Hrvatskog državnog arhiva. Treba napomenuti da spomenuta građa postoji u Nacionalnoj i sveučilišnoj biblioteci, ali samo u kompletima novina, a ne razvrstana prema interesnim područjima. Nitko u Hrvatskoj trenutačno ne razvija takvu interaktivnu bazu podataka tiskanih glasila."

Početak projekta

Sve je počelo naoko nedužno.

Početakom 2001. godine, tijekom 2002. i 2003. godine kao i nekoliko desetljeća ranije mala skupina dokumentalista u Vjesnikovoj novinskoj dokumentaciji skarama je izrezivala izabrane novinske članke.

Zatečeno je približno šest tisuća registratora u kojima su dokumenti grupirani u područja.

Početakom 2004. godine odredili smo prvu razinu našega zadatka: stvoriti pretraživu arhivu u elektronskom formatu s pretraživim sadržajem tekstova i atributima, dostupnu u svako vrijeme i na svakom mjestu, koju je moguće kreirati u što kraćem vremenu uz utrošak što manje ljudi, opreme i novaca.

Pri tom je trebalo imati na umu da obrađujemo devet (9) dnevnih novina, petnaest (15) tjednika/revija – novina. Trebalo je dnevno obraditi približno 500 stranica novina, ali i 700 članaka/dokumenata. Pošli smo od uvjerenja kako stvaramo novinski dokumentacijski centar koji organizira poslove prikupljanja, selekcije, obrade i pohrane građe u elektronsku bazu podataka. Bilo je potrebno izabrati postupke za pretraživanje i razvrstavanje dokumenata kao odgovor na posebne korisničke zahtjeve posebnih sadržaja. Dokumenti su grupirani prema područjima znanja.

Zajedničkim dogovorom definirani su osnovni deskriptivni elementi teksta novinskih članaka i standardizirane su ključne riječi koje su bile potrebne da se linearni tekst pohrani kao strukturirani **hipertekst** u elektroničku bazu podataka. Hipertekst / tekst nema ograničenja linearnosti, nego je tekst koji u sebi sadrži veze na druge tekstove (članke) ili druge dijelove istog teksta. Takvi dijelovi novinskih članaka su čvorovi ili nodes. Čvor obično sadrži jednu ideju, tj. pojam, a može sadržavati i podatak o vrsti informacije (npr. "sukob interesa" je standardizirana ključna riječ i predstavlja jedan čvor, odnosno vrstu informacije u području Unutarnja politika). Skup takvih čvorova čini novinsku elektronsku

bazu podataka, a čvorovi i njihove veze, unutarnje i vanjske, mrežnu strukturu između novinskih članaka. Unutarnje veze članka se nalaze unutar istog teksta, a vanjske veze članka su poveznice na neki drugi članak.

Ključne riječi usuglašavane su radom na posebnim projektima. Svaki dokumentalist je unutar svog područja odredio ključne riječi od relevantnog značenja za zadano područje ovisno o sadržaju novinskog članka. Teme unutar stručnih područja nisu razvrstane hijerarhijski, nego abecednim slijedom zbog bržeg snalaženja i pretrage novinskih članaka koja je potrebna pri korisničkim zahtjevima i tematskim upitima novinara kojima potrebne informacije uvijek moraju biti pripremljene i predane u vrlo kratkom roku.

Strukturiranje tekstova u bazi podataka uspostavljeno je korelacijskim odnosima pomoću tezaurusa, odnosa ekvivalancije, homonima, sinonima, hijerarhijskih i asocijativnih odnosa.

Struktura tezaurusa novinskih članaka se sastoji od odnosa među standardiziranim pojmovima ključnih riječi koji su ujedno pomoćno sredstvo za indeksiranje. Tezaurusom, tj. kontroliranim rječnikom za označavanje teksta, olakšava se pretraživanje relevantnih novinskih članaka. Prednosti tezaurusa u VND on-line bazi su što on obilježava područja znanja koja želimo pretražiti, omogućava generičku povezanost pojmova preko dokumentalističke atribucije i rješava mnoge semantičke probleme pobližim određivanjem specifičnih riječi. Tako je omogućena višekratna upotrebljivost teksta i neovisnost o programskoj podršci, platformi ili aplikaciji te je osigurana je precizna definicija i efikasnija obrada samog teksta.

Indeksiranje novinskih članaka za ulazak u elektronsku bazu podataka radi se prema tematski definiranim kriterijima u VND pregledniku.

Proces indeksiranja novina je određivanje bitnih karakteristika novinskih članaka, te je svaki dokumentalist unutar svog područja samostalno izradio stručne teme koje su relevantne za određeno područje.

Organizacija informacija u novinskoj elektronskoj bazi podataka (opis novinskog članka) podijeljena je na dva velika dijela (pripremni i stručni) i sastoji se od sljedećih elemenata:

1. Pripremna indeksacija VND on-line za unos tekstova u elektroničku bazu podataka: Autor, Datum, Izvor, Naslov, Opis, Stranica

OCR-irani tekst je programska podrška koja omogućava pretragu teksta po svim riječima ili full text search:

Optical Character Recognition je programska podrška koja omogućava prepoznavanje teksta na bazi slike dokumenta. Za OCR podsustav se koristi Abby FineReader 7.0 s podrškom za sve jezike. Klijentska aplikacija sadrži odabir slike, OCR postupak nad slikom, korekciju pogrešaka, opis dokumenta meta podatcima, te export dokumenta. Serverski dio je aplikacija izgrađena na temeljima Microsoft SharePoint portal Servera koja koristi WebStore podatkovno spremište za pohranu podataka (u XML obliku) i Microsoft SQL Server 2000 za pohranu meta podataka. Aplikacije za pre-

gled objavljenih podataka / meta podataka je prilagođeni portal koji omogućava svu potrebnu funkcionalnost u odabiru podataka i / ili pretraživanja podataka. Postoje tri tipa pretraživanja: jednostavno (po sadržaju), napredno (po sadržaju i dijelu meta podataka) i kombinirano pretraživanje (po sadržaju i svim meta podacima uz mogućnost izgradnje upita). OCR tekst vrlo često ima mnogo grešaka što će se vidjeti u primjerima iz prakse koji slijede. Ljudski faktor u pripremljenoj indeksaciji ima važnu ulogu u programskom ispravljanju tekstualnih grešaka, a dokumentalistička obrada također onemogućava greške unosom stručnih termina u polje ključnih riječi.

Područje znanja: unosi priprema, a obilježeno je unaprijed od strane dokumentalista;

Tip sadržaja: može biti članak ili stranica;

Status obrade: daje podatak da li je članak samo OCR-iran ili je i dokumentalistički obrađen;

Kreiranje članka: pripremljena indeksacija unosi i kreira članka u za to zadano polje, upućuje na ime i prezime osobe koja je novinski članak kreirala, OCR-irala, ispravila pogreške i unijela potrebne parametre;

Modificiranje članka: upućuje na ime i prezime dokumentalista koji je članak preuzeo i stručno obradio.

2. Atribucija je VND-ova stručna dokumentalistička obrada i sadrži:

Ključna osoba: u novinskom članku može biti nekoliko ključnih osoba, zato je programski omogućen unos i pretraga više osoba istovremeno;

Ključne riječi: međusobno su povezane unutarnjim i vanjskim čvorovima pomoću tezaurusa;

Medijska analiza: Negativno, Neutralno, Pozitivno

Novinska vrsta: Ankete i istraživanja, Feljton, Intervju, Izjava, Izvještaj, Kolumna, Komentar, Pisma, Portret osobe, Reportaža

Ostale osobe: su osobe koje se nalaze u tekstu i vezane su uz ključne osobe, slučajevite, afere ili događaje te ih je potrebno izdvojiti u posebno polje;

Područje znanja: moguće je istovremeno označiti nekoliko navedenih područja jer pojedini novinski članci u sebi sadrže problematiku iz više područja:

Društvo / Zdravstvo / Socijalna skrb, Ekologija, Ekonomija / Gospodarstvo / Privreda, Financije, Komunikacije / Mediji, Kultura i umjetnost, Odgoj i obrazovanje, Promet, Religija, Sport, Sudstvo / Kriminal, Svijet / Međunarodni odnosi / Vanjska politika, Unutarnja politika, Znanost;

Tema: može biti samo jedna, odnosi se na područje znanja u kojem se nalazi te omogućava sužavanje pretrage;

Institucija: podrazumijeva poduzeća, državne ustanove, Europska tijela, političke stranke itd. Za sve institucije programska podrška podržava i pretraživanje preko ključnih riječi, a ne samo preko zadanog polja Institucija. Zemljopisna odrednica: sve zemlje i gradovi.

Naša minimalistička norma morala je zadovoljiti sljedeće relacije i upite:

osoba

osoba tema

osoba tema zemljopisna odrednica

osoba tema zemlj. odr. vremenska odrednica

osoba tema zemlj. odr. vrem. odr. ključna riječ

osoba tema zemlj. odr. vrem. odr. ključna riječ institucija

osoba tema zemlj. odr. vrem. odr. ključna riječ institucija izvor

osoba tema zemlj. odr. vrem. odr. ključna riječ institucija izvor autor

osoba tema zemlj. odr. vrem. odr. ključna riječ institucija izvor autor novin-
ska vrsta

osoba tema zemlj. odr. vrem. odr. ključna riječ institucija izvor autor vrsta,
full text search

osoba tema zemlj. odr. vrem. odr. ključna riječ institucija izvor autor vrsta,
full text search medijska analiza

Prve recenzije: veljača 2005. godine

Istodobno smo u suradnji s našim razvojnim timom⁸ riješili potrebne definicije za eksport baza podataka na DVD nosačima. Naše su tadašnje strateške postavke zadatka glasile: na izabrani način prikupljene i organizirane skupine informacija osposobljenom analitičaru omogućavaju pronalaženje neuočljivih sadržaja/obavijesti i trendova. Bili smo dosljedni u tvrdnji da bazu podataka moramo organizirati kao javno dostupno znanje zato što uz primjerenu organizaciju *informacija* postaje i ostaje dugotrajni izvor ne samo novih obavijesti i tema, nego važan doprinos posebnim znanjima. Imali smo na umu i etičnost svog postupka jer smo znali da moramo voditi računa o vlastitim sposobnostima upotrebljavanja *informacije* tog novog/starog izvora moći.

Tada su kreirane naše prve e-zbirke: “Oluja, 5. kolovoza 1995 – vojno-redarstvena akcija” te “RAT U BIH – Sukobi Hrvata i Bošnjaka u Bosni i Hercegovini”.

Molbu za recenziju postignutoga poslali četvorici relevantnih stručnjaka: prof. dr. Damiru Borasu, prof. dr. Miroslavu Tuđmanu, prof. dr. Ivi Josipoviću i dr. Inoslavu Bešku.

Ukratko ćemo citirati njihove navode:

Prof. dr. sc. **Damir Boras** prodekan za znanost i međunarodnu suradnju Filozofskog fakulteta Sveučilišta u Zagrebu:

“Bilo mi je veliko zadovoljstvo pregledati građu Vjesnikove novinske dokumentacije koju ste mi stavili na uvid na dva programska CD-a koja ste pripra-

⁸ Za razvoj VND online iznimno je zaslužan Branimir Grilec, tada voditelj razvoja projekata u tvrtki eVision, a sada je vlasnik tvrtke Conscious d.o.o. Bile su korisne iscrpljujuće svade s tadašnjim šefom IT tima Vjesnik d.d., Robertom Mekinićem.

vili, 'Oluja, 5. kolovoza 1995 – vojno-redarstvena akcija' te 'RAT U BIH – Sukobi Hrvata i Bošnjaka u Bosni i Hercegovini', tekstova objavljeni u hrvatskim tiskanim glasilima.

Prema viđenim podacima, vaša je baza najkvalitetnija, i moglo bi se reći elitna baza dokumentacije koja omogućuje dostup do svih objavljenih podataka u glasilima koje pratite.

Neobično me veseli da se konačno u Hrvatskoj pojavila konzistentno i sveobuhvatno, te na standardan način organizirana građa novinskih tekstova koja predstavlja izvor ne samo za povijest i kulturu nego i za politiku, odnosno izvor podataka iz raznih izvora koji međusobno ukršteni omogućuju dobivanje kritičke i vjerodostojne slike o događajima koje opisuju. Velika je prednost to što se mogu vidjeti razni novinski izvori, razni autori tekstova, pogledi raznih država itd. U svakom slučaju organizacija građe u potpunost, pa čak i više od toga, odgovara međunarodnim standardima novinske i druge dokumentacije.

Posebno moram naglasiti da se kroz Vašu dokumentaciju afirmira i naš studij informacijskih znanosti jer one upravo to predstavljaju: organiziranje znanja. Očito se to vidi i u činjenici da zapošljavate mnogo naših studenata.

Dozvolite mi da Vam predložim suradnju u edukaciji naših studenata, budući da im je prema novom ECTS sustavu potrebna praksa baš u ustanovama ovakvog tipa, a smatram da bi u Vašoj dokumentaciji mogli s lakoćom postići vještine i kompetencije potrebne danas svim informacijskim stručnjacima, i to ne samo dokumentalistima, nego i arhivistima, bibliotekarima i ostalima."

Prof. dr. sc. **Miroslav Tuđman**, Odsjek za informacijske znanosti, Katedra za organizaciju znanja, Filozofski fakultet Sveučilišta u Zagrebu:

"Sustav organizacije i pretraživanja baze dokumenata "Oluja" i "Rat u BiH. Sukobi Hrvata i Bošnjaka u Bosni i Hercegovini" s dokumentalističkog stajališta zaslužuje najbolju ocjenu: prema kriteriju obrade sadržaja dokumenata, prema kriteriju pretraživanja, i prema vremenu pristupa pretraženim dokumentima.

Manji prigovori (naveli smo ih nekoliko) više su 'tehničke' naravi koje se odnose na prikaz sustava, tj. kako koristiti VND on-line baze dokumenata.

Osim što organizacija i pretraživanje ovih baze dokumenata zaslužuju najbolju ocjenu, ovo su, koliko je meni poznato, jedine hemeroteke koje se nude korisnicima na DVD nosačima. Zato je potrebno ocijeniti i procijeniti njihovu vrijednost i sa stajališta sadržaja dokumentima.

Vjesnikova novinska dokumentacija (VND) prva i jedina u Hrvatskoj kreira baze tekstova hrvatskih tiskanih glasila prema temama koje su ključne za interpretaciju suvremene hrvatske povijesti. Izbor tema na dva ponuđena DVD nosača primjer su važnosti organiziranja građe koja jest već sada relevantna u Haagu, ali koja će vrlo brzo biti relevantna i na procesima koje će ICTY preseeliti u nadležnost hrvatskom pravosuđu. Ovako organizirane hemeroteke, dostupne na DVD nosačima velikom broju korisnika – od novinara, znanstvenika,

odvjetnika, do javnih i kulturnih djelatnika – na najbolji način pomažu utvrđivanju podataka i dolaženju do istine o tim zbivanjima.

Projekt 'VND on-line. Baze dokumenata' u tehničkom smislu može se širiti u barem tri nove dimenzije: a) digitalizaciji članaka (tj. dostupnosti izvornih dokumenata u digitalnoj formi), b) on-line dostupnoj bazi podataka, c) obradi svih područja, tj. dnevnoj obradi ukupne nacionalne novinske produkcije.

Takav jedan projekt po svom je karakteru nužno nacionalni projekt. Kako po svojim ambicijama, obuhvatu, značenju, potrebama korisnika, ali i po cijeni koštanja."

Prof. dr. sc. Ivo Josipović, Pravni fakultet Sveučilišta u Zagrebu:

"Koristim priliku da Vam čestitam na kvalitetnom projektu koji ste pokrenuli. Mislim da je ovakav projekt posebno važan za profesionalce, novinare, državnu upravu, političare, stručnjake, znanstvenike i studente, te medijske kuće, znanstvene i stručne organizacije.

Mislim da je organizacija građe primjerena. Alati za pretragu su dobro postavljeni. Preporučam da se alati za pretragu prije stvaranja baze podataka ustroje konzultacijom stručnjaka i potencijalnih korisnika. Očekujem da će alati biti specifični s obzirom na temu i potencijalne korisnike. Sustav je funkcionalan i jednostavan za rukovanje.

Projekt je vrlo perspektivan i s obzirom na njegove potencijale, mislim da ga treba razvijati kao javni servis. Sastavni dio projekta treba biti povezivanje i s drugim bazama podataka. Uz to, trebalo bi ispitati potrebe Vlade i državne uprave. Daljnji potencijal projekta sadržan je u mogućnosti suradnje s gospodarstvom.

Suvremeno gospodarsko poslovanje traži stalnu dostupnost informacijama radi donošenja odgovarajućih poslovnih odluka. O važnosti projekta za znanstveni i stručni rad gotovo da je suvišno govoriti. Posebno, ako bi se uspostavila suradnja i razmjena podataka s odgovarajućim stranim bazama podataka.

Sumirajući, smatram da je riječ o projektu koji je iznimno vrijedan i koji zavrijeđuje apsolutnu podršku. Smatram da bi u nastavku projekta trebalo inzistirati na njegovom autonomnom ustroju i profesion-alizaciji. Uz to, bilo bi dobro naći strateške partnere u državnoj upravi, znanosti i gospodarstvu."

Dr. sc. Inoslav Bešker, sveučilišni predavač i novinar:

"Recenzirane dionice pokazuju razvedenu i za novinara veoma pogodnu organizaciju građe. Svaka pretraga se može organizirati po jednome od 11 kriterija ili po njihovoj kombinaciji, što je u praksi sasvim dovoljno za brzo snalaženje, te omogućava preciznost.

Organizacija po dionicama, kakva je ponuđena na recenziranim nosačima, izrazito je korisna i korisnicima izvan novinarstva, koji se usredotočuju samo na pojedine teme. Smijemo pretpostaviti da može bitno skratiti vrijeme istraživanja i proširiti tematski okvir kako studentima prava ili politologije, tako i djelatni-

cima u pravosuđu i politici – kao što bi druge teme našle druge korisnike na istim razinama bilo obrazovanja, bilo uprave.

Alati ponuđeni u tu svrhu su jednostavni, pa ih može rabiti svaki novinar s nižom srednjom razinom informatičke uvježbanosti. Pri smještanju cjelokupnog projekta na Internet koristan bi bio 'tutorial'.

Funkcionalnost sustava, kakav je demonstriran u recenziranim dionicama zavidno je visoka. Ako se uzmu u obzir ograničenja zadana tipom nosača informacija – ograničenja koja bi bila uklonjena dostupnošću istih informacija sa servera – funkcionalnost se može ocijeniti potpunom, kada je riječ o interesima istraživačkog novinara, koji u veoma kratkom vremenu može dobiti cjelovit uvid, a u razumnoj vremenu i detaljne podatke.

Budućnost tog projekta jamačno bi morala biti u njegovu širenju na cjelokupnu bazu podataka i na omogućavanju da ta građa bude dostupna online, 24/24 sata.

Puna djelotvornost te baze mogla bi se postići, i za novinare i za ine korisnike, kada bi se ona mogla relaciono umrežiti s drugim takvim bazama, kao što su Narodne novine, Državni arhiv, te slične ustanove, i kada bi se samo jednom potragom mogle obuhvatiti različite baze (kada bi, na primjer, potraga po ključnim riječima 'Gotovina Haag zločin' ponudila i međunarodne ugovore, i nacionalne zakone, i javnosti dostupne arhivske dokumente npr. iz Ureda Predsjednika).

Dionice projekta 'VND on-line' su uzorak javne baze podataka, informatički dostupne najširem krugu korisnika, ponajprije u javnom informiranju, te u stručnom obrazovanju i javnoj upravi.

Preporučuje se razvijanje projekta u javnu bazu podataka dostupno online, kao javni servis na nacionalnoj razini i od nacionalnog interesa, te njezino umreživanje s drugim nacionalnim bazama podataka."

Postignuti i novi ciljevi

Od tada iz baze podataka eksportirano je nekoliko desetina posebnih baza podataka. Među njima vrijedi izdvojiti prvi *tailor made product* – za Ured glasnogovornice Ministarstva pravosuđa RH kojim je do kraja 2006. godine provjeren medijski odjek procesa i rezultata provedbe reforme pravosuđa. Danas ta posebna baza podataka broji više od 60 tisuća dokumenata.

Posebna je cjelina projekt "100 javnih ličnosti hrvatske političke scene" nastao uoči parlamentarnih izbora 2003. godine sa 4.279 intervju političkih lidera iz svih hrvatskih političkih stranaka objavljenih u razdoblju od 1999. do izbora za Hrvatski sabor 2003. godine.

Uz bok tom projektu je građa na DVD nosaču "Predsjednički izbori 2004. godine". Od tog projekta započela je naša suradnja sa Studijem informacijskih znanosti Filozofskog fakulteta Sveučilišta u Zagrebu. Studenti Odsjeka postali su naši stalni gosti provodeći obveznu studentsku praksu radom na poslovima digitalizacije, unosa teksta u bazu podataka i – pripremajući svoje seminarske i

diplomske radove tako da sve što je napisano bude u službi razvoja baze podataka na različitim razinama zahtjeva. Jedan od njihovih radova je DVD nosač sa sadržajima intervjua hrvatskih ministara kulture. Taj skup dokumenata omogućava kontekstualnu provjeru kulturoloških i političkih, pragmatičnih i strateških razina promišljanja i odlučivanja osoba koje su odlučujuće utjecale na događaje što obilježavaju kulturu novije hrvatske povijesti.

Tijekom ljeta 2007. godine tom se procesu pridružio prijedlog dr. Jasminke Dobše, Fakultet organizacije i informatike, Varaždin, Sveučilište u Zagrebu da se u našoj bazi podataka testiraju metode i programske podrške. To je logičan izbor jer je na našoj bazi podataka moguće stvoriti vrlo kompleksan i relevantan *training set* potrebnih dokumenata. Predložila je suradnju HIB/IDC i Fakulteta organizacije i informatike kao podršku dodiplomskim radovima apsolviranih FOI.

Tako stvaramo HINA informacijsku bazu kao jedan od alata kojim se služi HINA u vlastitom organiziranju i HIB Informacijsko dokumentacijski centar koji povezuje interes korisnika sadržaja što ih HINA generira kao svoj proizvod osiguravajući pri tom dostupnost javnog znanja. Postupcima kojima se služi i poveznicama koje ostvaruje, HIB/IDC suvremenim načinom osigurava zaštitu i javnu upotrebu hrvatske kulturne baštine zabilježene u tiskanom i AV medijima. HIB/IDC svojim sustavom podržava kreiranje Informacijskog centra na Internetu pomoću umreženja s bazama u okruženju.

Na godišnjoj programskoj razini HIB/IDC obradit će milijun dokumenata, a prema posebnom programu obradit će bazu isječaka iz novina. Tijekom 2008. godine s našim suradnicima razvit ćemo SW za automatsku kategorizaciju sustava. Naše su pristupne odrednice općepoznate: utemeljenost na lematizaciji, svrstavanje tekstova u predefinirane kategorije (pronalaženje kategorija prema unaprijed naučenim pravilima, sposobnost sustava da "uči" pravila prema već napravljenoj kategorizaciji, sustav temeljem pravila samostalno određuje koje od predefiniranih kategorija odgovaraju pojedinom novom tekstu, strojno dobiveni rezultati postaju prijedlozi za humanog kategorizatora), pronalaženje imenovanih entiteta (u obradi sadržaja niz je entiteta koji karakteriziraju tekst i nužno ih je prepoznati, izdvojiti za dalju obradu: osobe, institucije, toponimi...). Rezultat će omogućiti približavanje cilju – bazi podataka i alatima, sustavima i sredstvima koji omogućavaju dubinsko pretraživanje podataka od interesa za korisnika tko god on bio i s kojeg god mjesta želi znati.

Zaključak

Tako će HINA informacijska baza podataka / Informacijsko dokumentacijski centar osigurati da u Hrvatskoj djeluje konzistentna i sveobuhvatna te na standardan način organizirana građa javnih glasila što je izvor ne samo za povijest i kulturu, nego i za svakog pojedinog hrvatskog građanina, gospodarstvo i politiku. Stvarajući takav javni servis pokazujemo ne samo da smo zreli za informacijski povezanu i organiziranu Europu, nego i prema nama samima u kulturnoj i

u političkoj sferi te da kao stabilno društvo organizirano djelujemo u području zaštite kulturnog identiteta, organizacije javnog znanja i dostupnosti znanja.

Literatura

- Castells, Manuel, Informacijsko doba, Golden Marketing, Zagreb 2000/2002, Internet galaksija, Jesenski i Turk, Zagreb; 2003.
- Hodonj, Željko; Prijedlog : Projekt pokretanja poduzeća – Dokumentacija Vjesnik, Vjesnik interno izdanje, 2002.
- Hrvatski državni arhiv; Pismo ministru kulture RH mr. Boži Biškupiću, 17. siječnja 2006.
- Hrvatski sabor, Državni proračun Republike Hrvatske, 2007.g., <http://www.nn.hr/clanci/sluzbeno/2006/3098.htm>,
- Hrvatski sabor, Odbor za informiranje, informatizaciju i medije; Izvješće o radu Upravnog vijeća HINA-e 2005/2006 g. <http://www.sabor.hr/Default.aspx?art=8080>
- Interpelacija o odgovornosti Vlade Republike Hrvatske za stanje Društva Vjesnik d.d. 2006. <http://www.sabor.hr/Default.aspx?art=8503>
- prof. dr. sc. Boras, Damir; prof. dr. sc. Tuđman, Miroslav; prof. dr. sc. Josipović, Ivo; dr. sc. Bešker, Inoslav ; recenzije *e-zbirki*: "*Oluja, 5. kolovoza 1995 – vojnoredarstvena akcija*" te "*RAT U BIH – Sukobi Hrvata i Bošnjaka u Bosni i Hercegovini*", veljača 2005.
- Vjesnik on line i Novinska dokumentacija, Vjesnik, interno izdanje, travanj 2003.