

iz naših knjižnica

Uređuje: Danko Škare

Harvestiranje Hrvatskog arhiva weba Nacionalne i sveučilišne knjižnice u Zagrebu

K. Holub* i I. Rudomino**

Nacionalna i sveučilišna knjižnica u Zagrebu
Hrvatske bratske zajednice 4, 10 000 Zagreb

O Hrvatskome arhivu weba

Nacionalna i sveučilišna knjižnica u Zagrebu u suradnji sa Sveučilišnim računskim centrom (Srce) 2004. osnovala je *Hrvatski arhiv weba*¹ kako bi preuzela i trajno pohranila građu s interneta koja predstavlja značajan dio hrvatske kulturne baštine, a temelji se na zakonskoj obvezi dostave obveznog primjerka građe objavljene na internetu. Arhiv je namijenjen za trajnu pohranu, čuvanje i javni pristup građi. Kako je ovoj vrsti sadržaja svojstvena nestalnost i dinamičnost (*web*-središta često mijenjanju svoj sadržaj, dizajn, domenu i sl.) velik je izazov i zahtjevna zadaća Arhiva sačuvati građu u što originalnijoj mjeri "*look & feel*" te ju učiniti dostupnom za sadašnje i buduće korisnike.

Nemoguća je misija sačuvati potpuni internetski (*web*) prostor jedne zemlje, stoga *Hrvatski arhiv weba* različitim pristupima pokušava sačuvati arhivsku sliku tj. repliku hrvatskog internetskog (*web*) prostora koja govori mnogo o razvoju društva, znanosti i kulturi. Do 2011. godine sadržaji s *weba* sakupljali su se samo selektivno, točnije, građa se odabirala i arhivirala prema *Kriterijima odabira*.² S posebnom se pažnjom sakupljaju sadržaji koji postoje samo na internetu te dokumentiraju društveni trenutak, trendove, popularna zbivanja, važne sportske, političke, kulturne i druge događaje. Ovakva vrsta sadržaja na *webu* javlja se u *news*-portalima, blogovima, mrežnim mjestima ustanova, udruga, događaja, klubova, znanstvenih projekata te u raznim vrstama časopisa i knjiga koji predstavljaju izvore informacija o suvremenoj kulturi, ali i o ekonomskim, društvenim i političkim trendovima. Svaka arhivirana publikacija bibliografski je obrađena te se može pronaći putem knjižničkog kataloga Nacionalne i sveučilišne knjižnice u Zagrebu kao i putem internetskih stranica *Hrvatskoga arhiva weba*, gdje se sadržaj može pretraživati preko bilo koje riječi iz naslova ili URL-a te s pomoću ključnih riječi. Također, sadržaj *Arhiva* može se pregledavati prema predmetnim kategorijama i po abecednom redu naslova (slika 1).

Harvestiranje nacionalne domene (.hr)

Međutim, kako bi upotpunila i još dodatno obogatila nacionalnu kolekciju arhiviranih sadržaja s *weba*, Nacionalna i sveučilišna knjižnica u Zagrebu se odlučila i za harvestiranje cijele nacionalne .hr domene. Harvestiranje se provodi jednom godišnje u odrede-

nom vremenskom razdoblju, a započinje popisom aktivnih domena dobivenih od CARNet-a. Prvo takvo harvestiranje provedeno je tijekom ljeta 2011. kada su prikupljena sva *web*-središta registrirana na domeni.hr. Prikupljeno je i spremljeno više od 56 milijuna datoteka veličine 3,1 TB. Sljedeće godine harvestiranje je provedeno u prosincu, a prikupljeno je 60 milijuna datoteka ukupne veličine 4.1 TB.

Svi sakupljeni (harvestirani) sadržaji od rujna 2013. dostupni su javnosti putem internetske stranice *Hrvatskog arhiva weba*.³

Pretraživati se može tako da se u tražilicu upiše točan URL arhivirane internetske stranice. Potom se u prikazanom kalendaru odabere godina i arhivirani primjerak tražene publikacije (slika 2).

Tematska harvestiranja

Prateći svjetske trendove u sakupljanju i očuvanju nacionalnih *web*-prostora odlučeno je proširiti opseg arhiviranja sadržaja HAW-a tematskim harvestiranjem kako bi se na jednom mjestu pohranili i okupili sadržaji na određenu temu te ih pružili na korištenje javnosti.⁴

Tematskim harvestiranjem prikupljaju se sadržaji relevantnih političkih, nacionalnih ili kulturnih događaja kao što su nacionalni izbori za Sabor, izbori za predsjednika Republike Hrvatske, povijesni sportski događaji, ali i neki nepredviđeni događaji poput prirodnih katastrofa (potresi, požari i sl.).

Do listopada 2013. provedeno je pet tematskih harvestiranja o aktualnim temama i događajima od nacionalnog značaja (slika 3):

1. Krajem 2011. godine *Hrvatski arhiv weba* proveo je prvo takvo harvestiranje na temu *Parlamentarni izbori 2011.* te su prikupljeni i pobrani svi relevantni sadržaji (internetske stranice političkih stranaka, portali i njihovi dijelovi koji su tematski pratili izbore) vezani uz ovaj politički i nacionalni događaj. Harvestiranjem je prikupljeno 140 URL-ova.

2. *Referendum o pristupanju Republike Hrvatske Europskoj uniji (22. siječnja 2012.)* bila je tema drugoga tematskog harvestiranja, gdje su prikupljeni članci i internetske stranice o ovom važnom događaju. Prikupljeno je 476 URL-ova.

3. *Izbori članova u Europski parlament iz Republike Hrvatske* održani su prvi put u Hrvatskoj te su odabrani sadržaji na temu

* Karolina Holub, kholub@nsk.hr

** Ingeborg Rudomino, irudomino@nsk.hr

¹ Hrvatski arhiv weba. URL: haw.nsk.hr/ (3. 10. 2013.).

² Kriterij odabira obveznog primjerka mrežne građe za obradu i arhiviranje. URL: haw.nsk.hr/kriterij_odabira (3. 10. 2013.).

³ Harvestiranje hrvatskog web prostora.

URL: haw.nsk.hr/harvestiranje-hrvatskog-web-prostora (2. 10. 2013.).

⁴ Tematska harvestiranja. URL: haw.nsk.hr/tematska-harvestiranja (2. 10. 2013.).

Napredno pretraživanje

Naslov
Naslov
URL
Ključne riječi

Traži

Hrvatski arhiv weba Nacionalne i sveučilišne knjižnice u Zagrebu zbirka je odabranih sadržaja preuzetih s interneta i pohranjenih na računalnom poslužitelju Knjižnice. Namijenjen je preuzimanju i trajnom čuvanju publikacija s interneta kao dijela hrvatske kulturne baštine. Arhivirani sadržaji mogu se pretraživati preko naslova, URL-a, ključnih riječi i predmetnih područja.

Pregledavanje predmetnih područja

Civilizacija i kultura	Društvo	Ekonomija i poslovanje
Filozofija i religija	Industrija i tehnika	Informacijske znanosti
Književnost i jezici	Masovni mediji i odnosi s javnošću	Obrana i nacionalna sigurnost
Obrazovanje	Organizacije, udruge, pokreti u Hrvatskoj	Politika
Povijest i geografija	Pravo i uprava	Priroda i okoliš
Prirodne znanosti	Psihologija	Sport i rekreacija
Turizam i putovanja	Umjetnost	Zdravlje i kultura življenja

Abecedno pregledavanje naslova

0-9 A B C Č Ć D Đ E F G H I J K L M
N O P Q R S Š T U V W X Y Z Ž

Slika 1 – Prikaz načina pretraživanja i pregledavanja u Hrvatskom arhivu weba

Hrvatski arhiv weba
Nacionalna i sveučilišna knjižnica u Zagrebu

English | Impressum | Naslovica NSK

Naslovica O arhivu weba Za nakladnike Dokumenti **Harvestiranje**

Harvestiranje hrvatskog web prostora Tematska harvestiranja

Obrazac za prijavu online publikacije

Zadnjih 5 arhiviranih

Operacija kataktae za Zulujujane

Review of psychology

Prvi vijenik

Radovi

Društvo povjesničara umjetnosti Rijeke, Istre i Hrvatskog primorja

Traženi pojmovi

izletna dragutin šela epidemiologija
gnac jstamj loc marulić
osmrtnice paska sopa lamet u

Copyright © Nacionalna i sveučilišna knjižnica u Zagrebu 2013. Sva prava pridržana.

Hrvatski arhiv weba
waybackmachine

http://www.irb.hr

Traži

http://www.irb.hr je arhiviran 4 puta počevši od srpanja 20, 2011.

1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-------------	------	------

SU							VI					VII					VIII										
1	2	3	4	5	6	7	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5						
6	7	8	9	10	11	12	6	7	8	9	10	11	12	6	7	8	9	10	11	12	6	7	8	9	10	11	12
13	14	15	16	17	18	19	13	14	15	16	17	18	19	13	14	15	16	17	18	19	13	14	15	16	17	18	19
20	21	22	23	24	25	26	20	21	22	23	24	25	26	20	21	22	23	24	25	26	20	21	22	23	24	25	26
27	28	29	30	31			27	28	29	30	31			27	28	29	30	31			24	25	26	27	28	29	30
30	31																										
SVI							LP					SRP					KOL										
1	2	3	4	5	6	7	1	2	3	4	1	2	3	4	5	1	2	3	4	5	6						
8	9	10	11	12	13	14	8	9	10	11	12	8	9	10	11	12	8	9	10	11	12	13					
15	16	17	18	19	20	21	15	16	17	18	19	15	16	17	18	19	15	16	17	18	19	20					
22	23	24	25	26	27	28	22	23	24	25	26	22	23	24	25	26	22	23	24	25	26	27					
29	30	31					29	30	31			29	30	31			28	29	30	31							
RUJ							LIS					STU					PRO										
1	2	3					1					1	2	3	4	5	1	2	3								

Srpanj 26, 2011
1 srijedak/a
• 13:25:41

Slika 2 – Prikaz stranice za pretraživanje harvestiranog sadržaja nacionalne (.hr) domene

Hrvatski arhiv weba
Nacionalna i sveučilišna knjižnica u Zagrebu

English | Impressum | Naslovnica NSK

Naslovnica
O arhivu weba
Za nakladnike
Dokumenti
Harvestiranje

Harvestiranje hrvatskog web prostora
Tematska harvestiranja

Obrazac za prijavu online publikacije

Zadnjih 5 arhiviranih

Operacija katarakte za Žutokljunce

Review of psychology

Pravni vjesnik

Radovi

Društvo povjesničara umjetnosti Rijeke, Istre i Hrvatskog primorja

Traženi pojmovi

bioetika dragutin šela epidemiologija

ignac jutarnji list marulić osmrtnice

paška sopa tamaro vadav

zamp

Tematska harvestiranja

Hrvatski arhiv weba povremeno pobire (harvestira) odabrane sadržaje s weba koji se odnose na aktualne teme ili događaje od nacionalnog značaja. Odabrani su najznačajniji sadržaji vezani uz određenu temu. Njihova poberivost ovisi o načinu izrade pojedinog web-sjedišta (više Preporuke http://haw.nsk.hr/preporuke_za_izradu_mreznih_publikacija) Klikom na naslov tematskog harvestiranja dobit ćete popis web-stranica čiji se sadržaj odnosi na odabranu temu.



Izbor zastupnika za Hrvatski sabor 2011.

Redoviti izbori zastupnika za Hrvatski sabor održavaju se svake četvrte godine. Hrvatski arhiv weba prikupio je i čuva sve javno dostupne sadržaje na temu izbora zastupnika za Hrvatski sabor. Pobrana su web-sjedišta političkih stranaka, blogovi, portali i video zapisi koji su pratili izbore.

Harvestiranje je provedeno od 2. do 8. prosinca 2011. godine. Prikupljeno je 140 URL-ova ukupne veličine 64 GB.



Referendum o pristupanju Republike Hrvatske Europskoj uniji (22. siječnja 2012.)

Hrvatski arhiv weba prikupio je i čuva sve javno dostupne sadržaje o državnom referendumu o pristupanju Republike Hrvatske Europskoj uniji koji je održan 22. siječnja 2012. S različitih su domena sakupljeni članci i web-stranice o ovom važnom događaju.

Harvestiranje je provedeno 16. listopada 2012. godine. Prikupljeno je 476 URL-ova ukupne veličine 929 MB.



Izbor članova u Europski parlament iz Republike Hrvatske 2013.

Izbori članova u Europski parlament iz Republike Hrvatske održani su po prvi put u Hrvatskoj 14. travnja 2013. i održavat će se svake četvrte godine. Hrvatski arhiv weba prikupio je i čuva sve javno dostupne sadržaje na temu izbora članova u Europski parlament. Pobrana su web-sjedišta političkih stranaka i portali koji su pratili izbore.

Harvestiranje je provedeno 1. svibnja 2013. godine. Prikupljeno je 55 URL-ova ukupne veličine 8,9 GB.



Lokalni izbori 2013

Redoviti izbori članova predstavničkih tijela jedinica lokalne i područne (regionalne) samouprave i izbori općinskih načelnika, gradonačelnika i župana te njihovih zamjenika održavaju se svake četvrte godine. Hrvatski arhiv weba prikupio je i čuva sve javno dostupne sadržaje na temu lokalnih izbora. Pobrana su web-sjedišta političkih stranaka, osobne stranice kandidata i portali koji su pratili izbore.

Harvestiranje je provedeno od 6. do 7. lipnja 2013. godine. Prikupljeno je 202 URL-a ukupne veličine 15 GB.



Pristupanje Republike Hrvatske u Europsku uniju (1. srpnja 2013.)

Republika Hrvatska pridružila se Europskoj uniji 1. srpnja 2013. kao 28. članica. Hrvatski arhiv weba prikupio je i pohranio javno dostupne sadržaje objavljene na internetu na temu pristupanja Republike Hrvatske Europskoj uniji. Pobrana su internetska glasila, portali gradova i lokalni mediji koji su pratili ovaj događaj.

Harvestiranje je provedeno od 16. do 17. srpnja 2013. godine. Prikupljeno je 116 URL-ova ukupne veličine 32 GB.

Copyright © Nacionalna i sveučilišna knjižnica u Zagrebu 2010. Sva prava pridržana.



Slika 3 – Prikaz stranice s tematskim harvestiranjima

izbora članova u Europski parlament. Arhivirana su web-sjedišta političkih stranaka i portali koji su pratili izbore te je prikupljeno 55 URL-ova.

4. Redoviti lokalni izbori članova predstavničkih tijela jedinica lokalne i područne samouprave održavaju se svake četiri godine. Prikupljeni su i sačuvani svi javno dostupni sadržaji na tu temu, sveukupno 202 URL-a.

5. Peto harvestiranje provedeno na temu povijesno značajnog događaja za Hrvatsku *Pristup Republike Hrvatske Europskoj uniji* 1. srpnja 2013. *Hrvatski arhiv weba* arhivirao je sadržaje s internetskih glasila, portala gradova i lokalnih medija koji su pratili ovaj događaj. Prikupljeno je 116 URL-ova.

Nove tematske zbirke Hrvatskog arhiva weba predstavljaju značajnu kolekciju arhiviranog sadržaja te su sada javno dostupne svakome tko želi "rekonstruirati" događaje vezane uz određenu temu.

Na ovaj način sačuvan i dostupan sadržaj namijenjen je različitim vrstama istraživanja, analizama i proučavanjima s ciljem otkrivanja, tumačenja i pojašnjavanja činjenica te šireg znanja o prošlim događajima, povijesti ili određenoj temi.

Hrvatski arhiv weba Nacionalne i sveučilišne knjižnice u Zagrebu time ispunjava svoju misiju da u što većoj mjeri čuva za buduću naraštaje i daje na korištenje našu "online-prošlost".