

Hrvatski arhiv weba - trinaest godina poslije

Karolina Holub i Inge Rudomino

Arhiviranjem weba u svijetu započela je međunarodna organizacija Internet Archive već 1996. godine. Ubrzo nakon toga nekoliko nacionalnih knjižnica počinje s pobiranjem weba i uspostavlja svoje arhive weba, među prvima su Australija, Švedska i Novi Zeland. Zbog potrebe okupljanja baštinskih ustanova posvećenih sakupljanju, čuvanju i pristupu online sadržajima je 2003. osnovan Međunarodni konzorcij za dugoročnu zaštitu internetskih sadržaja (International Internet Preservation Consortium, IIPC), kojega je i Hrvatski arhiv weba član od 2008. godine.

Temelj arhiviranju weba u Hrvatskoj je Zakon o knjižnicama, koji stupa na snagu 1997. I, između ostaloga, propisuje prikupljanje online publikacija kao obveznog primjerka. Već 1998. Nacionalna i sveučilišna knjižnica u Zagrebu započinje s katalogizacijom tada nove vrste građe (serijske publikacije, knjige), dok s katalogizacijom web-stranica započinje 2003. godine.

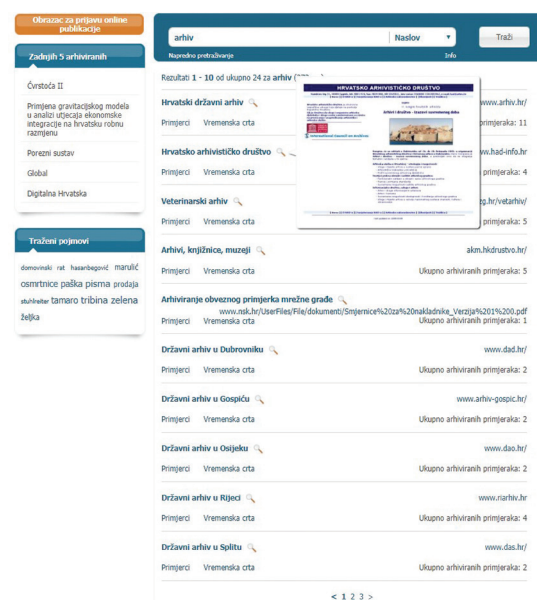
Hrvatski arhiv weba (HAW) Nacionalne i sveučilišne knjižnice u Zagrebu uspostavljen je u rujnu 2004. godine u suradnji sa Sveučilišnim računskim centrom Sveučilišta u Zagrebu (Srce). Hrvatski arhiv weba sustav je za preuzimanje i arhiviranje obveznog primjerka online građe s ciljem obrade, pohrane, trajnog čuvanja i osiguravanja pristupa arhiviranoj građi.

Gradi objavljenoj na webu svojstvena je česta promjena sadržaja i tehnologija izrade, ali i kratak vijek trajanja te je zbog izuzetno velike količine podataka i raznovrsnosti građe nužno imati kriterije koji jamče kvalitetu arhiviranog sadržaja. Knjižnica se odlučila za selektivan pristup arhiviranju weba, što podrazumijeva arhiviranje prethodno identificirane građe na temelju kriterija utvrđenih od same Knjižnice. Sve do 2011. godine sadržaji s weba prikupljali su se isključivo na taj način.

Grada koja se pohranjuje u HAW-u predstavlja dio naše nacionalne, kulturne i znanstvene baštine poput news portala (npr. Index.hr, Telegram, Jutarnji.hr itd.), blogova, web-sjedišta ustanova, udruga, događaja, klubova, znanstvenih projekata, ali i časopisa i knjiga. Uz sadržaje objavljene na .hr domeni arhiviraju se i sadržaji s drugih domena (.org, .net, .com itd.) ako su u skladu s Kriterijima odabira.

Programsku podršku sustavu za arhiviranje razvilo je Srce i redovno je nadograđuje kako bi se uspješno pratio brzi razvoj web-tehnologija i osigurala kvaliteta arhiviranih web-sjedišta prema principu „look and feel“.

Radni proces u selektivnom pristupu arhiviranja weba započinje identifikacijom i selekcijom, slijedi katalogizacija, arhiviranje te slanje obavijesti nakladnicima o pohrani i osiguravanje pristupa. Značajno je spomenuti da se radni proces temelji na svakodnevnoj interakciji između dva sustava, knjižničkog i sustava za arhiviranje, gdje se na dnevnoj bazi razmjenjuju podaci.

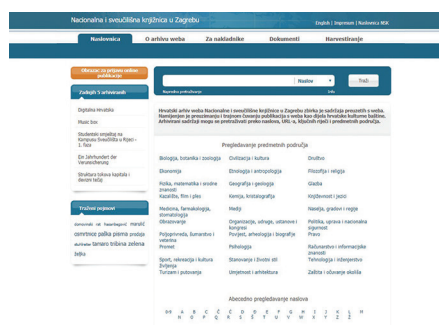


The screenshot shows a search interface for the HAW. At the top, there's a search bar with the text 'arhiv' and a search button. Below it, a list of search results is displayed, including 'Hrvatski državni arhiv', 'Hrvatsko arhivističko društvo', 'Veterinarski arhiv', 'Arhiv, knjižnice, muzeji', 'Arhiviranje obveznog primjerka mrežne građe', 'Državni arhiv u Dubrovniku', 'Državni arhiv u Gospiću', 'Državni arhiv u Osijeku', 'Državni arhiv u Rijeci', and 'Državni arhiv u Splitu'. Each result shows the number of items and a link to the full list.

Popis arhiviranih web-stranica

Većina arhivirane građe javno je dostupna i može se pretraživati po bilo kojoj riječi iz naslova i URL-a te po ključnim riječima, a omogućeno je i složeno pretraživanje. Također, sadržaj je moguće pregledavati prema predmetnim područjima i abecedno na <http://haw.nsk.hr/>. Do listopada 2017. u HAW je pohranjeno 6.528 arhiviranih naslova s 49.518 arhiviranih primjeraka.

Uz mnoge prednosti, poput prilagodbe arhiviranja na razini naslova i arhiviranja sadržaja na drugim domenama, ovakav pristup ima i svoje nedostatke. Zbog količine objavljenog sadržaja na webu kao i činjenice da knjižničari ne mogu promptno identificirati i kontinuirano pratiti novoobjavljene naslove, relevantan sadržaj nepovratno se gubi. Zbog toga se 2011. godine započelo s godišnjim pobiranjem (harvestiranjem) cijele nacionalne domene (.hr) koje za NSK provodi Srce. Harvestirani sadržaji su javno dostupni na web-stranicama Hrvatskog arhiva weba (<http://haw.nsk.hr/harvestiranje-hrvatskog-web-prostora>) i mogu se pretraživati upisivanjem potpunog URL-a.



The screenshot shows the homepage of the HAW website. It features a navigation menu with options like 'Nacionalna i sveučilišna knjižnica u Zagrebu', 'HAW | Internet | Nacionalna i sveučilišna knjižnica u Zagrebu', 'Arhiviranje obveznog primjerka', 'Za knjižnice', 'Diskusije', and 'Harvestiranje'. There's also a search bar and a list of featured items or categories.

Naslovnica Hrvatskog arhiva weba (HAW)

Pretraživanje godišnjih pobiranja .hr domene



Unesite potpunu web adresu (URL).
 Traži



Hrvatski arhiv weba
Nacionalna i sveučilišna knjižnica u Zagrebu

Iste se godine započelo i s trećim načinom prikupljanja sadržaja s weba – izgradnjom tematskih zbirki. Na ovaj se način arhiviraju sadržaji vezani uz aktualne teme i događaje od nacionalnog značaja, a čiji sadržaji relativno brzo nestaju s weba. Hrvatski arhiv weba do danas sadrži devet tematskih zbirki koje se mogu pregledavati putem <http://haw.nsk.hr/> tematska-harvestiranja.

Do listopada 2017. ukupan broj arhiviranog sadržaja cijelog Arhiva iznosi preko 40 TB.

Građa s weba predstavlja značajan izvor vrijednih informacija te ju je važno sačuvati i ponuditi budućim generacijama na korištenje. Arhiviranje weba postaje ključno za sva područja znanosti jer se danas sve veći dio znanstvenog života, ali i značajan dio društvenog i političkog života jedne zemlje odvija na webu i na raznim platformama kao što su blogovi, društvene mreže (Facebook, Twitter, Instagram i sl.). Ako takvi sadržaji nisu dokumentirani i adekvatno sačuvani neće biti dostupni za buduća istraživanja.

Nacionalna i sveučilišna knjižnica u Zagrebu vrlo je rano prepoznala važnost arhiviranja i čuvanja nacionalne kulturne i znanstvene baštine na webu te u najvećoj mogućoj mjeri pokušava biti u korak s najnovijim svjetskim trendovima u arhiviranju weba, kako bi se i ubuduće nastavilo s dosadašnjim radom, važnim za povijest i kulturu našega naroda. ■

Više pojedinosti o Hrvatskom arhivu weba pogledajte na:
<http://haw.nsk.hr/>



Predstavljanje HAW-a na tportalu: <https://www.tportal.hr/tag/hrvatski-arhiv-weba>

JESTE LI ZNALI?
4 festival hrvatskih digitalizacijskih projekata

Značajni projekti u 2014. godini

STRUČNI SKUP U POVODU DESETE OBLJETNICE HRVATSKOG ARHIVA WEBA

Dana 26. rujna 2014. godine stručnim skupom obilježena je deseta obljetnica arhiviranja weba u Hrvatskoj. Skup je organiziran kako bi se istaknula važnost novih usluga u digitalnome okruženju koje korisnicima pruža Nacionalna i sveučilišna knjižnica Zagrebu i time ostvaruje svoju informacijsku, obrazovnu, kulturnu i socijalnu ulogu u društvu. Također, cilj Skupa bio je potaknuti veće korištenje pohranjenih sadržaja, stjecanje novih znanja u vezi s područjem arhiviranja weba te podizanje svijesti o važnosti trajnoga čuvanja nove vrste baštine.

Hrvatski arhiv weba (HAW), sustav za prikupljanje i čuvanje sadržaja s interneta, započeo je s radom 2004. u suradnji sa Sveučilišnim računalnim centrom Sveučilišta u Zagrebu (Srce). Pohranjeni sadržaji javno su dostupni i mogu se pretraživati preko internetske stranice Arhiva (haw.nsk.hr).

Na Hrvatskom arhivu weba moguće je, primjerice, pronaći podatke i o svim političkim kampanjama u Hrvatskoj u posljednjem desetljeću.

Više informacija na mrežnoj stranici <http://haw.nsk.hr>

Fotografije: Jakob Gubišnik / Fotolaborator NINA