

Ana Garvas Delić

Hrvatska informacijsko-dokumentacijska referalna agencija – Hidra
Trg maršala Tita 3
Zagreb

Miroslav Milinović

Sveučilišni računski centar, Sveučilišta u Zagrebu – Srce
Josipa Marohnića 5
Zagreb

**PRIKUPLJANJE, ODABIR, OBRADA I POHRANA
DOKUMENATA SA SLUŽBENIH MREŽNIH STRANICA TIJELA
JAVNE VLASTI REPUBLIKE HRVATSKE**

UDK 004.4PPOPD+004.738.5:087.7(497.5)

Izlaganje sa znanstvenoga skupa

Hidra u okviru svoje osnovne djelatnosti prikuplja, obrađuje, pohranjuje i diseminira javnu službenu dokumentaciju i informacije Republike Hrvatske. Razvoj interneta u Hrvatskoj toliko je napredovao da se velik dio službenih publikacija i dokumenata objavljuje na nizu službenih web-sjedišta. Namjera je Hidre okupiti digitalnu građu iz ovog izvora i omogućiti građaninu – korisniku jednostavan pristup informacijama s jedne adrese: <http://www.hidra.hr>. Podjednako je važno očuvanje autentičnosti sadržaja, oblika i funkcionalnosti u svrhu dugoročne zaštite i korištenja. Okosnicu projekta Prikupljanje, odabir, obrada i pohrana dokumenata – POOPD čini Arhiv mrežnih dokumenata – AMD, sustav koji je nastao kao rezultat zajedničkog projekta Hidre i Sveučilišnoga računskog centra – Srca, započet u svibnju 2004, s ciljem uspostave sustava za preuzimanje i arhiviranje mrežno dostupnih dokumenata s odabranog skupa web-sjedišta. Sustav je razvijen na temelju iskustava Srca u području istraživanja weba (projekti Mjerenje web prostora – MWP i Digitalni arhiv mrežnih publikacija – DAMP) i utemeljen na programskoj podršci Open Source. Modularan je, proširiv, jednostavan za uporabu te dostupan putem web-sučelja.

Ključne riječi: *pobirač (robot), arhiv mrežnih dokumenata, kriterij odabira dokumenata, tijela javne vlasti*

1. Uvod

Zadatak je Hidre unapređivanje korištenja i promicanje javnih službenih podataka, informacija i dokumentacije Republike Hrvatske. Prema Zakonu o knjižnicama (NN 105/1997), svi nakladnici službenih publikacija dužni su besplatno i o svom trošku dostaviti Hidri jedan primjerak svojih izdanja. Prema prethodnom iskustvu, nikada nije u potpunosti ostvareno prikupljanje službenih publikacija. Uvođenje

interneta na području Hrvatske doprinijelo je dostupnosti javnih dokumenata i publikacija koja tijela javne vlasti objavljuju. Od 1997. se putem kataloga službenih publikacija, objavljenog i u papirnatom obliku i na mrežnim stranicama Hidre, upućivalo na mrežne adrese digitalnih dokumenata.

U 2003. godini pokušalo se pospješiti prikupljanje pregledavanjem internetskih stranica web-sjedišta određenih ustanova i spremanjem pronađenih sadržaja na računalo u Hidri. Posao se pokazao izrazito iscrpljujućim, teško ga je bilo sustavno obaviti i nadgledati promjene.

Istodobno, u Srcu se stjecalo iskustvo u području istraživanja weba na projektima *Mjerenje web prostora (MWP)* i *Digitalni arhiv mrežnih publikacija (DAMP)*, što je i dovelo do pokretanje suradnje s Hidrom. Polazeći od navedenih zadataka i činjenice brzog razvoja interneta u Hrvatskoj, pokrenut je projekt automatskog prikupljanja dokumenata sa službenih stranica tijela javne vlasti Republike Hrvatske – *Prikupljanje, odabir, obrada i pohrana dokumenata (POOPD)*.

Za potrebe projekta, u Srcu je razvijen sustav za Arhiviranje mrežnih dokumenata (AMD) sa zadatkom preuzimanja i arhiviranja mrežno dostupnih dokumenata s određenog skupa web-sjedišta. Projekt je započet u svibnju 2004., a AMD je za rad bio spreman u studenom 2004. Do srpnja 2005. prikupljeno je preko 32.000 dokumenata s više od 470 web-sjedišta. Novi alat donio je i novi pogled na radne zadatke u Hidri i novu organizaciju poslova s namjerom da se sadržaji što ažurnije obrađuju kako bi bili dostupni javnosti.

2. Razvoj sustava AMD

AMD je utemeljen na programskoj podršci Open Source, modularan je, proširiv i korisniku-administratoru jednostavan za uporabu. Čine ga:

- **pobirač (robot)** – podsustav za pobiranje dokumenata s web-sjedišta,
- **nadzornik** – podsustav za kontrolu pobiranja i arhiviranje rezultata pobiranja,
- **raspoređivač** – podsustav za raspoređivanje pobiranja,
- **skladište podataka** – baza podataka i posebno organizirani datotečni sustav (Slika 1) i
- **web-sučelje** za pristup arhivu i upravljanje sustavom (Slika 2).

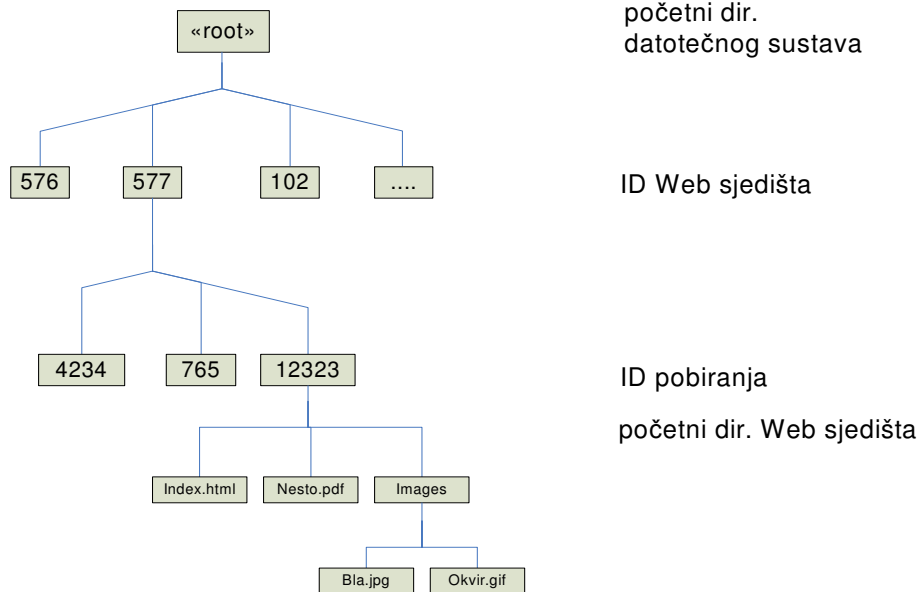
Sustavom AMD upravlja se preko web-sučelja koje ima nekoliko funkcija:

- upravljanje radom sustava: unos podataka o (novim) web-sjedištima; upravljanje pobiranjem (unos parametara) i ručni unos dokumenta,
- uvid u stanje sustava (pregled aktivnosti),
- uvid u rezultate pobiranja,
- pristup arhiviranim dokumentima,
- pregledavanje arhiva dokumenata i
- statistike prikupljanja i obrade.

Kako se u Hidri prikupljaju podaci o web-sjedištima tijela javne vlasti (NN 148/2004) čije djelovanje, odnosno publikacije, dokumentaciju i podatke prati i obrađuje, to je iskorišteno za inicijalno punjenje sustava. Odabrani su podaci iz Hidrih adresara tijela javne vlasti i političkih stranaka sljedećeg opsega: ID (identifikacijski broj) web-sjedišta, URL, stvarni naziv tijela, normirani naziv tijela, područje djelatnosti i klasifikacijska oznaka (prema Pojmovniku Eurovoc).

Obavljeno je inicijalno automatsko namještanje parametara pobiranja za sva web-sjedišta: frekvencija pobiranja (jednom mjesečno), dubina pobiranja (3) i vrste

dokumenata (.pdf, .doc .rtf, .xls i .ppt). Temeljem postavljenih uvjeta i parametara, raspoređivač uvrštava konkretno pobiranje u red, a pobirač izvršava pojedina pobiranja prema tom redu. U sustavu usporedo radi više pobirača, a njihov broj ovisi o potrebi, odnosno opsegu posla.

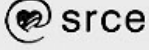


Slika 1. Organizacija datotečnog sustava

HIDRA - Microsoft Internet Explorer

Adresa: http://hidra.srce.hr/Index.php?show=publication_list&page=2

Google Search Web 641 blocked AutoFill

HIDRA Popis web sjedišta institucija Prealed queue-a Novo web sjedište Popis prikupljenih dokumenata Pobiranja koja čekaju brisanje Statistike 

Popis web sjedišta institucija

Uvjeti selektiranja zapisa:

Naziv:

Normirani naziv:

URL:

Klasifikacijska oznaka:

Prikupljanja u kojima je prikupljen bar jedan dokument: od (gggg-mm-dd) do

Dokumenti iz prikupljanja pregledani:

Dokumenti iz prikupljanja katalogizirani:

Filtriraj ▶

Naziv	URL	Klasifikacijska oznaka	Prikupljeni dokumenti	Zadnja obrada
Središnji državni ured za razvoju strategiju Vlade Republike Hrvatske	http://www.hrvatska21.hr	H11.1=2	0	2005-06-10 23:00:00
Središnji državni ured za upravljanje državnom imovinom Vlade Republike Hrvatske	http://www.vlada.hr/drzavna-imovina/index.html	H11.1=2	0	2005-07-10 23:00:00
Središnji državni ured za upravu Vlade Republike Hrvatske	http://www2.uprava.hr/	H11.1=2	2	2005-07-10 23:00:00
Nacionalni odbor za Habitat	http://www.mzopu.hr	H11.1=4	625	2005-06-10 23:00:00
Savjet za populacijsku politiku Vlade Republike Hrvatske	http://www.mobms.hr	H11.1=4	0	2005-06-10 23:00:00
Vijeće za djecu Vlade Republike Hrvatske	http://www.viiecezadjecu.hr	H11.1=4	0	2005-07-10 23:00:00
Vijeće za ekonomsku politiku Vlade Republike Hrvatske	http://www.hrvatska21.hr	H11.1=4	0	2005-06-10 23:00:00
Hrvatska informacijsko-dokumentacijska referalna agencija	http://www.hidra.hr	H11.1=5	176	2005-07-10 23:00:00

Prečac: do index.php?show=resource_list Internet

Slika 2. Prikaz web-sučelja sustava AMD

Treba naglasiti važnost osobnog udjela administratora sustava koji u nastavku namještaju parametre prepoznavanjem stvarne organizacije i učestalosti porasta broja dokumenata pojedinog web-sjedišta. Naglasimo da AMD podržava funkciju kontrole duplikata dokumenata. Naravno, osim automatske kontrole pojave duplikata dokumenata, potrebno je uključiti i stručno osoblje jer sustav pri provjeri duplikata ne analizira sadržaj pojedinih prikupljenih dokumenata, nego njihovu vrstu, veličinu i datum nastanka.

Skladište podataka AMD-a čine prikupljeni dokumenti u posebno organiziranom datotečnom sustavu kojem se pristupa kroz odgovarajuće web-sučelje. Datotečni sustav organiziran je prema web-sjedištima i broju pobiranja, kako je prikazano na Slici 1.

U bazi podataka bilježe se podaci:

- o web-sjedištima,
- parametri pobiranja (za svako web-sjedište),
- rezultati pojedinog pobiranja:
 - adresa za pristup arhiviranim dokumentima,
 - osnovni podaci o provedenom pobiranju (veličina, status itd.),
 - prikupljeni metapodaci (s početne stranice web-sjedišta) i
 - distribucija vrsta podataka za dohvaćene datoteke.

Svakom pobranom i pohranjenom dokumentu pridruženi su podaci koji ga potanko određuju:

- stvarni naziv ustanove,
- područje djelatnosti ustanove (prema Pojmovniku Eurovoc),
- klasifikacijska oznaka ustanove,
- URL dokumenta,
- URL dokumenta u arhivu,
- datum prikupljanja,
- vrsta dokumenta (format) i
- veličina dokumenta.

Osim toga, uz svako se pobiranje pohranjuje i zapis (log) o pobiranju koji omogućuje detaljan uvid u cjelovit proces pobiranja.

3. Obrada prikupljenih dokumenata i javni pristup prihvaćenim dokumentima

Novi pristup prikupljanju digitalne službene dokumentacije donio je niz saznanja i potreba za stalnom prilagodbom organizacije obrade prikupljenih sadržaja. Stručno osoblje zaduženo za administriranje sustavom krenulo je na edukaciju korištenja novih alata u toku probnog pobiranja dokumenata. Bilo je važno naučiti zadati parametre za podešavanje pristupa svakome pojedinom web-sjedištu. Podešavale su se dubine pobiranja i vrste dokumenata za pobiranje, ovisno o strukturi poslužitelja i sadržaju na pojedinom poslužitelju. Potrebno se pokazala stalna razmjena iskustava u radu među suradnicima Hidre i Srca.

Usporedo s pregledavanjem dokumenata, razvijaju se i dopunjavaju pravila za odabir dokumenata. Posebno je važno da svi koji rade na izboru dokumenata imaju stalan pristup ažurnim pravilima. Kriteriji za odabir dokumenata su sljedeći:

1. **autorstvo** – osnovni kriterij odabira; odabiru se samo dokumenti/publikacije kojih je **autor/izdavač** tijelo javne vlasti s čijih se mrežnih stranica prikuplja dokument/publikacija;

2. **sadržaj:**

- odabiru se **svi** dokumenti/publikacije **tijela državne vlasti** određeni drugim kriterijima (autorstvo, vrijeme, format itd.), osim efemerne građe (obraci i sl.),
- odabiru se **samo** dokumenti/publikacije **tijela s javnim ovlastima te ureda državne uprave u županijama i tijela jedinica lokalne i područne samouprave** određeni drugim kriterijima (autorstvo, vrijeme, format) koji se **izravno odnose** na rad samoga tijela;

3. **dodatni kriteriji:**

- isključuju se priopćenja za tisak, najave (za promocije knjiga i sl., izložbe), pristupnice, obrasci, formulari, rješenja o raznim predmetima itd.,
- ukoliko je identičan dokument dostupan i u formatima .doc i .pdf, .doc se isključuje (nakon provjere identičnosti dokumenta).

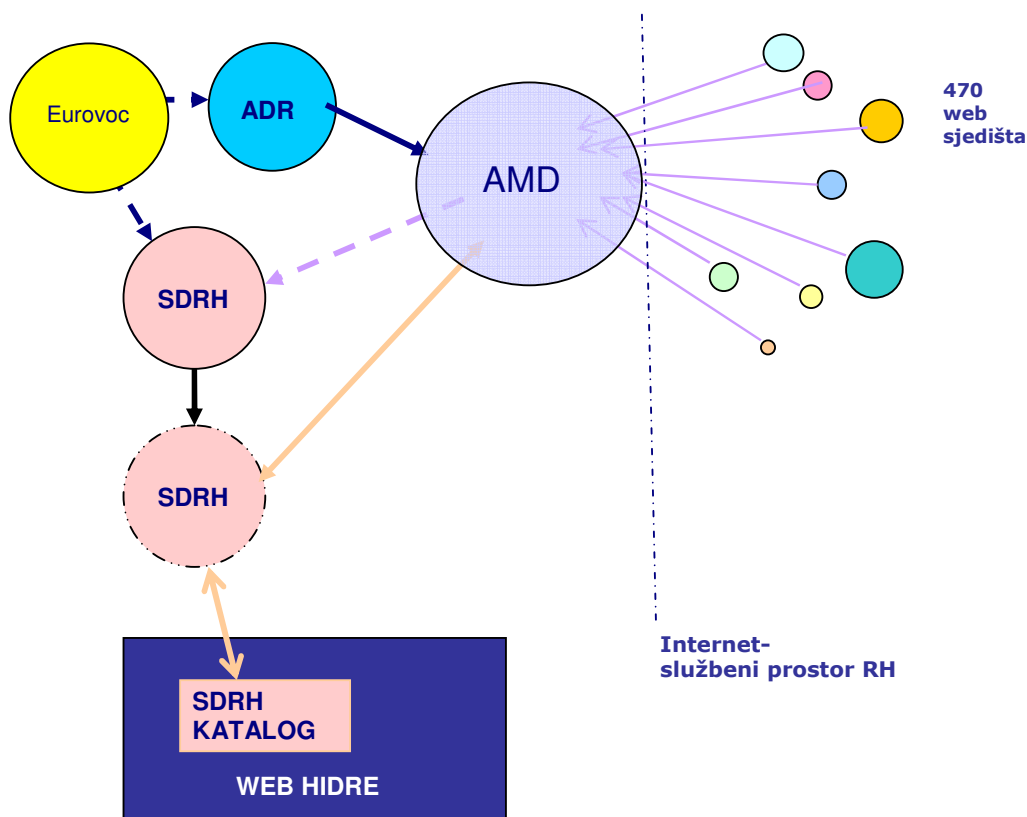
U **prvoj fazi obrade** prikupljenih digitalnih dokumenata prilazi se pregledavanju preko web-sučelja AMD-a. Prema važećim kriterijima za odabir dokumenata definiranim u Hidri, svaki prikupljeni dokument dobiva oznaku statusa prihvaćen (+) ili odbačen (-). AMD omogućuje pristup određenim prikupljenim dokumentima prema različitim uvjetima za filtriranje i sortiranja sadržaja. Tu su identifikacijski brojevi ustanove ili prikupljanja, datumi prikupljanja, klasifikacijske oznake, nazivi ustanova i statusi dokumenata.

Druga faza obrade katalogizacija je dokumenata koji imaju oznaku prihvaćen (+) u AMD-u. Katalogizacija se obavlja u bazi podataka Službene dokumentacije RH – SDRH. Bibliografska jedinica sadrži poveznicu na digitalnu inačicu dokumenta. Time su stvoreni preduvjeti trajne dostupnosti dokumenta preko internetskih stranica Hidre. Pristup se ostvaruje kroz Katalog službene dokumentacije Republike Hrvatske na adresi: <http://www.hidra.hr/dok/dok.htm> (Slika 3).



Slika 3. Katalog službene dokumentacije Republike Hrvatske

Sustav POOPD shematski je prikazan na Slici 4. Održavanje podataka u AMD-u obavlja se prijenosom podataka iz adresara usklađenog s normativnim podacima u Pojmovniku Eurovoc. Katalogizacija i predmetna obrada prikupljenih dokumenata obavlja se u SDRH-u. Pristup dokumentima ostvaren je preko WebPAC-a koji radi s kopijom SDRH-a i dostupan je na internetskim stranicama Hidre.



Slika 4. Shematski prikaz sustava POOPD

4. Rezultati POOPD-a

Tijekom rada na projektu POOPD, u 2005. godini postignuti su sljedeći rezultati:

- razvijen je sustav AMD,
- razvijeno je prikupljanje, te kriteriji odabira i obrade digitalne građe,
- postavljene su nove smjernice pristupa obradi digitalnih dokumenata, a
- u pripremi su smjernice za minimum standarda, kojeg će se morati pridržavati sva državna tijela pri objavljivanju svojih službenih publikacija.

AMD je instaliran na računalu Hidre, privremeno ugošćenom u Srcu. Obavljeno je inicijalno prikupljanje dokumenata s 422 web-sjedišta s rezultatom od 8.400 dokumenata. Pobirač se pokreće mjesečno i prikuplja nove dokumente. Do srpnja 2005. prikupljeno je preko 32.700 dokumenata ukupne veličine 10.700 MB. Tijek prikupljanja po mjesecima prikazan je u Tablici 1. Dokumenti se pregledavaju te se prihvaćeni dio upućuje na daljnju obradu – katalogizaciju i predmetnu obradu. Dosad je za obradu prihvaćeno oko 12 posto dokumenata. Ostali su, mada neobrađeni, trajno pohranjeni. Podaci o web-sjedištima četiri su puta dopunjeni i sada je aktivno 470 adresa.

Mjesec	Prikupljeno	Pregledano	Prihvaćeno	Odbaćeno	Nepregledano
--------	-------------	------------	------------	----------	--------------

2004-11	8422	8183	818	7365	239
2004-12	4599	4006	1506	2500	593
2005-01	3055	3055	327	2728	0
2005-02	3287	3212	271	2941	75
2005-03	2269	1886	238	1648	383
2005-04	2124	2064	109	1955	60
2005-05	5018	4767	553	4214	251
2005-06	1933	1636	155	1481	297
2005-07	1996	743	95	648	1253
Ukupno	32703	29552	4072	25480	3151

Tablica 1. Prikaz prikupljanja i obrade dokumenata po mjesecima, od početka rada AMD-a do srpnja 2005.

Hidra je AMD-om standardizirala i ubrzala proces prikupljanja digitalnih dokumenata koji su prethodno prikupljeni i pohranjivani ručno. Omogućeno je sustavno i cjelovito periodičko prikupljanje sa svih službenih web-sjedišta odjednom, te automatsko i jednostavno praćenje novih sadržaja. Brže se pregledavaju dokumenti koji zadovoljavaju unaprijed određene kriterije.

Usporedo se razvijalo i programsko rješenje AMD-a:

- razvijena je i instalirana nova inačica pobirača, čime je znatno povećan broj prikupljenih dokumenata,
- web-sučelje podešavalo se potrebama pristupa dokumentima i označivanju statusa obrade dokumenta,
- izrađen je modul za detaljnu dijagnostiku rada i statističku analizu sadržaja arhiva,
- dodan je modul za ručno arhiviranje dokumenata, a namijenjen je prevladavanju problema koji se mogu javiti kod prikupljanja važnih dokumenata; njime se u AMD dodaju ručno prikupljeni dokumenti, koji su nestali s poslužitelje pojedinih ustanova.

U Tablici 1 sažet je rezultat rada na odabiru i obradi prikupljenih dokumenata od studenog 2004. do sredine srpnja 2005. Isti podaci prikazani su prema klasifikacijskim skupinama ustanova u Tablici 2. U nastavku, dan je popis ustanova s čijih je adresa prikupljeno najviše dokumenata, bilo po veličini prostora (Tablica 3) ili po broju dokumenata (Tablica 4). Raspodjela prikupljenih dokumenata po vrstama prikazana je u Tablici 5.

Klasifikacijska oznaka	Ustanove	Broj prikupljenih dokumenata	Broj prihvaćenih dokumenata
H	državne ustanove RH	26480	2936

H1	<u>tijela državne vlasti</u>	9185	1962
H2	<u>ustanove javnog sektora</u>	15844	938
H3	<u>poduzeća od javnoga interesa</u>	1451	36
DH	diplomatski odnosi RH	150	12
L0	<u>jedinica lokalne i područne (regionalne) samouprave RH</u>	4719	1067
P	političke stranke	1073	57
	Ukupno	32423	4072 (12%)

Tablica 2. Prikaz broja prikupljenih i prihvaćenih dokumenata prema klasifikacijskim skupinama ustanova od početka rada AMD-a do srpnja 2005.

Naziv	Veličina arhiva (MB)	Broj dokumenata
Ministarstvo europskih integracija RH	223,86	550
Sveučilište u Zadru	234,44	455
Hrvatske šume d.o.o.	264,44	206
Hrvatska akademska i istraživačka mreža CARNet	371,57	714
Hrvatska lutrija d.o.o.	384,12	262
Hrvatski sabor	422,18	1304
Fakultet elektrotehnike i računarstva Sveučilišta u Zagrebu	444,65	1200
Fakultet strojarstva i brodogradnje Sveučilišta u Zagrebu	460,83	1747
Hrvatska narodna banka	559,24	1660
Vlada Republike Hrvatske	632,87	1377

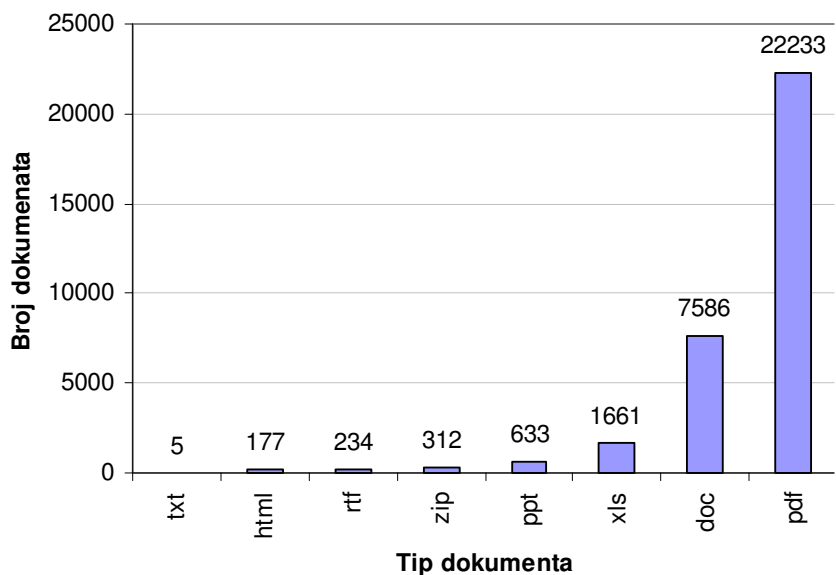
Tablica 3. Prikaz deset ustanova s najvećim arhivom prikupljenih dokumenata (prema uvjetima za prikupljanje) po veličini zauzetog prostora

Naziv	Broj dokumenata	Veličina arhiva (MB)
Ministarstvo zaštite okoliša, prostornog uređenja i graditeljstva RH	666	200,56
Ministarstvo znanosti, obrazovanja i športa RH	673	110,70
Državni ured za reviziju	677	180,65
Hrvatska akademska i istraživačka mreža CARNet	714	371,57
Ministarstvo kulture RH	903	126,40
Fakultet elektrotehnike i računarstva Sveučilišta u Zagrebu	1200	444,65
Hrvatski sabor	1304	422,18

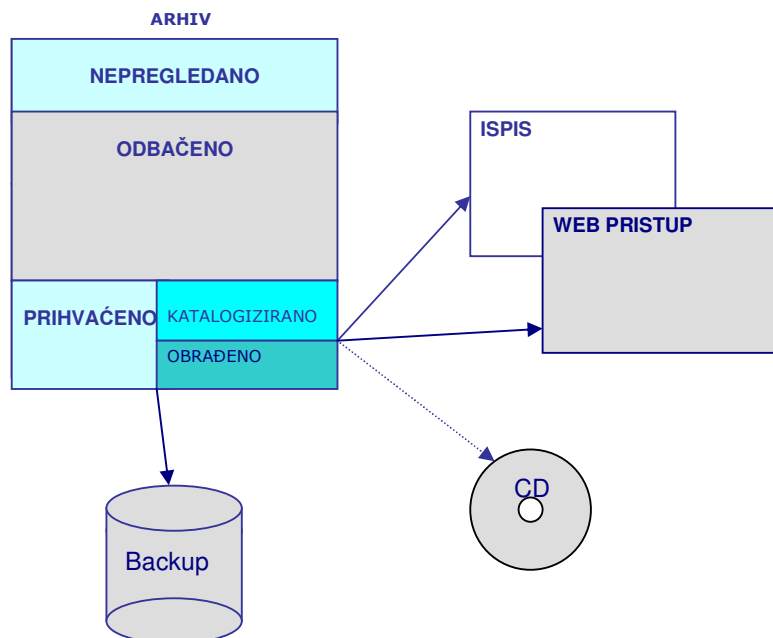
Vlada RH	1377	632,87
Hrvatska narodna banka	1660	559,24
Fakultet strojarstva i brodogradnje Sveučilišta u Zagrebu	1747	460,83

Tablica 4. Prikaz deset ustanova s najvećim arhivom prikupljenih dokumenata (prema uvjetima za prikupljanje) po broju dokumenata

Zbog osiguranja autentičnosti, pretpostavlja se izgradnja datotečnog sustava u arhivu u kojoj nema brisanja ni pojedinih dokumenata, ni ustanova u cjelini. Moguće su samo selekcije za izdvajanje dijelova za različite potrebe. Problem ustanova koje su prestale djelovati rješava se podešavanjem parametra za učestalost pobiranja. Na taj način postojeći dokumenti ostaju u arhivu, ali pobirač više ne obilazi web-sjedište te ustanove. Shematski prikaz arhiva sustava POOPD prikazan je na Slici 5.



Tablica 5. Prikaz broja prikupljenih dokumenata po vrstama od početka rada AMD-a do srpnja 2005.



Slika 5. Arhiv dokumenata

Tijekom pregledavanja i obrade dokumenata uočen je niz nepravilnosti i problema o kojima je odluke trebalo donositi u hodu:

- pojedina tijela javne vlasti RH ne koriste HR domenu za objavu mrežnih stranica,
- pojedine publikacije objavljene na mrežnim stranicama službenih tijela ne udovoljavaju osnovnim pretpostavkama elektroničkog izdavaštva, pa se tako susreću publikacije rascjepkane u nekoliko dokumenata, bez paginacije, naslova, autora, izdavača (službeno tijelo) i/ili godine objavljivanja,
- minimalno se koriste metapodaci i
- postoji niz web-sjedišta s nestandardnim rješenjima internetskih stranica koja otežavaju pristup dokumentima.

Zbog toga je potrebno ustanoviti minimum pravila koja bi se odnosila na oblik i organizaciju dokumenata, a kojih bi se trebala pridržavati sva tijela javne vlasti pri objavi na mrežnim stranicama.

U nastavku, planiramo izradu modula za arhiviranje odabranih web-sjedišta tijela javne vlasti kojim bi se sustavno pokrenula **izgradnja arhiva službenih internetskih stranica RH**. Isto tako, pred nama je razvoj sustava za trajno čuvanje sadržaja arhiva koji bi rizik gubitaka sadržaja sveo na najmanju moguću mjeru te osigurao njegov integritet i autentičnost.

Literatura

Baršić-Schneider, Ž. Pretpostavke i kriteriji za prikupljanje, odabir i pohranu dokumenata. 2004.

URL: <http://www.hidra.hr/hidra/10godina/hidra-srce101204.pdf> (15. 12. 2004.)

Bjelić M., Plavac J., Popović D. The Archive of Network Resources Collection, selection, processing and storage of documents acquired from the official websites of the public authority bodies of the Republic of Croatia. 2005.

URL: <http://www.hidra.hr/hidra/10godina/lida-hidra05-30.ppt> (15. 7. 2005.)

Cabinet Office e-Government Unit. Annex L. Archiving Websites. 2002.

URL: <http://www.cabinetoffice.gov.uk/e-government/resources/handbook/html/annexl.asp>. (15. 7. 2005.).

Cabinet Office e-Government Unit. Illustrated Handbook for Web Management Teams. 2002.

URL: <http://www.cabinetoffice.gov.uk/e-government/resources/handbook/html/htmlindex.asp>. (15. 7. 2005.).

Garvas Delić A. Arhiv Mrežnih Dokumentata – Prikupljanje, Odabir, Obrada i Pohrana Dokumentata sa službenih internetskih stranica državnih tijela RH. 2004.

URL: <http://www.hidra.hr/hidra/10godina/amd-poop.pdf> (15. 12. 2004.)

Gatenby P. Collecting and managing web resources for long –term access : web harvesting and guidelines to support preservation (ICABS Actions 3.3 and 3.4). 2004.

URL: <http://www.ifla.org/IV/ifla70/papers/026e-Gatenby.pdf>. (15. 7. 2005.).

Global Information Locator Service (GILS). URL: <http://www.gils.net/>. (15. 7. 2005.).

Hrvatska informacijsko-dokumentacijska referalna agencija – Hidra. Katalog službene dokumentacije Republike Hrvatske. 2004.

URL: <http://www.hidra.hr/dok/dokrhkat.htm> (15. 7. 2005.)

Hrvatska informacijsko-dokumentacijska referalna agencija – Hidra. HR vodič. 1997. URL: <http://www.hidra.hr/rh/rh.htm> (15. 7. 2005.)

Hrvatska informacijsko-dokumentacijska referalna agencija – Hidra. Pojmovnik Eurovoc. 2000. URL: <http://www.hidra.hr/eurovoc/eurovoc1.htm> (15. 7. 2005.)

Information Management Office-Australian Government. GovernmentOnline: the Commonwealth Government's strategy. 2002.

URL: <http://www.agimo.gov.au/publications/2000/04/govonline>. (15. 7. 2005.).

Milinović M. Arhiv mrežnih dokumenata. 2004.

URL: <http://www.hidra.hr/hidra/10godina/amd-10122004.pdf> (15.12.2004.)

Milinović M. Kako arhivirati Web? // Srce. Novosti (2004) br. 5.

URL <http://www.srce.hr/dokumenti/novosti/sn5.pdf> (10. 11. 2004.)

Milinović M. Koliko je velik Hrvatski WEB? // Srce. Novosti (2004) br. 2.

URL: <http://www.srce.hr/dokumenti/novosti/sn2.pdf> (10. 11. 2004.)

Milinović M., Topolšćak N. The Architecture of DAMP - A System for Harvesting and Archiving Web Publications, LIDA 2005, Dubrovnik (20. 5. – 3. 6. 2005.)

Milinović M., Willer M. DAMP: Sustav za preuzimanje i arhiviranje obveznog primjerka hrvatskih mrežnih publikacija, Posvetovanje »Radenci 2005«, Maribor (6. - 7. 4. 2005.)

Popis tijela javne vlasti // Narodne Novine (2004) 146.

URL: <http://www.nn.hr/clanci/sluzbeno/2004/2559.htm>. (15. 7. 2005.).

Popis tijela javne vlasti za 2005. godinu // Narodne Novine (2005) 39.
URL: <http://www.nn.hr/clanci/sluzbeno/2005/0767.htm>. (15. 7. 2005.).

The national archives. Electronical records management.
URL: <http://www.nationalarchives.gov.uk/electronicrecords/advice/default.htm>. (15. 7. 2005.).

The National Archives. Records of the UK government from Domesday to the present.
URL: <http://www.nationalarchives.gov.uk/>. (15. 7. 2005.).

Zakon o knjižnicama // Narodne Novine (1997) 105.
URL: <http://www.nn.hr/clanci/sluzbeno/1997/0767.htm>. (15. 7. 2005.).

Summary

COLLECTION, SELECTION, PROCESSING AND STORAGE OF DOCUMENTS FROM THE OFFICIAL WEBSITES OF THE PUBLIC AUTHORITY BODIES OF THE REPUBLIC OF CROATIA

As its standard basic activity Croatian Information Documentation Referral Agency (HIDRA) collects, processes, stores and disseminates publicly available official data, information and documentation of the Republic of Croatia. Due to the rapid development of the Internet and related ICT in Croatia the large part of the official documents and publications of the Republic of Croatia is published on the respective Web sites. It is HIDRA's intention to collect electronic resources available via Internet and to enable easy access to that information from one address <http://www.hidra.hr>. It is also important to preserve authenticity, format and functionality of these resources in order to achieve their long-term preservation and usability.

The basis of the project called POOPD is the Archives of the Networked Documents (AMD) system. AMD is the application system created at University Computing Centre of the University of Zagreb (Srce) as a result of the joint project with HIDRA. The project started in May 2004 with a primary goal to create a system for gathering and archiving networked documents from the selected set of Web sites. This system has been developed using Srce's experience in the area of Web technologies mainly through the results of the projects Web space measurement (MWP) and Digital archives for Web publications (DAMP). AMD is created using the Open Source software. It is modular, extensible, and simple to use via the Web interface.

***Key words:** gatherer, robot, networked documents archives, criteria for document selection, public authority bodies*