

# Dokument opozvan 1.10.2003.

HRVATSKA AKADEMSKA I ISTRAŽIVAČKA MREŽA - CARNet

## CDA0013

CROSS - Croatia Search Service

Ur.broj: 1-98-5-016

Kategorija: PREPORUKA

Trajanje: do opoziva

Datum nastanka: 21.05.1998.

Verzija: 1.0 (21.05.98.)

URL: <ftp://ftp.carnet.hr/pub/CARNet/docs/advisories/CDA0013.pdf>

### 1. CROSS i njegovi ciljevi

CROSS ("Croatia Search Service") je pretraživački mehanizam (eng. *search engine*), informacijski servis koji omogućuje pronalaženje dokumenata dostupnih na WWW poslužiteljima u .hr vršnoj Internet domeni.

Temeljni je cilj ovog servisa olakšavanje pronalaženja informacija o Hrvatskoj i u Hrvatskoj.

“O Hrvatskoj” znači povećanje kvalitete prezentacije ukupnog hrvatskog WWW informacijskog prostora, a “u Hrvatskoj” mogućnost pronalaženja pojedinih sadržaja lokalno, na Internetu u Hrvatskoj. Dva su pozitivna učinka korištenja sadržaja lokalno: brži dolazak do dokumenata i smanjenje mrežnog prometa između Hrvatske i Interneta.

### 2. Korištenje CROSS-a

Servis je dostupan putem World Wide Web-a na URL-u:

<http://cross.carnet.hr/>

Najveći dio sadržaja osim na hrvatskom postoji i na engleskom jeziku. Sučelje za postavljanje upita i prikaz rezultata pretraživanja se također temelji na WWW-u, a zasad postoje dva sučelja, tzv. "jednostavni upit", URL:

[http://cross.carnet.hr/index\\_hr.html](http://cross.carnet.hr/index_hr.html)

i "napredni upit", URL:

[http://cross.carnet.hr/advanced\\_hr.html](http://cross.carnet.hr/advanced_hr.html)

Upiti se mogu sastojati od riječi i fraza, mogu sadržavati jednostavne logičke funkcije (“AND” i “OR”) u kombinaciji sa zagradama koje određuju prioritet izvođenja logičkih operacija. Također je moguće pretraživati dokumente i prema strukturi, npr. tražiti riječi ili fraze u naslovu, podnaslovima, URL-u, imenu autora itd. Opis upitnog jezika i opcija pri postavljanju upita se nalazi na URL-u:

***[http://cross.carnet.hr/help\\_hr.html](http://cross.carnet.hr/help_hr.html)***

### 3. Informacije za webmastere

Da bi osigurao ažurnost informacija koje nudi, **CROSS** prikuplja dokumente s WWW poslužitelja iz vlastitog popisa u prosjeku jednom tjedno, obrađuje ih i sažima, te tako dobivene sažeteke dodaje u indeks, nakon čega su dostupni za pretraživanje. Ukoliko se dogodi da neki WWW poslužitelj nije dostupan u vrijeme prikupljanja podataka, on se ponovo posjećuje prilikom narednog prikupljanja.

Dodavanje poslužitelja, ili stranica, u popis **CROSS**-a moguće je putem WWW sučelja koje se nalazi na URL-u:

*[http://cross.carnet.hr/add\\_hr.html](http://cross.carnet.hr/add_hr.html)*

Osim samog URL-a kojem **CROSS** pristupa mogu se unijeti još neki parametri koji definiraju način pristupa **CROSS**-a stranicama: vrijeme između dohvata dviju stranica s istog poslužitelja, maksimalni broj stranica u jednom prikupljanju, najveća dubina do koje će **CROSS** dohvaćati stranice itd. Objašnjenje pojedinih parametra nalazi se na gore navedenom URL-u.

Kako prilagoditi svoj poslužitelj ili stranice za bolju prezentaciju na **CROSS**-u može se pročitati u "uputama za webmastere", na URL-u:

*[http://cross.carnet.hr/wmi\\_hr.html](http://cross.carnet.hr/wmi_hr.html)*

Upute zasad uključuju pregled metoda za kontrolu "WWW Robota", napomene oko prilagodbe za *cacheing*, te prijedloge za korištenje nekih mogućnosti HTTP/1.1 protokola.

Na temelju rada **CROSS**-a nastaje redovita mjesečna statistika, broja i vrste WWW poslužitelja u .hr Internet domeni koja je dostupna na URL-u:

*<http://cross.carnet.hr/stats/site/>*

### 4. Tehnička pozadina

**CROSS** je zasnovan na prilagođenom Harvest sustavu za pretraživanje Interneta. Projekt Harvest vodio je IRTF-RD (Internet Research Task Force - Resource Discovery) na University of Colorado, Boulder, SAD i završio sredinom 1996., od kada se Harvest razvija doprinosom pojedinaca iz Internet zajednice. Popis resursa o Harvestu na Internetu može se naći na URL-u:

*<http://cross.carnet.hr/harvest/>*

Distribucija najnovije inačice Harvest 1.5, koja se kompilira na University of Edinburgh zrcali se sa na **CROSS**-u (pristup je zasad omogućen samo putem HTTP protokola) na URL-u:

*<http://cross.carnet.hr/mirrors/ftp.tardis.ed.ac.uk/pub/harvest/>*

Voditelj projekta **CROSS** i sam aktivno sudjeluje u razvoju inačice Harvest 1.5. Harvest sustav je po svojoj prirodi distribuiran, te se njegovi moduli mogu nalaziti na zasebnim poslužiteljima, a u skladu s tim i Harvest poslužitelji mogu međusobno razmjenjivati informacije. Razvoj ovakvog distribuiranog sustava Harvest poslužitelja u Hrvatskoj omogućio bi kvalitetnije i brže pretraživanje informacijskog prostora te smanjenje mrežnog prometa unutar Hrvatske i opterećenja WWW poslužitelja uzrokovanog **CROSS**-ovim prikupljanjem podataka.

## 5. Planovi

Za naredno razdoblje planira se nastavak razvoja **CROSS**-a, sažeto naveden u slijedećim točkama:

- smanjivanje vremena potrebnog za dobivanje odgovora na upite
- omogućavanje pretraživanja Internet News grupa hr. hijerarhije s CARNet-ovog Internet News poslužitelja
- proširivanje područja rada na WWW poslužitelje koji sadržajem pripadaju hrvatskom informacijskom prostoru, a nalaze se izvan .hr vršne Internet domene
- izrada sučelja za jednostavno sužavanje pretraživanja na pojedine WWW poslužitelje kako bi webmasteri mogli koristiti **CROSS** za pretraživanje WWW stranica na vlastitim poslužiteljima
- razmatranje mogućnosti za razvijanje mreže Harvest poslužitelja u Hrvatskoj

## 6. O projektu

Projekt CARNet-a "**CROSS** - Croatia Search Service" je započeo 01.06.1997. godine, a vodi ga Hrvoje Stipetić. U prvih šest mjeseci rada na sadašnjoj adresi (od 01.11.1997. do 30.04.1998.) **CROSS** je odgovorio na više od 105.000 upita, dok je u istom razdoblju na WWW poslužitelju opsluženo preko 285.000 zahtjeva.

Projekt **CROSS** je nastavak rada istoimenog servisa, razvijenog uz podršku Hrvatskog Crvenog križa, koji je od sredine 1996. do listopada 1997. bio dostupan na URL-u: <http://cross.hck.hr/>.

E-mail kontakt adresa projekta **CROSS** je

*[cross@cross.carnet.hr](mailto:cross@cross.carnet.hr)*.