

WWW.HR:

First metadata-enabled service in Croatian Webospace

Igor Ljubi, Gordan Gledec

Department of Telecommunications
Faculty of Electrical Engineering and Computing
University of Zagreb

CUC 2001
September 24 – 26, 2001

WWW.HR briefly

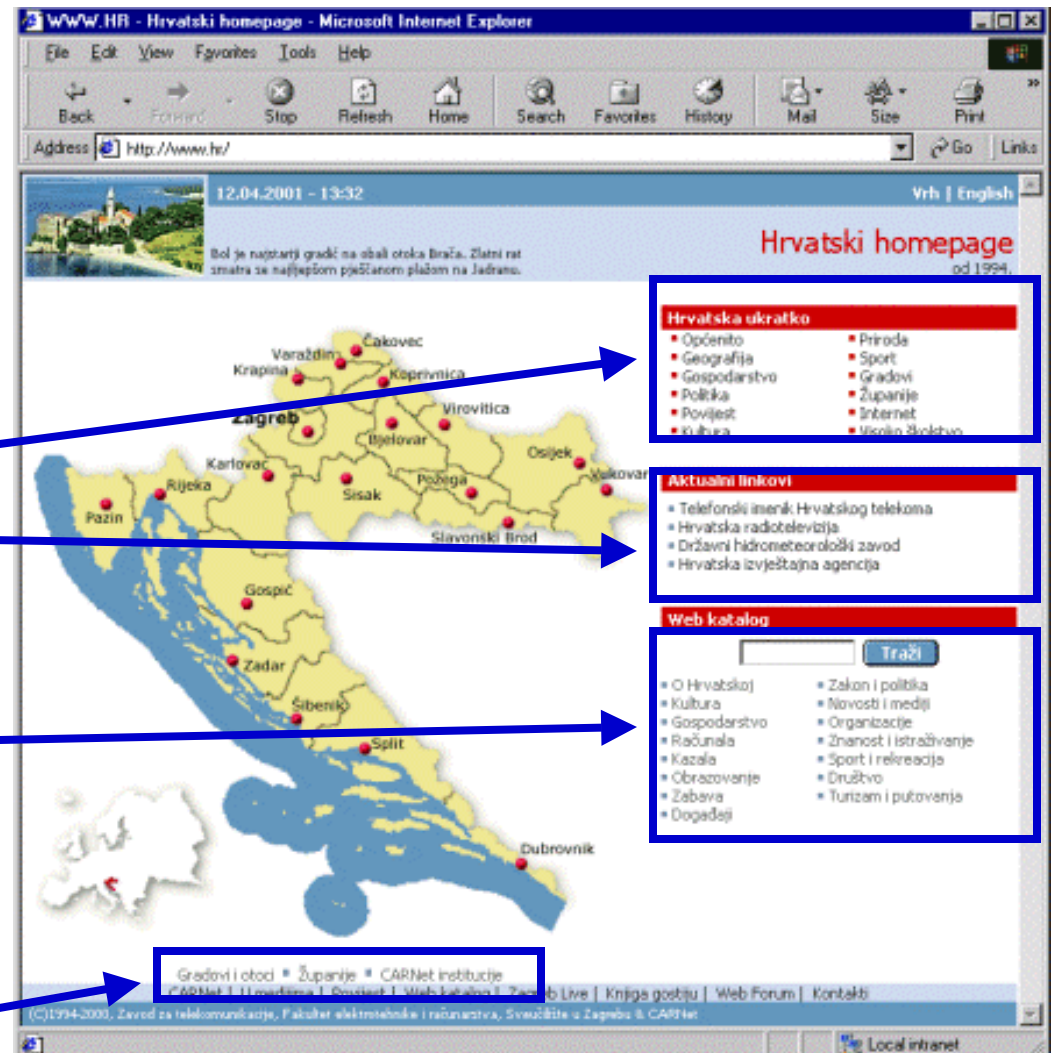
- Official “birthday” February 12th, 1994
- Registered as a “Croatian Homepage” with CERN’s Virtual Library
- In 2/1994, the number of WWW servers in the world was about 4,500
- Project supported by CARNet since 1996
- Awards: magazine PCChip *Top 5 portals in 1999*; magazine BUG *Top 50 in the year 2000*, “...probably the best catalogue of Croatian Web sites...”

Concept of the WWW.HR

- Web-based information service
- Includes two services:
 - General info on Croatia
 - Most important information on national history, tourism, economy, nature, geography, politics, arts, culture, sport, and Internet
 - Development phases: 1994-96, 1996-98, edition 1999, edition 2000, edition 2001
 - Directory of Croatian Web sites
 - Development through 1996, 1998-2000, 2000, 2001

General info on Croatia *Edition 2001*

- Touch-sensitive map
- Thirteen topics under About Croatia
- Useful links
- Main categories from the directory included in the home page
- Three touch-sensitive maps providing easier access to Croatian cities and counties



Directory of Croatian Web sites

Edition 2001

- As of 9/2001, the directory contains about 6,500 links
- Most frequently visited:
 - Tourism and Traveling
 - News, Media and Magazines
 - Education
 - Business and Economy
 - Art and Culture



Directory features

- Integrated, Web-based administration:
 - Webmasters submit their sites to the directory
 - Submitted sites must be thematically related to Croatia
 - Administrator checks the submission
 - Data fields from the submission form are inserted into the database
 - Webmaster receives an e-mail confirmation



The screenshot shows a web browser window titled "WWW.HR - Add URL - Microsoft Internet Explorer". The address bar contains "http://www.hr/www.hr/addurl.ar.html". The page features a navigation menu with "What's new?", "Useful info!", "Add URL!", "Create", and "Browse!". Below the menu is a search bar with a dropdown menu set to "all words", a search input field, a "Whole words" checkbox, and "Search" and "Help" buttons. The main content area is titled "Add URL" and contains two sections: "Webmaster info" and "URL info". Both sections include a red warning: "All the fields are required. Do not use uppercase. For Croatian alphabet, use ISO 8859-2 standard, if available." The "Webmaster info" section has fields for "Name and surname:", "E-mail:", and "Address:". The "URL info" section has a "URL:" field with "http://" entered and a "Site name:" field with instructions: "This is not a host name - use two or three words to name the content of your pages - the company name or similar. English."

Directory features (cont'd)

- static HTML pages, generated by Perl scripts
- URL and category databases kept separately
- Administration:
 - Editing URL properties
 - Cross-linking
 - Listing duplicate URLs, and checking status
 - Date of last change (if available)

Search capabilities

- Search by title or by content description
 - by keyword
 - using a Boolean expression (operators AND, OR, NOT)
- Full support for Croatian (ISO 8859-2) character set



Search capabilities (cont'd)

- All links in the directory are stored in a database
- A search request initiates a database query
- Database query returns a list of all links containing the search pattern(s), sorted by categories in which those links appear
- User can repeat the search using the CARNet's *Croatia Search Service* project's search engine (CROSS)

Why metadata?

- **Problem:** efficient search and retrieval of useful information from Web resources
- **Solution:** Use of metadata!
- **How:** Authors *must* add more information to their Web sites
- WWW.HR and CROSS experiences served as a foundation for CARNet's recommendation on metadata

<ftp://ftp.carnet.hr/pub/CARNet/docs/advisories/CDA0027.doc>

Dublin Core Metadata

- Dublin Core (DC) Metadata Initiative, 1995.
- DC Metadata Element Set (DCMES)
 - Content (Title, Subject, Description, Type Source, Coverage)
 - Intellectual property (Creator, Publisher, Contributor, Rights)
 - Instance (Date, Language, Format, Identifier)
- DCMES is not only for use on the Web - it may be used for all publishing forms
- CARNet recommends using a **subset of DCMES** in the Croatian Webospace

Selected tags

- CDA0027 recommends introduction of 8 DCMES tags:
 - `DC.Title`, `DC.Subject`,
`DC.Description`, `DC.Creator`,
`DC.Contributor`, `DC.Publisher`,
`DC.Date`, `DC.Language`
- Two synonyms were defined, to ease the transition to DCMES:
 - `Keywords` - sinonim za `DC.Subject`
 - `Description` - sinonim za `DC.Description`

Use of DC metadata in www.hr

- The idea is for WWW.HR to lead by example
- Metadata information has been added to all “Short info” pages

```
<META name="DC.Title" content="The Home page of the Republic of Croatia" >
```

```
<META name="DC.Publisher" content="FER, University of Zagreb and CARNet" >
```

```
<META name="DC.Creator" content="Igor Ljubi" >
```

```
<META name="DC.Date.Modified" content="2000-02-17" >
```

Use of DC metadata in www.hr (cont'd)

- DC metadata will soon be added to Directory pages, enrichening the information we provide
- encouraging our users to add DC metadata to their Web sites

Conclusions

- first step in improving search capabilities in www.hr was in cooperation with CARNet's Croatian Search Service (CROSS)
- as a second step, use of DC metadata will allow more efficient searching and information retrieval
- our future work includes adding metadata to the directory
- we encourage Webmasters to add DC metadata elements to their Web sites!