

WWW.HR: First metadata-enabled service in Croatian Webpace

Igor Ljubi, Gordan Gledec

University of Zagreb
Department of Telecommunications
Faculty of Electrical Engineering and Computing
Unska 3, HR-10000 Zagreb

E-mail: {igor, gordan}@tel.fer.hr

Abstract

Since the beginning of the project, in early 1994, WWW.HR has been gathering links to Web sites specifically related to Croatia. It grew to the biggest, and the most complete directory of Croatian Web sites. As of June 2001, there are more than 6200 Web sites in 15 main categories and 382 subcategories. WWW.HR also offers short info on Croatia, covering national history, economics, sports, art and culture, tourism etc.

As well as in the rest of the World Wide Web, searching, and recovering the useful information from the Web-based directory have emerged as two main problems. Using metadata is one way to deal with them. Experiences in creating and maintaining WWW.HR directory, along with those gathered by Croatian Search Service (CROSS), served as guidelines for Croatian Academic and Research Network's recommendation to use metadata in Croatian Webpace. WWW.HR will once again lead by example, and as the first step in implementing that recommendation on our site, metadata information according to the Dublin Core Metadata specification was added to the pages with short information about Croatia.

In this paper, introducing metadata in the pages about Croatia, and especially in the directory will be discussed as a way of enabling better search capabilities and thus adding value to this information service.

Introduction

Since its beginning, in early 1994, the WWW.HR project has been setting the standards for Croatian Web sites. From the start, it was designed to be a place that will provide all the necessary information about the Republic of Croatia. That part of the site is known as "Short info on Croatia". The second objective was to gather links to Web sites that are implicitly or explicitly related to Croatia. In the first few years it was a simple list, which grew so large that it became cumbersome for users to find the necessary information. In 1996 the list became a Directory of Croatian Web sites. For more details, an interested reader is referred to [1], [2], [3], [4].

This paper is organized as follows: Section 1 briefly covers the past of the project and explains its main features. Section 2 deals with metadata and why is it an important in developing a better service for our users. Section 3 presents the changes that have been made on our site to meet the need for metadata and states what yet has to be done and concludes the paper.

1. Seven years of leading Croatia's Internet community

The history of the WWW.HR project can be divided in 2 phases. First phase, from 1994 to 1996 was an unsupported, enthusiastic one, and from 1996 till now, as CARNet's project. Judging from usage statistics and user feedback, we believe that WWW.HR is one of the most popular Web sites in Croatian Web space.

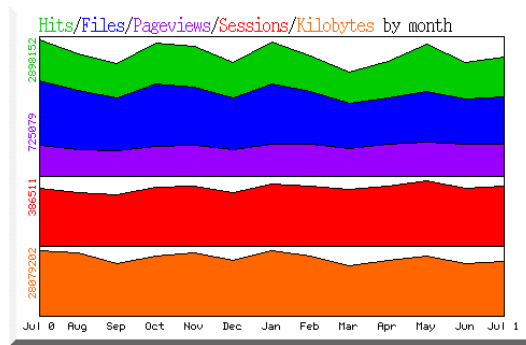


Figure 1. Statistics for Homepage of Croatia – July 2000 – July 2001

Seven years of evolution of the original idea to present Croatia on the World Wide Web resulted in the fact that it is now one of the most visited Web pages in Croatia. The Webmasters aim to have the most accurate information in the fields of interest. We also cooperate with governmental and non-governmental institutions in order to provide better information to our users. Sports and Culture pages were recently updated. Information in "Short info on Croatia" part of the site is being constantly refreshed, and new categories are constantly being added to the Directory. The Directory grows by approximately 250 links a month, and in August 2001 it contained more than 6500 links. All new entries into the Directory are first carefully evaluated by the site's Webmasters, who approve (or deny) the submission of the site into up to 5 appropriate categories. All the newly added sites are listed under "What's new?" category.

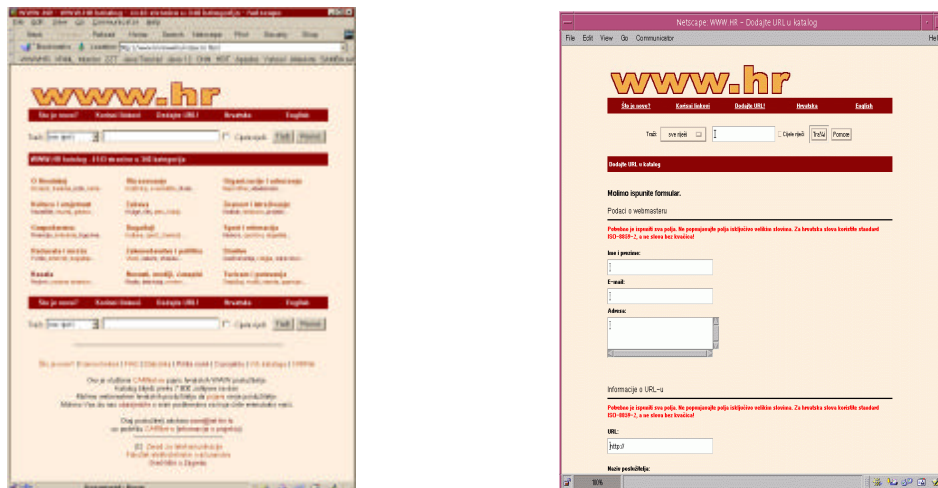


Figure 2. Directory of Croatian Web sites

2. The need for metadata

In 1994, when the project was started, there were about 4600 Web servers connected to the Internet. With the growth of Web's popularity, the number of servers has rapidly increased. Unfortunately, lack of the appropriate mechanisms to provide more accurate searching has resulted in many unusable links as a result of the users' search queries. It is not uncommon that users will receive hundreds of URL's, and will be unable to find the desired information. This is partly due to the inadequate way that Web documents are described. Research in this area shows that by simply improving search mechanisms the problem cannot be solved and the authors of the documents must play an important role. The use of metadata can be of assistance in providing better service in searching the documents presented on the WWW.

Metadata, or "data about data", is a set of attributes that describe the document, and the values of these attributes. Usually, the metadata record has the most common attributes that describe given document: author, title, publishing date, field it covers, and so on.

Many of the current WYSIWYG Web editors automatically add certain metadata to the documents they create, but the more in-depth knowledge of HTML is required to add information that will be useful to the searching robots. The current methods allow robots to more or less accurately guess about the content of the documents on the WWW. The result is, as it was stated before, almost complete chaos in the search results.

Based on the experiences of the WWW.HR and the Croatian Search Service, CROSS, CARNet has issued a recommendation for enhancing the searchability of the Croatian Web space by using metadata. As a de facto standard, Dublin Core Metadata Element Set was recommended to be used in Croatian Web space.

Dublin Core Metadata Initiative started with first workshop in Dublin, Ohio, USA, which gathered librarians, content experts and IT specialists. The joint efforts of these multidisciplinary groups resulted in consensus about the set of describing elements, know as Dublin Core Metadata Element Set (DCMES). DCMES is not designed specifically for the use on the Web, it can be used for all forms of publishing.

DCMES has 15 metadata elements. They can be divided into 3 groups (table 1.). Elements can relate to the content of the item, some to the item as intellectual property, still others to the particular instantiation, or version, of the item.

Content		Intellectual property	Instantion
Title	Source	Creator	Date
Subject	Relation	Publisher	Language
Description	Coverage	Contributor	Format
Type		Rights	Identifier

Table 1 - Dublin Core Metadata Element Set

The recommendation also defines the subset of DC elements that should be used in Croatian Web space. The maximum length of the element's values, that will WWW.HR and CROSS implement, is also defined.

3. Creating first Croatian metadata enabled Web service

From the previous section it is obvious that serious involvement from the creators of the documents on the WWW is needed in order for metadata concept to be successful. It is the authors who must adopt this concept, and apply it on their Web sites. It is a slow process, and someone has to be the driving force. In this case, WWW.HR decided to lead once again by example. As many times before, WWW.HR was the first Web service in Croatia that offered Dublin Core metadata enabled Web service, thus adding the value to it. The team implemented CARNet's recommendation for using DCMES.

Every page in "Short info on Croatia" part now has metadata information as a part of its content, as shown in the next example from the Croatian homepage:

```
<LINK rel="scheme.DC" href="http://purl.org/dc/elements/1.1/">
<META name="DC.Title" content="WWW.HR - Croatian Homepage">
<META name="DC.Description" content="The homepage of the Republic
of Croatia provides useful information about the country -
general info, tourist info, geography, history, politics,
culture, nature, sport, cities, internet, economy... It's been up
and running since February 1994.">
<META name="DC.Subject" content="Republic of Croatia, Croatian
Homepage">
<META name="DC.Creator" content="Department of
Telecommunications, FER">
<META name="DC.Publisher" content="ZZT, FER and CARNet ">
<META name="DC.Language" content="en">
```

Example 1. Metadata information on Homepage of the Croatia

It is clear that we manage to completely describe our pages about Croatia using just 6 metadata elements, and now the searching, and retrieving of our pages is simpler and more efficient by Dublin Core Metadata supported robots.

The next step is to add metadata information to our directory pages, and to encourage the authors of the Web sites listed in our Directory to do the same on their pages. From the previous example, it is clear that adding metadata is really not that complicated. It requires just a little knowledge and a bit of good will, but it can result in better searchability and consequently an increased traffic for the Web site. Hopefully, the Internet users will also benefit from this data, because they will spend less time in browsing the searching results, and will have more accurate results sooner.

The WWW.HR project team will continue to provide both the basic information about the Croatia, as well as the Directory of the Croatian Web pages. We plan to continue our support for the metadata initiative, and will try to improve our services with the latest technologies.

References

- [1] J. Juric, M. Matijašević, M. Mikuc, "*First experiences with the World-Wide-Web in Croatia*", 2nd International Conference on Telecommunication Information Systems, TIS '94, Žilina, Slovakia, November 1994.
- [2] J. Juric, M. Matijašević, M. Mikuc, "*Implementation of a Distributed Hypermedia Information System Based on the WWW Concept*" (in Croatian) Proceedings of SoftCOM '95 – Conference on software in telecommunications and computer networks, pp. 85 - 92, Split, Croatia, 1995
- [3] G. Gledec, J. Juric, M. Matijašević, M. Mikuc, "WWW.HR – Experiences with Web-server Development and Maintenance", Proceedings of the XXII. International Convention MIPRO 99, p. 83-86, Opatija, 1999.
- [4] I. Ljubi, G. Gledec, "WWW.HR – An entry point to the Croatian Cyberspace", Proceedings of 2nd CARNet Users Conference CUC 2000, on CD, 2000.
- [5] Dublin Core Metadata Element Set, Version 1.1. 1999-07-02
<<http://purl.org/dc/documents/rec-dces-19990702.htm>>

Short biography:

IGOR LJUBI is an assistant at the Department of Telecommunications, Faculty of Electrical Engineering and Computing, University of Zagreb. He received his B. Sc. at the Faculty of Electrical Engineering and Computing, University of Zagreb, in 1999. His research interests include Mobile Agents, Internet technologies, and Java. He is member of IEEE. He is an associate at the Department project [WWW.HR](#) - Homepage of Croatia since 1999.

GORDAN GLEDEC is an assistant at the Department of Telecommunications, Faculty of Electrical Engineering and Computing, University of Zagreb. He received his B. Sc. and M.Sc. degrees at the Department of Telecommunications in 1996. and 2000, respectively. His research interests include Internet technologies and Web services. He has been an associate at the WWW.HR project since July 1997.