

WWW.HR:

A tool for editing Dublin Core metadata information in the Croatian Web space

Gordana Stojšić, Igor Ljubi

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

CUC 2002

September 25 – 27, 2002

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

Dublin Core Metadata Element Set

- 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 Webspace

Available Metadata Generators

- State of the art was studied before designing
- Criteria:
 - form that has to be filled in order to generate metadata
 - extracting metadata from HTML
 - can convert DC metadata in other types of metadata
 - editing metadata
 - multi language support

Available Metadata Generators

(2)

	Form metadata	Metadata extraction from HTML	Converting schemes	Editing	Multi language support
<i>Nordic DC metadata creator</i>	+		+	+	
<i>DC-dot</i>		+	+	+	
Reggie		+	+	+	+
Klarity		+			

Available Metadata Generators

(3)

– Commercial tools:

- Metabrowser
 - Create, edit metadata, supports various schemes (DC, GILS, AGLS, EdNA)
 - <http://metabrowser.spirit.net.au/>
- Gen – Dublin Core Edition
 - GUI for creating metadata
 - Creates other type of metadata for search engines
 - Automated adding of created metadata
 - <http://bridges.state.mn.us/taggen.html>

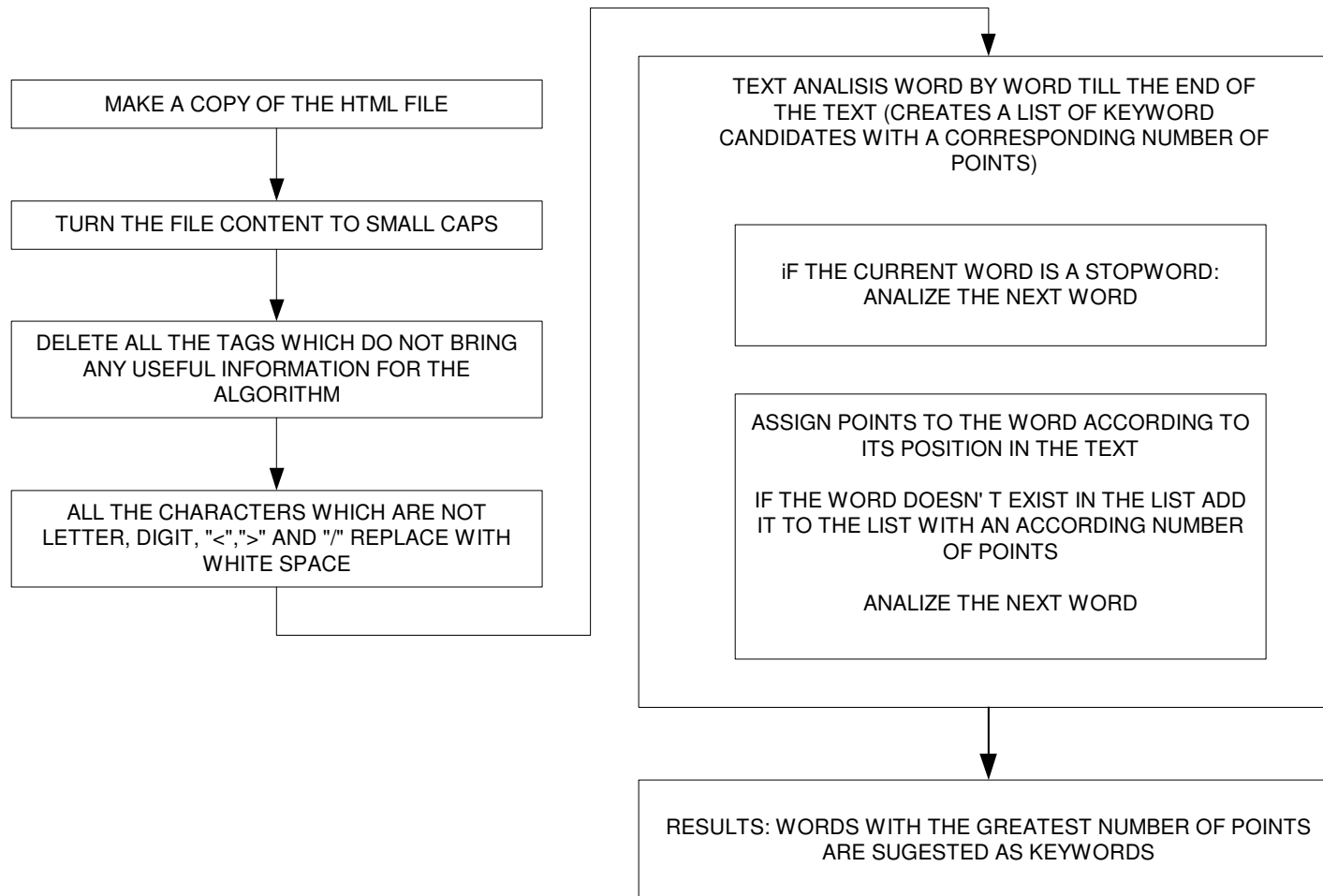
Requests on the tool

- Creating and editing of metadata
- Support for Croatian specific characters
- Dublin Core scheme using CDA0027 recommendation
- Automatic keyword generating
- Creating text file with generated metadata
- Functions for editing metadata

Requests on the tool (2)



Keywords extraction algorithm



Keywords extraction algorithm (2)

- Number of points assigned to a word for its appearance in the text:
 - Appearance in a <TITLE> tag: 23 points
 - Appearance in a <H1> tag: 15 points
 - Appearance in a <H2> tag: 10 points
 - Appearance in a <H3> tag: 6 points
 - Appearance in the rest of the text: 2 points

Application on the content of the WWW.HR

Seminar - [D:\Diplomski\Kodd\kodd\uzorci\korcula]

Datoteka Alati Pomoc

DC.Title KORČULA

DC.Subject korčula, stoljeća, korčuli, otoku, godine, tijekom, grad, otoka, stoljeću, zajedno, cijelom, sveti, korčule, grada

DC.Description O gradu i otoku Korčula, povijesni pregled

DC.Creator
Dodaj
Obrisi

DC.Contributor Gordana Stojšić

Dodaj
Obrisi

DC.Publisher FER, ZTT

DC.Language hr

DC.Date.Created **DC.Date.Modified** 2002-05-29 **DC.Date.Issued**

Application on the content of the WWW.HR (2)

- Generated metadata:

<META name="DC.Title" content="KORČULA">

<META name="DC.Description" content="O gradu i otoku Korčula, povijesni pregled">

<META name="DC.Subject" content="korčula, stoljeća, korčuli, otoku, godine, tijekom, grad, otoka, stoljeću, zajedno, cijelom, sveti, korčule, grada, pod, vlast, živjeli, tom, ">

<META name="DC.Creator" content="Gordana Stojšić">

<META name="DC.Publisher" content="FER, ZZT">

<META name="DC.Language" content="hr">

<META name="DC.Date.Modified" content="2002-05-29">

<META name="DC.Contributor" content="Maja Matijašević">

<META name="DC.Contributor" content="Gordan Gledec">

Conclusions

- WWW.HR – Croatian Web informational service
- New tool to add DC metadata to your pages
- Continue our efforts to enable better searchability of Croatian Web space