Croatian Web portals: from obscurity to maturity

Radovan Vrana
Faculty of humanities and social sciences, University of Zagreb
rvrana@ffzg.hr

Inge Rudomino
National and university library Zagreb
irudomino@nsk.hr
Introduction

• Web portals among the most popular type of Web sites in Croatia (according to Gemius ratings)

• From simple Web pages to very complex software structures

• The Croatian Web Archive – central point in Croatia for research of Web site development (for .hr domain)

• Research problem: tracking and understanding development of Web portals over time – growth in content and complexity of implemented Web technologies + design changes
The Croatian Web Archive

- Established as full service in 2004 in collaboration with the University of Zagreb University Computing Centre (Srce)

- Based on the concept of selective capturing of Web resources up to 2011 - by use of home made crawler called DAMP

- From 2011 harvesting of the Croatian Internet domain (.hr) was done annually .hr + building thematic collection (10 collections) - by use of HERITRIX

- Online access to the Web archive publicly available
Analysis of the Croatian Web portals

• Steps in the analysis:
  • Select **10 most popular / most visited Web portals in Croatia** (Gemius ratings)
  • Retrieve **two archived copies per year of portals from the Croatian Web Archive**
  • Test the archived copies of Web portals in Croatia against a group of parameters:
    • **Size in the archive** (of the whole portal) and on client / user side at the level of the starting page of a Web portal
    • **Loading time** (starting page)
    • **Size of content by types: HTML, CSS, JS and Images** (starting page)
    • **Type of a portal** (mostly news portals)
    • **Correct and complete copies of Web portals in the Croatian Web Archive**
    • Design changes in Web portals
  • Drawing the conclusion about changes made in the last 15 years (or less if the portal was not archived for such a long period of time)
Analysis

Web portals | Web archiving | Web technologies | Croatia
Analysis
Analysis

• Software used for preparation and analysis
  • Software used in the Croatian Web archive custom made by the University computing centre in Zagreb, Croatia
    • Selection of the archived versions of portals
  • Software used on the client / user side: Google chrome Web browser
    • Developer tools – simple yet powerful

• All other software discarded due to insufficient functions for the harvested Web site analysis
Results
Web content - HTML (in MB)
Web content - CSS (in MB)
Web content - JS (in MB)
Web content - images (in MB)
Web page loading time (in seconds)
Disk usage – client / user side (in MB)
Disk usage - complete copy of a Web portal at admin side (in MB)
Social networks links on Web portals
## Design, title and URL change

<table>
<thead>
<tr>
<th>Year</th>
<th>24sata</th>
<th>vecemjl.hr</th>
<th>tportal</th>
<th>dnevnik.hr</th>
<th>net.hr</th>
<th>poslovni.hr</th>
<th>novlist.hr</th>
<th>express.hr</th>
<th>telegram</th>
<th>rtl.hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006_1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008_2</td>
<td></td>
<td></td>
<td></td>
<td>new design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>title change, URL change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009_1</td>
<td>new design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009_2</td>
<td></td>
<td></td>
<td></td>
<td>title change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010_1</td>
<td>new design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010_2</td>
<td></td>
<td></td>
<td></td>
<td>new design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011_1</td>
<td></td>
<td></td>
<td></td>
<td>new design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011_2</td>
<td></td>
<td></td>
<td></td>
<td>new design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012_1</td>
<td></td>
<td></td>
<td></td>
<td>new design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012_2</td>
<td></td>
<td></td>
<td></td>
<td>new design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013_1</td>
<td>new design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015_1</td>
<td></td>
<td></td>
<td></td>
<td>new design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015_2</td>
<td></td>
<td></td>
<td></td>
<td>title change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016_1</td>
<td>new design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016_2</td>
<td></td>
<td></td>
<td></td>
<td>new design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017_1</td>
<td>new design</td>
<td></td>
<td></td>
<td>new design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017_2</td>
<td></td>
<td></td>
<td></td>
<td>new design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

• Web portals constantly adopting new technologies

• Growth in size of important elements of the Croatian Web portals

• Design changes to adapt to new types of devices

• Challenge: optimal design, size and technologies
Thank you!