Archiving the web editions of Croatian national newspapers – Slobodna Dalmacija readers' comments

Mirko Duć
Research Assistant, Department of Information Sciences, University of Zadar
Email: miduic@unizd.hr

Abstract
As one of the Legal deposit libraries in Croatia, National and University Library (NUL) in Zagreb is archiving web editions of major national newspapers. Except the NUL no other library or archiving institution in Croatia is involved in archiving of this content. The objective of this poster was to investigate how many web newspaper articles and their readers' comments in “Slobodna Dalmacija” national newspaper have been preserved or lost due to the existing system of archiving. 
I have compared two web sites which contain old issues of national newspaper „Slobodna Dalmacija” : NUL website and „Slobodna Dalmacija” newspaper web site. By comparing number of articles and readers' comments archived in a period of one month in heading “Šibenik” of “Slobodna Dalmacija” newspaper it was found out that archiving situation at these two web sites is very problematic because great majority of online content of the “Slobodna Dalmacija” newspaper is not archived.
To determine value of readers’ comments from 60 articles published in above-mentioned period I analyzed these readers’ comments using following elements: furthering knowledge; article evaluation; journalist evaluation; readers’ discussion; engagement call; various elements; advertisement; web links; insults. It was found out that there are considerable amount of valuable readers’ comments which aren’t archived.
The purpose of this research is to direct attention to the need of the more comprehensive archiving of web editions of national newspapers by the NUL or any other library/archiving institution.

Keywords: web archiving, Croatian national newspapers, readers’ comments, Slobodna Dalmacija, National and University Library
**Introduction**

As one of the Legal deposit libraries in Croatia, National and University Library (NUL) in Zagreb is archiving every printed issue of all major national newspapers. The NUL is also making the effort to archive web editions of these newspapers. However, in web archiving of these newspapers its efforts are limited – the NUL is archiving only some of the issues of some of the national newspapers. For example, only one or two issues of web editions of newspaper are archived in a week. Newspaper issues for other days of the week are not archived. Except the NUL no other library or archiving institution in Croatia is involved in archiving of web editions of Croatian national newspapers. Without archiving of all web issues of Croatian national newspapers one important part of content will be lost forever - readers' comments.

Almost every web edition of a newspaper article is supplemented by readers' comments. Often, there are a lot of comments. Their length measured in a number of words is often longer than the article to which they refer. Sometimes readers' comments are not very informative and often they are insulting but frequently they also give added informative value to the articles. They broaden perspective, give opinions and information about topic in question and sometimes they criticize biased view of the author or interviewed persons.

The topic of this poster will be to investigate how many web newspaper articles and their readers' comments have been lost due to the existing system of archiving.

The purpose of this poster is to direct attention to the need of comprehensive archiving of web editions of national newspapers by the NUL, or any other library/archiving institution.

**Method of research - summary**

I have compared archives of articles from two web sites which contain old issues of national newspaper „Slobodna Dalmacija“. At these two web sites (National and University Library's web archive and „Slobodna Dalmacija“ newspaper) old articles of „Slobodna Dalmacija“ newspaper can be found.

I have determined the number of readers' comments that were published online in a period of one month (from February 3, 2012 to March 3, 2012) in heading “Šibenik” of “Slobodna Dalmacija” newspaper. The day at which I made this inquiry was May 3, 2012.

I also determined how many articles and readers' comments were present at National and University Library's web archive, that were published in the same time period.

None of the readers’ comments and only one full-text article were found in NUL archive.

By comparing number of articles and readers' comments archived in analyzed period at these two web sites I found out that archiving situation of online content of the “Slobodna Dalmacija” newspaper is problematic because great majority of online content is not archived.

I also wanted to determine the value of readers' comments and to do that I analyzed readers' comments from 60 articles published in above-mentioned period. I analyzed these readers’ comments using following elements: furthering knowledge; article evaluation; journalist evaluation; readers’ discussion; engagement call; various elements; advertisement; web links; insults.

**Method of research – in more detail**

183 readers’ comments found in 60 articles of „Slobodna Dalmacija“ newspaper archive were analyzed. These articles were originally published in a period of one month (from February 3, 2012 to March 3, 2012) and they were published in the heading “Šibenik” which is filled with articles related to town of Šibenik and Šibenik-Knin County.

I choose articles from this heading is because I was living in Šibenik for a long time. For that reason I’m following regularly heading “Šibenik” in “Slobodna Dalmacija” newspaper and also news about Šibenik in other media. Awareness about situation and events in that region gives me more reliable position to judge value of readers’ comments in heading “Šibenik”.

The reason why the oldest article of which I have analyzed readers’ comments is published in February 3, 2012, is very simple. The day I analyzed heading “Šibenik” at “Slobodna Dalmacija” newspaper web site (May 3, 2012), there weren’t any other articles published before February 3, 2012. Only articles that were published in a last three month could be accessed. Every article that was published farther in the past is automatically removed from the web site.
I have chosen to analyze readers’ comments from 60 articles because I think that’s appropriate number of articles to get a good general picture about readers’ comments value. I found that for these 60 articles total number of readers’ comments was 183.

To evaluate these 183 readers’ comments I analyzed each readers’ comment by looking for positive, neutral and negative elements contained in it. Positive elements which I have used in analysis were:

- **furthering knowledge** element (this element is present in particular readers’ comment if the comment contains some additional information related to article content)
- **article’s evaluation** element (this element is present in particular readers’ comment if the comment contains reader’s evaluation of article’s content)
- **journalist’s evaluation** element (this element is present in particular readers’ comment if the comment contains readers’ evaluation of the work of article’s writer/journalist)
- **readers’ discussion** element (this element is present in particular readers’ comment if the comment is part of a discussion between readers)
- **humour** element (this element is present in particular readers’ comment if the comment is humorous)
- **engagement call** element (this element is present in particular readers’ comment if the comment contains readers’ engagement call to other readers, for example call to donate a money for hospital expenses of a person whose accident is described in article)

Neutral elements which I have used were:

- **other** element (this element is present in particular readers’ comment if the comment contains information which is not related to article’s content, for example, if it contains poems)
- **advertisement** element (this element is present in particular readers’ comment if the comment contains advertisement for some product or service)
- **web link** element (this element is present in particular readers’ comment if the comment contains web link)

I also used one negative element in a process of readers’ comment evaluation:

- **insult** element (this element is present in particular readers’ comment if the comment contains discussion between readers which deviates into verbal abuse)

I graded these elements in a following way:

If some readers’ comment contains information which I thought could contribute to better understanding of related article, than I would give a grade “one” to related element (in this case “furthering knowledge” element). The same readers’ comment could contain other elements. If the comment is, for example, humourous, I would give a grade “one” to “humour” element for that readers’ comment.

The same readers’ comment can also contain other elements, for example, negative “insult” element. In that case readers’ comment will have “insult” element graded with “one”.

Therefore, each readers’ comment can contain one or several elements (which are positive, neutral or negative) and each of these elements, when it is present in the comment, will be graded with “one”. Elements that are not contained in a particular readers’ comment are not graded.

There is a difference between positive, neutral and negative elements.

Positive elements mostly indicate readers’ comments which will enable better understanding of related articles.

Neutral elements indicate readers’ comments which are mostly neutral in that regard (with exception of some of the web links).

There is one negative element - **insult** element – which indicate readers’ comments that aren’t enabling better understanding of related article.

All these elements were not invented before analysis but were identified/constructed during the process of analysis.
Findings and Discussion

Exploration of the only two “Slobodna Dalmacija” newspaper public archives of online published articles, has identified troublesome archiving situation.

At “Slobodna Dalmacija” newspaper web site, only articles that were published in a last three month are accessible. Each article and its readers’ comments that was published farther away in the past is regularly removed from the web site.

Situation at National and University Library's (NUL) web archive isn’t much better. Here online published articles of “Slobodna Dalmacija” newspaper are extremely rarely archived in a full text version. In NUL archive only the inventories of published articles are available. In these inventories there are very few data about each published article: article title; few words from the article; mini picture related to article (only in some cases); date of publishing. Mostly there is no access to full text of “archived” articles nor readers’ comments. Therefore NUL web archive can only be used for rough orientation about “Slobodna Dalmacija” newspaper online published articles. Anyone who need to access full text of articles or readers’ comments, in great majority of cases will not find this content in NUL web archive.

Another finding is that readers’ comments contain quite a valuable content. In analyzed corpus of 60 articles there were 183 readers’ comments which contained 277 instances of either positive, neutral or negative elements.

Majority of readers’ comments contained positive elements (furthering knowledge; article evaluation; journalist evaluation; readers’ discussion; engagement call; humour) – 205 of the total of 277 instances of elements were instances of positive elements.

Among positive elements, highest number of instances was found for “article evaluation” element – 85.

There were 48 instances of “readers’ discussion” element; 40 instances of “furthering knowledge” element; 16 instances of “humour” element; 12 instances of “journalist evaluation” element; 4 instances of “engagement call” element.

24 of the total of 277 instances of elements were instances of neutral elements (“other” element; “advertisement” element; “web link” element).

Among neutral elements, highest number of instances was found for “other” element – 11 (almost all of the readers’ comment with this element were poems and all these poems were readers’ comments related to just one article).

There were 7 instances of “web link” element and 6 instances of “advertisement” element.

48 of the total of 277 elements were instances of negative element - “insult”.

There are 17,3 % instances of “insult” negative element with regard to total number of instances of all elements. There are 8,7 % instances of all the neutral elements in relation to total number of instances of all the elements. There are 74 % instances all the positive elements in relation to total number of instances of all the elements.

With the convincingly highest percentage of positive elements in a relation to the total number of instances of all the elements - 74 % - it is evident that instances of positive elements are undoubtedly prevailing.

As positive elements of readers’ comments are most valuable for better understanding of related articles, it could be concluded that readers’ comments are predominantly valuable for better understanding of articles.

Only one type of instances in readers’ comments is without almost any direct value for better understanding of articles: instances of “insult” negative element.

Concerning the fact that there are 17,3 % instances of “insults” negative element in analyzed readers’ comments, we can say that their presence is not negligible but at the same time their presence is fairly limited in relation to instances of all the other elements especially the most valuable instances of positive elements.
**Conclusion**

It could be concluded that there exists significant number of readers’ comments in “Slobodna Dalmacija” newspaper online published articles. Significant majority of these readers’ comments contain content that is valuable for better understanding of related articles. Therefore it would be recommendable to permanently archive all online articles together with their readers’ comments. Unfortunately, current situation with archiving of online published articles and their readers’ comments is rather desperate. Without some initiatives for comprehensive, permanent archiving the process of certain disappearance of this content will not be reversed.

* * *

**Curriculum Vitae – Mirko Duić**

Year of birth: 1980

- Student of postgraduate scientific study programme “Knowledge society and information transfer” (University of Zadar, from year 2009)
- Postgraduate degree in Journalism – MA (Faculty of Political Science, University of Zagreb, Croatia, 1998 – 2003)
- Postgraduate degree in Film and video editing - MA (Academy of Dramatic Art, University of Zagreb, 1998 – 2004)

Employed as the research assistant at the University of Zadar, Department of Library and Information Sciences.

**Participation at conferences, workshops, seminars, meetings**

Year 2012

- Summer School in User Studies – Ljetna škola o korisnicima (11 - 14 April, Zadar, conference participant and member of the organization team)
- Workshop for Mendeley software (22 February, Zadar)

Year 2011

- ”The Future of the Past of the Web” conference (7 October, London, UK)
- Summer school of manuscripts - Ljetna škola o rukopisima (26 – 30 September, Zadar; conference participant and member of the organization team)
- Summer school in publishing - Ljetna škola nakladništva (19 – 24 September, Zadar)
- Summer school of scientific communication - Ljetna škola znanstvene komunikacije (22 -26 August, Split, 4 ECTS)
- Workshop for max/MSP/jitter software (9-13 May, Zagreb)
• meeting „How to enable equal access to higher education to students with invalidity“ / „Kako studentima s invaliditetom omogućiti jednak pristup visokom obrazovanju“ (4 April, Zagreb, meeting was held as a part of the Tempus project: „Eduquality“)

• ISBD/XML meeting and seminar (25. February, Edinburgh)

Year 2010

• TEI Applied: Digital Texts and Language Resources (8-14 November, Zadar)
  o Understanding ODD: A full introduction to the TEI's own metalanguage (TEI@Oxford team) (workshop which was held as a part of TEI conference)
  o TEI for Transcriptions of Speech (workshop which was held as a part of TEI conference)

• IPRES 2010 - 7th International Conference on Preservation of Digital Objects (19-24. September, Austria)
  o IWAW - International Web Archiving Workshop (22-23 September, Austria, workshop which was held as a part of IPRES conference)
  o Expert User Group Forum for Heritrix Operators and Developers (24 September, Austria, seminar which was held as a part of IPRES conference)

• LIDA – Libraries in the digital age (25-28 May, Zadar; conference participant and member of the organization team; for the poster session of the conference I made poster: “British printed images and collections of prints from Croatia”)

Year 2009.

• Summer school of old book - Ljetna škola stare knjige (28 September – 2 October, Zadar; conference participant and member of the organization team)

• LIDA – Libraries in the digital age (25-30 May, Dubrovnik and Zadar; conference participant and member of the organization team)

Project

Croatian written heritage: theoretical and technological requirements for organisation and protection  (Research assistant)