Aleksandra Trtovac, Nataša Dakić

University Library "Svetozar Marković" Belgrade, Serbia

BRINGING HISTORICAL ADVERTISEMENTS INTO RESEARCH FOCUS – EUROPEANA NEWSPAPERS PROJECT

Abstract: The Europeana Newspapers Project will offer users worldwide over 18 million digitized pages of newspapers, of which 10 million will be available in full text. The basic idea of the project is to provide, through the Europeana portal, an integrated approach to newspaper collections from various European countries. The University Library "Svetozar Marković" participates in the project with more than 40 newspaper titles and over 400,000 digitized pages issued in the period from 1850 to 1945, mostly printed in Cyrillic. As an added value to the project, librarians of the The University Library "Svetozar Marković" analytically processed over 2,000 component parts (ca. 500 advertisements) from old Serbian newspapers so that the component parts are searchable through corresponding metadata and annotations. The emphasis is placed on the processing of daily news, stories, poems, short news, announcements, advertisements, obituaries, public calls, photos, etc. This paper will analyze the contribution to the project especially in the light of open access to the full text of newspaper advertisements. The content of this type of component parts fully reflects the spirit of the past, indicates development of certain industries, but also covers all aspects of cultural and social life since advertisements contain invitations to important events, manifestations, ceremonies, book promotions, but also job offers, offers for various types of services, selling furniture, food, medicine, land and property. Text within advertisements provides excellent material for researchers in the area of cultural history and theory of marketing, especially from the aspect of the history of advertising and marketing. Before the appearance of radio and television newspapers played a key role in the development of advertising and marketing.

Keywords: Europeana Newspapers, digitization, Serbian historical newspapers, advertisements, metadata, OCR

1. From the history of advertisements

Advertisement is a specific form of communication which its author uses when trying to convince the customers to purchase products or services. Modern advertising developed with mass production in the late 19th and early 20th centurirs, just at the beginning of mass production and branding. The product is advertised with the name, an image and numerous positive qualities that are aimed at the customer to always make a positive connection between the good qualities and certain brands.

In this article we will elaborate on historical printed promotions and advertisements published in Serbian newspapers in the second half of the 19th and the first half of the 20th centuries as a source of data for research in various fields. Through the project *Europeana Newspapers*, these advertisements are freely available in digitized form in full text.

2. Europeana Newspapers Project

Europeana Newspapers Project gathered 18 full partners, including some of the leading European national and scientific libraries, as well as LIBER, German company CCS¹ and 11 associated partners. Some of project partners are national libraries of France, UK, Germany, Poland, Netherlands, Slovenia, Croatia, Finland, Estonia, Iceland and university libraries of

¹ Content Conversion Specialists.

Innsbruck, Salford and Belgrade. The Europeana portal will provide to its users access to metadata and full text search and information retrieval of over 18 million digitized pages of newspapers. The project *Europeana Newspapers* lasts from 2012 until 2015.²

The project offers OCR (Optical Character Recognition), OLR (Optical Layout Recognition), NER (Named Entity Recognition) technologies as well as recognition of class pages for the purpose of improving search and information retrieval. Use of OLR is very important in the process of structuring newspaper articles. Optical layout and metadata recognition are based on the software docWorks³. General rule system enables recognition of words, text lines, text blocks, columns, illustrations, advertisements, tables and the following page types: title page, content page, illustration page (a page that has at least one illustration), a page that contains only advertisements. Structural analyses through classification of headlines including article continuation are possible. docWorks enables conversion of recognized articles metadata in METS XML format.

3. Role and participation of the University Library "Svetozar Marković"

The University Library "Svetozar Marković" is the oldest and largest academic library in Serbia, the central library of the University of Belgrade and the main library for all university and academic libraries in Serbia, as well as specialized libraries in academic institutions in Central Serbia. The University Library in Belgrade is considered to be one of the biggest promoters of open access in Serbia and has been involved in several projects and initiatives during the last decade that offer free data sharing. Participating in the project *Europeana Libraries: Aggregating digital content from Europe's libraries* implied digitizing two collections: the collection of Alexander the Great, dating from the 15th to the 19th centuries and the collection of Oriental manuscripts, mainly written in Arabic, Ottoman Turkish and Persian dating from the 13th to the 18th centuries.

The University Library was included in the *Europeana Newspapers Project* on the recommendation of LIBER based on good results achieved so far, mainly through participation in the project *Europeana Libraries*.

The University Library "Svetozar Marković", one of the full partners in the project, participates with more than 40 newspaper titles and over 400,000 digitized pages issued in the period from 1850 to 1945. All these digitized newspapers will be in public domain, since at the end of the project all newspaper issues will be over 70 years old, which is in accordance with the copyright law. Use of optical character recognition will enable data search by key words across entire text, including captions and explanations within pictures and photos.

3.1. Criteria for the selection of the digitized materials. The list of newspaper titles for which the University Library delivers metadata and allows full text search, points to the nature of the development of this type of publications. In addition to specialized newspapers – literary, agricultural, religious, military, political, and so on, there were also newspapers that brought daily news of general type and were used for information and entertainment. The overall impression of the lifestyle of the period, daily political news and advertisements are easily seen through Europeana by browsing the contents of digitized newspapers such as

² Europeana Newspapers Project. Available at: http://www.europeana-newspapers.eu/

³ docWorks is a software used by the most renowned libraries, publishing houses, and companies worldwide to digitize and convert their valuable holdings and archives for easy access, searchability, and long-term preservation. Available at: http://content-conversion.com/?lang=en#docworks-2

⁴ University Library "Svetozar Marković". Available at: http://ubsm.bg.ac.rs/engleski/strana/69/misija-vizija-i-vrednosti

Vreme, Pravda, Srpsko kolo, Beogradske opštinske novine, Nedelja, Zvezda, Male novine, Peštansko-budimski skoroteča, Srpska pčela,⁵ etc.

Since many historical Serbian newspapers are already digitized and they are an integral part of the Digital National Library of Serbia⁶ and the Digital Library of the Matica Srpska Library,⁷ it was important to establish good criteria for selection of digitized material for further processing and refinement of the text.

The selection criteria were:

- Titles that were not yet digitized in the National Library of Serbia and could not be found on the portal Digital National Library of Serbia.
- Availability of good digital copies which could enable quality text processing, especially OCR, and provide quality information to the end users.
- Technical limitations (poor microfilm copies, missing editions, incomplete supplements). This criterion implies that incomplete volumes or bad microfilm copies were not included. We tried to complete volumes through lending issues from the network of university libraries and in several cases we succeeded, so those issues were digitized afterwards.

3.2. Processing of newspapers within the Project. For content provided by the University Library "Svetozar Marković", OCR was planned for all 408,000 pages, and it will be done by the University of Innsbruck. In May 2013 the first 200,000 pages were delivered to the University of Innsbruck for OCR. Since the newspapers delivered by the University Library to Europeana are mostly printed in Cyrillic, including the old Cyrillic alphabet, which contains Russian letters, we were very pleased with the results of the OCR testing done by the University of Innsbruck with over 90% success in character recognition. The characters were not recognized in ca. 3%. The biggest problems presented the Old Serbian Cyrillic characters. The OCR problems appeared mostly on pages with poor scanning quality. However, for the purpose of evaluating the software created at the University of Salford we sent 50 pages and the results were excellent. The University of Salford developed the tool Aletheia – An Advanced Document Layout and Text Ground-Truthing System for Production Environments. This tool provides Automatic Layout Analysis and OCR using Tesseract OCR Engine v 3.028. By using the Aletheia in two testing iterations the OCR was corrected in ca. 99.8% (in a sample of 50 tested pages).

The use of the OLR and NER technologies was not originally planned for materials provided by the University Library "Svetozar Marković". But, there is a possibility for students of the Department of Library and Information Science at University of Belgrade to engage in the work related to the named entity recognition for Serbian language (regarding personal names, place names and names of organizations). The use of named entity recognition in Serbian language was discussed in detail during the workshop in Belgrade, held

-

⁵ University Library "Svetozar Marković" - Digital collection of the newspapers. Available at: http://ubsm.bg.ac.rs/cirilica/zbirka/novina/

⁶ Digital National Library of Serbia - Digital collection of the newspapers. Available at: http://serbia-forum.mi.sanu.ac.rs/Webbook.jsp?setLang=en

⁷ Digital Library of the Matica Srpska Library - Digital collection of the newspapers. Available at: http://digital.bms.rs/ebiblioteka/publications/index/collection:10

⁸ Tesseract OCR Engine v 3.02. Available at: https://code.google.com/p/tesseract-ocr/

on 13th and 14th June ⁹. Apart from German, English and Dutch, it was decided that NER should be made for French, Lithuanian and Serbian language.

Improved by the University of Innsbruck, the METS/ALTO¹⁰ format (format for bibliographic, administrative and technical metadata) will be used for data entry and metadata display of each issue of provided newspapers. It was agreed to export records for each newspaper title from the University Library database COBISS in Dublin Core and to deliver this metadata set to the University of Innsbruck for further metadata processing through the METS/ALTO, together with digital objects.

4. Added value to the *Europeana Newspapers Project* – analytical processing of component parts in Serbian newspapers

At the beginning of the project, the University Library thought how to improve detection, visibility and accessibility of the content of the old Serbian newspapers to the Europeana users. The diversity of component parts of historical newspapers was large – daily news, stories, poems, short news, announcements, advertisements, obituaries, public calls, photos, etc. In the light of discovering newspaper content, as an added value to the project, the librarians of the University Library "Svetozar Marković" analytically processed over 2,000 component parts (ca. 500 advertisements) from old Serbian newspapers so that the records in the University Library database could be searched through the corresponding metadata and annotations. The emphasis was placed on the processing of short news, announcements, public calls, obituaries, photos and advertisements. All these component parts were processed with relevant bibliographic metadata and annotations in a bibliographic data format and can be automatically exported to XML, Dublin Core or MARC21.

To demonstrate the content and structure of these records presented are two examples of export from COBISS bibliographic database in Dublin Core for advertisements since they will be further analyzed. Notes and subject headings on advertisements were entered in the bibliographic database in Serbian and English, so the labels "description" and "subject" show the kind of advertisement.

```
1. Ice selling
<?xml version="1.0" encoding="UTF-8"?>
<dc:collection xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dcterms="http://purl.org/dc/terms/">
<dc:dc>
<dc:dc>
<dc:title>Лед за лето / Стефан Шул. </dc:title>
<dc:creator>Šul, Stefan</dc:creator>
<dc:type>serial component part</dc:type>
<dc:type>text</dc:type>
<dc:type xsi:type="dcterms:DCMIType">Text</dc:type>
<dc:date xsi:type="dcterms:W3CDTF">1867</dc:date>
<dcterms:isPartOf xsi:type="dcterms:isPartOf>
```

⁹ Europeana Newspapapers Workshop "Refinement & Quality Assessment" was held in Belgrade and discused the techniques for refining and assessing the quality of digitized historical newspapers. Available at: http://www.europeana-newspapers.eu/focus-on-newspaper-refinement-quality-assessment-in-belgrade/

¹⁰ Based on METS and ALTO formats developed in Library of Congress. Available at: http://www.loc.gov/standards/mets/ and http://www.loc.gov/standards/alto/

```
<a href="mailto:de:relation"></a>In: Земунски гласник. - Земун: Игњат К. Сопрон, 1867/68-1969. - ISSN
2217-7612. - Бр. 6 Год. 1 (1867), стр. 45</dc:relation>
<dcterms:bibliographicCitation>Земунски гласник, Земун: Игњат К. Сопрон, 1867/68-
1969, Бр. 6, Год. 1 (1867), стр. 45</dcterms:bibliographicCitation>
<dc:format>crp. 45.</dc:format>
<dc:description>Оглас за продају леда за лето.</dc:description>
<dc:description>Објављено у листу "Земунски гласник" који је излазио недељом ујутру
од 1867. до 1869. у Земуну и поред чланака из књижевности, доносио је вести локалног
карактера. </dc:description>
<dc:description>Advertisment for sale of ice for the summer. </dc:description>
<dc:description>Published in the "Zemunski glasnik" newspaper which was published on
Sunday mornings in the period between 1867 and 1869 in Zemun and, beside articles on
literature, it brought the news of a local character.</dc:description>
<dc:subject xml:lang="sr">led</dc:subject>
<dc:subject xml:lang="sr">prodaja</dc:subject>
<dc:subject xml:lang="sr">oglasi</dc:subject>
<dc:subject xml:lang="sr">Srbija</dc:subject>
<dc:subject xml:lang="sr">Zemun</dc:subject>
<dc:subject xml:lang="sr">19v</dc:subject>
<dc:subject xml:lang="en">ice</dc:subject>
<dc:subject xml:lang="en">sale</dc:subject>
<dc:subject xml:lang="en">advertisements</dc:subject>
<dc:subject xml:lang="en">Serbia</dc:subject>
<dc:subject xml:lang="en">19c</dc:subject>
<dc:identifier>39141647</dc:identifier></dc:dc>
</dc:collection>
       2. Singer sewing machines
<?xml version="1.0" encoding="UTF-8"?>
<dc:collection xmlns:dc="http://purl.org/dc/elements/1.1/"</pre>
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dcterms="http://purl.org/dc/terms/">
<dc:dc>
<dc:title>Сингер-ове шиваће машине...</dc:title>
<dcterms:alternative>Varying form of title: Друштво за продају шиваћих машина Борн и
Ko</dcterms:alternative>
<dc:type>serial component part</dc:type>
<dc:type>text</dc:type>
<dc:type xsi:type="dcterms:DCMIType">Text</dc:type>
<dc:date xsi:type="dcterms:W3CDTF">1911</dc:date>
<dcterms:isPartOf xsi:type="dcterms:URI">urn:ISSN:Y000-4383</dcterms:isPartOf>
<dc:relation>In: Нова искра. - Београд: Р. J. Одавић, 1899-1911. - ISSN Y000-4383. - Год.
10, бр. 3 (1911), кор. лист</dc:relation>
<dcterms:bibliographicCitation>Нова искра, Београд: Р. J. Одавић, 1899-1911, Год. 10,
бр. 3 (1911), кор. лист</dcterms:bibliographicCitation>
<dc:format>кор. лист.</dc:format>
<dc:description>Наслов са почетка текста.</dc:description>
<dc:description>Реклама трговине за продају машина за шивење Сингер.
</dc:description>
```

<dc:description>Објављено у листу "Нова искра" који је излазио од 1899. до 1911. у Београду и доносио чланке из области књижевности, науке, уметности и културне историје. </dc:description>

<dc:description>Trade company advertistment for Singer sewing machines. </dc:description> <dc:description>Published in the "Nova iskra" newspaper which was published in the period between 1899 and 1911 in Belgrade and brought articles on literature, science, art and cultural history.</dc:description>

```
<dc:subject xml:lang="sr">mašine za šivenje</dc:subject>
```

- <dc:subject xml:lang="sr">reklame</dc:subject>
- <dc:subject xml:lang="sr">trgovine</dc:subject>
- <dc:subject xml:lang="sr">Beograd</dc:subject>
- <dc:subject xml:lang="sr">Srbija</dc:subject>
- <dc:subject xml:lang="en">trade company</dc:subject>
- <dc:subject xml:lang="en">Singer (company)</dc:subject>
- <dc:subject xml:lang="en">sewing machines</dc:subject>
- <dc:subject xml:lang="en">advertising</dc:subject>
- <dc:subject xml:lang="en">Belgrade</dc:subject>
- <dc:subject xml:lang="en">Serbia</dc:subject>
- <dc:identifier>39062543</dc:identifier></dc:dc>
- </dc:collection>

5. Advertisements in historical Serbian newspapers

Advertisements in Serbian press appeared in the early 19th century, modeled on the newspapers from Central Europe, mainly from Austria. Although their primary purpose was commercial in nature, contents of advertisements and corresponding illustrations can be seen a source of information regarding new inventions, technological innovation, industrialization, various craft activities, services and products of that time. The real development of this form of news in Serbian press can be traced back to the appearance of Novine serbske in January 1834. By the mid-19th century, there was no other Serbian newspaper that delivered advertisements, but then advertisements started appearing in newspapers Podunavka and Šumadinka whose pages will be available through the project Europeana Newspapers. Since the seventies of 19th century there was an interesting phenomenon in Serbian press of advertising and marketing dentists and doctors, as well as institutions in the field of economy – banking, credit and insurance agencies. The eighties of the 19th century were the era of the best equipped advertisements in Serbian press. Visuals in illustrated advertisements contributed to having quite ordinary household items won space in advertisements and thus becoming desirable to the readers of the newspaper. The promotion of products whose brands are still recognizable like Frank coffee, Bar-Lock typewriters, Singer sewing machines, etc. begun in the late 19th century. In addition to luxury framed and illustrated advertisements in the late 19th and early 20th centuries, the Serbian newspapers offer the so-called classifieds which advertised sales of houses and apartments, estates, second hand furniture, clothes, etc., as well as business deals of all types.

An issue of the newspaper *Male novine* with advertisements for various things, services, job offers, houses or land rent and sale on the front and on the last page provides another example. Issue 72, from March 13, 1895 had the first and the last page dedicated to advertisements. It should be noted that the advertisements were not sorted by any logical order and schedule, and were related to various products, services and jobs. The first page

<dc:subject xml:lang="sr">Singer (kompanija)</dc:subject>

brings advertisements for: sales of children's dresses, sea fish and honey store, production and selling of clay stoves and fireplaces, the inn *Serbian Crown*, fish oil in "the pharmacy of the Red Cross", renting houses, shops and flats, recruiting salesman of the *Male novine*, and the last page for a job for pharmacy assistants, medical clinic J. Stanković, ironing service, clock and jewelers shop Leopold Slezinger, Adolf Rešovski's cafeteria supplies, etc.





Figure 1: Examples from *Male novine*, vol. 10, issue 244, 1896 (left) and *Nova iskra*, vol. 10, issue 4, 1911 (right) – advertisements for (among other things) Singer sewing machines and Bar-Lock typewriters





Figure 2: Example from *Male novine* – first and last page

5.1. Importance of advertisements for users and researchers. Free access to millions of pages of European newspapers, as well as over 400,000 pages of Serbian newspapers will deliver to the users an immense quantity of information. The information contained in

advertisements ranges from entertainment and everyday life to important data for a variety of social, economic, cultural, political, marketing, and linguistic researches.

The content of this type of component parts fully reflects the spirit of the past, indicates development of certain industries, but also covers all aspects of cultural and social life since advertisements contain invitations to important events, manifestations, ceremonies, book promotions. Announcements of cultural events indicate that people in cities and towns were very active in attending theater performances and visits of foreign theater troops, going to the exhibitions and promotions of books, but also in attending open air plays by traveling entertainers, and subscribing to newspapers and magazines. Job offers (teacher, teacher of foreign languages, director, property manager, administrator, journalist, seller, photographer, etc.) as well as offers for various types of services (house renting, swimming pool, restaurant, insurance, candy shop, dance school, etc.), or advertisements for selling furniture, music instruments, books, food, medicine, domestic animals, land and property (houses, vineyards) show that various types of craft shops, factories, service industries in Serbia appeared at a given time, and that people learned foreign languages, children took private lessons in various subjects, healthy life was essential to the population, they practiced different sports, sold and bought real estate. Advertisements also reflect political, economic and social conditions.

Text within advertisements can be used as excellent material for researchers in the area of cultural history and theory of marketing, especially for the study of the history of advertising and marketing. More and more consumers expressed their social status by purchase in the second half of the 19th and beginning of 20th century. The social importance of consumer goods was increased. At the same time mass production and modern marketing were developed. Advertisers usually try to increase interest in the advertised product by its branding. This changed paradigm is the key moment of the development of the advertising industry because in addition to the need for the product and its use value, the decision of the customer is now influenced by emotions – packaging design, confidence in the seller or company. Some illustrated advertisements show packaging of certain items, and this aspect can be researched. One of the most important characteristics of advertisement is the way it reaches the consumers of advertised product or service.

In addition to these research possibilities based on the content and layout of advertisements in old Serbian newspapers, the significance of study of the history of print and typography needs to be mentioned. Layout of advertisements in Serbian newspapers remained unchanged since the mid-19th and mid-20th centuries. History of graphic processing in newspapers can be traced from the rich illustrations for advertising furniture, clothing and footwear, primarily luxury dresses, through the influence of Art Nouveau, to reduced layout in small advertisements.

Newspaper advertisements offer a huge potential for studying Serbian language, style, grammar and spelling. The fact that words in newspapers and advertisements will be automatically recognized and searchable through OCR provides tremendous opportunities for researchers of Serbian language and Cyrillic, keeping in mind that some titles were printed in old Serbian Cyrillic alphabet, which contains Russian letters (*Zemunski glasnik*, *Peštansko-budimski skoroteča*). Advertisements show the development of language. Names and addresses of institutions and individuals, addressing and communication between people, important for linguistic research, are also inevitable for the study of social customs and contacts.

6. Conclusion

Because of their value and importance it is interesting, wise, appropriate and useful to present the advertisements that existed in many daily and weekly European newspapers in 19th and 20th centuries, including Serbian newspapers. Before the appearance of radio and television, newspapers played a key role in the development of advertising and marketing.

The fact that the immense number of over 18 million digitized newspaper pages with listings and advertisements will be available in open access will significantly affect the availability of these materials, and therefore bring new users to the Europeana portal. Accordingly, Serbian advertisements in the newspaper collection of the University Library "Svetozar Marković" will be more used, and at the same time, through a variety of topics and areas they will open a number of possibilities for various users.

The project *Europeana Newspapers* will enable everybody to travel back through time and explore the marketing role of newspapers. On the other hand, we hope that Serbian and European newspapers will be useful to all kinds of researchers, and facilitate their access to information, mainly regarding general history, culture and identity of Europe through a single point of access: Europeana.

References

- [1] ALTO standard. Available at: http://www.loc.gov/standards/alto/ (2014-02-10)
- [2] Chambers, S., Schallier, W. *Bringing research libraries into Europeana: establishing a library-domain aggregator*. Liber Quarterly, vol. 20, no. 1, 2010. Available at: http://liber.library.uu.nl/index.php/lq/article/view/7980/8291 (2013-07-14)
- [3] docWorks. Available at: http://content-conversion.com/?lang=en#docworks-2 (2014-02-14)
- [4] Europeana Newspapers Project. Available at: http://www.europeana-newspapers.eu/ (2013-07-10)
- [5] Europeana Newspaparers Workshop "Refinement & Quality Assessment". Available at: http://www.europeana-newspapers.eu/focus-on-newspaper-refinement-quality-assessment-in-belgrade/ (2014-02-08)
- [6] Janc, Z. Oglasi u staroj srpskoj štampi 1834-1915. Beograd, Muzej primenjene umetnosti, 1978.
- [7] METS standard. Available at: http://www.loc.gov/standards/mets/ (2014-02-10)
- [8] Moren, E. *Uvod u istraživanje reklame*. Treći program, Radio Beograd, 2,4(1979), 454-462.
- [9] Ostojić, P. Propaganda. Beograd, Prosveta, 1940.
- [10] Pekárek, A., Willems, M. *The Europeana Newspapers a gateway to European newspapers online*. Progress in Cultural Heritage Preservation: proceedings of 4th International Conference, EuroMed 2012, Limassol, Cyprus, October 29 November 3, 2012. Lecture Notes in Computer Science, vol. 7616, pp. 654-659, Berlin, Heidelberg, Springer, 2012.
- [11] Reilly, S. K. *Building Bridges: from Europeana Libraries to Europeana Newspapers*. IFLA 2012, Session 119 Users and portals: digital newspapers, usability, and genealogy. Available at: http://conference.ifla.org/conference/past/ifla78/119-reilly-en.pdf (2013-08-25)
- [12] Todić, M., Pištalo, V. Kupite nešto i ovde. Beograd, Službeni glasnik, 2010.
- [13] *University Library "Svetozar Marković"*. *Digital collection of the newspapers*. Available at: http://ubsm.bg.ac.rs/cirilica/zbirka/novina/ (2013-08-20)

<u>aleksandra@unilib.bg.ac.rs</u> dakic@unilib.bg.ac.rs