

Zoran Velagić, Tomislav Jakopec
Josipa Selthofer
Faculty of Philosophy, Osijek

UTILIZING COMPARATIVE ADVANTAGES OF NEW PLATFORMS IN DIGITAL COLLECTION *EDICIJA*: “DUAL APPROACH” IN PRESENTING THE OLD PRINTS

Abstract: The aim of the paper is to outline the strategy of dual presentation of old prints in digital environment. After a brief introduction regarding content development, the paper focuses on steps and technologies concerning platform development. The content development strategy stems from the presumptions that two main approaches in old print presentation, one presenting only the text usually obtained through the OCR software, and the other presenting the print page in image file formats, deprive users either of original context of text presenting (a print page with specific typography, margin size, position of page numbering and so on) or of the advantages of full text searchability. These two deficiencies can be solved by “dual” approach in content presenting, based on simultaneous view to both the image (a page) and edited, searchable text. Regarding platform development, *Edicija* has been developed on open source technologies. It runs on operating system Linux, web server Apache, Relational database management system MySQL and server side programming language Hypertext preprocessor (PHP). *Edicija* implements web 2.0 principles using AJAX technology. The main goal on the technology level is to achieve visibility and functionality of the collection regardless of device, operating system or browser. The application of aforementioned strategies resulted with *Edicija* – digital library of Croatian printed heritage, available at <http://web.ffos.hr/EDICIJA>

Keywords: Edicija, dual approach, digital collection, tags, old prints.

1. Introduction

Digital library *Edicija* stems from the efforts of developing new, more efficient models of old prints presentation in a digital environment. The project was commenced with a main goal to facilitate access to Croatian heritage printed in vernacular by 1800. Though many of the reprints of such works are available, the Croatian book history is still deprived of systematic endeavours in publishing of preserved printed heritage in organized way (chronologically, thematically, etc.), which has resulted by marginalization of authors whose works were found to be aesthetically inconsequential, but also by a slow development of book history as a modern science. New findings on the possibilities of digitization of printed works and common trend of heritage digitization incited preliminary researches that consisted of composing comprehensive bibliography with special accent on recently discovered works, investigating the most useful form of web presentation with particular interest for the needs of book researchers and finally building priorities of publishing in digital collection (taking into account interdisciplinary importance of the work, accessibility of the work for the researchers, etc.). The conceived library was supposed to provide researchers with different comparative advantages in comparison with classical printed material: quicker insight and greater searchability regarding title, author, dialect, printer, year and town of publishing, etc., and particularly by key words search throughout the collection. Additionally, in order to offer the context of the current state of the research, every work in collection should be bibliographically elaborated and supplemented by critical foreword, notes on preserved copies, and basic literature dealing with it. In

time, the virtual library would gather a number of works enabling a comparative approach and detailed analysis of all aspects of cultural and social role of written word. It would make it easier to study the history of Croatian books which, according to modern scientific trends, includes interdisciplinary and comparative elaboration of the work's content, analysis of its physical characteristics (typography, paragraphs, notes, indexes, etc.), analysis of the role of participants in the process of book manufacturing (author, censor, sponsor, master printer, distributor, reader) etc.

2. Objectives of content development

The investigations that preceded to content development strategies included identifying experiences and restraints of diverse old prints digitization projects in order to utilize comparative advantages of new platforms by identifying and applying more effective and inventive ways in presenting the old prints in digital environment, outlining consequences of media convergence on modern scientific research, and finally identifying crucial points in developing the models of digital presentation of old prints and outlining basic steps in content and platform development.

Available digitization projects mostly transfer the contents from printed to electronic media, without using the comparative advantages of the latter ones. Seemingly, the computer screen is used simply as a book page that presents – a book page. Digitization of heritage is, in such a way, shrunk to digital photocopying. Another approach, text digitization (OCR), tears off the contents from the original printed media and the reader cannot get the feeling of the text's original. Both approaches hold huge restraints if we consider a digital collection as a tool for scientific research. The first one, which in a way equates electronic documents with facsimiles of paper originals neglects common advantages of digital platforms, reduces the digitization process to photocopying, shows up the format problem (vertical – horizontal) etc. The second, that treats electronic documents as text files only, deprives users of original context (print page layout, typography, margin size, position of page numbering, use of white space, foliation...), narrows research to inner aesthetics of the book, separates content from the platform, hinders some aspects of modern book history research and so on.

Two crucial questions from the early phase of digital content development, how value may be added to print originals and how value can be added to digitized contents in more innovative ways,¹ are in *Edicija* answered by “dual” or “binary” approach in content presenting and by dividing information into multiple layers. In other words, *Edicija* develops different approach, offering new, more powerful search possibilities, by using dual presentation of old prints – based on simultaneous looking for both the image (a page) and edited, searchable text – and introducing the tags searchable at the level of entire collection.

In order to accomplish such a new model of text presentation and to add additional features described in the introduction, the contents of collection is organized into multiple layers² which can be accessed according to the users' needs. The first two layers offer an interactive (simultaneous) image of the book page scan and OCR text of the single page. A larger image could be obtained in a separate layer; the same as the complete text of the book. The users are free to download content in different formats: ePub, TEI encoded xml, pdf, zip

¹ Compare Armstrong, C. J. – Lonsdale, Ray. The publishing of electronic scholarly monographs and textbooks, 1998., <http://www.ukoln.ac.uk/dlis/models/studies/elec-pub/elec-pub.htm>

² Organization of content through the multiple layers is proposed in texts of Robert Darnton. Compare Darnton, R. *A Historian of Books, Lost and Found in Cyberspace*, <http://www.historians.org/prizes/gutenberg/rdarnton.cfm> and Darnton, R. *The New Age of the Book*. // The New York Review of Books, Volume 46, Number 5 (March 18, 1999), <http://www.nybooks.com/articles/546>

and html. Additional layers offer information about the book, the state of current research, and insight into secondary literature.

The next objective of the content development consisted of increasing the search capabilities by tagging the key terms in each text. The tags present innovative approach of *Edicija* by enabling intertextual searching throughout the entire digital collection regardless the orthographic or dialectal differences in the early modern Croatian literature. They could be personalized, meaning that researchers of different aspects of old prints could define and implement different tags. Currently, the tags are giving information about crucial terms, personal and geographic names, symbols and similar topics. They offer readymade information to researchers: for example, a historian dealing with fraternities through the tags gets information of any mention of any fraternity in the entire digital collection.

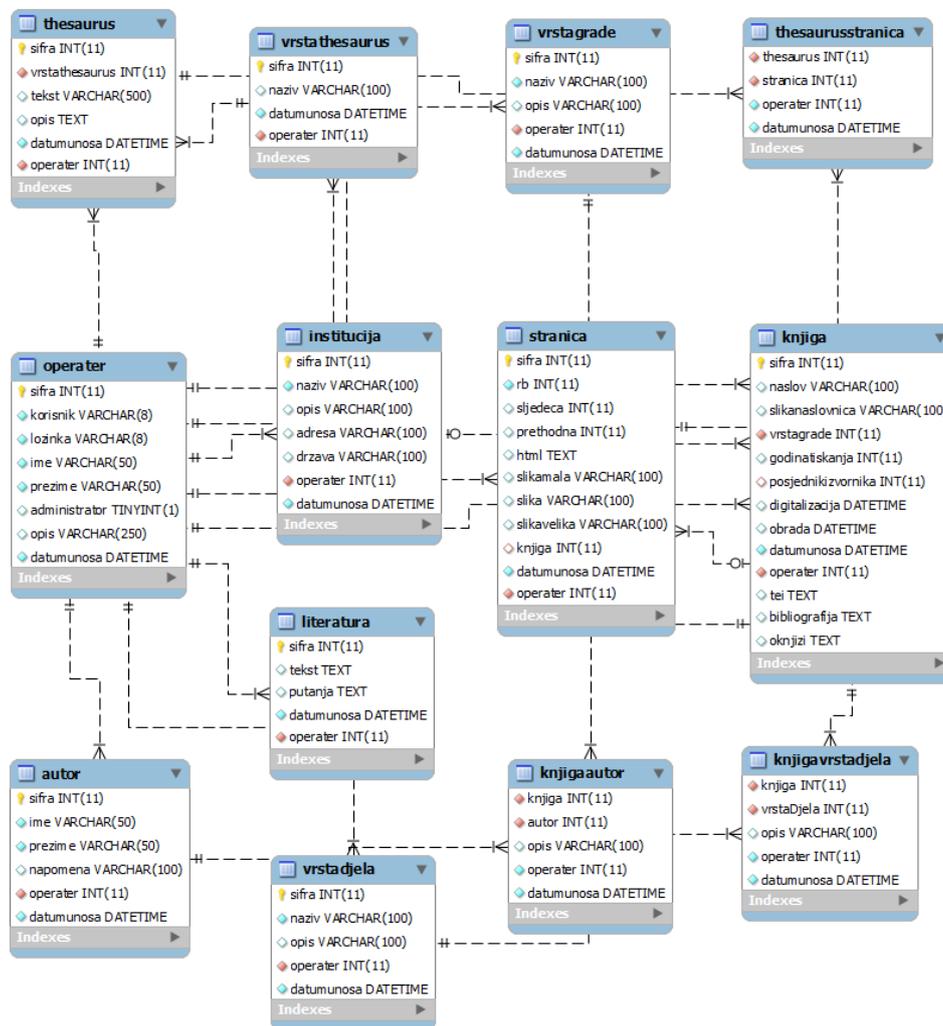


Diagram 1.

3. Technologies of platform development

Edicija has been developed on open source technologies. It uses an operating system Linux, web server Apache and relational database management system MySQL (ERA³ shown in dia-

³ Elmasri R, Navanthe S: Fundamentals of Database Systems (4th Edition)// Addison Wesley, 2004

gram 1). Server side programming language is Hypertext preprocessor (PHP). *Edicija* implements web 2.0 principles using AJAX technology. The main goal on the technology level is to achieve visibility and functionality of collection regardless of device, operating system platform and browser. Modern users expect RIA (Rich Internet Applications) with AJAX capabilities, but RIA application (web site) brings encapsulation of data towards searching engines (Google, Bing, Yahoo...) which cannot index content on web sites, meaning that the application could be invisible in the vast web environment.

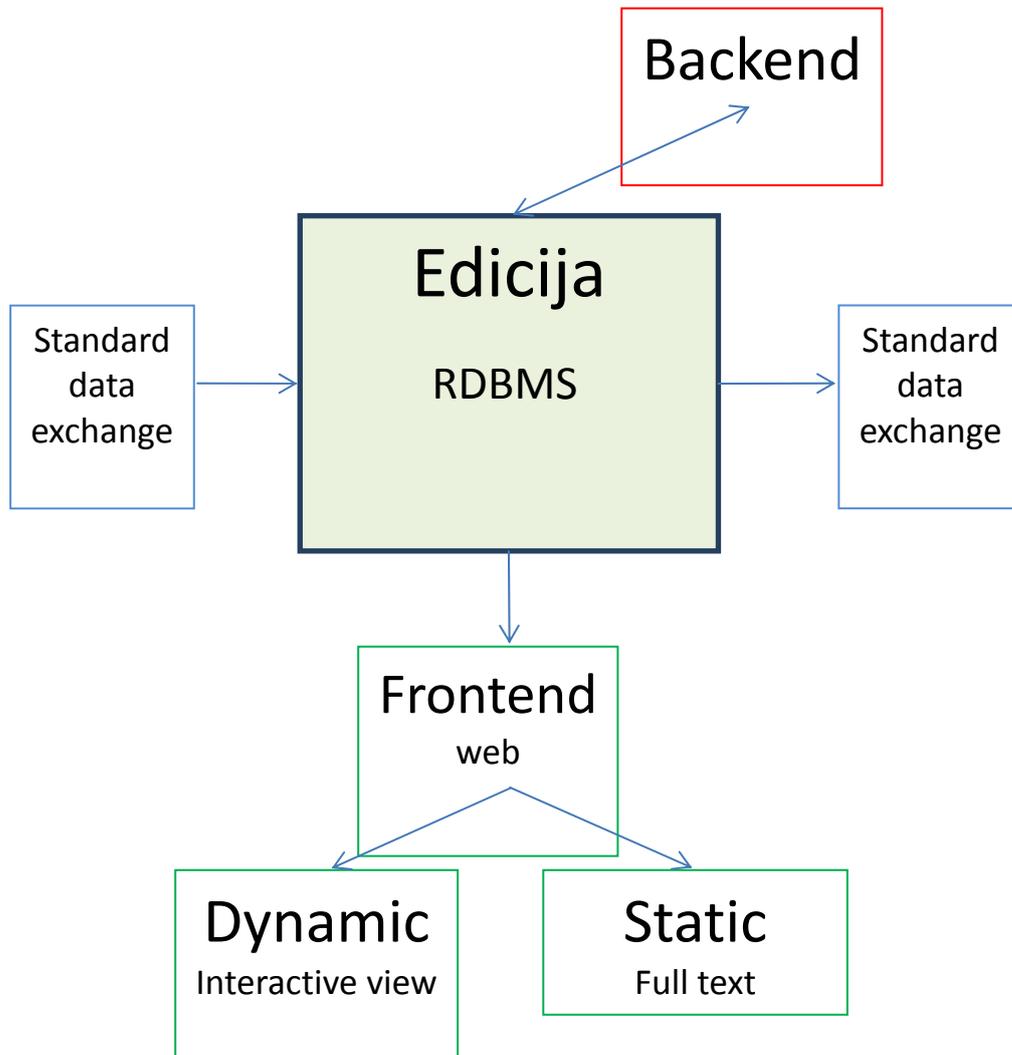


Diagram 2

The solution (Diagram 2) to such issue is a static web site, but a static web site neglects users as it does not support RIA standards. Thus, on the level of technology, *Edicija* combines two approaches by creating a relational database. Objects are placed in relation in order to achieve high manipulation capabilities. Content is managed with backend application (create new book, add page, insert tags, and so on). It will be possible to import content from standard data exchange and export content to standard data exchange. What is important to an end user is, naturally, front end. User expects RIA and search engine expects static web, and *Edicija* can offer both. From relational database both could be generated: interactive content for the

end user and static text for search engines. “Generate static content” is oxymoron – but static content could be generated with URL rewrite technology and *Edicija* is using it.

6. Conclusion

By applying dual approach model and developing the tags system *Edicija* solves few important issues regarding the book research in digital environment:

- “dual” approach in content presenting offers a simultaneous access to both the image (printed page) and edited, searchable text,
- tagging the key terms increases the search capabilities in texts regularly written in different ways as an outcome of the early modern non-standardized orthography,
- tags provide the search results at the level of collection (OCR text), with simultaneous view to the context of original presentation (the print page),
- with additional layers collection keeps high standards of digital collection development (bibliographic records, standard browsing capabilities...) and provides additional materials and research tools.

Edicija – digital library of Croatian printed heritage is aimed at the highly demanding users – the scientific community. By aforementioned innovation it tries to attract scientific community for, as any other text or text collection, it could exist “(...) only because there is a reader to give it meaning.”⁴

References

- [1] Armstrong, C. J. – Lonsdale, Ray. *The publishing of electronic scholarly monographs and textbooks*, 1998., <http://www.ukoln.ac.uk/dlis/models/studies/elec-pub/elec-pub.htm> (2011-04-11)
- [2] Chartier, R. *Forms and Meanings. Texts, Performances, and Audiences from Codex to Computer* (Philadelphia: University of Pennsylvania Press, 1995): 2
- [3] Darnton, R. *A Historian of Books, Lost and Found in Cyberspace*, <http://www.historians.org/prizes/gutenberg/rdarnton.cfm> (2011-04-11)
- [4] Elmasri R, Navathe S: *Fundamentals of Database Systems* (4th Edition)// Addison Wesley, 2004
- [5] Darnton, R. *The New Age of the Book*. // *The New York Review of Books*, Volume 46, Number 5 (March 18, 1999), <http://www.nybooks.com/articles/546> (2011-04-11)

zvelagic@ffos.hr
tjakopec@ffos.hr
jselthofer@ffos.hr

⁴ Chartier, R. *Forms and Meanings. Texts, Performances, and Audiences from Codex to Computer* (Philadelphia: University of Pennsylvania Press, 1995): 2