Преглед НЦД 12 (2008), 75–81

Saša Malkov,
Nenad Mitić,
Žarko Mijajlović
(Faculty of Mathematics,
University of Belgrade)

NIKOLA TESLA ONLINE CLIPPING LIBRARY PROTOTYPE[1]

Abstract: Nikola Tesla Museum (Belgrade, Serbia) possesses a unique clipping library, collected by the famous scientist himself. The Digitizing Group at Faculty of Mathematics and the museum analyzed the problem of clipping library online publication. A prototype digitization procedure and a prototype Web site are developed.

Key words: Nikola Tesla, online library, digitization.

1. Introduction

Nikola Tesla (1856–1943), American scientist of Serbian origin, usually presented himself as electrical engineer and inventor. His professional results and unique personality made him one of the most celebrated scientists of his time. As a public person, Tesla used to speak or publish comments on current topics, not only in science and technology, but also on different social problems, from food quality to war prevention. His work, personality and, above all, his visionary approach to any topic, very often resulted in public disapprovals. Many scientists and journalists in different parts of the world almost competed in publishing papers on Nikola Tesla and his spoken or written statements, either with acceptance or disapproval, and either with meaningful or irrational criticism [1].

Today we often concentrate on Nikola Tesla's results, ignoring the circumstances of his time and environment, but any in-depth analysis of his work and results absolutely must consider all available information on acceptance and disapprovals of his work by his contemporaries. What is amazing is that great work in providing such information did Nikola Tesla himself.

During his life, Nikola Tesla, with support of his secretary, collected journals and newspapers clippings covering his work and statements. A number of collected clippings were organized in 57 books, but many clippings are preserved in boxes and not sorted yet. It is assumed that less than a half of the clippings are organized. The value of the clipping library collected by the famous inventor is even increased by a lot of comments, written by hands on sides of these clippings.

According to the last wish of the scientist, all the documents and Tesla's personal things are transferred in Belgrade in 1949. Nikola Tesla Museum [1] was founded in 1952, to preserve and publish the valuable inheritance and to support research on Nikola Tesla in all parts of the world. The museum possesses exceptional collections, including: above 160.000

[1] This work was supported under project No 6201 by Ministry of Science and Environment of Republic of Serbia.

original documents, above 2000 books and journals, above 1200 historical technical exhibits, above 1500 photographs, above 1000 plans and drawings and many technical objects, instruments and apparatus. Nikola Tesla Clipping Library is one of the most interesting collections in the museum, probably the unique of this kind in the world.

2. Digitization

The intention to publish the clipping library online is not new. However, the limited funds and the richness of the museum collections and the plentiful of regular preservation activities caused the delay. In 2005th, Museum and Faculty of Mathematics approached the problem. The Digitizing Group at Faculty of Mathematics was involved in problem analysis and prototype development. The project is divided in two major parts: content digitization and online library publishing.

As one of initial steps, the digitization team scanned one book with 405 clippings. While it could be easier to digitize unbound materials, it would require the significant resource preservation work and clippings classification to select the representative sample. Book No. 19 is selected, because it contains clippings from the very beginning of the 20th century, and the clippings quality is very variable. Many different factors influenced the clippings quality, from paper material and thickness, to ink robustness and the compatibility of ink and paper.

Some kinds of paper are preserved in almost original form, or suffered only the slight change of color. On the other side, some clippings are highly deteriorated by structure weakening and threading or extreme color change. While some inks remained almost shiny, others lost the darkness in different degrees. In some cases ink is drawn by paper threads, making text almost unreadable. The worst shape is “achieved” by clippings whose paper became dark, ink became light and soothed. Many clippings are wrapped when pasted in books, some are even partially torn, or clipping parts are completely missing.

Many different digitization methods are tried. Because of the book hard covers and the decision not to unbind the book during the prototype phase, no flat scanning method was efficient enough. The initial digitization was performed using high quality digital cameras. Pages were photographed using many shots. Smaller clippings were photographed jointly, but larger clippings were photographed individually.

Three digital versions of each clipping are prepared: digital documents in PDF [2] and DjVu formats [3] and extracted plain text. PDF and DjVu are widely used formats for digital documents. They are similar in capabilities. Because of different image processing, slight visual differences are observable in resulted documents. However, they seam to be equally qualified for the purpose.

The plain texts were to be obtained using OCR techniques. Many different OCR software packages are tested, but with no generally satisfactory result. We already emphasized the low quality of the clippings. However, there are additional problems, like highly stylized fonts used at the beginning of the century and different languages. Generally, the result of OCR was highly inaccurate and plain texts had to be manually corrected. During the prototype phase only the few clippings were manually corrected, to provide enough data for other analyses, like searching based on clippings content.

3. Publishing

The publishing team had to design and develop the library database, tools for automatic import of digitized content, tools for manual corrections of individual clippings data, library Web site and appropriate search engine.

3.1. Data. The first topic in online library development was the data organization. Each clipping is presented in many different forms:

- as digital document in PDF format;
- as digital document in DjVu format;
- as plain text, intended for clipping text searching, further research and referencing;
- as metadata, including: clipping caption, journal name and volume/number, publishing date, keywords, language, comments written next to the clipping, clipping book and page number.

After analyzing pros and contras it was decided to keep all data, including digital documents and plain text, in a database. Contemporary database systems feature powerful data access methods and very high efficiency. While it could be necessary to organize digital documents as files because of even higher efficiency, the decision was to check if the database based solution is efficient enough or not.

The database management system IBM DB2 [4] was initially selected because of good experience of group members in some previous projects. Different editions are considered, but after IBM published DB2 Express-C as a free community edition, including all features significant for the project, the decision was very simple.

*Data Model***.** Database model is designed with clipping as a collection of resources (digital documents, including plain text) with all available metadata.

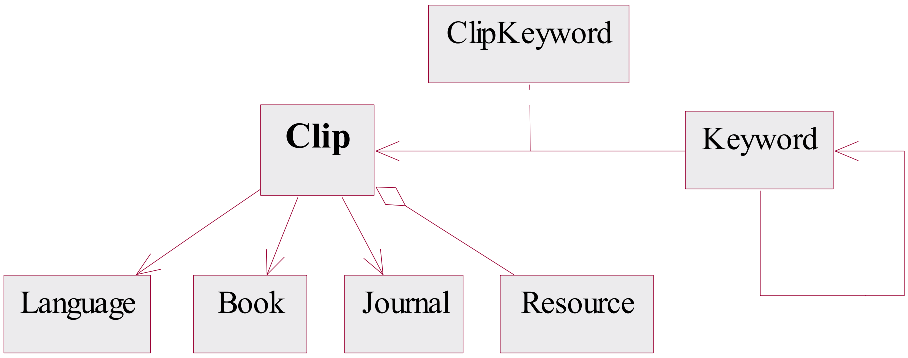


Figure 1. Simplified database model

The repeatable information (languages, books, journals, keywords) are extracted, but other information (captions, publishing dates, page numbers, comments) are kept as clippings' content. This simple design allows for both later extensions and simple application of efficient searching algorithms.

Data Load. During the digitization process, the clippings metadata are collected in a spreadsheet file. For the most of the clippings metadata include caption, journal, publishing date and language. Other metadata is available only for a few of the clippings.

The import tools are developed for automatic full data load. Import tools consist of metadata loader, digital documents loader, metadata indexer and plain text indexer. Tools for manual editing are developed, too, to allow for manual error corrections.

3.2. Library Web site. When the project was initiated, members of the group were involved in research of applications of functional programming languages in Web development. Thus, the choice of development methods and tools was easy.

Wafl. Wafl [5] is functional programming language, customized for Web development. The most of its concepts are designed to be domain independent, in order to provide the general applicability. It is strongly typed language with implicit static type checking. The strong static type checking is very important element of the error prevention process. The automatic type inference makes programs shorter and encourages the writing of polymorphic functions and data types.

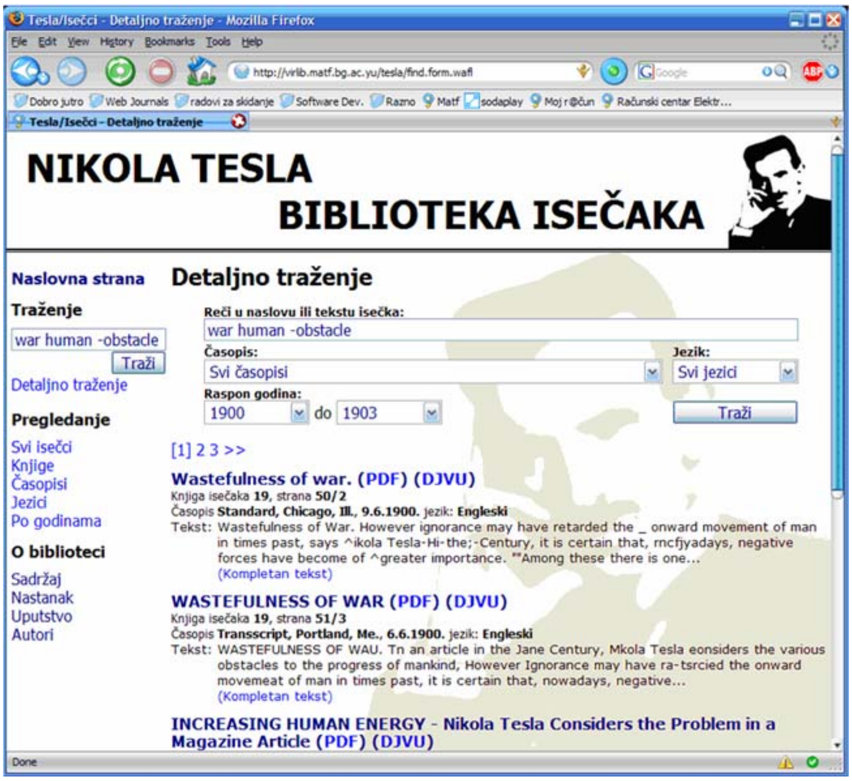
The page templates allow the development of programs for automatic construction of HTML and XML documents in very intuitive way, while preserving the functional language semantics. Page templates allow for easy separation of service logic development and user interface development. The modularization of graphical design aspects is fully supported.

Wafl programs access a database for both reading and updating. Read only access is provided by query functions. Query function body is defined using SQL select statements. Each query returns a list of rows. Wafl extends the transaction concept to cover not only database data, but also Web session data.

Implementation. The prototype library implementation was targeted primarily to provide rich data access features. Developed interface allows browsing by books, journals, languages or years. Two searching interfaces are implemented.

Simple search interface consists of single edit control and provides the searching based on clippings captions, keywords and plain text. Any word may be prefixed with control symbol + or –.

Advanced search interface is extended by journal, language and time period. Figure 2 represents both search interfaces.


Tesla/Isečci - Detaljno traženje - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://virlib.matf.bg.ac.yu/tesla/find.form.wafl
Google
Dobro jutro
Web Journals
radovi za skidanje
Software Dev.
Razno
Matf
sodaplay
Moj r@čun
Računski centar Elektr...
Tesla/Isečci - Detaljno traženje
Done

NIKOLA TESLA
BIBLIOTEKA ISEČAKA

Naslovna strana

Traženje

war human -obstacle

Traži

Detaljno traženje

Pregledanje

Svi isečci
Knjige
Časopisi
Jezici
Po godinama

O biblioteci

Sadržaj
Nastanak
Uputstvo
Autori

Detaljno traženje

Reči u naslovu ili tekstu isečka: war human -obstacle

Časopis: Svi časopisi

Jezik: Svi jezici

Raspon godina: 1900 do 1903

Traži

[1] 2 3 >>

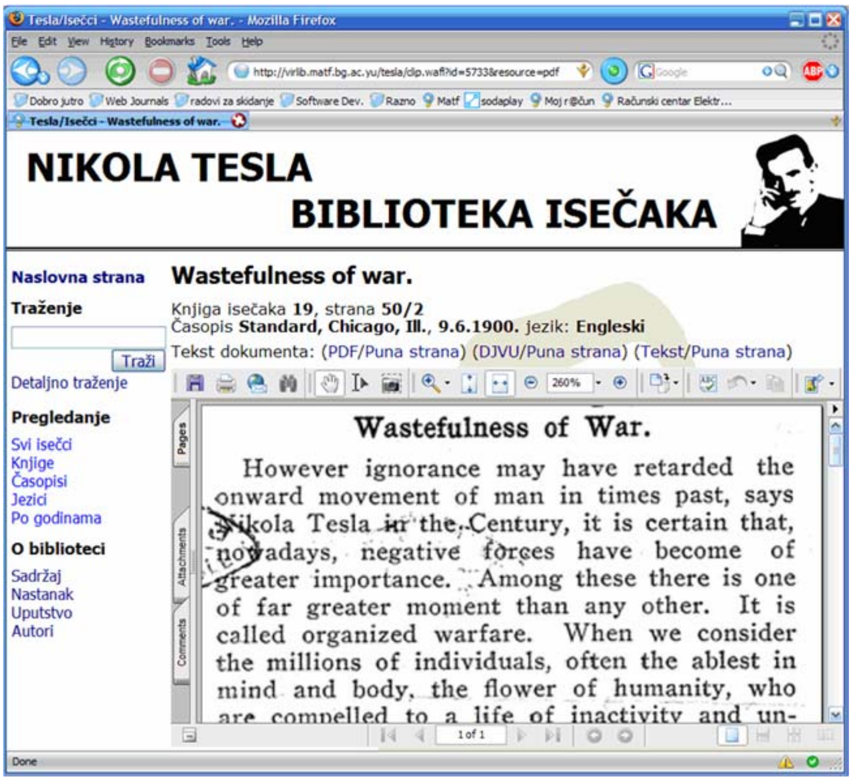
Wastefulness of war. (PDF) (DJVU)
Knjiga isečaka **19**, strana **50/2**
Časopis **Standard, Chicago, Ill., 9.6.1900.** jezik: **Engleski**
Tekst: Wastefulness of War. However ignorance may have retarded the _ onward movement of man in times past, says ^ikola Tesla-Hi-the;-Century, it is certain that, rncfjyadays, negative forces have become of ^greater importance. ""Among these there is one...
(Kompletan tekst)

WASTEFULNESS OF WAR (PDF) (DJVU)
Knjiga isečaka **19**, strana **51/3**
Časopis **Transscript, Portland, Me., 6.6.1900.** jezik: **Engleski**
Tekst: WASTEFULNESS OF WAU. Tn an article in the Jane Century, Mkola Tesla eonsiders the various obstacles to the progress of mankind, However Ignorance may have ra-tsrcied the onward movemeat of man in times past, it is certain that, nowadays, negative...
(Kompletan tekst)

INCREASING HUMAN ENERGY - Nikola Tesla Considers the Problem in a Magazine Article (PDF) (DJVU)

Figure 2. Advanced searching

Clippings are presented in both PDF (Figure 3) and DjVu (Figure 4) formats using appropriate browser plugins.


Tesla/Isečci - Wastefulness of war. - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://virlib.matf.bg.ac.yu/tesla/clip.wafl?id=5733&resource=pdf
Dobro jutro
Web Journals
radovi za skidanje
Software Dev.
Razno
Matf
sodaplay
Moj r@čun
Računski centar Elektr...
Tesla/Isečci - Wastefulness of war.
260%
Pages
Attachments
Comments
1 of 1
Done

NIKOLA TESLA
BIBLIOTEKA ISEČAKA

Naslovna strana
Traženje
Traži
Detaljno traženje
Pregledanje
Svi isečci
Knjige
Časopisi
Jezici
Po godinama
O biblioteci
Sadržaj
Nastanak
Uputstvo
Autori

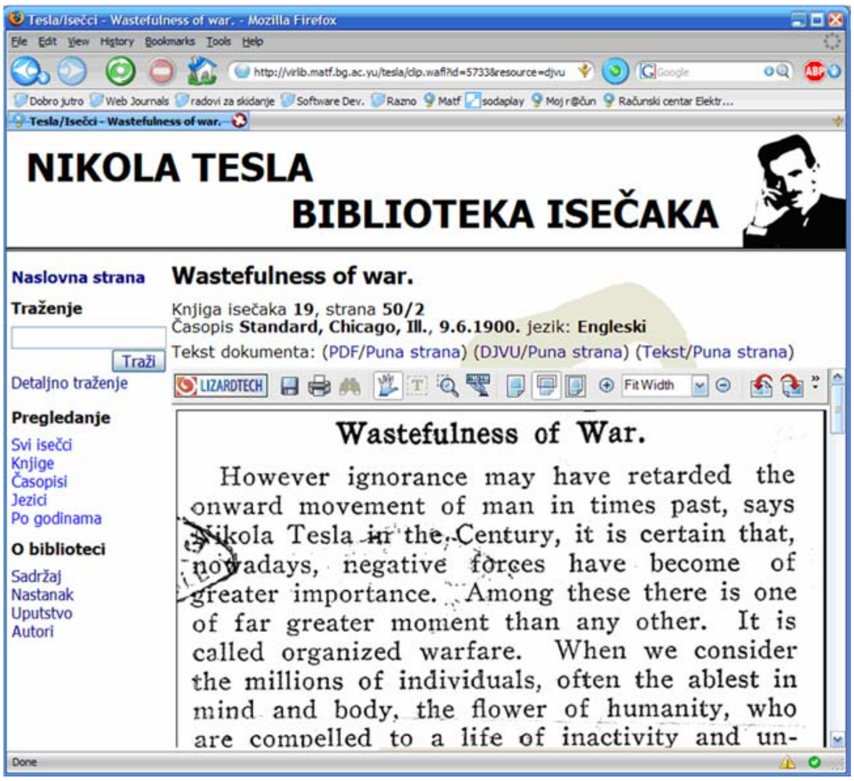
Wastefulness of war.

Knjiga isečaka 19, strana 50/2
Časopis Standard, Chicago, Ill., 9.6.1900. jezik: Engleski
Tekst dokumenta: (PDF/Puna strana) (DJVU/Puna strana) (Tekst/Puna strana)

Wastefulness of War.

However ignorance may have retarded the onward movement of man in times past, says Nikola Tesla in the Century, it is certain that, nowadays, negative forces have become of greater importance. Among these there is one of far greater moment than any other. It is called organized warfare. When we consider the millions of individuals, often the ablest in mind and body, the flower of humanity, who are compelled to a life of inactivity and un-

Figure 3. A digitized clipping in PDF format.


Tesla/Isečci - Wastefulness of war. - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://virlib.matf.bg.ac.yu/tesla/clip.wafl?id=5733&resource=djvu
Dobro jutro
Web Journals
radovi za skidanje
Software Dev.
Razno
Matf
sodaplay
Moj r@čun
Računski centar Elektr...
Tesla/Isečci - Wastefulness of war.
LIZARDTECH
FitWidth
Done

NIKOLA TESLA
BIBLIOTEKA ISEČAKA

Naslovna strana
Traženje
Traži
Detaljno traženje
Pregledanje
Svi isečci
Knjige
Časopisi
Jezici
Po godinama
O biblioteci
Sadržaj
Nastanak
Uputstvo
Autori

Wastefulness of war.

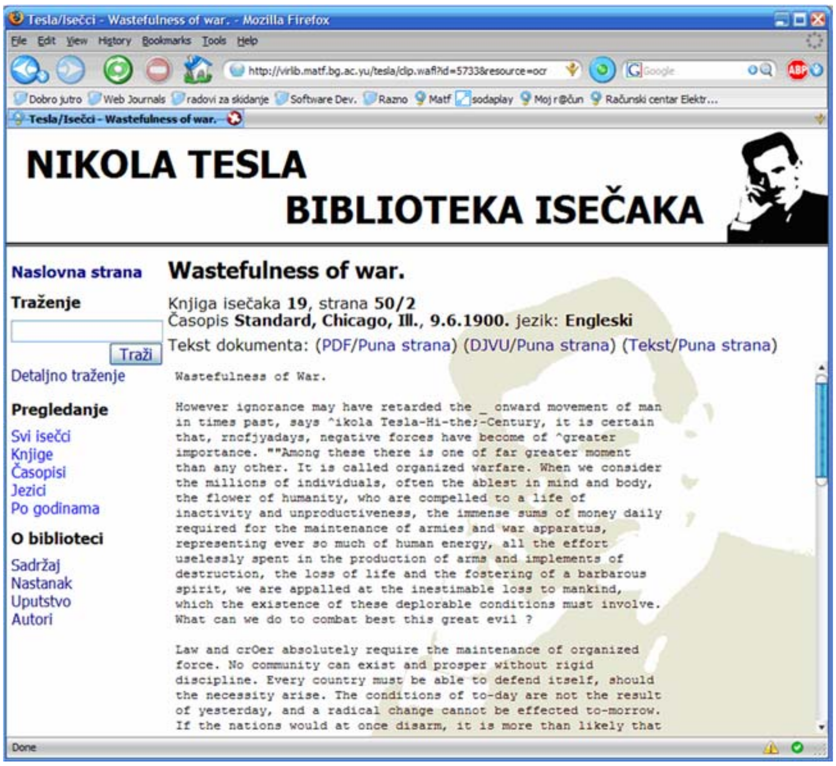
Knjiga isečaka 19, strana 50/2
Časopis Standard, Chicago, Ill., 9.6.1900. jezik: Engleski
Tekst dokumenta: (PDF/Puna strana) (DJVU/Puna strana) (Tekst/Puna strana)

Wastefulness of War.

However ignorance may have retarded the onward movement of man in times past, says Nikola Tesla in the Century, it is certain that, nowadays, negative forces have become of greater importance. Among these there is one of far greater moment than any other. It is called organized warfare. When we consider the millions of individuals, often the ablest in mind and body, the flower of humanity, who are compelled to a life of inactivity and un-

Figure 4. A digitized clipping in DjVu format.

For deeper research, the full extracted text is available for both reading and copying (Figure 5).


Tesla/Isečci - Wastefulness of war. - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://virlib.matf.bg.ac.yu/tesla/clip.wafl?id=5733&resource=ocr
Google
Dobro jutro Web Journals radovi za skidanje Software Dev. Razno Matf sodaplay Moj r@čun Računski centar Elektr...
Tesla/Isečci - Wastefulness of war.
Naslovna strana
Traženje
Traži
Detaljno traženje
Pregledanje
Svi isečci
Knjige
Časopisi
Jezici
Po godinama
O biblioteci
Sadržaj
Nastanak
Uputstvo
Autori
Done

NIKOLA TESLA
BIBLIOTEKA ISEČAKA

Wastefulness of war.

Knjiga isečaka **19**, strana **50/2**
Časopis **Standard, Chicago, Ill., 9.6.1900.** jezik: **Engleski**

Tekst dokumenta: (PDF/Puna strana) (DJVU/Puna strana) (Tekst/Puna strana)

```
Wastefulness of War.

However ignorance may have retarded the _ onward movement of man
in times past, says ^ikola Tesla-Hi-the;-Century, it is certain
that, rncfjyadays, negative forces have become of ^greater
importance. ""Among these there is one of far greater moment
than any other. It is called organized warfare. When we consider
the millions of individuals, often the ablest in mind and body,
the flower of humanity, who are compelled to a life of
inactivity and unproductiveness, the immense sums of money daily
required for the maintenance of armies and war apparatus,
representing ever so much of human energy, all the effort
uselessly spent in the production of arms and implements of
destruction, the loss of life and the fostering of a barbarous
spirit, we are appalled at the inestimable loss to mankind,
which the existence of these deplorable conditions must involve.
What can we do to combat best this great evil ?

Law and crOer absolutely require the maintenance of organized
force. No community can exist and prosper without rigid
discipline. Every country must be able to defend itself, should
the necessity arise. The conditions of to-day are not the result
of yesterday, and a radical change cannot be effected to-morrow.
If the nations would at once disarm, it is more than likely that
```

Figure 5. A clipping plain text.

4. Conclusions

After the content preparation and implementation of prototype clipping library, the results are tested in both capabilities and performances. The following conclusions are made:

- The quality of digitized resources using high quality digital cameras is satisfactory for online publishing for the most of the clippings. However, for archival purposes and many of the clippings it is crucial to use flat scanners.
- For highest quality digitization it would be necessary to unbind the clipping books, or at least the pages with clippings pasted too close to the bindings.
- No OCR software proved to be fully adequate. While almost every tested professional level OCR software performed very well with well preserved clippings, the fact is that the most of the clippings are not in good enough shape to rely on the software. It is suggested to test all the software again with higher resolution digitized clippings images, obtained using flat scanning.
- The applied data organization model, using database system for all resources, fully satisfies the projected load and performances.
- The selected database system, IBM DB2 Express-C, proved to be more than capable for the task;
- The Web development platform (Linux Apache with Wafl programming language) provided for simple and efficient development.
- The applied searching algorithms proved to be sufficient for the projected load.

The future activities will include both digitization using flat scanners and testing OCR software on these new digitized documents.

References

[1] Nikola Tesla Museum, http://www.tesla-museum.org/
[2] PDF Reference, *Adobe PDF Technology Center*, http://www.adobe.com/devnet/pdf/pdf_reference.html
[3] Lizardtech DjVu Reference, *Lizardtech*, http://djvu.org/docs/DjVu3Spec.djvu
[4] IBM DB2 Web Site, *IBM Corporation*, http://www.ibm.com/db2
[5] S. Malkov, WAFL – *Functional Programming Language for Development of Web Applications*, Master Thesis, University of Belgrade, Faculty of Mathematics, 2002.

Саша Малков,
Ненад Митић,
Жарко Мијајловић
(Математички факултет, Београд)

ПРОТОТИП ДИГИТАЛНЕ БИБЛИОТЕКЕ ИСЕЧАКА НИКОЛЕ ТЕСЛЕ

Музеј Николе Тесле (Београд, Србија) поседује јединствену колекцију исечака, које је познати научник лично сакупљао. Група за дигитализацију Математичког факултета и Музеј су анализирали могућности дигиталног публиковања библиотеке исечака. Дефинисан је прототип поступка дигитализације и развијен је прототип Веб локације.

smalkov@matf.bg.ac.yu