Editing and Representation of Ancient Russian Semiographic Chants on the Web

Andrey Philippovich, Marina Danshina(✉), and Irina Golubeva(✉)

Bauman Moscow State Technical University, Moscow, Russia
{aphilippovich, mdanshina, igolubeva}@it-claim.ru

Abstract. This paper presents a web service for input of ancient Russian manuscript into electronic form. Researchers, historians and musicians can process musical manuscripts written in Znamenny Notation via this input system. In order to represent semiographic chants, we used a special font-face, which was developed during the project "Computer Semiography". Our objective was to study features of ancient Russian musical chants and create a system for editing and playing of these chants.

Keywords: Musical information technologies · Znamenny Notation · Semiography · Ancient musical manuscripts · Representation of chants · Song patterns · Decoding semiotic systems

1 Introduction

Thousands of ancient manuscripts containing musical notation are among the monuments of the Russian culture. Most of them have been written using special musical notation systems. Znamenny Chant has been the oldest type of Russian liturgical singing since the XI century. This notion was received from the general naming used for "Znamena" or "znamia" (distinctive mark, flag, banner), due to its particular type of musical notation (also sometimes called "Kryuk", hook, or bend). This system of musical notation takes its beginning from the early Byzantine non-linear notation also called Neume Notation. Neumes and Znamena are the precursors of modern-day musical symbols.

Semiography is a system of symbolic notation. In Russian scientific tradition, semiography also refers to the direction of research of ancient musical notations, like Znamenny Chants [1, 2]. During the past 10 years, we conducted extensive research and contributed novel developments in this direction based on methods of computational linguistics and semiotics [3–7]. Our project, "Computer Semiography" [8], seeks to provide – as nearly as possible – a lossless data representation of Znamenny Chant manuscripts. We anticipate that such representation will be useful to musicologists in Russia and around the world for sharing and analysis of ancient chant manuscripts. Additionally, we foresee that such sharing could very well take place primarily over the Internet in the form of a web-based scientific information system; that is to say, a database with special services for editing and analyzing Znamenny Chants.

D.I. Ignatov et al. (Eds.): AIST 2014, CCIS 436, pp. 66–77, 2014.
DOI: 10.1007/978-3-319-12580-0_6

2 Features of Ancient Russian Musical Chants

The writing system of Znamenny Chants differ from the more familiar and current use of a 5-line staff (i.e. linear notation). There is no strong correspondence between the pitch of each sign and the duration of each note when playing or notating a melody. Therefore, transferring melodies from Znamenny manuscripts to that of a modern-day five-line notation is a very complex task. Difficulties associated with decoding znamenny symbols exist in the fact that znamenny letters do not adhere to conventional notation, wherein melodies are expressed by specific semiographic signs, as opposed to something else [9].

Fig. 1. Example of a Znamenny Chant.

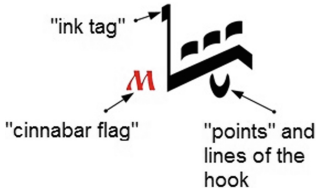

Fig. 2. Structure of a semiographic sign.

Semiographic signs possess a complex structure of main symbols and additional elements which aid a singer. Main symbols usually have a large black hook or a black stroke, several smaller black "points" and "commas", and lines near the hook or crossing through it. Additional elements – "cinnabar flag" (red color) and "ink tag" (black color) – show the pitch range of sounds and the character of its melody (Fig. 2).

Some semiographic signs may represent one, two or several notes, and some even a whole melody of more than 10 notes with a complicated rhythmic structure. The Znamenny Chant system has 12 pitches which are shared across four vocal ranges. Ink tags indicate only vocal ranges, while the cinnabar flags symbolize the highest pitch in a sequence of notes corresponding to the sign. Some cinnabar flags also indicate the nature for the notation's execution (referred to as indicatory flag).

There are several types of semiographic manuscripts, which have different structures:

- *Ordinary Chants* (Fig. 1).
- *Semiographic alphabet (primer)* – provides verbal descriptions of the semiographic signs, or matches with one or more notes (Figs. 3 and 4).
- *Collections of song patterns ("popevkas")* – different parts of a chant (stanza), which have sequences of semiographic signs that need to be translated according to special rules (Fig. 5). These patterns often have unique names and textual instructions.

- *Double noticed chants* – manuscripts (Fig. 6) which are given in direct correspondence with Linear and Znamenny Notation, and from which you can extract some knowledge regarding the translation of song patterns and semiographic signs in context, including and depending on the words with which they were matched.

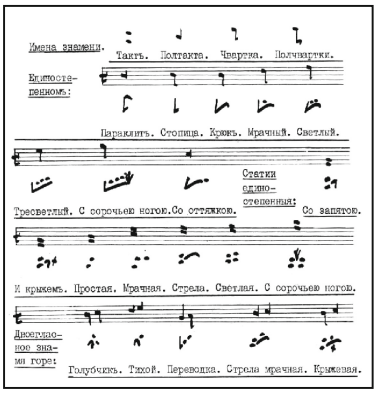

Fig. 3. Semiographic alphabet (primer) with linear notation.

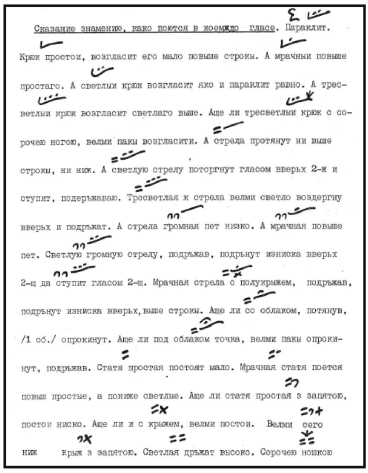

Fig. 4. Semiographic alphabet (primer) with text descriptions.

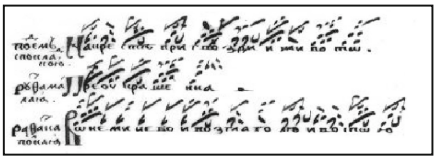

Fig. 5. Fragment of a collection of song patterns ("popevkas").

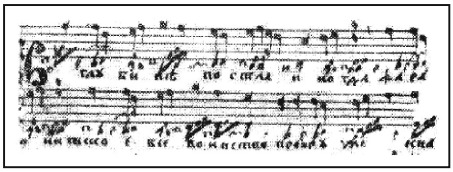

Fig. 6. Fragment of a double noticed chant ("dvoznamennik").

3 Methods for the Visualization of Znamenny Chants

Currently, a few other projects and websites share electronic versions of Znamenny Chants with internet users. One of the biggest internet resources in this field is the "Fund for Znamenny Chants".[1] This project was opened in 2003, and so far, it has managed to enter more than 10 thousand chants into Znamenny and linear notations. Developers of this project decided to convert each manuscript into separate chants.

[1] "Fund for Znamenny Chants", http://znamen.ru/.

This allows users to choose certain chants before viewing them; however, books cannot be viewed in their entirety by browsing from one page to the next. Additionally, chants are visualized without binding to its primary source (i.e. its scanned version).

Primary sources – more than 3,000 ancient manuscripts – are stored in state libraries, but access is restricted and only possible after obtaining a special permit. And copying a manuscript is complicated, as you have to get special permission and are only allowed to copy individual pages; not the entire manuscript. So though you are able to see how Znamenny Chants look and find information on the topic, getting access to any particular source is almost impossible for most people.

Furthermore, viewing is not the only problem. Most important is the research of manuscripts. Working with a picture or a highly fragmented book is uncomfortable because it's difficult to find information. For example, the implementation of statistical analysis takes time, as this work is quite monotonous and requires concentration. Using computers, however, does facilitate this task. There are two main abilities to represent Znamenny Chants on the Internet:

- Scan ancient manuscripts and place them on a website in graphic formats.
- Input chants in symbolic form and store them in a database with special services for viewing, searching and analyzing.

The former method is the easiest way, because it is only necessary to organize the storage of images in a specific location (e.g. a graphic database or a folder on a computer) and load them out for display to the user. There are many ready-made solutions for this, so it isn't very difficult. But this method has one main disadvantage: the inability to search through chants, as you can only look directly at pages. Additionally, more detailed queries would require special indexing.

The latter method requires developing special fonts or even software, and using them to enter chants into a text document or database. This method – unlike the former – allows the ability to search within a chant, itself; consequently, various researches could be performed.

Furthermore, the second variant takes up less space, as compared to the first. As an example, 70 pages of a book in picture format (e.g. *.bmp) could take up to 260–300 Mb, while the same page typed in a text editor might require only 2.5–3 Mb. On the other hand, a lot of effort must go into the development of fonts and to input the chants into a database. Further, it would require checking and proofreading entered chants, as people make a lot of mistakes when typing. The second method could also be accomplished in two different ways: local editing of chants using well-known software (e.g. Microsoft Word) or using special software (e.g. local or web service-based).

During this project, we tried and compared both of these variants; for which the results are represented in our publications and on the website for this project.[2] The result of our experiment: we have provided a special web service with dynamically-loadable chants and the possibility to play melodies. Users can view, edit, and play songs online. This web service is currently in development; however, initial results can already be

[2] Project, "Computer Semiography", http://it-claim.ru/semio.

viewed and tested. The data of this service is used in a statistical machine translation system developed during the project, called "Computer Semiography" [3–5, 7].

4 Chant Input Technology Based on a Text Processor

At the beginning of this project, we developed a special true-type font ("Andrew Semio") to import Znamenny Chants via Microsoft Word, along with the use of the "Irmologion" font set [10] for inputting the lyrics. Our Znamenny font went through several stages of improvements, ranging from experimental design to the study of an ergonomic component for this font. As a result, similar semiographic signs are located on the same letter, but can be changed with a different font style (e.g. normal, bold, and italic).

Chant input technology is shown in Fig. 7. Time expenses were for the first part ("Oktoikh") of the "Circle of Ancient Church Znamenny Chants" [11]. The scope of this book was 115 pages. The first step of this technology was to scan the raw chants. Processing a single page took about 2.5 min, with a total of 5 h to process an entire book. Importing one page of a chant takes about an hour. After importing a chant, checking for errors takes an additional 45 min. The end result is a book typeset in MS Word tables. Then the tables can be presented online and analyzed.

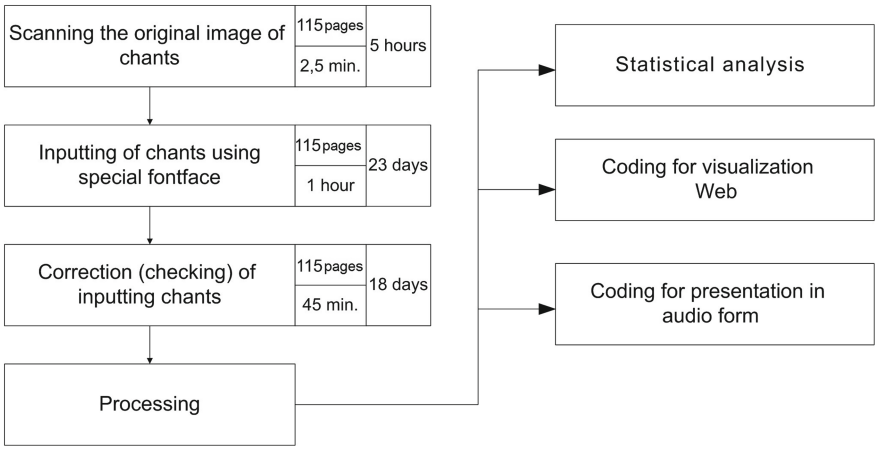

Fig. 7. Input technology for Znamenny Chants.

As a comparison, Fig. 8 shows an example of a left page from "Circle of Ancient Church Znamenny Chants" and on the right, the same page typed by using specially designed fonts. It can be observed how it is necessary to develop a font containing all ornaments, as well as a font for the title input at the first level and a font to enter the initial letters. This is necessary in order minimize the task, as well as to differentiate between the original and the electronic versions.

Fig. 8. Scanned and typed versions of a Znamenny Chant.

5 Web Services for Chant Input and Its Representation

5.1 Font Representation

There are two ways to ensure font availability. The first method would be to install the necessary fonts on a user's computer. This method is simple, but those users having additional fonts installed on their computer may be unable to perform such an installation or do it incorrectly.

The second method is an automatic replacement of server-side images and text output to a screen. There are several ready-made solutions; PCDTR (PHP + CSS Dynamic Text Replacement) [13], FLIR (Facelift Image Replacement) [18], sIFR (Scalable Inman Flash Replacement) [19]. This is a more sophisticated way; slower perhaps, but the advantage is that the user does not need to take any additional measures.

In our project for converting chants to images, we implemented it by using a specially developed technology, "PCDTR" [13]. It's a JavaScript-free version of the Dynamic Text Replacement method, originally created by Stewart Rosenberger [14]. It allows a user to take a standard HTML web page and dynamically create images to replace and enhance page headings using only PHP and CSS. These scripts have been adapted to this problem and can successfully be used as a replacement technique.

The development of fonts and the use of them for visualization purposes would allow conducting various research projects; and in the future, this may help towards decoding Znamenny Chants. Therefore, after a stage for the correction of entered

chants, it will be necessary to have the most authentic data; ideally, faultless input and a minimum difference from the original document. For this purpose, it is necessary not only to improve the fonts, but also the technology. We offer the following options:

- MS Word templates.
- OCR software (Optical Character Recognition) [12].
- Other third-party programs.

5.2 Coding Semiographic Signs

The structure of ancient musical manuscripts significantly affects the coding methods data structure, coding methods and features of editing and representation services.

The first variant for storing Znamenny Chants was developed through the use of an XML format. Each sign is encoded using special attributes, utilizing the form shown in Fig. 9:

```
▼<ROWDATA>
   <ROW Znam="а" Slog="ко" Stil=" обычный" VPom="м" DPom=""/>
   <ROW Znam="Ар" Slog="кэч" Stil=" обычный Italic" VPom="в" DPom=""/>
   <ROW Znam="а" Slog="но" Stil=" Bold" VPom="п" DPom=""/>
   <ROW Znam="а" Slog="му" Stil=" Bold" VPom="п" DPom=""/>
   <ROW Znam="а" Slog="т" Stil=" Bold" VPom="п" DPom=""/>
   <ROW Znam="а" Slog="от" Stil=" Bold" VPom="п" DPom=""/>
   <ROW Znam="а" Slog="шу" Stil=" Bold" VPom="п" DPom=""/>
   <ROW Znam="а" Slog="ся" Stil=" Italic" VPom="п" DPom=""/>
```

Fig. 9. Fragment of an XML file.

Because this is unreadable, it was recoded as an original hymn using a special conversion table (Fig. 10).

As a result, we obtained a sequence of identifiers which could be applied to the translation rules in linear notation. First, the longest rules are applied, and then the shorter ones. Finally, rules from the alphabet are applied, so as not to replace semiographic signs. Thus, from the original chant notation, signs are recoded into linear notation. To play back the chant's two codes which were generated, music and time codes are stored in a separate file and read by a Flash application.

5.3 Editing and Playing Chants

A user can review a chant, compare it with its translated linear notation version and conduct visual analysis, as well as analysis by listening to its melody. Also, the user can edit the alphabet and reconvert a chant (Figs. 11 and 12).

The user selects from a list and clicks on the chant "Send" button. Afterwards, the screen displays a Znamenny Chant notation. When you click on "Transcode", the chant is converted from a semiographic notation into a linear one. The result is then shown on the screen and its melody code is generated. After displaying the original and translated

chants, the matching strings are stored, which is convenient for later comparison. In order to listen to the resulting melody, the user must click on the "Play" button.

On the left part of the screen, there is a block of links to the basic data. After clicking on a link, the user is provided with a view of the alphabets and the initial chant which was created.

Fig. 10. Fragment of the coding table

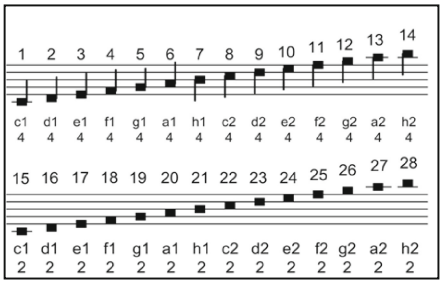

Fig. 11. Example of a code identifier matching notes.

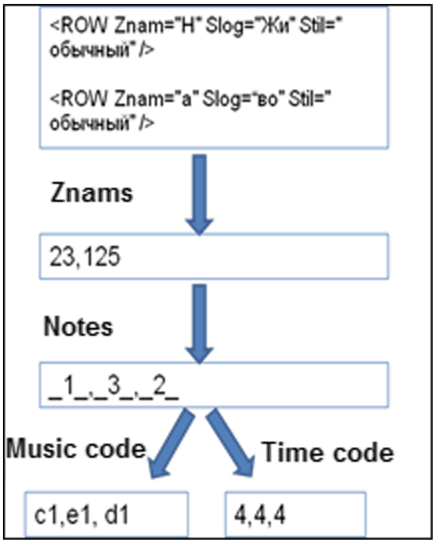

Fig. 12. Scheme of a chant conversion.

A user can review a chant, compare it with its translated linear notation version, and conduct visual analysis, as well as analysis by listening to its melody. The user can also edit the alphabet and reconvert the chant. A screenshot of the application is shown in Fig. 13. On the left part of the screen, there is a block of links to basic data. Clicking on a link, the user is provided a view of the alphabets, as well as the initial chant which was created.

Despite the fact that a developed music editor can encode songs from Znamenny into linear notation, a system of replacement rules still hasn't been worked out; namely, an alphabet translation. The current version of the alphabet is a test version built on the alphabet of a double noticed chant, ("dvoznamennik") "Tikhon – Macarius", and is not yet complete. In order to develop a high quality alphabet, studies should be undertaken with other double noticed chant manuscripts and collections of songs patterns ("popevkas"). Additionally, manuscripts are needed to be submitted in electronic form and entered into a database; whereby, the need for a special editor for this purpose would still be necessary.

The goal of our project's development is not only to obtain a system of transfer rules, but also to obtain new knowledge about deciphering chants. Applying these new

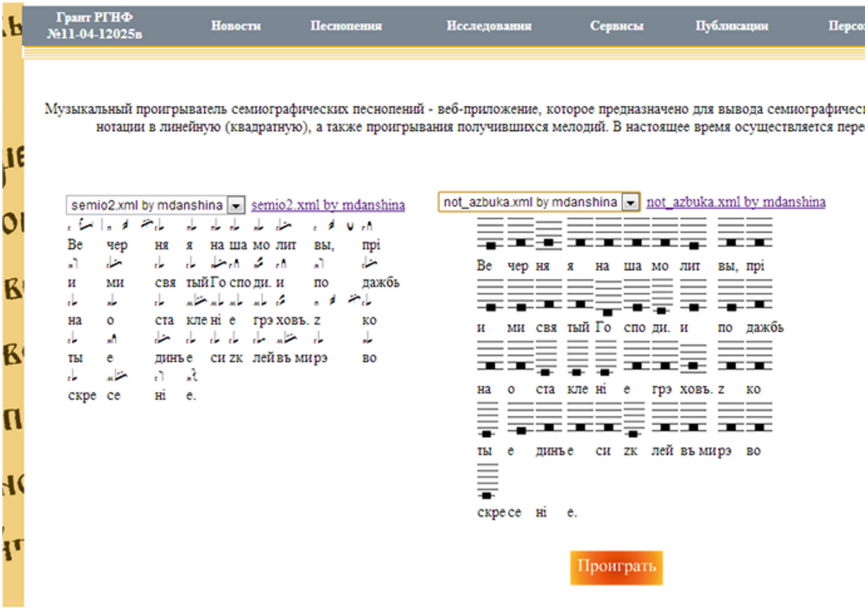

Fig. 13. Screenshot of a web service for playing Znamenny Chants.

rules to a manuscript, one can make a conclusion about the completeness of the rules and new laws and then acquire other important information.

While developing this service, the usability was also taken into account. Even though the Internet *is* widespread, it is not always available. And though the storage and viewing of manuscripts in MS Word is more convenient for viewing and listing, MS Word does not allow the processing of chants. Therefore, it is important to provide interaction both in offline *and* online operating modes (Fig. 14). The user can type a manuscript in Word, and by pressing a button, send this material to a database. To the contrary, utilizing a database by means of a web application, it is then necessary to unload pages in MS Word.

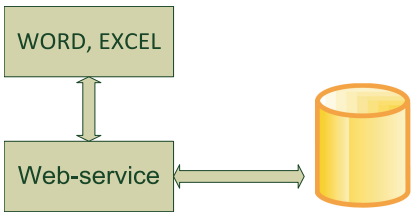

Fig. 14. Scheme of interaction for offline and online modes.

5.4 Service for Inputting Znamenny Chants into a Database

For inputting chants into a database, a special program was developed. First, the user selects a manuscript book that he is going to import, then a particular page is opened. For convenience, a page of manuscript is shown on the screen as an image during the chant's input process [15].

To input a particular semiographic sign ("znamia"), users are asked to select a group to which the sign matches, and then a particular "znamia". Division into groups is conditional: semiographic signs are grouped by similarity to their outline. The program is allocated a total of 212 banners that are divided into 6 groups; each of which has one or more subgroups. The maximum number of subgroups is 7.

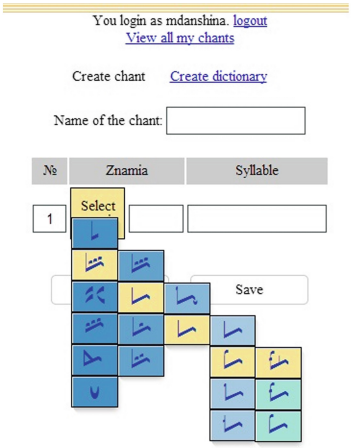

Fig. 15. Scheme example of the application for inputting chants.

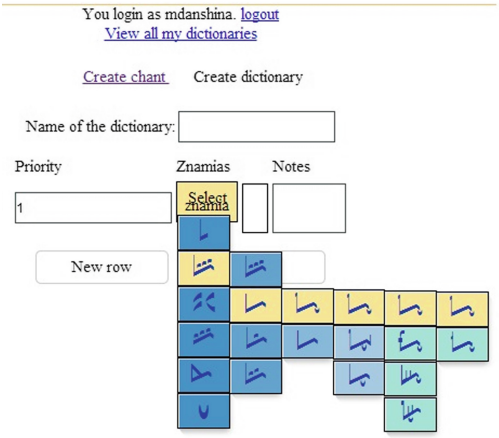

Fig. 16. Scheme example of the application for inputting a translation dictionary.

By means of this application, a registered user can not only input chants, but also a dictionary containing translation rules, which can then be used further in the musical editor. During the input of a chant, the user has to choose a separate znamia and specify a syllable which corresponds to it (Fig. 15). And should a dictionary be used/applied, the user would need to choose one or even several banners and specify to what notes they correspond (Fig. 16).

6 Conclusion

Our developed system is among the first in the field of Semiography and Znamenny Chants; as such, with the exception of Russia, there are no other like-systems with which to make a comparison. In other words, to the best of our knowledge, our system has no analogies. However, issues of visualization and the processing of ancient manuscripts on the web are relevant to other domains [16, 17].

We developed and compared several methods for the representation of ancient Russian musical manuscripts and took into consideration features of chants, as well as creation of a web service for inputting Znamenny Chants. This permits the input of chants into a database and performs a faster statistical analysis of different types of manuscripts.

As a result, the database contains 29,376 records from "Circle of Ancient Church Znamenny Chant", 234 records of the appendix for "Circle of Ancient Church Znamenny Chants", 10,897 records of double noticed chants (dvoyeznamennik) and 16,914 records of the manuscript with a collection of song patterns (popevkas).

In order to listen to a resulting melody, a new web service has been developed. Any registered user can create their own semiographic alphabet with rules for decoding znamenny signs (hooks) and their sequences (patterns).

References

1. Preobrazhenskij, A.: Ocherk istorii cerkovnogo penija v Rossii (1915) - in Russian
2. Brajnikov, M.: New Monuments of the Znamenny Chant, Leningrad (1967) - in Russian
3. Danshina, M.V.: Using methods of machine translation for analysis of ancient music manuscripts. Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh 16, 263–266 (2013) - in Russian
4. Philippovich, A.Yu., Danshina, M.V.: The frequency analysis of verbal and melodic associations of Znamenny Chants. In: Challenges of Information Society and Applied Psycholinguistics - Proceedings of the X International Congress of the International Society of Psycholinguistics, pp. 251–252. RUDN - Institute of Linguistics RAN - MIL, Moscow (2013) - in Russian
5. Golubeva, I.V., Philippovich, A.Yu.: Syntactic analysis musical texts. Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh 16, 257–262 (2013) - in Russian
6. Danshina, M.V.: The method of selection, storage and processing of popevkas on musical manuscripts. In: Information Technology and Textual Heritage. materialy IV mezhdunar. nauch. konf., Petrozavodsk, 3–8 Sentyabrya 2012, pp. 79–81 (2012) - in Russian
7. Philippovich, A.Yu., Golubeva, I.V.: Research of syntax of semiographic chants. In: Proceedings of the Higher Education Institutions. Problems printing and publishing, vol. 6, pp. 147–163 (2012) - in Russian
8. Philippovich, A.Yu., Smoliakov, B.G.: Computer semiography. In: Kniga i mirovaya tsivilizatsiya: Materialy XI Mezhdunar. nauch. konf. po problemam knigovedeniya, Moskva, 20–21 April 2004, vol. 1, pp. 398–401. Nauka, Moscow (2004) - in Russian
9. Smoliakov, B.G.: About melodic ambiguity in Russian bezlineynoy stolpovoy notation. In: Sergievskie chteniya -1. O russkoy muzyke : nauch. tr. Sb., vol. 4, pp. 33–40. MGK im. P.I. Chaykovskogo, Moscow (1993) - in Russian
10. Irmologion fontset. http://en.wikipedia.org/wiki/Irmologion_(typeface)
11. Krug tserkovnago drevnyago znamennago peniya v shesti chastyakh. Obshchestvo lyubiteley drevney pis'mennosti, Saint-Petersburg (1884) - in Russian
12. Vylomova, E.A.: Recognition System of semiographic chants. In: Intellektual'nye tekhnologii i sistemy. Sbornik uchebno-metodicheskikh rabot i statey aspirantov i studentov, Moscow, vol. 9, pp. 58–70. NOK «CLAIM» (2007) - in Russian
13. Technology PCDTR. http://www.artypapers.com/csshelppile/pcdtr/

14. Rosenberger, S.: Dynamic Text Replacement. http://alistapart.com/article/dynatext
15. Danshina, M.V.: IPSM - application for inputting and processing semiographic chants. In: Information Technology and Textual Heritage. materialy III mezhdunar. nauch. konf., Izhevsk, 3–8 Sentyabrya 2010, pp. 79–82 (2010) - in Russian
16. Doumat, R., Egyed-Zsigmond, E., Pinon, J.-M.: A web 2.0 archive to access, annotate and retrieve manuscripts. Multimed. Tools Appl. **64**(1), 97–117 (2013)
17. Doumat, R., Egyed-Zsigmond, E., Pinon, J.-M.: Digitized ancient documents what's next. Document Numérique **12**(1), 31–51 (2009)
18. Facelift Image Replacement. http://mawhorter.net/flir
19. sIFR 2.0: Rich Accessible Typography for the Masses. http://www.mikeindustries.com/blog/sifr/