
Textograf: A Web Application for Manuscript Digitization

Boris Orekhov

nevmenandr@gmail.com

National Research University Higher School of Economics

Russia

Fekla Tolstoy

6975991@gmail.com

The Leo Tolstoy State Museum, Russia

Textograf is a web-based app for the digitization of manuscripts. *Textograf* is intended to automate the work of the textual critic of any handwritten text. *Textograf* allows you to select a specific set (or “layer”) of edits to a text, to compare different iterations of the text, and ultimately to visualize the correlation between the final text and the process that stood behind it. *Textograf* is able to work with prose and poetry, with manuscripts and printed sources. A description of the project can be found on the *Textograf* [web page](#).

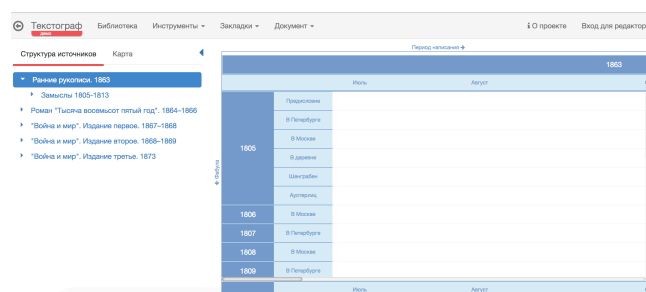
The app allows you to upload an image of the manuscript, to enter the transcript of the manuscript, and to match specific parts of the transcript with where they appear on the page of the manuscript. You can also mark on the transcript the different stages of edits that were made. The app also enables you to see and highlight all text that was deleted, all text that was added, and all text that was initially deleted and subsequently restored.

All documents (scanned images and texts) are stored in a so-called “library”. Access to the documents depends on the documents’ status – “public” or “private”. Private documents are visible only to the editor, public documents can be accessed by everyone.

Textograf allows its user to create complex documents that include manuscript images and their text versions. Text fields are searchable, allowing the user to find search results in the manuscript itself. Every manuscript image can be archived according to customized categories (whose handwriting, where it is currently stored, size of paper, pen or pencil, published or not published, etc).

Textograf allows you to download documents from the library in TEI format. Texts can then be ordered in accordance with the user’s version of the text. Often the text has a number of different versions, and the user has to find the right one and to collect all the corresponding manuscripts. *Textograf* makes this possible.

Different types of source documents connected to the work (outlines, synopses, the manuscripts themselves, printed editions) can be shown as an infographic, demonstrating visually the correlation between the different source documents. For example, this is a map of the early stages of Leo Tolstoy’s work on *War and Peace* [displayed on the Textograf site](#).



The app was initially developed using the manuscripts of Leo Tolstoy: 200 pages (out of 5,000) from *War and Peace*.

The map has two axes: the real time of Tolstoy’s work (starting from 1863) and the fictional place and contents of the novel so you can see which episodes Tolstoy started with and how each part of the text came together.

As of today, the *Textograf* app is unique. It allows the user to work with manuscripts and texts in automatic or semi-automatic mode, meaning that the user can focus on the creative aspects of this work and become immersed in the manuscripts. Before the creation of *Textograf*, many of these functions had to be performed manually, such as establishing the link between text and image, and identifying the sequence of writing.

In the process of developing the application, we had to convert the terms and methods of textual criticism in terms of electronic resources. For example, we had to establish the meaning of the word “layer”, or “edit”, and to show how to visualize it on the screen.

When we talk about the influence different inputs, we mean that any text has a complex history. For example, a poem may obtain its final form as a result of a long history of changes from one version to another. Thus, what stays in the archives will be several manuscripts that contain similar, but not identical versions of the same text. Earlier versions will influence the

later ones. The "text map" shows the full history and evolution of the text.

Textograf is designed to work with any paper-based sources. It could be both manuscripts and printed editions. *Textograf* is a computer programme editor and is designed to prepare in electronic form both critical and diplomatic editions.

Textograf can work with the manuscripts of any writers, be it Leo Tolstoy, William Faulkner, Oscar Wilde, Marcel Proust or anyone else.

As already mentioned, *Textograf* is a web-based application and cannot be installed on a computer. Integration with other tools is possible only through the files that can be downloaded from the application.