

DOI: 10.17516/1997-1370-0703
УДК 026.0

“Volume 91”: an Electronic Index to the Complete Works of Leo Tolstoy

Boris V. Orekhov*

*National Research University “Higher School of Economics”
Moscow, Russian Federation*

Received 11.08.2020, received in revised form 11.09.2020, accepted 09.12.2020

Abstract. The collected works of Leo Tolstoy were printed and published in 90 volumes of some 46,000 pages between 1928 and 1958. This paper, however, is not about the 90 volumes themselves, but about Volume 91 of this edition, a supplement volume containing indexes of works and proper names, from both the fictional works and the many volumes containing Tolstoy’s letters. “Volume 91” is a web application based on the digitised index of proper names for the 90-volume collection of Tolstoy’s works (<http://index.tolstoy.ru/>). The digitised data features additional properties, which can be explored by the enthusiast as well as the specialist. This paper not only presents a new tool for literary scholars, but generalizes and shows how this kind of resources can be used to gain new insights into larger text collections.

Keywords: index, book studies, web application, digital literary studies.

Research area: philology.

Citation: Orekhov, B.V. (2020). “Volume 91”: an electronic index to the complete works of Leo Tolstoy. *J. Sib. Fed. Univ. Humanit. Soc. Sci.*, 13(12), 2049–2055. DOI: 10.17516/1997-1370-0703.

1. Introduction

Thanks to A.F. Losev, the aphorism "Das Buch ohne Index ist kein Buch" ("A book without an index is not a book") attributed to historian T. Mommsen is widely known. Despite the fact that it is hardly possible to agree with this statement as it is, it reflects the attention paid to this component of the apparatus of the edition by professionals. "There were orders of the State Committee for Publishing (Goskomizdat), obliging to make indexes to books of more than 20 sheets" (Grishunin, 1998). These orders are not accidental: texts published in books often (especially in research practice) become the object of not linear reading, but of selective search, in which the index provides important navigational assistance. According to the well-known bibliologist A.E. Mil'chin: "The one who searches for information in a book without an index is like a person who searches for a friend's apartment in a multi-storey building without knowing its number. The reader is forced to study the contents of dozens of pages until he finds the text fragments he is looking for, just as a person looking for an apartment is forced to knock on dozens of doors before he finds the one his friend lives in" (Mil'chin, 2009). It is for obvious reasons that the index acquires its particular value in multi-volume editions. If we have 90 volumes in the edition (that is true for the complete works of Leo Tolstoy (Tolstoi 1928–1964)), the index becomes a key tool to work with the text.

2. Data

Indexes to *the Complete Works* of Leo Tolstoy are presented in a separate volume (Ukazateli, 1964) and combine separate indexes to individual volumes of the collection. The book is divided into 4 sections:

- 1) an alphabetical index of works,
- 2) an alphabetical index of addressees,
- 3) an alphabetic index of proper names,
- 4) a chronological index of works.

Thanks to the large-scale crowdsourcing project "All Tolstoy in one click", in 2014 *the Complete Works* were digitized, currently all texts are available on tolstoy.ru. Simultaneously with the texts, an additional volume with indexes was subject to digitization.

3. Digital index problem

Despite the fact that the indexes were in a digital form, it was necessary to use them in the same way as they are used in the paper form: first, it is necessary to find a person in the list of proper names, to find the volume number and the page on which this person is mentioned, and then to find the volume and the corresponding page on the web-site. At the same time, the digital form can facilitate these routine operations, if you can make them in a specialized web application. In addition, the index itself contains a large amount of additional research-valuable information that is difficult to retrieve manually, but it becomes available through technical means.

However, what the configuration and design of such a digital service should look like is not obvious. Simple mechanical transfer of practices common to other areas of knowledge to the digital support of the humanities is not necessarily effective. It would be productive to develop an application architecture based on the specificity of the data that professionals deal with, the data from a philological index in this case. In such a situation, resource development itself becomes not so much an engineering problem but a scientific one, the center of which is to identify the most conceptually loaded sides of the available data and to find ways to adequately present them in the service.

During 2016-2017, the scientific team of the National Research University Higher School of Economics and the State Museum of Leo Tolstoy worked on the creation of an electronic resource based on the data from the index to *the Complete Works*.

4. Data Approach Concept

The concept was to use a minimum of manual labor associated with markup and data extraction and to rely on automatic index processing. We based our approach on the fact that, on the one hand, significant human resources have already been invested in the creation of the index, on the other hand, an attempt to repeat these efforts will lead to a significant delay in the application development. At the same time, these were precisely philological considerations that were at the core of the application architec-

ture. To these we attribute a) the idea of close interaction between the text and the reference apparatus, b) emphasis on the structure of the text, c) attention to the source. It is the criticism of the source and the ranking of information depending on the source considerations that often distinguishes the human view on the data from the natural sciences. A good tone in philological science is considered to be a reference to the first publication of a cited text. Such considerations are not typical to mathematical or physical scientific etiquette. It seems not by chance.

The results of the work show that the chosen strategy has proved to be effective; the structured nature of the index implies a comparative ease in computer data processing.

The application was built and deployed at index.tolstoy.ru. At the moment it functions on the Internet and is publicly available. Its architecture embodies the above mentioned nuclear elements of the philological approach to data.

Let's consider them sequentially.

5. Structure of the web application and its corresponding conceptual areas

5.1. Digital index

Web application "Volume 91" helps to achieve several goals at once.

As a basic principle, the index retains its original functionality, namely the index of names. The integration of the text and references is an important factor of practical research work. The index in the digital form is much more convenient than the paper version. Instead of manual search of the names in the list the user has an opportunity to use a search line and, having entered a name, to receive the reference to the page on which this name is mentioned. One search query can result in several different items. These items are ranked from the most common in Tolstoy to less frequent. This ranking for items is made in the interests of the user: it is assumed that the frequently mentioned name is most likely to be searched for.

So, with the query "Siberia" we find 5 results; the actual name of the region is mentioned in *the Complete Works* on 193 pages. In addition, we found the "Sibir" newspaper and the names of several low-frequency place

names. The basis of this arrangement of the search results is the hypothesis that the user is interested in the Siberian region of Tolstoy, and they will see it in the first place.

The search incorporated in the system is the so-called "fuzzy". That means that by entering a combination of the letters "ava", for example, the results show both "Poltava" and "Bavaria", "Abdulla al-Mamun Zuravardi".

This functionality, not available in the book version, is critical for the cases when the reader does not exactly remember what name he needs.

If the user is not sure what he wants to see, and he needs a list of all the names from the index (there are more than 16 thousand of them), then this list is also available. Due to several circumstances, it is also more convenient than a paper one. In a book form, if we learn an indication on the page where the necessary person is mentioned, we should take the book from a shelf and open it on the necessary page. "Volume 91" shows us all the references (as in the paper index), but the web app additionally provides all the hyperlinks to the corresponding pages, and we can click on the link to go to the place in the digital version of *the Complete Works*.

The user has the so-called "word cloud", that is, a graphic representation of the frequency of mentioned names in *the Complete Works*. This cloud does not contain all the names from the index, but only the most frequent ones, otherwise the cloud would become unreadable. Thanks to the cloud, we can estimate the frequency of references to various names in books. The paper version of the index does not actually provide for such an assessment.

5.2. Data for the network analysis

The Index provides a useful tool to study Leo Tolstoy's life and work, for example, in the framework of the approach that is now gaining popularity. The approach is commonly called social network analysis. Yet, the method itself is wider and more complex. It also analyzes objects which have no relation to social life (i.e., linguistic or literary criticism data).

The essence of the method is the representation of some entities (in this case, personal-

ities from the index) and the relationships between them in the form of a scientific model. The entity appears in the model as a node of the graph, and the connection is its edge. This model is promising for calculating the most significant network nodes (i.e., using centrality calculation algorithms) and distributing nodes across different "subnets", clusters within a large graph.

"Volume 91" shows us Leo Tolstoy's social network, where not only the writer's connections with other people are recorded (this can be seen, for example, in the correspondence), but also the connections between people. This network is based on the same index: the fact of the joint occurrence of some names on a page in a 90 volume edition becomes the basis for drawing the edges of the graph, which is realized as a connection within the network. The Indian "Bhagavad-gita" appears on the pages of *the Complete Works* 5 times, while 43 more names are mentioned on the same pages. These names are not random. They all relate to a single topic and form an "Indian cluster": "Gītopadesa", "Dhammapada", "Vāmana Purāṇa", and Ramakrishna Sri Paramagamza. However, Tolstoy is also interested in these texts as the texts of philosophical knowledge. Therefore, these names are related to such names and titles as "The Bulletin of Theosophy", Xenophon, Montaigne, Montesquieu, Pascal, Skovoroda, and Socrates.

These networks offer a great opportunity for scientific methods to cover the diversity of Tolstoy's interests and ideas. You can evaluate the scope on the page depicting the whole network of names. It presents a panoramic picture, which gives the idea of the general trends and the largest thematic clusters. For each individual name there is also a small graph showing the most significant links to other names.

It is noteworthy that the thematic rubrics in the Index have already attracted the attention of book history experts: "Each of the references to the book pages in the index covers one or another aspect of the topic indicated by the name of the search subject in the heading of the section. However, fragments of the book material, which the heading (cell) of the index refers to, as a whole, form a collection

of material from the book about the object of interest to the reader. The heading of the index is, therefore, a selection of materials, the content of which is replaced by the page numbers" (Mil'chin, 2009).

5.3. *Proper name and text structure*

Another useful tool for studying Tolstoy's work is heatmaps of names. In this part of our concept, emphasis on the structure of the text is crucial for philology. Proper names are an important constructive component of a literary text (Sokolova, 2011): "Proper names included in the structure of a work of art are directly related to its content. Studying them in literary onomastics follows first of all from the need for a deeper understanding of a work of art".

Proper names attract the reader's attention more than other words. It is due to the capitalization. In addition, names are usually heavily loaded (cf. the analysis of associations for Eugene Onegin's name in Yury Lotman's commentary to the novel). Finally, the names have special phonetics, which is extremely meaningful in the words of foreign origin. All these effects create a special layer of the text structure that is important for the perception of a work of art.

The web-application has a page that shows the frequency of mentioning the proper names in Tolstoy's texts, this frequency (or a lack of it) is color-coded: the more names the text involves, the warmer the color in a range from cold blue to hot red is. So, in the very first volume of the edition, where *Detstvo* ("Childhood") and youthful experiments were published, in the middle of a rather calm blue background, a red splash suddenly occurs around page 269. You can visit this place in the edition and find out that a number of European cities (Rome, Naples, Dresden, and Berlin) are listed there one after another. They are listed in the way to create an effect of meaningless flickering of these places in front of the reader's eyes.

5.4. *Criticism of the source*

The index allows you to study another important cultural subject with a difficult fate, associated with Leo Tolstoy indirectly. It is a publication entitled "The Complete Works" by

L.N. Tolstoy. The editorial work itself is worthy of study. The difficulties of extra-philological nature that had to be overcome while working on this edition are outlined in a special book (Osterman, 2002). But the conversation about this is not over yet. The index contains indirect data that makes it possible to shift the focus to the book history aspects of the edition, or better to say, to engage in criticism of the source of Tolstoy's texts.

Volume 91 allows you to follow the change in editorial principles. On a special page, you can see the difference in the ratio in the proper names mentioned in Tolstoy's text and in the editorial comments. This reflects (often due to time and external reasons) the amount of time and work spent on a commentary or simply the availability of a commentary in this volume. Volume 13 with draft revisions of *Voina i Mir* ("War and Peace") has no commentaries (and this can be evidenced by the number of names in Tolstoy's text which is more than the number of the names in the comments), whereas Volume 47 (diaries and notebooks) required so many detailed comments that it became the most detailed in the entire 90-volume set. Behind all these numbers is the specificity of the data contained in a particular volume as well as, at the same time, the history of the preparation of the books for publication.

6. Regional studies:

Leo Tolstoy and the Yenisei province

Here is an example of how a digital index can serve as a starting point for local history research.

Let us use the "Yenisei" query in order to get all the references to the Yenisei province in all its forms (Table 1).

Items found in the index are not grouped in a single alphabetical nest. This means that the search by paper index would imply looking through the entire list of 16 thousand records. There is no other way to find all the six relevant records. The digital version helps us to retrieve all the necessary items.

"The Tale of Life and Deeds..." is mentioned in *the Complete Works* in two volumes: Volume 63 and Volume 75. These are the volumes of letters dating 1880-1886 and 1904-1906. In the first case, the letter is to P.I. Birjukov. Tolstoy asks him to find a book about Daniil Achinsky. The writer calls either the book itself or its hero's deed something of "a great importance for a historic event". In the second case, Tolstoy addresses a similar request ("find Daniil Achinsky, please") to I.I. Gorbunov-Posadov, a writer.

"Notes on the Yenisei prison" appears in Volume 53. These are diaries and notebooks of 1895-1899. As it follows from the editorial comments in the same volume, this book is included in the list that Tolstoy studied in the course of his work on *Voskresenie* ("Resurrection").

The Yenisei province and Turukhansk are mentioned in volume 74 which contains letters from 1903. In his letter to the Yenisei governor N.A. Aigustov Tolstoy asks for easier living conditions for the exiled peasant Athanasius Ageev, who was convicted of blasphemy.

Volume 68 (letters of 1895) mentions the village of Iudino, where, according to the

Table 1. Results of the "Yenisei" query retrieved by the index

№	Item in the index	Number of references
1	"The tale of life and deeds of the deceased old man Daniil, who asceticized in the Siberian country in the Yenisei province, within the city of Achinsk"	4
2	Krivoshapkin, M.F. "Notes on the Yenisei prison"	2
3	Yenisei province	1
4	Iudino Yenisei province	1
5	Tes' Yenisei province	1
6	Turukhansk Yenisei province	1

writer, Timofei Mikhailovich Bondarev lived. Tolstoy asks S.A. Vengerov about Bondarev who was exiled to Siberia and settled in Iudino for "subbotnichestvo". Volume 79 (letters of 1909) mentions the village of Tes' as the place of residence of G.N. Vetvinov's sister, the defendant, in whose fate the writer took part.

As we can see, the picture is quite complete. The Yenisei province appears in Tolstoy's biographical (not artistic and journalistic) texts only in connection with penitentiary occasions (Daniil Achinsky also ended up in Siberian prison). Obviously, this was the image of the region: this is the area of exile and prisons, and Tolstoy, well aware of this, is still trying to use his authority to make the fate of people (including strangers), who found themselves in a difficult situation, easier.

7. Conclusion

The application retains all the features of a traditional index. At the same time it expands

its potential due to computer tools of information manipulation, search engines and visualization. "Volume 91" should be useful both to the general reader and to a specialist who studies Tolstoy's texts.

In some ways, the web application gives us a kind of Tolstoy's mental map, his intellectual universe. Yet, even such a perspective turns out to be a giant system which is difficult to see at a glance. The digital mode allows us to manipulate the model, retrieve information from it, and present it to the user.

The development of an application based on the specificity of the data structure and the approach to the data adopted to the subject area (philology) can yield useful results. Our experience has shown that the subject area of philology, sensitive to the status of the apparatus, the structure of the text and the critique of the source, can provide a conceptual vector for the design of digital services, useful for a wide range of studies.

References

- Grishunin, A.L. (1998). *Issledovatel'skie aspekty tekstologii [Research aspects of textology]*. Moscow, Nasledie, 416 p.
- Mil'chin, A.E. (2009). *Spravochnik izdatelia i avtora [Handbook of the publisher and the author]*. Moscow, Izd-vo Studii Artemiia Lebedeva, p. 1084.
- Mil'chin, A.E. (2012). *Kak nado i kak ne nado delat' knigi. Kul'tura izdaniia v primerakh [How to make books and how not to. Publication culture in the examples]*. Moscow, NLO, 352 p.
- Osterman, L. (2002). *Srazhenie za Tolstogo. Istoriia izdaniia polnogo sobraniia sochinenii [The Battle for Tolstoy. History of the publication of the Complete Works]*. Moscow, Grant, 296 p.
- Sokolova, M.V. (2011). Funktsional'no-stilisticheskaia nagruzka imeni sobstvennogo v hudozhestvennom tekste [Functional-stylistic load of the proper name in fiction]. In *Vestnik Cheliabinskogo gosudarstvennogo universiteta [Bulletin of the Chelyabinsk State University]*, 33 (248), Filologiya. Iskusstvovedenie [Philology. Art Criticism]. Issue 60, 182-184.
- Tolstoi, L.N. (1928-1964). *Polnoe sobranie sochinenii: V 90 t. Iubileinoe izdanie [Complete Works: In 90 vol. Anniversary Edition]*. Moscow-Leningrad, Gos. Izd-vo.
- Ukazateli k Polnomu sobraniuu sochinenii L.N. Tolstogo: Alfavitnyi ukazatel' proizvedenii. Alfavitnyi ukazatel' adresatov. Alfavitnyi ukazatel' imen sobstvennykh. Khronologicheskii ukazatel' proizvedenii [Indexes to the Complete Works of Leo Tolstoy: Alphabetical index of works. Alphabetical index of addressees. Alphabetical index of proper names. Chronological index of works]* (1964). Moscow, Khudozh. Lit., 667 p.

«91-й том»: возможности электронного указателя к полному собранию сочинений Л. Н. Толстого

Б.В. Орехов

Национальный исследовательский университет

«Высшая школа экономики»

Российская Федерация, Москва

Аннотация. Собрание сочинений Льва Толстого было опубликовано в 90 томах объемом около 46 000 страниц в период с 1928 по 1958 год. Но наша статья о 91-м томе этого издания, томе-приложении, содержащем указатели произведений и имена собственные как из художественных произведений, так и из других томов, содержащих письма и дневники Толстого. «91-й том» – это веб-приложение, основанное на оцифрованном указателе имен собственных для 90-томного собрания сочинений Толстого (<http://index.tolstoy.ru/>). Оцифрованные данные имеют дополнительные свойства, которые могут быть полезны как энтузиастам, так и специалистам. В этой статье делается попытка не только представить новый инструмент для литературоведов, но и обобщить, как можно использовать такого рода ресурсы для получения нового знания из крупных текстовых коллекций.

Ключевые слова: указатель, книговедение, веб-приложение, цифровое литературоведение.

Научная специальность: 10.01.00 – литературоведение.