

Russian Virtual Library (RVL)

Some aspects of visualization of literature text and comments to it

Vladimir Litvinov, presenting author
Russian Federal Nuclear Center,
456770, Grechishnikova 7—2, Snezhinsk, Chelyabinsk reg., Russia,
tel.: +7-35172-32363, e-mail:litvinov@snezhinsk.ru

Igor Pilshchikov,
Institute of World Culture, Moscow State University,
RVL Editor-in-Chief
e-mail: igor@rvb.ru

Abstract:

The paper covers basic aspects of functioning of Russian Virtual Library and application of conventional mark-up languages for demonstration of intrinsic text structure.

Introduction

Russian Virtual Library (<http://www.rvb.ru/>) is an Internet resource that was developed to satisfy the needs of a wide range of users studying Russian literature (from scholars to high school students). Critical editions of Russian literature texts are the major part of the site content.

Two major problems are solved within the site development:

1. Authentic reproduction of the text and comments to it following the scholarly edition;
2. Marking-up the text in a meaningful way, similar to those used in TEI scheme [1] and presentation of this mark-up in a way reproducing layout of original book.

1. Visualization of bibliographic description of the text

All texts are marked-up in HTML 4.0, thus author of the text, title and brief reference to reproduced edition are written into the <title> element of each file. While re reader browses along the text, he sees numbers of pages and full bibliographic description is given in the footer of each text and in navigation pages (index of the author).

2. Visualization of text structure

Basic assumption in text mark-up is the following: formats of text presentation developed within the traditional publishing industry implicitly include vast information about text structure. These templates of text presentation deserve adequate explication in terms of text theory and recording in text mark-up language.

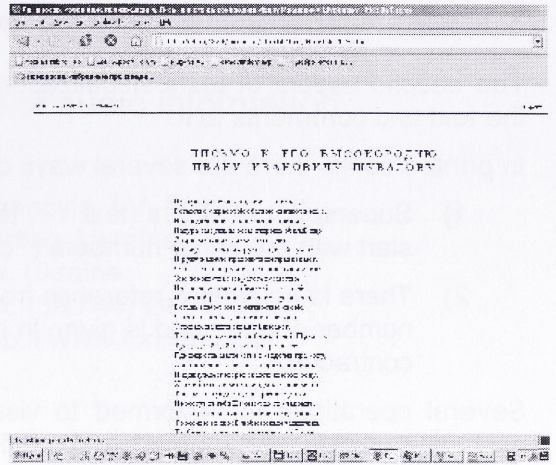
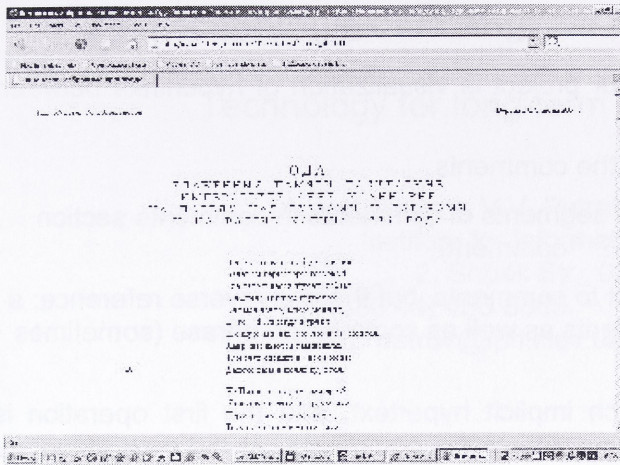


Figure 1: Poem written in iambic four-foot verse (tetrameter) (left) has wider left margin than the one written in iambic six-foot verse (hexameter) (right).

Major intention of SGML (XML) languages, that is unification of the artistic text itself and metadata (encoded in mark-up language elements) in marked-up text is implemented within RVL project using HTML 4.0 language [2], and visualization of this mark-up is governed by CSS 2.1 rules [3].

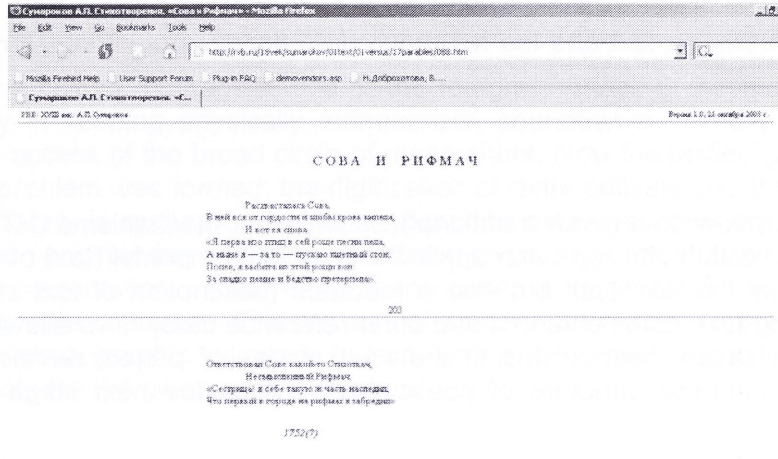


Figure 2: Poem written in iambic hexameter is designated as `<div class="versusia6">`, and in iambic tetrameter as `<div class="versusia4">` (HTML 4.0), and respectively in CSS language for the first case we have the rule `{margin-left: 20%;}` and for the second — the rule `{margin-left: 30%;}`.

Values of this margin may be varied, but mutual position of the poems written in different meters is preserved as in printed books and thus text visualization becomes meaningful for the reader. All verse meters, their variants and other meaningful text elements are defined in a similar way. On one hand this makes perception of the text more easy and on another hand simplifies text processing in view of making the indexes of verse forms and/or concordances. When the text is marked up this way, it can be easily converted to other formats to be given to the reader. Major tool for such markup is Xmarkup utility

(Russian description is given at http://www.rvb.ru/soft/xmarkup/xmarkup_170.htm)

Examples of the text marked-up this way may be found, for instance, following the links from the Karamzin's (Russian writer of 18th century): <http://www.rvb.ru/18vek/karamzin/1bp/toc.htm>

3. Visualization of comments to the text

The second aspect of texts visualization within our project is visualization of hyperlinks between the text and comments to it.

In printed books there are several ways of links to the comments:

- 1) Superscript numbers (text ¹ ... text ²), and segments of comments in comments section start with the same numbers (¹ comment ... ² comment),
- 2) There is no explicit reference from the text to comments, but there is reverse reference: a number of page and is given in the comments as well as commented phrase (sometimes contracted).

Several operations is performed to visualize such implicit hypertext, and the first operation is assigning unique alphanumerical identifiers to segments of the comments. At the next stage we define which segments of the text are commented and links are semi-automatically generated to the comment file and IDs.

For example in the novels of Fedor Dostoevskiy: “Crime and Punishment” — ca. 200 links to comment items, “The Possessed”— ca. 300 links, “Karamazov Brothers”— ca. 600 links.

Links are assigned the appearance traditional for web-resources (in CSS 2.0 rules) and when the reader sees the text highlighted in blue color, he realizes that there is a hyperlink and when he clicks, he sees the needed comment.

Hypertext links between text and the reference data (indexes of various kinds) are visualized in a similar way.

Conclusion

We use quite conventional (even traditional) visualization mechanisms (HTML 4.0 + CSS 2.0) for text visualization within our approach applied to traditional material (text of literature), however we visualize not only the text itself but also a metatext (description of text structure) and hypertext (system including text, comments to it and other reference data). This allows development of more advanced visualization mechanisms at the next stages of project development. Most probably visualization of phonetic structure of poetic texts will be the next stage of applied technology development.

References

1. Text Encoding Initiative. The XML Version of the TEI Guidelines — <http://www.tei-c.org/P4X/>
2. HTML 4.01 Specification — <http://www.w3.org/TR/REC-html40/>
3. Cascading Style Sheets, level 2 revision 1. CSS 2.1 Specification — <http://www.w3.org/TR/CSS21/>