Classical Russian Literature and Digital Humanities

The lecture will focus on digital libraries and collections of texts that offer comprehensive body of materials pertaining to classical Russian literature and folklore. In particular, attention will be given to their functional capabilities (information search and retrieval) and methods of text presentation (text, image, image+text), as well as to available reference tools (concordances, dictionaries, encyclopedias, bibliography). Additionally, digital resources designed for the analysis of Russian literary texts will be discussed: Russian language corpora, tools for linguistic analysis (such as morphological analyzer), and programs of digital-computational verse analysis.

Lecture outcomes

Learn / Get an idea
– how to effectively use the functionality of different search engines (such as Google, Yandex, etc.),
– how to search Russian texts written in different spelling systems (contemporary and pre-revolutionary orthography),
– how to effectively retrieve information from digital libraries,
– how to take advantage of digital corpora and massive collections of texts,
– how to use linguistic and reference tools for your study and research.

1. The use of search systems.
Query languages of specific engines (Google, Yandex, Rambler, Mail.ru).
Morphological analyzers. Lemmatized search (morphologically-based techniques) and word form search.
Search of Russian texts written in different spelling systems (contemporary and pre-revolutionary orthography).

Project history.
Basic principles.
The structure of FEB-web.
Digital Scholarly Editions.
Data composition.
Search functions.
Display and export of information.
Digitization, markup, and formatting process.
English-language interface (beta-version).

3. FEB-web 2.0. From a “marked-up text” to “image+text” presentation.
New digitization, markup, and formatting process.
Structure of text and structure of edition.

4. The Russian National Corpus (RNC).
Project history.
Basic principles.
The structure of RNC.
The Main Corpus and the subcorpora.
The Spoken Speech Corpus.
The Corpus of the contemporary Russian press.
The Poetic Corpus.
5. RNC: functional capabilities.
Lexical and grammatical search.
Распределение по годам (cf. Google Books Ngrams)
N-gramms (not to be confused with Google Books Ngrams!)
Search of poetic forms.
English-language interface (beta-version).

6. The English-Russian and Russian-English parallel corpora at RNC.
New parallel corpora projects:
The Russian-French Poetical Corpus
A Corpus of Translations of the Tale of Igor’s Campaign.
Corpora for lexicographers: a problem.

7. Other important digital libraries of Russian literature and collections of texts.
Lib.ru
Russian Virtual Library (RVB)
ImWerden.de
aleph.rsl.ru — электронные версии
Overview of the histories and the present state of the projects.

8. Russian books in the Google Books project.
Advantages and disadvantages of Google Books as discussed by librarians and academicians.
Google Books  Ngram Viewer
Problems with metadata. Gbooks = starieknigi.info
Search problems connected with OCR
Various types of Russian text recognition strategies used in this project, and the resulting heterogeneity.
The ways to overcome the information chaos and to use this huge but under-organized resource for study and research.

9. Concordances:
FEB-web: The Griboedov Concordance and the Lomonosov Concordance ().
RVB: Word indexes. The wordTabulator utility.
Dictionaries and encyclopedias:
   FEB-web
   Yandex.Slovari
   Other resources
The Abby Linguo dictionaries on Yandex.Slovari.
Каталоги РГБ (алф), РНБ (карточный и машиночитаемый) и НБ МГУ (карточный и машиночитаемый)

Web resources
7. GBooks “Old Russian Books” http://starieknigi.info

Scholarly papers and collections of essays

