

The National Library Bibliographic Classification (BBK) as a Base for Subject Search in the Integrated RSL Digital Library. Project presentation.

Lavrenova Olga, Russian State Library, Moscow

1. **The integrated RSL Digital Library** contains about 805000 titles, including the Dissertation Digital Library (690000 full texts).
2. The full text search is possible in the most of RSL digital recourses. A **high quality of subject search** is meanwhile accessible only provided that a system can take into account semantic relationships between search characteristics (hierarchy, synonymy, association). The project's purpose is ensuring of these conditions.
3. **The RSL uses a full version of the Library Bibliographic Classification BBK** (about 120000 classification numbers, Classification Scheme code *rubbk* in MARC 21). All catalogue cards in the main RSL catalogues contain BBK classification numbers. Most bibliographic records in the OPAC contain BBK as well. In addition, the Library uses free (uncontrolled) key words in the OPAC. The full BBK version classification numbers are in use in some big Russian libraries as well. There exist some condensed versions for other libraries.
4. As is well known, some libraries install in their OPACs various machine-readable classification schemes in their initial forms. But it is not very easy for users to construct a search characteristic if a classification structure is complicated. Many present-day users would like to get an outcome quickly and "on default".

The RSL has chosen **2 variants of technologies to use semantic relationships by information retrieval on an existent classification base:**

- **Inclusion interpretations (or wording) of BBK classification numbers into every bibliographic record or in metadata embedded in full texts (XML structure, e.g.).** An interpretation of a BBK classification number looks like a caption hierarchy, i.e. an hierarchical sequence or a string of captions of all hierarchical levels in the classification number. It allows to take into account all hierarchical relationships and automatically on default by information retrieval. The technology has been embedded in the RSL OPAC from the very beginning (since 1998) and improves "on default" subject search capabilities to a great extent.
- **Installation in open access an «operational» classification variant related with OPAC.** This «operational» classification will include all complex classification numbers from all bibliographic records generated during a cataloguing process and caption hierarchy for every complex classification number.

Example 1. Subject search in the Dissertation Digital Library (the real work of an end user). Request: dissertations or full abstracts on "Volga language group of Finno-Ugric languages" - . "Volga & language & group".

One record of search result (translated in English, SH =Russian letter III):

<i>Concise record</i>	
Class. number	/SH 166.32-211 /2 <i>rubbk</i>
Author	/Ignatyeva Elisaveta
Title	/ <i>The Derivation of Negative in the Mari language</i>
Publication	<i>Yoshkar-Ola 2004</i>
BBK caption hierarchy	<i>Philological sciences. Belles-lettres -- Linguistics -- Languages of the world -- Finno-Ugric languages -- Volga language group -- Mari languages (Mari, Cheremis) -- Grammar -- Morphology -- Word-formation</i>
E-address	http://dlib.rsl.ru/rsl01002000000/rsl01002732000/rsl01002732180/rsl01002732180.pdf

Conclusion: the Title and the full abstract don't include *Volga language group*. Users can find the full abstract only by the caption hierarchy.

The classification number structure (9 levels):

SH Philological sciences. Belles-lettres
+ *SH1 Linguistics*
++ *SH16 Languages of the world*
+++ *SH166 Finno-Ugric languages*
++++ *SH166.3 Volga language group*
+++++ *SH166.32 Mari languages (Mari, Cheremis)*
++++++ *SH166.32-2 Grammar*
+++++++ *SH166.32-21 Morphology*
+++++++ *SH166.32-211 Word-formation*

5. Such **caption hierarchies (caption strings) are available and function by search in OPAC** records from 1998 year only.

The aim is “decoding” of classification numbers in all remaining records for books (more than 7 millions), i.e. they should get caption hierarchies. We shall have in result more than 90% of records for books and dissertations in the OPAC for hierarchical search as a whole.

The General Classified Catalog guide cards were completely **digitized** for that in RSL.

A special software generates hierarchical trees in 2 forms like in Examples 2 and 3 on the base of that data.

Example 2. A fragment of the hierarchical tree of BBK classification numbers and their captions (like data form on the catalogue guide cards). (SHCH=Russian letter Ш).

SHCH Art. Study of Art
+*SHCH 3 Separate music varieties and musical performance*
++ *SHCH 31 Music*
+++ *SHCH 315 Instrumental music*
+++ +*SHCH 315.3/9 Music instruments. Study of instruments*
+++++ *SHCH 315.31 Ancient instruments*
+++++ +*SHCH 315.32 Folk instruments*
+++++ +*SHCH 315.4 Keyboard instruments*
+++++ +*SHCH 315.41 Clavichord. Clavecin. Spinet*
+++++ +*SHCH 315.42 Piano*

We use this data form **for data editing by software and experts**.

Example 3. Fragment of a hierarchical tree of BBK classification numbers and caption hierarchies (caption strings) from example 2.

SHCH 315.3/9 Art. Study of Art -- Music -- Separate music varieties and musical performance -- Instrumental music -- Music instruments. Study of instruments

And so on ...

SHCH 315.4 Art. Study of Art -- Music -- Separate music varieties and musical performance -- Instrumental music -- Music instruments. Study of instruments -- Keyboard instruments

SHCH 315.41 Art. Study of Art -- Music -- Separate music varieties and musical performance -- Instrumental music -- Music instruments. Study of instruments -- Keyboard instruments -- Clavichord. Clavecin. Spinet

SHCH 315.42 Art. Study of Art -- Music -- Separate music varieties and musical performance -- Instrumental music -- Music instruments. Study of instruments -- Keyboard instruments -- Piano

That kind of data was done already for all BBK sections and passed to experts.

6. We are going later **to build in the trees** all different classification numbers with the caption strings from our OPAC. For example (3 new levels for the example 3):

SHCH 315.42-03(0)5-3,0 Art. Study of Art -- Music -- Separate music varieties and musical performance -- Instrumental music -- Music instruments. Study of instruments -- Keyboard instruments -- Piano – Piano music history -- -- 19th century -- Themes and figures

7. A new task for the next year is embedding of that classification numbers with caption strings into **database of the «operational» classification system**. The database would serve **for the direct hierarchical search** and for search on free word combinations. The relevant amount of bibliographic records for every classification number (caption) would be indicated by search. It is very important to allocate the data base in the open access for processing by well-known Internet search systems.

8. **The classification numbers with caption strings are also the base for “decoding” of classification numbers** from “retro-records”.

A check-out of the technology on the base of an experimental software and many random classification numbers have made it clear that it is an effective means for that purpose.

At the input - E693.32-739.1,0 from a bibliographic record. The software compares it with numbers in database, finds out **E693.32-73** and takes out the string for this part of the classification number:

Life sciences-- Zoology -- Animals taxonomy. -- Chordata. Chordates -- Vertebrata. Vertebrates. Vertebrates zoology -- Pisces. Fish. Ichthyology -- Physiology, biophysics and biochemistry -- Physiology

Quite enough indeed.

Examples of full “decoded” numbers: *R410.150.11 (Pericarditis), SH5(2=P)75-65я44 (Poetry, collection), E693.363.99 (Dolphins), V192.18,09 (Mathematics. Iteration method) etc.*

We know what kind of **problems** is waiting us ahead. But we have skilled experts to solve them.