Software & Tools

Making Indices for VINITI's "Mathematics" Abstract Journal

S. A. Strelkov and G. R. Epshtein

Abstract

The process of automatically making the indices for the abstract journal *Mathematics* is considered. The basis for creation of the indices is the database of the bibliography description sources and the text abstracts, which are typeset in the russification LaTeX 2.09.

An IBM PC is applied for working with a database and printing camera-ready copies. *EmTeX*, *MakeIndex* and Perl scripts are used. Program *DviSpell* is applied for generating the Sort Keys.

Introduction

The principal task of the Printing and Publishing Plant of the All-Russia Institute of Scientific and Technical Information (VINITI) is to complete processing of rough copies, and to prepare the originals for typesetting, for the issue of scientific information publications.

Camera-ready copies for abstract journals are produced with the help of TeX. The abstract journal *Mathematics* is typeset in LaTeX 2.09 with NFSS release 1. The russification format file `rlplain.fmt`, permitting the application of Russian letters in TeX-commands, is used. XCM* fonts, the basis of Russian font forms, was developed in IHEP (Protvino) [1]. Typesetting, compilation of tex files, and output of `.dvi` files to a printer or to the screen take place on an IBM PC, thus emTeX programs are used.

We introduce our method for the accumulation of databases of the abstracts along with preparation of a camera-ready copy of the journal, *Mathematics*. During this process the abstracts and bibliographical descriptions of sources are typeset in the fields with the mnemonic entries.

```
{{{}}}\НомерРеферата 5
{{{}}}\СИД 6019838
{{{}}}\ЗаглавиеОсновное Mathematik kommt vor
{{{}}}\Автор Schmidt G\"unter
{{{}}}\Объем 61--74
{{{}}}\ЯзыкТекста Нем.
{{{}}}\ВидДокумента 1
{{{}}}\РубрШифр 51.009
{{{}}}\Процент 51:37
{{{}}}\ПОД АКК изучение геометрии
{{{}}}\Процент АКК геометрия окон
```

```
{{{}}}\Процент АКК декартова схема
{{{}}}\РусПеревод Математика выходит вперед.
{{{}}}\Реферат Обсуждается вопрос о приципах
привлекательности темы изучения геометрии в
архитектурном оформлении окон в готических
церквах. Далее указываются различные подходы
к решению задач такого рода (схема,
основанная на геометрических местах,
декартова схема).
{{{}}}\АвторРеферата В. Рыжков
{{{}}}\КонецФайла
```

The operators type in several abstracts in one text file, for example, the file `4n131.tex` contains the abstracts since number 131. The files that pertain to one issue of a journal are assembled in one directory, for example, `d:\4n\4n*.tex`. These files can then be used to collect the current number of a journal, to bring information into a database, and/or to construct the various indices.

The developed software is intended to formulate a collection of current, cumulative indices for an issue of *Mathematics*:

1. Short Authors Index;
2. Full Authors Index;
3. Classified Index;
4. Subject Index;
5. Index of Periodicals;
6. Index of Scientific Forums;
7. Index of Proceedings;
8. Index of Persons.

We shall consider the process of formation of the cumulative short authors index. The following procedures are usually used for a construction of an index [2]:

1. Construction of a raw index (`.idx` file) with the help of LaTeX.
2. Running the program *MakeIndex* for building an alphabetized index (.ind file).
3. Reading of the index (.ind file) by LaTeX to give the final typeset result.

In our case, the index construction process is carried out automatically with the help of additional converter programs written in the Perl 4.036 language. The program *DviSpell* [3] and appropriate Perl scripts are used for generating Sort Keys.

The operators, working in Norton Comander on an IBM PC, shift by the pointer to the file with the `.idx` extension (for example, author.idx) and press `<Enter>`.

Batch file mkidx.bat is started and all the necessary components of an alphabetized index are produced. In a couple of minutes, the author.tex file,

with an index completely ready for inclusion in the main file of a number of the abstract journal, will be formed in the current directory.

Generating and Transforming the Raw Index

At first, the raw index can be created in the usual manner through LaTeX with use of the command `\makeidx`. Here it is necessary to note that in an abstract journal a number of the abstracts in the given issue (short number) for the references in an index are used instead of a numbers of pages. An issue number (1-12) and an issue code (А, Б, В, Г, .93.) are specified in a semi-annual index also. A complete number of the abstract looks like: 11Б129 or 4.93.1297. In the abstract journal style file `rjmat.sty` the respective alterations are made for an insertion of short numbers of the abstracts in a raw index, which has a special aspect:

```
Lewis Ted    0410001щщщ
Schmidt G\"unter    0410005щщщ
Протасова~Л.~А.    0410012щщщ
Горбачук~М.~Л.    0410013щщщ
Ішлінський~О.~Ю.    0410013щщщ
```

Here the raw index, with complete numbers of the abstracts (file `author_f.idx`), is shown, which is extracted from files-portions assembled in the directory of issues.

The program *Mkidxmat.pl*, in the Perl language, was written for this purpose. This program is called as follows:

```
perl.exe mkidxmat.pl file_dir file_mode -d
```

The first parameter is the name of the file, containing names of directories subject to handling.

Example of the `file_dir`:

```
d:\1a
d:\2a
d:\3a
d:\4a
```

The second parameter is the name of the file, containing names of conditions of handling and names of output files adequate for these conditions. Numbers of the abstracts must also be specified as short or complete.

Format of a line:

```
<mode name>:<full or short>:<output file name>
```

Example of the file `file_mode`:

```
author:short:author_s.idx
author:full:author_f.idx
```

Apart from files described as outputs, the file `mkidxmat.log` is created with work protocols. The non-blank third parameter generates output of additional debug information in the protocol file.

Use DviSpell for Generating the Sort Keys

A feature of the indices for abstract journals is the usage of both the Russian and Latin alphabets, with almost all possible accents. The accents are mainly used in surnames of the authors. For a set of some letters, a mathematical mode is used.

When typesetting a large volume of journals, a tool for an automatic construction of sort keys is required. Because the operators do not know the English language, standard tools for construction of keys, available in the program *MakeIndex* 2.11 [6, 7] and 3.08 [8], do not provide a satisfactory outcome in such situations. This circumstance has induced the authors to use E. Mattes' program, *DviSpell* [3], for automatic construction of sort keys.

DviSpell is the table-controlled program intended for transformation of `.dvi` files into readable text files. This program is an effective, flexible tool for discernment and transformation of symbols with accents. The tables controlling the work of *DviSpell* are concentrated in files with the extension `.dsi`. The program *DviPrep* translates `.dsi` files into a binary file of parameters with the extension `.dsb`.

The `.dvi` file to be transformed with the help of *DviSpell* is a result of compilation of the file `dvimake.tex`, which looks like this:

```
\documentstyle[russian]{article}
\begin{document}
\parindent=0mm
Lewis Ted    0410001щщщ
Schmidt G\"unter    0410005щщщ
Протасова~Л.~А.    0410012щщщ
Горбачук~М.~Л.    0410013щщщ
Ішлінський~О.~Ю.    0410013щщщ
A    \vspace{8dd} 0\par
...................
Z    \vspace{8dd} 0\par
А    \vspace{8dd} 0\par
......................
Я    \vspace{8dd} 0\par
\end{document}
```

This file is produced from the file `author.idx` by simple transformations, adding at the end of the file lines such as

Ю   `\vspace{8dd} 0\par`.

These lines become first on the letter in the sorting file.

Note: In the examples of files the long lines have been broken into several short lines.

DviSpell converts the characters found in the `.dvi` file into symbolic names using the layout table and the font table. The resulting sequence of symbols is converted into another sequence of symbols,

using the active conversion table. The result is printed using the active output table.

The main idea of the application *Dvispell* for a construction of sort keys is for use on two letter code by the lower case latin letters in a table of output. If the order of the characters, letters with accents, and other symbols to be included in the sorting is known, a table of output of the symbols exceeding 256 is created. A fragment of such a table is represented below.

```
(output makeindx
(WORDSPACE ' )
(NEWLINE   "0a)
(NEWPAGE   ' )
(Z#A     ('a 'a)) % capital A
....................................
(Z#Ishrt ('a 'l)) % capital short I
....................................
(Z#YU    ('b 'g)) % capital Yu
(Z#YA    ('b 'h)) % capital Ya
....................................
)
```

In the tables of fonts and the tables of transformations are added descriptions of Russian fonts of sets LH* [5] and XCM* [1]. In a table of the letters the Russian letters and the additional accented characters (a-cedilla, d-cedilla, e-cedilla, e#dieresis, E#dieresis, i-caron, and I-caron) are added. The names of the Russian letters in the tables begin with Z#. This is to prevent incorrect operation of the substitution algorithm used in *DviSpell*.

As a result, the binary file of parameters – russianz.dsb – taking into account a given order of accented letters and unaccented characters, is obtained. Namely, the accented letter directly follows the letter in a lexicographic order of names of accents.

For a production of indices in technological process the simplified scheme of a construction of sort keys, realized in the file russianm.dsb is used. Letters are not case-sensitive and the Russian letters go before Latin. The accents do not influence sort keys, i.e., the conversion table of accents is empty. In the output table makeidx these will not be transformed. In a table of the letters the Russian letters are added, and all the accented letters are eliminated.

DviSpell is called as follows:

```
dvispell.exe -y 0.8 -o makeidx -d russian.dsb
            -v dvimake.dvi author.eee
```

The value of the parameter -y must equal 0.8, to prevent lines in the output file author.eee from running together.

Running the MakeIndex Program

With the help of programs in Perl we shall transform the files author.idx and author.eee to an input file author.jjj for the sorting program *MakeIndex*:

```
\indexentry{clcecwcics ctcecd@
Lewis Ted
\markboth{LEW}{LEW}}{0410001}
\indexentry{csccchcmcicdct cgcucnctcecr@
Schmidt G\"unter
\markboth{SCH}{SCH}}{0410005}
\indexentry{arasaqauaaataqacaa an aa@
Протасова~Л.~А.
\markboth{ПРО}{ПРО}}{0410012}
\indexentry{adaqasabaaazavam ao an@
Горбачук~М.~Л.
\markboth{ГОР}{ГОР}}{0410013}
\indexentry{cibaanciapatbeamakal aq bg@
Ишлінський~О.~Ю.
\markboth{ИШЛ}{ИШЛ}}{0410013}
\indexentry{ca@A \vspace{8dd}}{0}
....................................
\indexentry{cz@Z \vspace{8dd}}{0}
\indexentry{aa@A \vspace{8dd}}{0}
....................................
\indexentry{be@Я \vspace{8dd}}{0}
```

The sort keys are located in the left part of the command \indexentry{, in the right part (after @) is the surname of the author extracted from author.eee. In the fields of the command \markboth the first three decoded letters from the sort key are located. This is necessary for construction of the page headers of a cumulative index.

For the programs *MakeIndex* 3.08 or 2.11, the *EMX* 0.9b compiler is used. For *MakeIndex* 3.08, enlarged values of parameters are set at:

```
#define LONG_MAX      1024
#define STRING_MAX    256
#define ARABIC_MAX    7
```

The command line to start *MakeIndex* looks like this:

```
makeidx32.exe -s mkind.ist
              -o author.ind author.jjj
```

The sorted file author.ind will be transformed with the help of the programs in Perl to the file author.tex, ready for insertion in the main file:

```
\chapter{
}{АВТОРСКИЙ УКАЗАТЕЛЬ}
{А\,в\,т\,о\,р\,с\,к\,и\,й%
\,у\,к\,а\,з\,а\,т\,е\,л\,ь{}
\dsection{}{}{Авторский указатель{}
\begin{multicols}{5}
\raggedright
\leftskip=8dd  \parindent=-8dd
\par
\vspace{0dd}
```

```
{\bf А}блялимов~С.~Б.%
\markboth{АБЛ}{АБЛ} 4А94\par
Абрамов~С.~А.%
\markboth{АБР}{АБР} 4А275\par
............................
\vspace{8dd}
{\bf Ю}рьев~Д.~В.%
\markboth{ЮРЬ}{ЮРЬ} 4186\par
{\bf Я}нчевский~В.~И.%
\markboth{ЯНЧ}{ЯНЧ} 4А273 \par
\end{multicols}
\rule{0dd}{3mm}
\begin{multicols}{5}
\raggedright
\leftskip=8dd
\parindent=-8dd
\vspace{0dd}
{\bf A}bdesselam~B.%
\markboth{ABD}{ABD} 4A364\par
Aberbach Ian~M.%
\markboth{ABE}{ABE} 4A341\par
............................
\vspace{8dd}
{\bf Z}\'adori L\'aszl\'o%
\markboth{ZAD}{ZAD} 4A204\par
ZsilinszkyL\'aszl\'o%
\markboth{ZSI}{ZSI} 4A87\par
............................
\.Zukowski Tomasz%
\markboth{ZUK}{ZUK} 4A454\par
Zulli Louis%
\markboth{ZUL}{ZUL} 4A431\par
\endinput
```

Conclusions

In this paper we have only touched on the main
ideas of construction of an index file. The actual
process involves a great many more intermediate
steps, including transformations of files, filtration
and checks. When the typesetting is done with
essentially no errors, the process of construction of
indices is carried out automatically.

References

[1] Н.Л. Глонти, С.В. Клименко, В.К. Малы-
шев, А.В. Самарин, Б.Б. Филимонов :
Метапроект кирилловского алфавита для
печатающих устройтв с высоким разреше-
нием: Препринт ИФВЭ 90-96, – Протвино,
1990.

[2] M. Goossens, F. Mittelbach, and A. Samarin :
The LaTeX Companion, Addison-Wesley, Read-
ing, MA, 1994. : Chapter 12. Index Generation.
pp. 345-370.

[3] Eberhard Mattes: file dvispell.doc 1.0b 16-Aug-
1995, file dvisprep.doc 1.0a 04-Jun-1995.

[4] Eberhard Mattes : file emxdoc.doc 0.9b 10-Dec-
1995.

[5] O. Lapko : MAKEFONT as part of *CyrTUG*-
emTeX package, Proceedings of the Eighth Eu-
ropean TeX Conference, Gdańsk, pp. 110–114,
1994.

[6] Pehong Chen and Michael A. Harisson : Index
preparation and processing. Software-Practice
and Experience, 19(9): 897-915, September
1988.

[7] Pehong Chen. : MakeIndex manual page.

[8] Joachim Schrod and Gabor Herr : MakeIndex
Version 3.0. Reference, Draft version, August
1991.

⬦ S. A. Strelkov
The Keldysh Institute of Applied
Mathematics RAS
Miusskaya Sq., 4
Moscow A-47, Russia, 125047
strelka@applmat.msk.su

⬦ G. R. Epshtein
The Printing and Publishing Plant
of the All-Russia Institute
of Scientific and Technical
Information
October Street, 403
Lyubertsy, Moscow region, Russia,
140010